



Study Reports

P4-C4-001

Objectives 1–5

DARWIN EU® - Enabling pregnancy and mother-child research in DARWIN EU® Data Network - DARWIN EU® PeriNet

25/06/2026

Final Version

Public

CONTENTS

OBJECTIVE 1.....	9
LIST OF ABBREVIATIONS.....	11
1. TITLE	13
2. DESCRIPTION OF THE STUDY TEAM.....	13
3. ABSTRACT	14
4. AMENDMENTS AND UPDATES	17
5. MILESTONES	17
6. RATIONALE AND BACKGROUND	17
7. RESEARCH QUESTION AND OBJECTIVES.....	17
8. RESEARCH METHODS.....	18
8.1. Study design	18
8.2. Study setting and data sources	18
Table 1. Description of the selected data sources.....	20
8.3. Statistical methods	24
8.3.1. Main summary measures	24
8.4. Deviations from the protocol	24
9. RESULTS.....	24
9.1. Participants.....	24
9.2. Main results	25
9.2.1. Results from questionnaire 1 (objective 1a, 1b, and 1c).....	25
Table 2. Details of data sources and availability of pregnancy related information (results from questionnaire 1).	25
9.2.2. Counts from the DARWIN EU® Portal (objective 1a, 1b, and 1c)	27
Table 3. Sum of person counts of all searched pregnancy and infant related concept IDs in each data source, obtained from the DARWIN EU® Portal.....	28
Table 4. Pregnancy related concept IDs in selected data sources among individuals from 0 to 9 years of age, obtained from the DARWIN EU® Portal.....	30
Table 5. Pregnancy related concept IDs in each data source recorded at any time among individuals older than 50 years of age, obtained from the DARWIN EU®	31
Table 6. Pregnancy related concept IDs in each data source recorded at any age in males, obtained from the DARWIN EU® Portal.	33
9.2.3. Results from questionnaire 2 (objective 1d)	35
10. DISCUSSION	35
10.1. Strengths and limitations of the research methods.....	36
11. CONCLUSIONS	37
12. REFERENCES	38
13. ANNEXES.....	39
ANNEX I. Description of data sources.....	39
ANNEX II. Questionnaires for Data Partners on pregnancy data and Mother-Child Linkage details.....	49
ANNEX III. Codes for pregnancy and infant information searches.....	55
Table S1. List of pregnancy related codes for search.	55

Table S2. List of infant related codes for data search.	73
ANNEX IV. Detailed results	89
Table S3. Detailed comments of availability of pregnancy related information in data sources (questionnaire 1).	89
Table S4. Results from Questionnaire 1.	89
Table S5. Record counts (RC) and Person Counts (PC) of pregnancy-related concepts explored in the DARWIN EU® Portal.	89
Table S6. Record counts (RC) and Person Counts (PC) of pregnancy-related concepts explored in the DARWIN EU® Portal, by sex.	89
Table S7. Record counts (RC) and Person Counts (PC) of pregnancy-related concepts explored in the DARWIN EU® Portal, by age.	89
Table S8. Record counts (RC) and Person Counts (PC) of infant-related concepts explored in the DARWIN EU® Portal.	89
Table S9. Record counts (RC) and Person Counts (PC) of infant-related concepts explored in the DARWIN EU® Portal, by sex.	90
Table S10. Record counts (RC) and Person Counts (PC) of infant-related concepts explored in the DARWIN EU® Portal, by age.	90
Table S11. Results from Questionnaire 2.	90
ANNEX V. Glossary.	91
OBJECTIVE 2.	94
LIST OF ABBREVIATIONS.	96
1. TITLE.	98
2. DESCRIPTION OF THE STUDY TEAM.	98
3. ABSTRACT.	100
4. AMENDMENTS AND UPDATES.	103
5. MILESTONES.	103
6. RATIONALE AND BACKGROUND.	103
7. RESEARCH QUESTION AND OBJECTIVES.	103
8. RESEARCH METHOD.	104
8.1. Study design.	104
8.2. Follow-up.	104
8.3. Study population with inclusion and exclusion criteria.	105
8.4. Study setting and data sources.	105
Table 1. Data sources.	105
8.5. Study period.	108
8.6. Variables.	109
8.6.1. Outcome.	109
8.6.2. Covariates.	109
8.7. Study size.	109
8.8. Data transformation.	109
8.9. Statistical methods.	110
Figure 1. Study process and relationship with each study objective.	110
Table 2. Episode metrics to assess internal consistency and plausibility of derived pregnancy episodes.	111

Table 3. Metrics and sources for reference data to assess the external validity of the algorithm's output at the national level.	114
8.10. Deviations from the protocol	115
9. RESULTS	116
9.1. Main results	116
9.1.1. Algorithm scoping review (objective 2a)	116
9.1.2. <i>PregnancyIdentifier</i> R package	117
9.1.3. Input variables (objective 2b)	118
Table 4. Number of individuals with availability of pregnancy algorithm input concepts across 15 data sources at Run 1 (<i>CohortDiagnostics</i>).	119
9.1.4. Multi-data source implementation and iterative improvements (objective 2c)	120
Table 5. Cross-database quality control metrics for inferred pregnancy episodes (15 data sources).	121
Table 6. Summary of quality control (QC) metric improvements across runs.	123
Table 7. Summary of the runs describing the main issues and corresponding changes to address them.	124
9.1.5. Internal consistency and plausibility checks in the data sources grouped by settings (primary care/secondary care/claims/registry)	124
Table 8. Run 3 summary metrics across participating data sources.	125
9.1.6. Primary care.....	125
Netherlands: Integrated Primary Care Information (IPCI).....	125
Spain: Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público (BIFAP).	126
Spain: The Information System for Research in Primary Care (SIDIAP)	126
United Kingdom: Clinical Practice Research Datalink GOLD (CPRD GOLD)	127
9.1.7. Secondary care	127
Finland: Tampere University Hospital patient cohort (FinOMOP-TaUH Pirha).....	127
France: Clinical Data Warehouse of Bordeaux University Hospital (CDW Bordeaux).....	127
Hungary: Semmelweis University Clinical Data (SUCD).....	128
Spain: Institut Municipal Assistència Sanitària Information System (IMASIS)	128
9.1.8. Claims	129
Germany: InGef Research Database (InGef RDB).....	129
9.1.9. Registry-based	129
Croatia: Croatian National Public Health Information System (NAJS).....	129
Denmark: Danish Data Health Registries (DK-DHR)	129
Finland: Finnish Care Register for Health Care (FinOMOP-THL).....	130
Norway: Norwegian Linked Health Registry data (NLHR@UiO).....	130
Sweden: Health Impact - Swedish Population Evidence Enabling Data-linkage (HI-SPEED)	130
Table 9. Run 3 overview of episode-level QC metrics by data source (grouped by setting).....	131
Table 10. Population-level benchmarking by data source (grouped by setting).	132
Table 11. PET comparison summary by data source (grouped by setting).	134
10. DISCUSSION	135
10.1. Key results	135
10.2. Strengths and limitations of the research methods.....	135
10.3. Interpretation	137
10.4. Generalisability	140
11. CONCLUSION	140
12. REFERENCES	141

13. ANNEXES.....	143
ANNEX I. Description of data sources.....	143
ANNEX II. Fitness for use assessment.....	163
Overview of the two Norwegian CDM datasets available.....	164
ANNEX III. Operational and reporting considerations.....	167
ANNEX IV. List of stand-alone documents.....	168
Table S1. National statistics table utilised as reference for examining external validity of pregnancy algorithm.	168
Table S2. A scoping literature review to assess published pregnancy algorithms.	169
ANNEX V. Glossary.....	183
OBJECTIVE 3.....	186
LIST OF ABBREVIATIONS.....	188
1. TITLE	189
2. DESCRIPTION OF THE STUDY TEAM	189
3. ABSTRACT	190
4. AMENDMENTS AND UPDATES	192
5. MILESTONES	192
6. RATIONALE AND BACKGROUND	192
7. RESEARCH QUESTION AND OBJECTIVES.....	193
8. RESEARCH METHOD	193
8.1. Data partners.....	193
Table 1. Description of the selected data sources.....	194
8.1.1. Data Selection.....	194
8.2. Description of mapping options	194
8.3. Evaluation of mapping options.....	195
Table 2. Evaluation Criteria.....	196
8.4. Creation of recommendations	196
8.5. Support PET mapping for DPs agreeing to map	196
9. RESULTS.....	196
9.1. Description of mapping options (Objective 3a).....	196
Figure 1. Overview of the OMOP CDM v5.4 with added episode table.	197
9.1.1. Mapping options for pregnancy episodes.....	197
9.1.2. Mapping options for pregnancy-related variables.....	198
9.1.3. Mapping options for parent-child linkage.....	199
Table 3. Fields of the PET table relevant for linkage between mother and child.	199
Table 4. Fields of the OMOP FACT_RELATIONSHIP table with an example of a bi-directional linkage of mother and child.	199
9.2. Evaluation of described mapping options (Objective 3b)	200
9.2.1. Quantitative Assessment.....	200
Table 5. Expected ease of Implementation for each option, scored by the four DPs.....	200
Table 6. Mapping options scored by Extensiveness, Implementation, and Ease of analysis.	201
9.2.2. Qualitative Assessment	202
Table 7. Assessment of Pregnancy Episodes solutions.....	202
Table 8. Assessment of Pregnancy-related Variables solutions.	203
Table 9. Assessment of Parent-Child Linkage solutions.	203

9.3. Perinatal pregnancy recommendation (Objective 3c)	204
Figure 2. Screenshot of the Pregnancy Episode guidance page.	205
9.4. Supporting DPs in study participation (Objective 3d)	206
Figure 3. Screenshot of the PET Table Overview page.	206
Table 10. Number of checks implemented across PET by table and level.	207
10. DISCUSSION	207
10.1. Key results	207
10.2. Strengths and limitations of the research methods.....	209
11. CONCLUSION.....	210
12. REFERENCES	211
13. ANNEXES.....	212
ANNEX I. Description of data sources.....	212
ANNEX II. Overview of mapping options	217
Table S1. Pregnancy Episodes.....	217
Table S2. Pregnancy-Related Variables.	218
ANNEX III. EPISODE and EPISODE_EVENT Tables	220
Table S3. EPISODE Table.	220
Table S4. EPISODE_EVENT Table.	223
ANNEX IV. Reviewer reflections	226
OBJECTIVE 4.....	231
LIST OF ABBREVIATIONS.....	233
1. TITLE	235
2. DESCRIPTION OF THE STUDY TEAM	235
3. ABSTRACT	236
4. AMENDMENTS AND UPDATES	240
5. MILESTONES	240
6. RATIONALE AND BACKGROUND	240
7. RESEARCH QUESTION AND OBJECTIVES.....	241
8. RESEARCH METHODS.....	241
8.1. Study design	241
8.2. Follow-up	241
Figure 1. Graphical depiction of the study design for Objective 4c, assessing pregnant individuals.....	242
8.3. Study population with inclusion and exclusion criteria.....	243
8.4. Study setting and data sources	243
Table 1. Data sources.....	244
8.5. Study period	245
8.6. Variables	245
8.6.1. Covariates	245
Table 2. Covariates included in the characterisation of pregnant individuals.	249
Table 3. Covariates included in the characterisation of infants with mother-child linkage.....	251
8.7. Study size	251
8.8. Data transformation	251
8.9. Statistical methods	252

8.9.1. Main summary measures	252
8.9.2. Main statistical methods	252
8.9.3. Missing values.....	252
8.10. Deviations from the protocol	253
9. RESULTS.....	254
9.1. Participants.....	254
9.1.1. Pregnancy episodes	254
Table 4. Study attrition of pregnancy episodes, by data source.	255
9.1.2. Infants with mother-child linkage	256
Table 5. Study attrition of infants with mother-child linkage, by data source.....	256
9.2. Main results	257
9.2.1. Characterisation of pregnancy episodes, overall and by year (Objectives 4a & 4b).....	257
Table 6. Descriptive statistics of pregnancy episodes, as recorded in the PET.	258
Figure 2. Number of pregnancy episodes by calendar year of pregnancy start (in blue), and number of infants with mother-child linkage by year of infant observation start (in red).....	261
Figure 3. Percentage of pregnancy episodes starting in each calendar year by pregnancy outcome.	262
9.2.2. Maternal characteristics (Objective 4c)	263
Demographics, health-related behaviours, and baseline comorbidities	263
Table 7. Demographics and health-related behaviours of pregnant individuals (assessed at pregnancy episode level).....	264
Table 8. Baseline comorbidities of pregnant individuals prior to pregnancy start (assessed at pregnancy episode level).....	266
Pregnancy complications.....	266
Table 9. Characteristics of pregnant individuals in terms of pregnancy complications (assessed at pregnancy episode level).....	267
Figure 4. Percentage of pregnancy complications, overall and by trimester.....	268
Medication use based on ATC 3 rd level groups.....	268
Table 10. Top 10 pharmacological subgroups (ATC 3 rd level groups) used during pregnancy.....	270
Large-scale characterisation	272
Table 11. Top 10 conditions identified through large-scale characterisation during pregnancy.....	272
Table 12. Top 10 ingredients identified through large-scale characterisation during pregnancy.	273
9.2.3. Infant characteristics (Objective 4d)	275
Table 13. Characteristics of infants with mother-child linkage.	275
Table 14. Breastfeeding, assessed in the SIDIAP data source.	276
10. DISCUSSION	276
10.1. Key results	276
10.2. Strengths and limitations of the research methods.....	278
10.3. Interpretation	280
10.3.1. Results in context	280
10.3.2. Perinatal data capture	282
10.3.3. Implications for future research in DARWIN EU®.....	284
10.4. Generalisability	285
11. CONCLUSION.....	286
12. REFERENCES	287
13. ANNEXES.....	290
ANNEX I. Description of data sources.....	290

Norway: Norwegian Linked Health Registry data (NLHR)	290
Spain: Institut Municipal Assistència Sanitària Information System (IMASIS)	291
Spain: The Information System for Research in Primary Care (SIDAP)	292
Sweden: Health Impact - Swedish Population Evidence Enabling Data-linkage (HI-SPEED)	293
United Kingdom: Clinical Practice Research Datalink GOLD (CPRD GOLD)	294
ANNEX II. Fitness for use assessment.....	296
Table S1. Overview of the two Norwegian CDM datasets available.	297
ANNEX III. Concept sets	300
Table S2. List of codes to identify healthcare visits.....	300
Table S3. List of codes used to assess smoking status.	300
Table S4. List of condition definitions.....	301
Table S5. List of medicines definitions.	370
Table S6. List of pregnancy complications (not included in the PET).	384
Table S7. Infant covariates.	395
ANNEX IV. Operational and reporting considerations	396
ANNEX V. Supplementary tables and figures	398
Figure S1. Graphical depiction of the study design for Objective 4d.	398
Figure S2. Percentage of pregnancy episodes starting in each calendar year by mode of delivery.	399
ANNEX VI. Glossary	400
OBJECTIVE 5	403
LIST OF ABBREVIATIONS	405
1. TITLE	406
2. DESCRIPTION OF THE STUDY TEAM	406
3. ABSTRACT	407
4. AMENDMENTS AND UPDATES	409
5. MILESTONES	409
6. RATIONALE AND BACKGROUND	409
7. RESEARCH QUESTION AND OBJECTIVES.....	409
8. RESEARCH METHODS.....	410
8.1. Definition of variables (objective 5a)	410
8.2. Development of software to extract aggregated counts (objective 5b)	410
8.3. Extraction of aggregated counts (objective 5c).....	411
8.4. Creation of new feasibility dashboard (objective 5d)	411
9. RESULTS	411
9.1. Definition of variables (objective 5a).....	411
Table 1. Overview of type I variables needed to inform future feasibility assessments of perinatal studies included in the onboarding packages.	411
Table 2. Overview of variables for feasibility of pregnancy and perinatal studies to be aligned with DPs.	412
9.2. Development of software to extract aggregated counts (objective 5b).....	412
Table 3. List of analyses implemented in the <i>DashboardExport</i> package. The 'stratum' column contains the OMOP CDM field names that the analysis stratifies by.....	413
9.3. Execution of aggregate calculation software (objective 5c)	413
9.4. Creation of dashboard with aggregates for feasibility (objective 5d)	413

Figure 1. Screenshot of the Perinatal section of the Network Dashboard in the DARWIN EU® Portal.	414
Figure 2. Screenshot DARWIN EU® Portal showing available familial relationships per data partner. Note that only results from data partners in PeriNet are shared here.	415
Figure 3. Screenshot DARWIN EU® Portal showing pregnancy person and record counts stratified by mode of delivery, and by outcome.	415
Figure 4. Screenshot of the DARWIN EU® Portal showing a boxplot of the age at start of pregnancy.	416
Figure 5. Screenshot DARWIN EU® Portal showing the questionnaire section for a data partner on Pregnancy availability and Parent-child linkage.	417
10. DISCUSSION	418
10.1. Key results	418
10.2. Strengths and limitations of the research methods	418
11. CONCLUSION	419
12. REFERENCES	420
13. ANNEXES	421
ANNEX I. Glossary	421



Study Report

P4-C4-001

DARWIN EU® - PeriNet

OBJECTIVE 1

**Identifying data partners in the DARWIN
EU® Data Network with available data on
pregnancy, pregnancy start and end
dates, pregnancy outcomes, and mother
and child linkage**

14/04/2026

Version 4.0

Authors: Talita Duarte-Salles, Anum Zahra, Julieta Politi, Berta Raventós, Maxim
Moinat, Natasha Yefimenko

Public

Study title	DARWIN EU® - PeriNet Objective 1: Identifying data partners in the DARWIN EU® Data Network with available data on pregnancy, pregnancy start and end dates, pregnancy outcomes, and mother and child linkage
Study report version	V4.0
Date	14/04/2026
EUPAS number	EUPAS1000000812
Active substance	None
Medicinal product	None
Research question and objectives	To identify data partners in the DARWIN EU® Data Network with available data on pregnancy, pregnancy start and end dates, pregnancy outcomes, and mother-child linkage. The specific objectives were: 1a. To identify data partners in the DARWIN EU® Data Network with available structured data on pregnancy (pregnancy start and end date, pregnancy outcomes) mapped to the OMOP CDM or available at source but still to be mapped. 1b. To identify data partners without structured data on pregnancy but where pregnancy algorithms could be implemented to derive such information. 1c. To identify data partners in the DARWIN EU® Data Network with mother-child linkage. 1d. To obtain detailed information on availability, completeness and extensiveness of pregnancy-related data and mother-child linkage at source from each participating data partner.
Countries of study	Belgium, Croatia, Denmark, Estonia, Finland, France, Germany, Hungary, Italy, Norway, Portugal, Spain, Sweden, the Netherlands, the United Kingdom
Authors	Talita Duarte-Salles (t.duarte@darwin-eu.org) Anum Zahra (a.zahrar@darwin-eu.org) Natasha Yefimenko (n.yefimenkonosova@darwin-eu.org) Julieta Politi (j.politi@darwin-eu.org) Berta Raventós (b.raventos@darwin-eu.org) Maxim Moinat (m.moinat@darwin-eu.org)

LIST OF ABBREVIATIONS

Acronyms/term	Description
BIFAP	Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público
CC	Coordination Centre
CDM	Common Data Model
CDW Bordeaux	Clinical Data Warehouse of Bordeaux University Hospital
CPRD GOLD	Clinical Practice Research Datalink GOLD
DARWIN EU®	Data Analysis and Real-World Interrogation Network
DK-DHR	Danish Data Health Registries
EBB	Estonian Biobank
EHR	Electronic Health Records
EMA	European Medicines Agency
EMDB-ULSEDV	Egas Moniz Health Alliance database - Entre o Douro e Vouga
EMDB-ULSGE	Egas Moniz Health Alliance database - Gaia E Espinho
EMDB-ULSRA	Egas Moniz Health Alliance database - Baixo Vouga (Região de Aveiro)
ETL	Extract, Transform, and Load
EU	European Union
EUPAS	EU Post-Authorisation Studies Register
FinOMOP-ACI Varha	Auria Clinical Informatics
FinOMOP-HUS	Hospital District of Helsinki and Uusimaa
FinOMOP-TaUH Pirha	Tampere University Hospital patient cohort
FinOMOP-THL	Finnish Care Register for Health Care
GP	General Practitioner
HI-SPEED	Health Impact - Swedish Population Evidence Enabling Data-linkage
ICD	International Classification of Diseases
IMASIS	Institut Municipal Assistència Sanitària Information System
InGef RDB	InGef Research Database
IPCI	Integrated Primary Care Information
IQVIA DA Germany	IQVIA Disease Analyzer Germany
IQVIA LPD Belgium)	IQVIA Longitudinal Patient Database Belgium
NCR	The Netherlands Cancer Registry
NAJS	Croatian National Public Health Information System
NLHR	Norwegian Linked Health Registry data
NLHR@UiO:PERINATAL	Norwegian Linked Health Registry data - Perinatal and Pregnancy
NNRD	National Neonatal Research Database
OHDSI	Observational Health Data Sciences and Informatics
OMOP	Observational Medical Outcomes Partnership

Acronyms/term	Description
PC	Person count
PET	Perinatal Expansion Table
POLIMI	Research Repository @Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico
SIDIAP	The Information System for Research in Primary Care
SNDS	Système National des Données de Santé
SNOMED	Systematized Nomenclature of Medicine
SUCD	Semmelweis University Clinical Data
UKBB	UK BioBank
ULSM-RT	Unidade Local de Saúde de Matosinhos Realtime Database
VID	Valencia Health System Integrated Dataset

1. TITLE

DARWIN EU® - PeriNet Objective 1: Identifying data partners in the DARWIN EU® Data Network with available data on pregnancy, pregnancy start and end dates, pregnancy outcomes, and mother and child linkage.

2. DESCRIPTION OF THE STUDY TEAM

Study team role	Names	Organisation
Principal Investigator	Talita Duarte-Salles	Erasmus MC
Co-Principal Investigators	Julieta Politi Anum Zahra	Erasmus MC
Data Scientist	Aniek Markus Maarten van Kessel Ger Inberg Ioanna Nika Cesar Barboza Ross Williams	Erasmus MC
Study Manager	Natasha Yefimenko	Erasmus MC
Network Operations Lead	Maxim Moinat	Erasmus MC
Epidemiologist	Berta Raventós	Erasmus MC

3. ABSTRACT

Title

DARWIN EU® - PeriNet Objective 1: Identifying data partners in the DARWIN EU® Data Network with available data on pregnancy, pregnancy start and end dates, pregnancy outcomes, and mother and child linkage.

Rationale and background

Pregnant individuals are often excluded from clinical trials, creating an evidence gap on the safety and effectiveness of medicines during pregnancy. As a result, pregnancy is a key focus for post-authorisation safety monitoring and risk management. To enable robust analyses within DARWIN EU®, systematic assessment of pregnancy-related data of structure, coding, and completeness is needed in line with the European Medicines Agency Data Quality Framework.

The DARWIN EU® network's adoption of the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM) supports standardised data harmonisation. However, its application in perinatal research has been limited given the lack of common definitions and methods to derive pregnancy episodes, outcomes, and mother-child linkages.

This study aimed to identify data partners across the DARWIN EU® Data Network with valuable information on perinatal health, evaluating data availability, mapping of this information to the OMOP CDM, and assessing mother-child linkage capabilities. Findings informed the network's readiness to support perinatal-related research and additional mappings needed.

Research question and objectives

Research question

Which data partners in the DARWIN EU® Data Network have available data on pregnancy, pregnancy start and end dates, pregnancy outcomes and mother and child linkage, and in what format?

Objectives

To identify data partners in the DARWIN EU® Data Network with available data on pregnancy, pregnancy start and end dates, pregnancy outcomes, and mother-child linkage.

The specific objectives were:

- 1a. To identify data partners in the DARWIN EU® Data Network with available structured data on pregnancy (pregnancy start and end date, pregnancy outcomes) mapped to the OMOP CDM or available at source but still to be mapped.
- 1b. To identify data partners without structured data on pregnancy but where pregnancy algorithms could be implemented to derive such information.
- 1c. To identify data partners in the DARWIN EU® Data Network with mother-child linkage.
- 1d. To obtain detailed information on availability, completeness and extensiveness of pregnancy-related data and mother-child linkage at source from each participating data partner.

Methods

Data sources

1. Belgium: IQVIA Longitudinal Patient Database Belgium (IQVIA LPD Belgium)
2. Croatia: Croatian National Public Health Information System (NAJS)
3. Denmark: Danish Data Health Registries (DK-DHR)

4. Estonia: Estonian Biobank (EBB)
5. Finland: Auria Clinical Informatics (FinOMOP-ACI Varha)
6. Finland: Hospital District of Helsinki and Uusimaa (FinOMOP-HUS)
7. Finland: Tampere University Hospital patient cohort (FinOMOP-TaUH Pirha)
8. Finland: Finnish Care Register for Health Care (FinOMOP-THL)
9. France: Clinical Data Warehouse of Bordeaux University Hospital (CDW Bordeaux)
10. France: Système National des Données de Santé (SNDS)
11. Germany: InGef Research Database (InGef RDB)
12. Germany: IQVIA Disease Analyzer Germany (IQVIA DA Germany)
13. Hungary: Semmelweis University Clinical Data (SUCD)
14. Italy: Research Repository @Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico (POLIMI)
15. Norway: Norwegian Linked Health Registry data (NLHR)
16. Norway: Norwegian Linked Health Registry data - Perinatal and Pregnancy (NLHR@UiO:PERINATAL)
17. Portugal: Egas Moniz Health Alliance database - Entre o Douro e Vouga (EMDB-ULSEDV)
18. Portugal: Egas Moniz Health Alliance database - Gaia E Espinho (EMDB-ULSGE)
19. Portugal: Egas Moniz Health Alliance database - Baixo Vouga (Região de Aveiro) (EMDB-ULSRA)
20. Portugal: Unidade Local de Saúde de Matosinhos Realtime Database (ULSM-RT)
21. Spain: Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público (BIFAP)
22. Spain: Institut Municipal Assistència Sanitària Information System (IMASIS)
23. Spain: The Information System for Research in Primary Care (SIDIAP)
24. Spain: Valencia Health System Integrated Dataset (VID)
25. Sweden: Health Impact - Swedish Population Evidence Enabling Data-linkage (HI-SPEED)
26. The Netherlands: Integrated Primary Care Information (IPCI)
27. The United Kingdom: Clinical Practice Research Datalink GOLD (CPRD GOLD)
28. The United Kingdom: UK BioBank (UKBB)

Statistical analysis

The results of the screening questionnaire were described narratively and through the use of simple frequency counts and percentages. A minimum cell count of 5 was used when reporting results, with any smaller count reported as "<5" and zero counts as "0". Detailed pregnancy-related person counts in the DARWIN EU® portal, which contains pregnancy and infant related conditions from OMOP mapped data, was also used for objective 1d.

Results

Of data sources onboarded in DARWIN EU® at time of feasibility assessment (i.e., 15th July 2024), a total of 14 data partners with access to 16 data sources from 12 countries were selected and accepted to participate in the study. Selection of data sources was based on data suitability, responses to questionnaire 1 (27 data sources), responses to questionnaire 2 (15 data sources), detailed pregnancy-related person counts in the DARWIN EU® portal (16 data sources) and confirmation to study participation. 10 out of 15

data sources (66%) reported (in questionnaire 2) that a pregnancy start date is available in their database while only 5 of these data sources currently had it mapped to the OMOP CDM (33%). 8 out of 15 data sources (53%) reported to have mother-child linkage available, while 4 data sources currently had it mapped to OMOP CDM (27%).

Most pregnancy related variables (e.g., Gestational age, last menstrual period, delivery, stillbirths, mode of delivery) had already been frequently mapped to the OMOP CDM. However, timing specific pregnancy variables (e.g., last menstrual period, derived pregnancy start date) were less consistently mapped to the OMOP CDM across data sources that replied to have this information available in their source data in questionnaire 2.

Discussion

Based on the results, 14 (of 28) data partners with access to 16 data sources from 12 countries were identified with self-reported structured pregnancy data either mapped or mappable to the OMOP CDM and with pregnancy information that potentially could be derivable through the application of pregnancy algorithms. Several data partners with existing or potential mother-child linkage were identified. Data sources appeared to contain relevant detailed pregnancy-related information although considerable heterogeneity of pregnancy information and granularity may be evident that could impact on the performance of a pregnancy algorithm. Insight into the level of detail and OMOP mapping status regarding pregnancy related information in these data sources emphasises the need for further harmonisation and mapping efforts to support future cross-database pregnancy research.

4. AMENDMENTS AND UPDATES

None.

5. MILESTONES

Study deliverable	Timelines (planned)	Timelines (actual)
Final Study Protocol	December 2025	December 2025
Draft Study Report	January 2026	February 2026
Final Study Report	March 2026	23/03/2026

6. RATIONALE AND BACKGROUND

Observational healthcare data provide a key opportunity to address existing evidence gaps in perinatal health as data sources may capture clinically relevant information on pregnancy, perinatal outcomes, and infant health. However, the availability, granularity, and research-readiness of such data vary considerably across data sources. A systematic assessment of pregnancy-related data, including its structure, coding, and completeness, is essential to support robust pharmacoepidemiological analyses in DARWIN EU® and to align with the European Medicines Agency (EMA) Data Quality Framework, for example, in terms of extensiveness, relevance, and reliability.[3]

The adoption of the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM) within the DARWIN EU® network provides a standardised approach for data harmonisation and analysis. Nevertheless, the use of OMOP CDM for perinatal research has been limited due to structural constraints in defining pregnancy episodes and key characteristics, such as gestational length, pregnancy outcomes, and mother–infant linkages. The recent development of the OMOP Perinatal Expansion Table (PET) addresses some of these limitations by introducing a standardised structure for pregnancy and infant data.[4] Successful utilisation of the PET, either through mapping at Extract, Transform, and Load (ETL) stage or through the creation of a temporary PET table through the use of pregnancy algorithms, however, depends on the underlying data capture from source data.[5–7]

This data characterisation and descriptive study aimed to systematically evaluate pregnancy-related data across data sources participating in the DARWIN EU® network. This report corresponds to the first objective of the “DARWIN EU® - Enabling pregnancy and mother-child research in the DARWIN EU® Data network – DARWIN-PeriNet”. By assessing data availability, coding practices, and linkage capabilities in the DARWIN EU® Network, the study provided an overview of the network’s capacity to support pregnancy-related research and additional mappings needed.

7. RESEARCH QUESTION AND OBJECTIVES

Research question

Which data partners in the DARWIN EU® Data Network have available data on pregnancy, pregnancy start and end dates, pregnancy outcomes and mother and child linkage, and in what format?

Objectives

To identify data partners in the DARWIN EU® Data Network with available data on pregnancy, pregnancy start and end dates, pregnancy outcomes, and mother-child linkage.

The specific objectives were:

1a. To identify data partners in the DARWIN EU® Data Network with available structured data on pregnancy (pregnancy start and end date, pregnancy outcomes) mapped to the OMOP CDM or available at source but still to be mapped.

1b. To identify data partners without structured data on pregnancy but where pregnancy algorithms could be implemented to derive such information.

1c. To identify data partners in the DARWIN EU® Data Network with mother-child linkage.

1d. To obtain detailed information on availability, completeness and extensiveness of pregnancy-related data and mother-child linkage at source from each participating data partner.

8. RESEARCH METHODS

8.1. Study design

A descriptive study was conducted using routinely collected health data from 27 data sources from 15 countries across Europe.

Objective 1a, 1b and 1c

For objectives 1a, 1b, and 1c, a first selection of data partners was made through a feasibility assessment based on information provided by data partners as part of a short screening questionnaire on the self-reported availability of data on pregnancy start and end dates and mother-child linkage. This was supplemented with selected information available in the DARWIN EU® Portal (a database catalogue containing metadata of each of the DARWIN EU® data source, including aggregated counts for data already mapped to the OMOP CDM, which is used internally to inform feasibility assessments) on pregnancy start and end dates and select pregnancy outcomes.

Based on the information provided by data partners in this first questionnaire and data from the DARWIN EU® Portal, we selected data partners to participate in objective 1d for further in-depth assessment.

Objective 1d

Selected data partners were invited to participate in objective 1d based on the results from objectives 1a, 1b, and 1c. For objective 1d, additional information was obtained through a second, more in-depth questionnaire.

This second, in-depth questionnaire covered information about how pregnancies are registered and/or derived in terms of sources, start and end date, pregnancy related variables (such as gestational age, pregnancy-outcomes), mother-child linkage, strengths, and limitations in data sources.

The responses to the in-depth questionnaire were described narratively and through use of simple frequency counts and percentages. The output of objective 1 was used to inform objectives 2, 3, and 4 of the DARWIN EU® PeriNet project.

8.2. Study setting and data sources

The feasibility assessment was performed considering all data sources onboarded in DARWIN EU® as of time of feasibility request sent by EMA to the DARWIN EU® coordinating centre (i.e., 15th July 2024).

Neonatal data sources and disease specific registries onboarded in DARWIN EU® at the time of receiving the feasibility assessment were regarded as out of scope and were not considered for inclusion. These data sources involved: the National Neonatal Research Database (NNRD), and the Netherlands Cancer Registry (NCR).

This is the complete list of included data sources:

1. Belgium: IQVIA Longitudinal Patient Database Belgium (IQVIA LPD Belgium)

2. Croatia: Croatian National Public Health Information System (NAJS)
3. Denmark: Danish Data Health Registries (DK-DHR)
4. Estonia: Estonian Biobank (EBB)
5. Finland: Auria Clinical Informatics (FinOMOP-ACI Varha)
6. Finland: Hospital District of Helsinki and Uusimaa (FinOMOP-HUS)
7. Finland: Tampere University Hospital patient cohort (FinOMOP-TaUH Pirha)
8. Finland: Finnish Care Register for Health Care (FinOMOP-THL)
9. France: Clinical Data Warehouse of Bordeaux University Hospital (CDW Bordeaux)
10. France: Système National des Données de Santé (SNDS)
11. Germany: InGef Research Database (InGef RDB)
12. Germany: IQVIA Disease Analyzer Germany (IQVIA DA Germany)
13. Hungary: Semmelweis University Clinical Data (SUCD)
14. Italy: Research Repository @Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico (POLIMI)
15. Norway: Norwegian Linked Health Registry data (NLHR)
16. Norway: Norwegian Linked Health Registry data - Perinatal and Pregnancy (NLHR@UiO:PERINATAL)
17. Portugal: Egas Moniz Health Alliance database - Entre o Douro e Vouga (EMDB-ULSEDV)
18. Portugal: Egas Moniz Health Alliance database - Gaia E Espinho (EMDB-ULSGE)
19. Portugal: Egas Moniz Health Alliance database - Baixo Vouga (Região de Aveiro) (EMDB-ULSRA)
20. Portugal: Unidade Local de Saúde de Matosinhos Realtime Database (ULSM-RT)
21. Spain: Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público (BIFAP)
22. Spain: Institut Municipal Assistència Sanitària Information System (IMASIS)
23. Spain: The Information System for Research in Primary Care (SIDIAP)
24. Spain: Valencia Health System Integrated Dataset (VID)
25. Sweden: Health Impact - Swedish Population Evidence Enabling Data-linkage (HI-SPEED)
26. The Netherlands: Integrated Primary Care Information (IPCI)
27. The United Kingdom: Clinical Practice Research Datalink GOLD (CPRD GOLD)
28. The United Kingdom: UK BioBank (UKBB)

Information on data sources included in the feasibility assessment can be found in [Table 1](#) and [Annex I](#).

Table 1. Description of the selected data sources.

Country	Name of Data source	Health Care setting	Type of Data	Number of active individuals ¹	Calendar period covered by each data source
Belgium	IQVIA LPD Belgium	Primary care: GPs	Outpatient General Practitioner Care	189k	2005–2024
Croatia	NAJS	Primary care: GPs, secondary care: specialists (ambulatory or hospital outpatient care), hospital inpatient care	Registry	4.3M	2017–2024
Denmark	DK-DHR	Community pharmacists, secondary care: specialists (ambulatory or hospital outpatient care), hospital inpatient care	Registry	5.98M	1995–2024
Estonia	EBB	Primary care: GPs, community pharmacists primary care specialists (paediatricians)	Registry	212k	2004–2024
Finland	FinOMOP-ACI Varha	Secondary care: specialists (ambulatory or hospital inpatient care), hospital inpatient care	EHR	181k	2004–2024
Finland	FinOMOP-HUS	Secondary care specialists (ambulatory or hospital outpatient care), hospital inpatient care	EHR	616k	2004–2024
Finland	FinOMOP-TaUH Pirha	Secondary care specialists (ambulatory or hospital outpatient care), hospital inpatient care	EHR	263k	2007–2024
Finland	FinOMOP-THL	Primary care: GPs, primary care specialists (e.g., paediatricians), secondary care specialists (ambulatory or	Registry	5.7M	2011–2024

Country	Name of Data source	Health Care setting	Type of Data	Number of active individuals ¹	Calendar period covered by each data source
		hospital outpatient care), hospital inpatient care			
France	CDW Bordeaux	Secondary care specialists (ambulatory or hospital outpatient care), hospital inpatient care	EHR	247k	2005–2024
France	SNDS	Primary care: GPs, primary care specialists (e.g., paediatricians), secondary care specialists (ambulatory or hospital outpatient care), hospital inpatient care, claims data	Claims data	1.9k	2006–2024
Germany	InGef RDB	Primary care: GPs, primary care specialists (e.g., paediatricians), secondary care specialists (ambulatory or hospital outpatient care), hospital inpatient care, claims data	Registry	7.67M	2014–2024
Germany	IQVIA DA Germany	Primary care: GPs, primary care specialists (e.g., paediatricians)	Outpatient General Practitioner Care	4.48M	1992–2024
Hungary	SUCD	Hospital inpatient care, claims data	Registry	227k	2010–2024
Italy	POLIMI	Secondary care specialists (ambulatory or hospital outpatient care), hospital inpatient care,	EHR	104k	2006–2024
Norway	NLHR	Primary care: GPs, primary care specialists (e.g., paediatricians), secondary care specialists (ambulatory or hospital outpatient	Registry	5.55M	2008–2024

Country	Name of Data source	Health Care setting	Type of Data	Number of active individuals ¹	Calendar period covered by each data source
		care), hospital inpatient care, claims data			
Norway	NLHR@UiO:PERINATAL	Primary care: GPs, primary care specialists (e.g., paediatricians), hospital inpatient care	Registry	1.94M	1967–2024
Portugal	EMDB-ULSEDV	Primary care: GPs, secondary care specialists (ambulatory or hospital outpatient care), hospital inpatient care	EHR	101k	1999–2024
Portugal	EMDB-ULSGE	Primary care: GPs, secondary care specialists (ambulatory or hospital outpatient care), hospital inpatient care	EHR	130k	2004–2024
Portugal	EMDB-ULSRA	Primary care: GPs, secondary care specialists (ambulatory or hospital outpatient care), hospital inpatient care	EHR	89.5k	1997–2024
Portugal	ULSM-RT	Primary care: GPs, secondary care specialists (ambulatory or hospital outpatient care), hospital inpatient care	EHR	76k	1998–2024
Spain	BIFAP	Primary care: GPs, community pharmacists, primary care specialists (paediatricians), hospital inpatient care	Registry	23.2M	2001–2024
Spain	IMASIS	Secondary care specialists (ambulatory or hospital outpatient care), hospital inpatient care	EHR	103k	1990–2024
Spain	SIDIAP	Primary care: GPs, hospital inpatient care	EHR	5.95M	2006–2024

Country	Name of Data source	Health Care setting	Type of Data	Number of active individuals ¹	Calendar period covered by each data source
Spain	VID	Primary care: GPs, secondary care specialists (ambulatory or hospital outpatient care), hospital inpatient care	EHR	5.58M	2009–2024
Sweden	HI-SPEED	Primary care: GPs, secondary care specialists (ambulatory or hospital outpatient care), hospital inpatient care	Registry	10.6M	2015–2024
The Netherlands	IPCI	Primary care: GPs	Outpatient General Practitioner Care	1.33M	2006–2024
The United Kingdom	CPRD GOLD	Primary care	EHR	2.83M	1988–2024
The United Kingdom	UKBB	Primary care: GPs, primary care specialists (paediatricians), secondary care specialists (ambulatory or hospital outpatient care), hospital inpatient care	Registry	500k	2006–2024

¹Defined as the maximum number of individuals in observation in the last 6 months of data.

GP=General Practitioner, k=thousands, M=millions, EHR=Electronic Health Records

IQVIA Longitudinal Patient Database Belgium (IQVIA LPD Belgium), National Public Health Information System (NAJS), Danish Data Health Registries (DK-DHR), Estonian Biobank (EBB), Auria Clinical Informatics (FinOMOP-ACI Varha), Hospital District of Helsinki and Uusimaa (FinOMOP-HUS), Tampere University Hospital patient cohort (FinOMOP-TaUH Pirha), Finnish Care Register for Health Care (FinOMOP-THL), Clinical Data Warehouse of Bordeaux University Hospital (CDW Bordeaux), Système National des Données de Santé (SNDS), InGef Research Database (InGef RDB), IQVIA Disease Analyzer Germany (IQVIA DA Germany), Semmelweis University Clinical Data (SUCD), Norwegian Linked Health Registry data (NLHR), Norwegian Linked Health Registry data - Perinatal and Pregnancy (NLHR@UiO:PERINATAL), Egas Moniz Health Alliance database - Entre o Douro e Vouga (EMDB-ULSEDV), Egas Moniz Health Alliance database - Gaia E Espinho (EMDB-ULSGE), Egas Moniz Health Alliance database - Baixo Vouga (Região de Aveiro) (EMDB-ULSRA), Unidade Local de Saúde de Matosinhos Realtime Database (ULSM-RT), Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público (BIFAP), Institut Municipal Assistència Sanitària Information System (IMASIS), The Information System for Research in Primary Care (SIDAP), Valencia Health System Integrated Dataset (VID), Health Impact – Swedish Population Evidence Enabling Data-linkage (HI-SPEED), Integrated Primary Care Information (IPCI), Clinical Practice Research Datalink (CPRD) GOLD, UK BioBank (UKBB)

8.3. Statistical methods

8.3.1. Main summary measures

Two questionnaires were sent to data partners: a first screening questionnaire (objective 1a, 1b, and 1c) and a second, more in-depth questionnaire (objective 1d) which was sent to a subset of data sources based on feedback from the first questionnaire. Questionnaires 1 and 2 can be found in [Annex II](#) and were developed through consultation between the DARWIN EU® CC and EMA to define relevant questions and data elements necessary for pregnancy related real world data studies. These covered information related to pregnancy, infant, and mother-child linkage.

The results of both questionnaires were described narratively and through the use of simple frequency counts and percentages.

To complement findings from the questionnaires, we examined information available in the DARWIN EU® Portal (for data already mapped to the OMOP CDM) on selected pregnancy and infant outcomes. The DARWIN EU® Portal is a database catalogue containing metadata of each of the DARWIN EU® data sources and is used internally to inform feasibility assessments. We used ATLAS (version 2.14.1)[8], a web-based open-source analysis tool from OHDSI, to build the concept list with pregnancy and birth as keywords. A total of 630 pregnancy related codes and 540 infant related codes were searched for person and record counts in the data sources. Codes used to explore this information are detailed in [Annex III](#). Person and record counts indicating the presence of these codes in the entire data source were explored overall, and by sex and age groups (0–9, 10–19, 20–29, 30–39, 40–49, 50–59, 60–69, 70–79 years).

A minimum cell count of 5 was used when reporting results, with any smaller count reported as “<5” and zero counts as “0”. Counts for concepts explored in the DARWIN EU® Portal were reported as displayed in this tool, with counts being rounded to the nearest multiple of 100 (e.g., a count of 100 means that a concept occurs between 6 and 100 times; a count of 200 corresponds to 101–200, etc.).

8.4. Deviations from the protocol

Deviation number	Protocol version	Date	Section of study protocol	Deviation	Reason
1	V2.0	12/12/2025	8.2. Data sources	Data source HI-SPEED not included to the list of data sources mentioned in protocol	Added HI-SPEED in study later as it was onboarded in January 2025.

9. RESULTS

All results obtained as part of this study are presented in the [Annex IV](#), including results from the two complete questionnaires and counts explored in the DARWIN EU® Portal. To facilitate the reporting of results, only the most relevant findings are presented in this section.

9.1. Participants

A total of 28 data sources were considered for assessment of availability of pregnancy related information ([Table 1](#)), using questionnaire 1 ([Annex II](#)) and DARWIN EU® Portal person and record counts. Based on this information, we invited 15 data partners with access to 21 data sources from 13 countries to participate in the study and to reply to a second questionnaire to better understand the characteristics of the available data.

From those data partners invited to participate in the study, 14 data partners (including 16 data sources) from 12 countries accepted to participate in the study. Data sources that did not accept to participate in the study were POLIMI, FinOMOP-ACI Varha, FinOMOP-HUS, EMDB-ULSEDV, and EMDB-ULSRA.

The responses provided by the data sources to the questionnaires, along with the DARWIN EU® Portal person and record counts, are summarised below.

9.2. Main results

9.2.1. Results from questionnaire 1 (objective 1a, 1b, and 1c)

From the 27 data sources that replied to questionnaire 1, 17 reported having data on pregnancy start and end date available, one data source reported having this information partially available, and four reported that they could potentially have this information added to their source data through linkage to other data sources in their region/country. From these 17 data sources, six data sources reported that they had information on start and end of pregnancy already mapped to the OMOP CDM, whereas for 11 data sources, these data were reported as being available in source format only. Four data sources (NLHR, NLHR@UiO:PERINATAL, SIDIAP, and HI-SPEED) indicated that pregnancy data was mapped to the OMOP CDM using the PET.

Regarding mother-child linkage, 13 data sources reported to have this linkage available, 2 reported this linkage to be under development, and one data partner (FinOMOP) reported that this linkage could be potentially included in their data sources through linkage with external data.

The results from questionnaire 1 are shown in [Table 2](#), with comments from data partners in [Annex IV Table S3](#). The raw results as received by data partners can be found in [Annex IV Table S4](#).

Table 2. Details of data sources and availability of pregnancy related information (results from questionnaire 1).

Country	Data source	Pregnancy start/end date	Mother-Child linkage	Mapped to the OMOP CDM
Belgium	IQVIA LPD Belgium	No	No	No
Croatia	NAJS	Yes	Yes	No
Denmark	DK-DHR	Yes	Yes	No
Estonia	EBB	No	No	No
Finland	FinOMOP-ACI Varha	Potentially*	Potentially*	No
Finland	FinOMOP-HUS	Potentially*	Potentially*	No
Finland	FinOMOP-TaUH Pirha	Potentially*	Potentially*	No
Finland	FinOMOP-THL	Potentially*	Potentially*	No
France	CDW Bordeaux	Yes	No	Yes
France	SNDS	Yes	Yes	No
Germany	InGef RDB	Yes	No	No
Germany	IQVIA DA Germany	No	No	No
Hungary	SUCD	Yes	Yes	No
Norway	NLHR	Yes	Yes	Yes (PET)
Norway	NLHR@UiO:PERINATAL	Yes	Yes	Yes (PET)
Portugal	EMDB-ULSEDV	Yes	Yes	No

Country	Data source	Pregnancy start/end date	Mother-Child linkage	Mapped to the OMOP CDM
Portugal	EMDB-ULSGE	Yes	Yes	No
Portugal	EMDB-ULSRA	Yes	Yes	No
Portugal	ULSM-RT	Yes	Yes	No
Spain	BIFAP	Yes	Under development**	No
Spain	IMASIS	Yes	Under development**	No
Spain	SIDIAP	Yes	Yes	Yes (PET)
Spain	VID	Yes	Yes	Yes (PET not implemented)
Sweden	HI-SPEED	Yes	Yes	Yes (PET)
The Netherlands	IPCI	No	No	No
The United Kingdom	CPRD GOLD	No	No	No
The United Kingdom	UKBB	Partial***	No	No

PET=Perinatal Expansion Table

IQVIA Longitudinal Patient Database Belgium (IQVIA LPD Belgium), National Public Health Information System (NAJS), Danish Data Health Registries (DK-DHR), Estonian Biobank (EBB), Auria Clinical Informatics (FinOMOP-ACI Varha), Hospital District of Helsinki and Uusimaa (FinOMOP-HUS), Tampere University Hospital patient cohort (FinOMOP-TaUH Pirha), Finnish Care Register for Health Care (FinOMOP-THL), Clinical Data Warehouse of Bordeaux University Hospital (CDW Bordeaux), Système National des Données de Santé (SNDS), InGef Research Database (InGef RDB), IQVIA Disease Analyzer Germany (IQVIA DA Germany), Semmelweis University Clinical Data (SUCD), Norwegian Linked Health Registry data (NLHR), Norwegian Linked Health Registry data - Perinatal and Pregnancy (NLHR@UiO:PERINATAL), Egas Moniz Health Alliance database - Entre o Douro e Vouga (EMDB-ULSEDV), Egas Moniz Health Alliance database - Gaia E Espinho (EMDB-ULSGE), Egas Moniz Health Alliance database - Baixo Vouga (Região de Aveiro) (EMDB-ULSRA), Unidade Local de Saúde de Matosinhos Realtime Database (ULSM-RT), Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público (BIFAP), Institut Municipal Assistència Sanitària Information System (IMASIS), The Information System for Research in Primary Care (SIDIAP), Valencia Health System Integrated Dataset (VID), Health Impact – Swedish Population Evidence Enabling Data-linkage (HI-SPEED), Integrated Primary Care Information (IPCI), Clinical Practice Research Datalink (CPRD) GOLD, UK BioBank (UKBB)

*Potentially: Refers to data sources that could get this data from linkage to other data sources.

**Under development: Refers to data sources are working on creating a mother-child linkage.

***Partial: Refers to a data source with availability of pregnancy start date but pregnancy end date not available as it is considered sensitive.

9.2.2. Counts from the DARWIN EU® Portal (objective 1a, 1b, and 1c)

Information available in the DARWIN EU® Portal was used to assess availability, completeness, and extensiveness of pregnancy- and infant- related data already mapped to the OMOP CDM. Person and record counts of pregnancy- and infant-related concept IDs in each data source in all individuals, including stratified by age and sex, can be found in [Annex IV \(Tables S5 to Table S10\)](#).

To facilitate the visualisation of results of the 28 data sources assessed, we summed person counts of all searched pregnancy- and infant-related concept IDs in [Table 3](#) and added the number of concepts explored that had >5 counts in each data source. The sum of person counts should be interpreted with consideration of the rounding of numbers described in [Section 8.3.1](#). Additionally, it is important to note that, all person counts obtained for each data source were summed. However, a single person may have multiple codes related to pregnancy. Therefore, this number should not be interpreted as the actual count of distinct subjects.

Person counts for pregnancy related concepts ranged from 14,500 in IQVIA LPD Belgium to 4,858,400 in BIFAP, and person counts for infant related concepts ranged from 4,800 in IQVIA LPD Belgium to 6,070,200 in NLHR. Among concepts explored (n=630 pregnancy-related; n=540 infant-related), the proportion of concepts with >5 counts ranged from 13% to 27% in most data sources. Proportions were particularly high in BIFAP (42.5% pregnancy-related concepts; 32.6% infant-related concepts) and CPRD GOLD (33.8% and 48.9%, respectively). In some data sources, proportions were lower than 5%, including IQVIA DA Germany, IMASIS (2% of infant-related concepts), and IQVIA LPD Belgium (4.8% of infant-related concepts).

Table 3. Sum of person counts of all searched pregnancy and infant related concept IDs in each data source, obtained from the DARWIN EU® Portal.

Country	Data source	Number of cumulative persons ³	Pregnancy related concepts ¹		Infant-related concepts ²	
			Sum of Person Counts	Recorded concepts, N (%) ⁴	Sum of Person Counts	Recorded concepts, N (%) ³
Belgium	IQVIA LPD Belgium	1.06	14,500	30 (4.8%)	4,800	196 (19.6%)
Croatia	NAJS	4.79M	1,375,900	115 (18.3%)	455,100	121 (22.4%)
Denmark	DK-DHR	8.59M	2,927,700	138 (21.9%)	2,708,000	113 (20.9%)
Estonia	EBB	212k	272,700	112 (17.8%)	52,600	59 (10.9%)
Finland	FinOMOP-ACI Varha	850k	182,900	107 (17.0%)	128,400	117 (21.7%)
Finland	FinOMOP-HUS	2.92M	519,500	114 (18.1%)	827,400	127 (23.5%)
Finland	FinOMOP-TaUH Pirha	862k	184,400	99 (15.7%)	152,500	105 (19.4%)
Finland	FinOMOP – THL	6.62M	1,832,300	121 (19.2%)	1,121,000	145 (26.9%)
France	CDW Bordeaux	2.41M	200,600	92 (14.6%)	252,700	124 (23.0%)
France	SNDS	10.6M	96,100	100 (15.9%)	125,100	122 (22.6%)
Germany	InGef RDB	10.7M	702,700	103 (16.3%)	1,232,700	118 (21.9%)
Germany	IQVIA DA Germany	43.7M	1,945,300	101 (16.0%)	391,600	11 (2.0%)
Hungary	SUCD	2.49M	198,700	92 (14.6%)	446,500	103 (19.1%)
Italy	POLIMI	1.72M	225,900	83 (13.7%)	119,100	133 (24.6%)
Norway	NLHR	6.11M	3,300,700	116 (18.4%)	6,070,200	31 (5.7%)
Norway	NLHR@UIO:PERINATAL	1.94 M	3,248,100	110 (17.5%)	4,427,200	57 (10.6%)
Portugal	EMDB-ULSEDV	575k	44,700	87 (13.8%)	96,400	84 (15.6%)
Portugal	EMDB-ULSGE	728k	113,200	122 (19.4%)	68,200	92 (17.0%)
Portugal	EMDB-ULSRA	514k	80,100	119 (18.9%)	97,100	89 (16.5%)
Portugal	ULSM-RT	620k	75,400	114 (18.1%)	82,900	80 (14.8%)

Country	Data source	Number of cumulative persons ³	Pregnancy related concepts ¹		Infant-related concepts ²	
			Sum of Person Counts	Recorded concepts, N (%) ⁴	Sum of Person Counts	Recorded concepts, N (%) ³
Spain	BIFAP	30.3M	4,858,400	268 (42.5%)	1,550,600	176 (32.6%)
Spain	IMASIS	1.79M	261,800	141 (22.4%)	94,600	11 (2.0%)
Spain	SIDIAP	8.89M	2,862,200	133 (21.1%)	830,400	103 (19.1%)
Spain	VID	6.0M	2,887,700	202 (32.1%)	173,500	123 (22.8%)
Sweden	HI-SPEED	12.1M	1,888,900	106 (16.8%)	647,500	90 (16.7%)
The Netherlands	IPCI	3.06M	307,300	30 (4.8%)	12,300	104 (19.3%)
The United Kingdom	CPRD GOLD	17.5M	2,715,900	213 (33.8%)	2,519,200	264 (48.9%)
The United Kingdom	UKBB	502k	78,900	153 (24.3%)	43,500	132 (24.4%)

¹Results for 630 pregnancy-related codes searched for person counts in the data sources (see [Annex III](#)). All person counts obtained for each data source have been summed. A single person may have multiple codes related to pregnancy; therefore, this number should not be interpreted as the actual count of distinct subjects.

²Results for 540 infant-related codes searched for person counts in the data sources (see [Annex III](#)). All person counts obtained for each data source have been summed. A single person may have multiple codes related to the infant; therefore, this number should not be interpreted as the actual count of distinct subjects.

³Defined as total number of individuals ever observed in the data source.

⁴Number of concepts with >5 counts in the DARWIN EU® Portal, relative to the total number of codes searched.

k=thousands, M=millions, IQVIA Longitudinal Patient Database Belgium (IQVIA LPD Belgium), National Public Health Information System (NAJS)

Danish Data Health Registries (DK-DHR), Estonian Biobank (EBB), Auria Clinical Informatics (FinOMOP-ACI Varha), Hospital District of Helsinki and Uusimaa (FinOMOP-HUS), Tampere University Hospital patient cohort (FinOMOP-TaUH Pirha), Finnish Care Register for Health Care (FinOMOP – THL), Clinical Data Warehouse of Bordeaux University Hospital (CDW Bordeaux), Système National des Données de Santé (SNDS), InGef Research Database (InGef RDB), IQVIA Disease Analyzer Germany (IQVIA DA Germany), Semmelweis University Clinical Data (SUCD), Research Repository @Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico (POLIMI), Norwegian Linked Health Registry data (NLHR), Norwegian Linked Health Registry data - Perinatal and Pregnancy (NLHR@UiO:PERINATAL), Egas Moniz Health Alliance database - Entre o Douro e Vouga (EMDB-ULSEDV), Egas Moniz Health Alliance database - Gaia E Espinho (EMDB-ULSGE), Egas Moniz Health Alliance database Baixo Vouga (Região de Aveiro) (EMDB-ULSRA), Unidade Local de Saúde de Matosinhos Realtime Database (ULSM-RT), Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público (BIFAP), Institut Municipal Assistència Sanitària Information System (IMASIS), The Information System for Research in Primary Care (SIDIAP), Valencia Health System Integrated Dataset (VID), Health Impact – Swedish Population Evidence Enabling Data-linkage (HI-SPEED), Integrated Primary Care Information (IPCI), Clinical Practice Research Datalink (CPRD) GOLD, UK BioBank (UKBB)

Assessment of the data extensiveness was done by stratification (age and sex) of pregnancy related concept IDs counts in each data source (Table S6 and Table S7). Results stratified by age 0 to 9 years (Table 4) and >50 years (Table 5) showed counts for concepts related to pregnancies, such as *finding related to pregnancy* (concept ID: 444094), *complication of pregnancy, childbirth and/or puerperium* (concept ID: 435875), and *high-risk pregnancy* related codes (concept ID: 4188598, 45765728) with the highest person counts in these age groups. Counts were higher among younger age groups (0–9 years) than in older (>50 years), suggesting availability of pregnancy-related data in those children possibly due to information linked to their mothers.

Similarly, results stratified by sex (Table 6) showed that pregnancy-related concept codes recorded in the data sources in males, such as *finding related to pregnancy* (concept ID: 444094), *complication of pregnancy, childbirth and/or puerperium* (concept ID: 435875), and *Supervision of high-risk pregnancy* (concept ID: 45765728), were frequently recorded in males. This also showed that this information might be available in those individuals with mother-child linkage available, and therefore, also including male individuals.

Table 4. Pregnancy related concept IDs in selected data sources among individuals from 0 to 9 years of age, obtained from the DARWIN EU® Portal.

Country	Data source ¹	Concept name ²	Person counts
Croatia	NAJS	Supervision of high risk pregnancy	28,600
		Uterine scar from previous surgery affecting pregnancy	8,800
		Gestational diabetes mellitus	8,800
		Anaemia in mother complicating pregnancy, childbirth AND/OR puerperium	4,600
		Symptomatic disorders in pregnancy	4,100
		Pregnancy-induced hypertension	2,800
		Post-term pregnancy	2,300
		Twin pregnancy	2,200
		Diabetes mellitus during pregnancy, childbirth and the puerperium	1,500
		Supervision of pregnancy with history of infertility	1,400
		Finding related to pregnancy	1,200
Denmark	DK-DHR	Supervision of high risk pregnancy	16,800
Finland	FinOMOP-HUS	Complication of pregnancy, childbirth and/or puerperium	10,800
Finland	FinOMOP-THL	Complication of pregnancy, childbirth and/or puerperium	6,400
Germany	IQVIA DA Germany	Supervision of high risk pregnancy	2,300
Norway	NLHR@UiO:PERINATAL	Finding related to pregnancy	4,300
Portugal	EMDB-ULSEDV	Finding related to pregnancy	3,600
Portugal	EMDB-ULSGE	Third trimester pregnancy	6,500
		Finding related to pregnancy	3,600
Portugal	EMDB-ULSRA	Finding related to pregnancy	4,000
Portugal	ULSM-RT	Finding related to pregnancy	3,100
		Third trimester pregnancy	1,200

Country	Data source ¹	Concept name ²	Person counts
Spain	BIFAP	High risk pregnancy	8,700
		High risk pregnancy due to history of preterm labour	8,600
		Abdominal pain in pregnancy	5,800
		Twin pregnancy	2,400
		Symptomatic disorders in pregnancy	1,800
		Finding related to pregnancy	1,600
		Premature pregnancy delivered	1,000
Spain	IMASIS	Finding related to pregnancy	2,500
Spain	SIDIAP	Finding related to pregnancy	1,600
		Psychosocial problems related to unwanted pregnancy	1,200
Sweden	HI-SPEED	Complication of pregnancy, childbirth and/or puerperium	6,600
The Netherlands	IPCI	Gestational age	16,400
		Finding related to pregnancy	5,600
		Pregnancy	5,000

¹Data sources not shown in this table did not show any counts of searched concept IDs in this population subgroup.

²Concepts related to pregnancy with more than 1,000 person counts. All person counts obtained for each data source have been summed. A single person may have multiple codes related to pregnancy; therefore, this number should not be interpreted as the actual count of distinct subjects.

National Public Health Information System (NAJS), Danish Data Health Registries (DK-DHR), Hospital District of Helsinki and Uusimaa (FinOMOP-HUS), Finnish Care Register for Health Care (FinOMOP-THL), IQVIA Disease Analyzer Germany (IQVIA DA Germany), Norwegian Linked Health Registry data - Perinatal and Pregnancy (NLHR@UiO:PERINATAL), Egas Moniz Health Alliance database - Entre o Douro e Vouga (EMDB-ULSEDV), Egas Moniz Health Alliance database - Gaia E Espinho (EMDB-ULSGE), Egas Moniz Health Alliance database - Baixo Vouga (Região de Aveiro) (EMDB-ULSRA), Unidade Local de Saúde de Matosinhos Realtime Database (ULSM-RT), Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público (BIFAP), Institut Municipal Assistència Sanitària Information System (IMASIS), The Information System for Research in Primary Care (SIDIAP), Health Impact – Swedish Population Evidence Enabling Data-linkage (HI-SPEED), Integrated Primary Care Information (IPCI).

Table 5. Pregnancy related concept IDs in each data source recorded at any time among individuals older than 50 years of age, obtained from the DARWIN EU®.

Country	Data sources ¹	Concept name ²	Person counts
Belgium	IQVIA LPD Belgium	Diabetes mellitus during pregnancy, childbirth and the puerperium	2,800
		Finding related to pregnancy	2,800
		Pre-existing hypertension complicating pregnancy, childbirth and puerperium	2,000
		Ectopic pregnancy	1,300
		Infectious disease in mother complicating pregnancy, childbirth AND/OR puerperium	1,200
Croatia	NAJS	Aspiration curettage of uterus for termination of pregnancy	3,500
		Supervision of high risk pregnancy	1,100
		Choriogonadotropin (pregnancy test) [Presence] in Urine	1,100
		Pre-existing hypertension complicating pregnancy, childbirth and puerperium	1,000

Country	Data sources ¹	Concept name ²	Person counts
		Urine pregnancy test	1,000
Denmark	DK-DHR	Supervision of high risk pregnancy	5,100
Finland	FinOMOP-TaUH Pirha	Choriogonadotropin (pregnancy test) [Presence] in Serum or Plasma	5,300
Finland	FinOMOP-THL	Finding related to pregnancy	1,900
		Pregnancy	1,200
Germany	IQVIA DA Germany	Gestational diabetes mellitus	2,600
		Pregnancy with abortive outcome	1,700
		Supervision of high risk pregnancy	1,300
		Pregnancy	1,300
		Disorder of pregnancy	1,200
Norway	NLHR	Pre-existing type 2 diabetes mellitus in pregnancy	5,800
		Symptomatic disorders in pregnancy	5,300
		Supervision of high risk pregnancy	1,600
		Pre-existing diabetes mellitus in pregnancy	1,200
		Pregnancy not yet confirmed	1,000
Norway	NLHR@UIO:PERINA TAL	Finding related to pregnancy	1,400
Portugal	EMDB-ULSGE	Finding related to pregnancy	1,000
Portugal	EMDB-ULSRA	Finding related to pregnancy	2,100
Spain	BIFAP	High risk pregnancy	18,600
		High risk pregnancy due to history of preterm labour	18,000
		Symptomatic disorders in pregnancy	7,600
		Abdominal pain in pregnancy	4,700
		Pregnancy-related glycosuria	3,000
		Finding related to pregnancy	1,500
		Gestational diabetes mellitus	1,000
		Illegitimate pregnancy	1,000
		Wanted pregnancy	1,000
Spain	IMASIS	Choriogonadotropin.beta subunit (pregnancy test) [Presence] in Urine	2,800
Spain	SIDIAP	Finding of gestational age	3,100
		Psychosocial problems related to unwanted pregnancy	3,000
		Finding related to pregnancy	1,700
		Measure of pregnancy	1,000
Spain	VID	Finding related to pregnancy	2,900
Sweden	HI-SPEED	Sequela of complication of pregnancy childbirth and the puerperium	3,700
		Disorder of pregnancy	3,600

Country	Data sources ¹	Concept name ²	Person counts
		Supervision of high risk pregnancy	1,800
		Pre-existing hypertension complicating pregnancy, childbirth and puerperium	1,400
		Twin pregnancy	1,000
The Netherlands	IPCI	Gestational age	2,300
United Kingdom	CPRD GOLD	Last menstrual period -1st day	11,700
		Urine pregnancy test, by visual colour comparison methods	7,600
		Not pregnant	5,800
		Pregnancy advice for patients with epilepsy	2,900
		Pregnancy advice for patients with epilepsy not indicated	2,600
		Urine pregnancy test	1,300
United Kingdom	UKBB	Last menstrual period -1st day	3,500

¹Data sources not shown in this table did not show any counts of searched concept IDs in this population subgroup.

²Concepts related to pregnancy with more than 1,000 person counts. All person counts obtained for each data source have been summed. A single person may have multiple codes related to pregnancy; therefore, this number should not be interpreted as the actual count of distinct subjects.

IQVIA Longitudinal Patient Database Belgium (IQVIA LPD Belgium), National Public Health Information System (NAJS), Danish Data Health Registries (DK-DHR), Tampere University Hospital patient cohort (FinOMOP-TaUH Pirha), Finnish Care Register for Health Care (FinOMOP-THL), IQVIA Disease Analyzer Germany (IQVIA DA Germany), Norwegian Linked Health Registry data (NLHR), Norwegian Linked Health Registry data - Perinatal and Pregnancy (NLHR@UiO:PERINATAL), Egas Moniz Health Alliance database - Gaia E Espinho (EMDB-ULSGE), Egas Moniz Health Alliance database - Baixo Vouga (Região de Aveiro) (EMDB-ULSRA), Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público (BIFAP), Institut Municipal Assistència Sanitària Information System (IMASIS), The Information System for Research in Primary Care (SIDRAP), Valencia Health System Integrated Dataset (VID), Health Impact – Swedish Population Evidence Enabling Data-linkage (HI-SPEED), Integrated Primary Care Information (IPCI), Clinical Practice Research Datalink (CPRD) GOLD, UK BioBank (UKBB)

Table 6. Pregnancy related concept IDs in each data source recorded at any age in males, obtained from the DARWIN EU® Portal.

Country	Data sources ¹	Pregnancy related mapped data ²	Person counts
Croatia	NAJS	Supervision of high risk pregnancy	13,800
		Gestational diabetes mellitus	4,600
		Uterine scar from previous surgery affecting pregnancy	4,500
		Anaemia in mother complicating pregnancy, childbirth AND/OR puerperium	2,300
		Symptomatic disorders in pregnancy	2,100
		Pregnancy-induced hypertension	1,500
		Post-term pregnancy	1,100
Denmark	DK-DHR	Supervision of high risk pregnancy	1,300
Finland	FinOMOP-HUS	Complication of pregnancy, childbirth and/or puerperium	2,000

Country	Data sources ¹	Pregnancy related mapped data ²	Person counts
Finland	FinOMOP-TaUH Pirha	Choriogonadotropin (pregnancy test) [Presence] in Serum or Plasma	2,300
		Pregnancy	8,900
		Finding related to pregnancy	2,700
Finland	FinOMOP-THL	Complication of pregnancy, childbirth and/or puerperium	2,700
Germany	IQVIA DA Germany	Supervision of high risk pregnancy	1,400
Norway	NLHR@UiO:PERINATAL	Finding related to pregnancy	6,100
Portugal	EMDB-ULSEDV	Finding related to pregnancy	1,800
Portugal	EMDB-ULSGE	Third trimester pregnancy	3,400
		Finding related to pregnancy	2,100
Portugal	EMDB-ULSRA	Finding related to pregnancy	1,900
Spain	BIFAP	High risk pregnancy	10,000
		High risk pregnancy due to history of preterm labour	9,900
		Abdominal pain in pregnancy	7,600
		Symptomatic disorders in pregnancy	7,500
		Pregnancy-related glycosuria	2,400
		Finding related to pregnancy	1,700
		Twin pregnancy	1,200
		Concern about body image related to pregnancy	1,100
		Wanted pregnancy	1,000
		Illegitimate pregnancy	1,000
Spain	SIDIAP	Psychosocial problems related to unwanted pregnancy	1,400
		Finding related to pregnancy	1,500
Sweden	HI-SPEED	Complication of pregnancy, childbirth and/or puerperium	2,200
The Netherlands	IPCI	Gestational age	10,800
United Kingdom	CPRD GOLD	Pre-pregnancy education	1,100

¹Data sources not shown in this table did not show any counts of searched concept IDs in this population subgroup.

²Concepts related to pregnancy with more than 1,000 person counts. All person counts obtained for each data source have been summed. A single person may have multiple codes related to pregnancy; therefore, this number should not be interpreted as the actual count of distinct subjects.

National Public Health Information System (NAJS), Danish Data Health Registries (DK-DHR), Tampere University Hospital patient cohort (FinOMOP-TaUH Pirha), Finnish Care Register for Health Care (FinOMOP-THL), IQVIA Disease Analyzer Germany (IQVIA DA Germany), Norwegian Linked Health Registry data - Perinatal and Pregnancy (NLHR@UiO:PERINATAL), Egas Moniz Health Alliance database - Gaia E Espinho (EMDB-ULSGE), Egas Moniz Health Alliance database - Baixo Vouga (Região de Aveiro) (EMDB-ULSRA), Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público (BIFAP), Institut Municipal Assistència Sanitària Information System (IMASIS), The Information System for Research in Primary Care (SIDIAP), Health Impact – Swedish Population Evidence Enabling Data-linkage (HI-SPEED), Integrated Primary Care Information (IPCI), Clinical Practice Research Datalink (CPRD) GOLD

9.2.3. Results from questionnaire 2 (objective 1d)

A total of 15 data sources completed the second questionnaire ([Annex II](#)), covering the availability, derivation, granularity, sources of data recording, and mapping status of key pregnancy variables. Results of this questionnaire are shown in [Annex IV Table S11](#).

Fourteen of the fifteen data sources reported that pregnancies can be identified in their data sources, including administrative health records, medical birth registries, hospital discharge data, and primary care records. Overall, 10 out of 15 data partners reported that a pregnancy start date is available in their data source (BIFAP, DK-DHR, EMDB, FinOMOP-THL, HI-SPEED, IMASIS, NAJS, NLHR, SIDIAP, SUCD). Among these, six (BIFAP, DK-DHR, FinOMOP-THL, IMASIS, NAJS, SUCD) indicated that the pregnancy start date is available in the source data but not yet mapped to the OMOP CDM. Five data partners reported using an algorithm to derive the pregnancy start date (HI-SPEED, DK-DHR, CPRD GOLD, BIFAP, and IPCI). In these cases, the pregnancy start date is not directly recorded in the source data but is inferred using combinations of available recorded pregnancy-related information, such as gestational age, pregnancy end date, or registry-based rules. Some data sources provided references (see [Annex IV Table S11](#)) or documentation describing the applied algorithms, while others reported locally developed approaches.

The date of last menstrual period was reported to be available in 12 out of 15 data sources, although 9 indicated that it has not yet been mapped to the OMOP CDM. Gestational age was also available in 11 data sources (BIFAP, CPRD GOLD, DK-DHR, EMDB, InGef RDB, IPCI, FinOMOP-THL, HI-SPEED, IMASIS, NAJS, NLHR), this information is typically derived from clinical records or registries and may be recorded at delivery or during antenatal care. The methods used to determine gestational age varied across partners and included ultrasound-based estimates (6 data sources), registry-derived values (1 data source), or clinical records (5 data sources).

Most data sources reported having information on pregnancy outcomes, such as live birth (15 data sources), stillbirth, miscarriage, and termination of pregnancy (13 data sources), as well as other pregnancy-related variables, such as mode of delivery (13 data sources), pregnancy complications, such as maternal death (11 data sources), or antenatal risk factors, such as alcohol consumption (10 data sources).

Eight data sources reported data availability on breastfeeding (SUCD, FinOMOP-TaUH, FinOMOP-THL, NAJS, SIDIAP, CPRD GOLD, IPCI, EMDB), which was already mapped to the OMOP CDM in some cases (SIDIAP, CPRD GOLD, IPCI, EMDB).

Five data sources reported availability of father-child linkage in their data (HI-SPEED, DK-DHR, NAJS, FinOMOP-THL, NLHR). From these, only HI-SPEED and NLHR reported to have already mapped this information in their OMOP CDM.

10. DISCUSSION

We identified data partners from the DARWIN EU® Data Network that reported containing data on pregnancy and mother-child linkage through two questionnaires that were shared with data partners, together with the information already mapped to the OMOP CDM available in the DARWIN EU® Portal.

Three data partners reported to have information on pregnancy start and end date, as well as mother-child linkage, already mapped to the OMOP CDM (HI-SPEED, NLHR, and SIDIAP). Ten data partners reported to have pregnancy start and end date as well as mother-child linkage available in their source but only partially or not yet mapped to the OMOP CDM (NAJS, DK-DHR, FinOMOP (potentially able to include mother-child linkage), SNDS, SUCD, EMDB, ULSM, IMASIS (mother-child linkage under development), BIFAP (mother-child linkage under development), and VID).

Additional information collected in a second questionnaire showed that several data partners have pregnancy-related data, including on pregnancy outcomes, pregnancy risk factors (such as alcohol consumption), breastfeeding, or father-child linkage. While core events, such as delivery and pregnancy

outcomes, were more frequently mapped to the OMOP CDM, more granular information reflecting gestational age or other time-related aspects of the pregnancy (e.g., last menstrual period, derived pregnancy start date) were less consistently mapped across data sources.

Results from the DARWIN EU® Portal indicated heterogeneity in the extensiveness of the pregnancy- and infant-related information available in the data network. Further stratification of counts by age and sex revealed potentially implausible records. However, pregnancy-related data recorded in young or male individuals may reflect data originating from mother-child linkage pairs; while information observed in older women may represent historical pregnancy records, as age in the DARWIN EU® Portal is defined based on most recent data cut-off.

In summary, results from the two questionnaires, together with counts available in the DARWIN EU® Portal, highlighted that most data sources have access to pregnancy-related information, often from high-quality registries or linked healthcare data. Nevertheless, there is considerable heterogeneity in data granularity, derivation methods, and OMOP mapping status. Use of algorithm-derived pregnancy timing variables is common, underscoring the importance of clear documentation and harmonised approaches for multi-database pregnancy research.

Finally, this was the list of data partners included in DARWIN EU®-PeriNet: HI-SPEED, NLHR, SIDIAP, NAJS, DK-DHR, FinOMOP, SUCD, EMDB, IMASIS, BIFAP, CDW Bordeaux, InGef RDB, IPCI, and CPRD GOLD.

10.1. Strengths and limitations of the research methods

All data sources onboarded to the DARWIN EU® data network participated in the feasibility assessment at the time that the feasibility request was received from EMA (15th July 2024). The data sources onboarded at a later stage were not included in the current study but could be considered to be part of PeriNet in the future. Some data partners were in consideration to be included in the study but were ultimately not included due to different reasons: for SNDS, data were not readily available for assessment at the cut-off date; VID only had information for females at the time; ULSM was excluded given that EMDB already had access to four data sources in Portugal; and POLIMI, FinOMOP-ACI Varha, FinOMOP-HUS, EMDB-ULSEDV, and EMDB-ULSRA chose not to participate in the study this time. This does not mean that these data sources may not be considered to have pregnancy-related data potentially suitable for future studies related to pregnancy.

Availability, completeness, and extensiveness of data on pregnancy related information were examined using two questionnaires and assessment of detailed person counts of existing standardised OMOP CDM codes in DARWIN EU® Portal. This allowed for capturing local knowledge of pregnancy related data, identify discrepancies or gaps and providing reliable characterisation of data readiness across participating data partners for future pregnancy studies.

Some of the pregnancy-related information presented in this report was collected using questionnaires self-reported by the data partners, and participation was voluntary. Therefore, some variability in responses may have reflected differences in question interpretation and in respondents' familiarity with the OMOP CDM and local data structures. Additionally, the questionnaires captured information at a single point in time and may not reflect future improvements. As an example, at the time of submitting this report, IMASIS has implemented the PET, which was not available at the time the questionnaires were shared. However, the purpose of this assessment was to understand current data gaps and availability with the aim of implementing improvements in future, and omissions may have been minimised by use of complementary data from the DARWIN EU® Portal.

The DARWIN EU® Portal contains the occurrence of records for certain pregnancy- and infant-related concepts populated each time a data partner (or a new data release) is onboarded. These counts are aggregated and only reflect the number of times a certain code appears in the data source (or in how many subjects) and are rounded, and therefore, it does not allow for further interpretation. In addition, and for

data privacy reasons, counts in the Portal are rounded to the nearest 100, which should be considered when interpreting results.

The observed counts for pregnancy specific codes among the very young (0 to 9 years) as well as among males, are likely to be attributed to records related to the pregnancies that resulted in their birth. Most of these codes correspond to findings related to pregnancy, or codes related to high-risk or pregnancy complications. While we expect these concepts to be recorded in the maternal records, it is plausible that these are also recorded in infant records to monitor specific risks or conditions identified during pregnancy. A further assessment through mother-child linkage (where available) could provide insights into this. In addition, we also found pregnancy-specific concepts among individuals aged >50 years, many of which were relative to high-risk pregnancies. However, age was defined at time of conducting the feasibility assessment in the DARWIN EU® Portal, meaning that pregnancy-related information could have been recorded earlier in individuals' life. Finally, while some pregnancies might have occurred after age >50, the extent to which these counts represent actual pregnancy episodes warrants further investigation.

11. CONCLUSIONS

We identified data partners in the DARWIN EU® Data Network with available data on pregnancy, pregnancy start and end dates, pregnancy outcomes, and mother-child linkage. Ten data partners reported to have pregnancy start and end date recorded, although six had not yet mapped this information to the OMOP CDM. We identified eight data partners with existing or potentially available data on mother-child linkage. The availability and granularity of pregnancy-related data were characterised using questionnaire responses supplemented by detailed person counts in the DARWIN EU® Portal. There was heterogeneity in the level of pregnancy information granularity captured by these data sources due to different derivation methods, including algorithms. Insight into the level of detail and OMOP CDM mapping status regarding pregnancy related information in these data sources emphasises the need for further harmonisation and mapping efforts to support future cross-database pregnancy research, which will be considered further in Objective 3. Further efforts are still needed to better understand and evaluate the quality of these data partners considering dimensions such as extensiveness, reliability, and relevance, which will be considered in Objectives 2 and 4.

12. REFERENCES

1. EMA. Guideline on Risk Management Systems for Medicinal Products for Human Use (Rev. 2). EMA/838713/2011. European Medicines Agency; 2012.
2. FDA. Postapproval Pregnancy Safety Studies Guidance for Industry. U.S. Food and Drug Administration; 2019.
3. EMA. Data Quality Framework for DARWIN EU® and Real-World Evidence Studies. EMA/158264/2023. European Medicines Agency; 2023.
4. Estiri H, et al. Perinatal Expansion Table (PET): Structuring Pregnancy and Infant Data in the OMOP Common Data Model. JAMIA. 2023;30(7):1228–1238.
5. Matcho A, et al. Inferring Pregnancy Episodes and Outcomes within a Network of Observational Databases. PLoS One. 2018;13(2):e0192033.
6. Li Q, Andrade SE, Cooper WO, et al. Algorithms for Identifying Pregnancy Episodes in Claims Data. Pharmacoepidemiol Drug Saf. 2013;22(8):890–897.
7. Blacketer C, et al. Developing and Validating OMOP-CDM Pregnancy and Mother–Infant Linkage Algorithms in US Claims Databases. Drug Saf. 2022;45(9):961–974.
8. The Book of OHDSI: OHDSI; 2021.

13. ANNEXES

ANNEX I. Description of data sources

IQVIA Longitudinal Patient Database Belgium (IQVIA LPD Belgium)

Belgium Longitudinal patient data (LPD) is collected from GP prescribing systems and contains patient records on all signs and symptoms, diagnoses, and prescribed medications. The information recorded allows patients and doctors to be monitored longitudinally. Data are recorded directly in the LPD from doctors' surgeries in real-time during patient consultations via a practice management software system. It is used in studies to provide various market insights, such as treatment trends, patient pathway analysis, and treatment compliance. The panel of contributing physicians (a stable 300 GPs) is maintained as a representative sample of the primary care physician population in Belgium, according to three criteria known to influence prescribing: age, sex, and geographical distribution. Currently, the database is covering 1.1 M cumulative patients and covers from 2012 through to the present. The panel consists of a stable 300 GPs that are geographically well spread. The total number of active GPs in Belgium is 15,602. The regional geographical spread of physicians in the LPD data is also representative of the distribution across the country: 57% GPs in the North (compared to 54% nationally), 31% in the South (33% nationally), and 12% in Brussels (13%). The provider of the data has more than 2,250 GPs under contract, so in case of a drop out, a replacement is easily found. Drugs obtained over the counter by the patient outside the prescription system are not reported. No explicit registration or approval is necessary for DARWIN EU[®] studies.

Croatian National Public Health Information System (NAJS)

The National Public Health Information System (Nacionalni javnozdravstveni informacijski sustav - NAJS) is an organised system of information services by Croatian Institute of Public Health (CIPH). NAJS enables the collection, processing, recording, management, and storage of health-related data from healthcare providers, as well as the production and management of health information. NAJS contains medical and public health data collected and stored in health registries and other health data collections, including cancer registry, mortality, work injuries, occupational diseases, communicable and non-communicable diseases, health events, disabilities, psychosis and suicide, diabetes, drug abuse, and others. NAJS obtained an umbrella approval for DARWIN EU[®] studies.

Danish Data Health Registries (DK-DHR)

Danish health data is collected, stored, and managed in national health registers at the Danish Health Data Authority, covering the entire population. This enables the study of disease development and treatment over time. There are no gaps in terms of gender, age, or geography in Danish health data, as mandatory reporting is conducted on all patients from birth to death in all hospitals and medical clinics. Personal identification numbers enable the linking of data across registers, allowing access to data on all Danes throughout their lives, regardless of whether they have moved around the country. High data quality, due to standardisation, digitisation, and documentation, means that Danish health data is not based on interpretation. The Danish Health Data Authority is responsible for maintaining and developing national health registers, as well as for establishing and maintaining standards and classifications within the Danish healthcare system. Legislation ensures balance between personal data protection and use. In the present data base, we have access to the following registries for the entire Danish population of 5.9 million persons from 1/1/1995: The central Person Registry (CPR), The National Patient Registry (LPR), The Register of Pharmaceutical Sales (LSR), The National Cancer Register (CAR), The Cause of Death registry (DAR), The Clinical Laboratory Information Register (LAB), COVID-19 test and vaccination Registries (SSI-OVD, SSI-DDV), The complete Vaccination registry (DDV_all). All data registered from January 1, 1995, is included.

DK-DHR obtained an umbrella approval for DARWIN EU[®] studies.

Estonian Biobank (EBB)

The Estonian Biobank (EBB) is a population-based biobank of the Estonian Genome Center at the University of Tartu (EGCUT). Its cohort size is currently close to 200,000 participants ("gene donors" ≥ 18 years of age), which closely reflects the age, sex, and geographical distribution of the Estonian population. Estonians represent 83%, Russians 14%, and other nationalities 3% of all participants. Genomic GWAS analyses have been performed on all gene donors. The database also covers health insurance claims, digital prescriptions, discharge reports, information about incident cancer cases, and causes of death from national sources for each donor.

Auria Clinical Informatics (FinOMOP-ACI Varha)

The data covers the patient register at the Hospital District of Southwest Finland (HDSF), containing Turku University Hospital, which is one of the five university hospitals in Finland. It covers the public specialist health care and most emergency health care in the area of Southwest Finland (Varsinais-Suomi) for all demographic groups. The data is utilized for scientific research from the data lake in the HDSF under the Finnish legislation (The Act on Secondary Use of Health and Social Data). The most relevant data domains are patients, visits, inpatient episodes, diagnoses, laboratory results, procedures, medication, pathology, radiology, radiotherapy, chemotherapy, obstetrics, and narrative patient reports, however there are also other data domains available.

Hospital District of Helsinki and Uusimaa (FinOMOP-HUS)

The HUS data lake is a comprehensive, integrated data source derived in real-time from all patients who visit the HUS hospitals and receive treatment. HUS is responsible for specialized healthcare in Finland's Uusimaa region and treatment of many rare and severe diseases that are nationally centralized to HUS. HUS' catchment area covers a population of about 2.2 million people: 2.4 million outpatient clinic visits, 670,000 patients treated, 84,000 surgical procedures, 710 specialist care emergency clinic visits per day. All visits, examinations, laboratory test, procedures, and treatments are recoded in the HUS IT systems and integrated into the data lake. The data lake stores decades of clinical information in digital format, and data from both past and current source systems are available. Systems providing data into the data lake include: CGI Uranus and Epic Apotti (EHR: visits, diagnoses, medication, etc.), Opera (procedure records), Kemokur and Beacon (cancer-specific medications), Marela (hospital pharmacy), Multilab/Mylab+ (laboratory system), Qpati/Mylab+ (pathology records system). HUS Acamedic is a dedicated, accredited, secure analytics environment for research utilizing these data.

Tampere University Hospital patient cohort (FinOMOP-TaUH Pirha)

TaUH Research Database (Finland) includes all specialties/all patient groups treated in the Tampere University Hospital, secondary and tertiary care given in the region including given clinical and pathology diagnoses, diagnostic and therapeutic procedures, laboratory findings, radiology and pathology reports, medication given in the hospital and electronic prescriptions and continuous medical records (free text) including discharge letters since 2007.

Finnish Care Register for Health Care (FinOMOP-THL)

The THL database covers both public and private, primary, and specialised inpatient and outpatient health care encounters in Finland, starting from 2011. The entire public sector and private inpatient encounters have been included since 2011, while private outpatient encounters, including occupational care, are included since 2020. The main content of the THL CDM is The Finnish Care Register for Health Care (fi:Hoitoilmoitusrekisteri, HILMO). It is a continuation of the former Hospital Discharge Register, which originally gathered data on patients discharged from hospitals. The Care Register has comprehensive data on the use of services and service users from Finnish public inpatient and outpatient primary and specialised care nationwide. Since 1998, the register has covered both public outpatient and inpatient specialised care and private inpatient care (TerveysHilmo). Since 2011, the register has covered public

primary care (AvoHilmo). Since 2020, the register has covered private outpatient care and occupational care. In addition, the CDM also contains the vaccination data from the Finnish National Vaccination Register, and positive COVID-19 test results from the Finnish National Infectious Diseases Register, which is maintained by THL. The CDM is currently produced from the above-mentioned and limited to observation periods commencing after 1/1/2011. The National Population registry is also used as a source for the CDM database. The National Population registry data forms the basis for forming the patient population. This ensures an up-to-date location (municipality of residence) of patients, as well as complete death occurrences (although not the cause of death). Using the complete population as a basis for the person table also serves to facilitate calculations on a population level, e.g., incidence rates. The current CDM population comprises all persons having been alive and residing in Finland since the beginning of 2011.

Clinical Data Warehouse of Bordeaux University Hospital (CDW Bordeaux)

The clinical data warehouse at Bordeaux University Hospital, in France, comprises electronic health records for over 2 million patients, with data collection beginning in 2005. The hospital complex comprises three main sites and has a total of 3,041 beds (as of 2021). The database currently holds information about the person (demographics), visits (inpatient and outpatient), conditions and procedures (billing codes), drugs (outpatient prescriptions and inpatient orders and administrations), measurements (laboratory tests and vital signs), and dates of death (in-hospital and out-of-hospital deaths).

Système National des Données de Santé (SNDS)

The SNDS comprehensively collects individual characteristics, outpatient healthcare prescriptions, and procedures reimbursed. It includes information on the presence of long-term chronic diseases (LTD) (Affection de Longue Durée, ALD) eligible for 100% reimbursement of healthcare expenditures, when requested by a patient's general practitioner (GP). All of this information is linked to data concerning public and private hospital stays. Hospital stay diagnoses and LTD diagnoses are coded according to the International Classification of Diseases 10th revision (ICD 10). Causes of death (ICD 10) are collected by the Epidemiological Center for the Medical Causes of Deaths (Inserm-CépiDc) and itemized in the SNDS.

InGef Research Database (InGef RDB)

The InGef database comprises anonymised longitudinal claims data of about 10 million individuals across more than 50 statutory health insurance providers (SHIs) throughout Germany. Data are longitudinally linked over a period of currently ten years. Patients can be tracked across healthcare sectors. All patient-level and provider-level data in the InGef research database are anonymised to comply with German data protection regulations and German federal law. German SHI claims data available in the InGef database includes information on demographics (year of birth, gender, death date if applicable, region of residence on administrative district level); hospitalizations; outpatient services (diagnoses, treatments; specialities of physicians); dispensing of drugs; dispensing of remedies and aids; and sick leave and sickness allowance times. Additionally, costs or cost estimates from SHI's perspective are available for all key cost elements. All diagnoses in Germany are coded using the International Classification of Diseases, version 10 in the German Modification (ICD-10-GM). The persistence (membership over time) is relatively high in the InGef database: During a 5-year period (2009-2013), 70.6% of insurance members survived and remained insured with the same SHI without any gaps in their observational time. Persons leaving one of the participating SHIs and entering another participating SHI can be linked during yearly database consistency updates and are thus not lost over time. The InGef database is dynamic in nature, i.e., claims data are updated in an ongoing process and new SHIs may join or leave the database. By law, only the last 10 years of data are allowed to be used. At every new release, this window shifts, dropping older data and adding new data.

InGef RDB obtained an umbrella approval for DARWIN EU[®] studies.

IQVIA Disease Analyzer Germany (IQVIA DA Germany)

Germany DA is collected from extracts of patient management software used by GPs and specialists practicing in ambulatory care settings. Data coverage includes 39.6 M cumulative persons. Patients visiting more than one provider are not cross identified for data protection reasons and are therefore recorded as separate in the system. Dates of service range from 1992 through to the present. Observation time is defined by the first and last consultation dates. Germany has no mandatory GP system and patients have free choice of specialist. Drugs are recorded as prescriptions of marketed products. No explicit registration or approval is necessary for DARWIN EU[®] studies.

Semmelweis University Clinical Data (SUCD)

Semmelweis University is the largest provider of health care services in Hungary. Most of the departments cater to the most serious cases and patients requiring complex treatment, thus making the university a national healthcare provider. The overwhelming majority of patient data originates from Hungary, primarily from the central region of the country, including Budapest and Pest County. The database contains approximately 2 million individual patients across all care settings of the University since 2011. The hospital information system (MedSolution) is an integrated IT system that provides functional support for inpatient and outpatient care processes, serving as a unified platform for various diagnostic areas. In certain areas, it supports the registration of medications. It supports all kinds of hospital work processes from admission to discharge. The outpatient module serves as a platform for registering activities related to care episodes within outpatient specialist care. During care provision, data related to the patient's health state, diagnosis, documentation of requested examinations and medical consultations, prescribed medication, final reports, and performed interventions are recorded. The functions of the inpatient module support care provision within inpatient settings. It documents the patient's health state at admission and during the hospital stay, along with the anamnesis, diagnosis, performed examinations and interventions, hospital final reports, and provided medication in certain areas of care provision, such as chemotherapy. Among other modules, the diagnostic module registers the requested laboratory and imaging examinations and records the laboratory results.

Integrated Primary Care Information (IPCI)

The database collects data in the Netherlands. It was started in 1992 by the Department of Medical Informatics of the Erasmus University Medical Center in Rotterdam, the Netherlands. The current database contains patient records from 2006 onwards, when the database's size began to increase significantly. IPCI is a nationwide Dutch database. However, it mainly covers the central part of the country, including the most densely populated and non-urban areas. The IPCI database contains data from records of general practitioners' (GPs) practices. It contains information on all patients registered with GPs responsible for non-emergency care and referrals. More than 99% of the Dutch population has health insurance, and almost all citizens are registered with a general practitioner. Over 12 months, around 78% of the population has at least one contact with their GP. IPCI included around 350 GP practices out of approximately 5,000 in the country (~7%). The demographic composition of the IPCI population mirrors that of the general Dutch population in terms of age and sex.

IPCI obtains data from computer-based patient records. Patient-level data includes demographic information, complaints and symptoms, diagnoses, laboratory test results, lifestyle factors, and correspondence with secondary care, such as referral and discharge letters. Dutch GPs use the International Classification of Primary Care (ICPC-1) coding for complaints, symptoms, and diagnoses, an international standard developed and updated by the World Organisation of Family Doctors (WONCA) International Classification Committee.

Extensive quality control steps are performed before each data release. These include comparing patient characteristics between practices and checking for abnormal temporal data patterns in practices.

Additional checks include over 200 indicators related to population characteristics (e.g., reliability of birth and mortality rates) and medical data (e.g., availability of prescription durations, completeness of laboratory results, presence of hospital letters and prescriptions, proportion of patients with blood pressure measurements, etc.).

IPCI is not linked with other databases. Vital status (death date and cause) is collected based on GP records. The main limitation is that IPCI is limited to GP records, and although it contains information on referrals and discharge letters, it may not capture specific hospital information. The database profile was described by de Ritter et al. (2022). The database description is also available at ipci.nl.

Norwegian Linked Health Registry data (NLHR)

Norway has a universal public healthcare system consisting of primary and specialist healthcare services, covering a population of approximately 5.4 million inhabitants. Many population-based health registries were established in the 1960s with use of unique personal identifiers facilitating linkage between registries. Data in these health registries are used for health analysis, health statistics, improving the quality of healthcare, research, administration, and emergency preparedness. Data has been harmonized across the following registries: the Medical Birth Registry of Norway (MBRN), the Norwegian Prescription Registry (NorPD), the Norwegian Patient Registry (NPR), Norway Control and Payment of Health Reimbursement (KUHR), the Norwegian Surveillance System for Communicable Diseases (MSIS), the Norwegian Immunisation Registry (SYSVAK), the National Death Registry, and the National Registry (NR). The linkage between the registries was facilitated using project-specific person IDs generated from unique personal identification numbers assigned at birth or immigration for all legal residents in Norway. In brief: MBRN stores information about the pregnancy, the mother, father, and child. NPR records diagnosis in secondary care (e.g., hospital). KUHR contains information about diagnosis and contact in primary care (e.g., GPs and outpatient specialists). NorPD recorded all medications dispensed outside of hospitals. MSIS collects test results of communicable diseases (e.g., SARS-CoV-2). SYSVAK recorded vaccinations. Notifications of pregnancies and pregnancy outcomes are mandatory in Norway > gestational week 12. However, the UiO Pregnancy algorithm improves the detection and dating of early non-live pregnancy outcomes that would have gone unnoticed if relying solely on the medical birth registry information by linking MBRN records with hospital and primary care records. (Nordeng et al., 2024) This data source has 2 CDMs available (NLHR@UiO CDM and NLHR@UiO:PERINATAL). For this study, the NLHR@UiO CDM will be used.

Norwegian Linked Health Registry data - Perinatal and Pregnancy (NLHR@UiO:PERINATAL)

Norway has a universal public health care system, consisting of primary health care services and specialist health care services. Many population-based health registries were established in the 1960s, with use of unique personal identifiers facilitating linkage between registries. The mandatory national health registries were established to maintain national functions. They are used for health analysis, health statistics, improving the quality of healthcare, research, reimbursement purposes, administration, and emergency preparedness. The medical birth registry of Norway (MBRN) is a population-based registry containing information on all births in Norway since 1967 (approximately 55,000 births per year). It is based on mandatory notification of all live and non-live births and abortions occurring from 12 weeks of gestation onwards, including miscarriages, terminations of pregnancy, and stillbirths. MBRN includes demographic information of the mother and father, the mother's health before and during pregnancy, complications during pregnancy and delivery, and length of pregnancy, as well as information about the infant, including congenital malformations and other perinatal outcomes. The Norwegian health care system is founded on the principles of universal access, decentralization, and free choice of provider. All pregnant women are entitled to free maternity follow-ups with their midwife, general practitioner, and/or obstetrician. Nine pregnancy consultations, including fetal diagnostics and ultrasounds, are offered, with the first consultation being at around weeks 6-12 of pregnancy in primary care. Women may also be offered more consultations, if needed, due to high-risk pregnancies, disease, or complications in pregnancy. Ultrasound examinations

are performed in specialist outpatient care or hospitals, and elective terminations are performed in hospitals (secondary care). We harmonized data from the following registries: the Medical Birth Registry of Norway (MBRN), the Norwegian Prescription Registry (NorPD), the Norwegian Patient Registry (NPR), Norway Control and Payment of Health Reimbursement (KUHR). Linkage between the registries was facilitated using project-specific person IDs generated from unique personal identification assigned at birth or immigration for all legal residents in Norway. In brief: • MBRN stores information about the pregnancy, the mother, father, and child. • NPR records diagnosis in secondary care (e.g., hospital). • KUHR contains information about diagnosis and contact in primary care (e.g., GPs and outpatient specialists) • NorPD recorded all medications dispensed outside of hospitals.

Egas Moniz Health Alliance database - Entre o Douro e Vouga (EMDB-ULSEDV)

The Clinical Academic Center Egas Moniz Health Alliance (CAC-EMHA) integrates several Portuguese institutions—the University of Aveiro and 4 Local Health Units (Aveiro, Entre Douro e Vouga, Vila Nova de Gaia/Espinho, and Matosinhos). More than 1 million clinical records of patients are included. The CAC-EMHA has defined main problems for intervention, aligned with the needs of public health and considering the clinical and scientific differentiation of its professionals, in the following areas: a) cardiovascular and respiratory; b) muscle and bone; c) infection and resistance; d) neurosciences. Unidade Local de Saúde de Entre Douro e Vouga (ULSEDV) is an integrated public medical care centre comprising of primary, secondary, and tertiary healthcare. It fully serves approximately 274,000 patients of the municipalities of Santa Maria da Feira, Arouca, São João da Madeira, Oliveira de Azeméis, Vale de Cambra, Ovar, and Castelo de Paiva. The ULSEDV includes 32 primary care centres assisted by three hospitals (Hospital de São Sebastião, Hospital São João da Madeira, and Hospital São Miguel).

Egas Moniz Health Alliance database - Gaia E Espinho (EMDB-ULSGE)

The Clinical Academic Center Egas Moniz Health Alliance (CAC-EMHA) integrates several Portuguese institutions - the University of Aveiro and 4 Local Health Units (Aveiro, Entre Douro e Vouga, Vila Nova de Gaia/Espinho and Matosinhos). More than 1 million clinical records of patients are included. The CAC-EMHA has defined main problems for intervention, aligned with the needs of public health and considering the clinical and scientific differentiation of its professionals, in the following areas: a) cardiovascular and respiratory; b) muscle and bone; c) infection and resistance; d) neurosciences. Unidade Local de Saúde de Gaia e Espinho (ULSGE) is an integrated public medical care centre comprising both primary, secondary, and tertiary healthcare. It fully serves the population of the municipalities of Gaia and Espinho, that amounts to approximately 350.000 patients. The ULSGE includes 32 primary care centres assisted by three hospitals (Hospital Eduardo Santos Silva, Hospital Distrital Vila Nova de Gaia, Hospital Nossa Senhora da Ajuda) and one specialized rehabilitation center (Centro de Reabilitação do Norte).

Egas Moniz Health Alliance database - Baixo Vouga (Região de Aveiro) (EMDB-ULSRA)

The Clinical Academic Center Egas Moniz Health Alliance (CAC-EMHA) integrates several Portuguese institutions—the University of Aveiro and 4 Local Health Units (Aveiro, Entre Douro e Vouga, Vila Nova de Gaia/Espinho, and Matosinhos). More than 1 million clinical records of patients are included. The CAC-EMHA has defined main problems for intervention, aligned with the needs of public health and considering the clinical and scientific differentiation of its professionals, in the following areas: a) cardiovascular and respiratory; b) muscle and bone; c) infection and resistance; d) neurosciences. Unidade Local da Região de Aveiro (ULSRA) is an integrated public medical care centre, comprising of primary, secondary, and tertiary healthcare. It serves approximately 390,000 patients from most of the municipalities of Aveiro. The ULSRA includes 41 primary care centres, assisted by four hospitals (Hospital Dr. Francisco Zagalo, Hospital Visconde de Salreu, Hospital Distrital de Águeda, Hospital Infante D. Pedro).

Unidade Local de Saúde de Matosinhos Realtime Database (ULSM-RT)

The database is comprised of clinical information of patients admitted at 1 public hospital centre, located in Portugal, and 14 connected primary healthcare centres providing full coverage to the municipality of Matosinhos. It includes administrative, sociodemographic, and clinical data (visits, measurements, drugs, procedures, observations) of over 700,000 patients with ages ranging from 0 to 100 years. Of these, 200,000 patients have complete follow-up in primary care with complete data in the mentioned domains. The hospital centre includes departments of most medical specialties, comprising data from surgery, outpatient, ward, accident and emergency, and ICU areas. Data spans back to 1998, with the highest density starting from 2014.

Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público (BIFAP)

BIFAP is a longitudinal, population-based data source of medical patient records from the Spanish National Health Service, comprising several participating Regions throughout Spain. The population currently included represents 36% of the total Spanish population. The Spanish National Health Service provides universal access to health services through the Regional Healthcare Services. Primary care physicians, including general practitioners (GPs) and paediatricians, play a central role. They act as gatekeepers of the system, exchanging information with other levels of care to ensure continuity. Most (98.9%) of the population is registered with a primary care physician, and, in addition, the majority of drug prescriptions are written at this level. BIFAP comprises a collection of databases linked at the individual patient level. The main one is the Primary care Database, given the central role of primary care physicians in the Spanish National Health Service. There are additional important structural databases, such as the medicines dispensed at community pharmacies and the patients' hospital diagnoses at discharge, linked to BIFAP. Linkage to SARS-CoV-2 diagnostics test and COVID-19 vaccination registries are also included. Additional databases are also linked for a subset of patients (hospital pharmacy, cause of death registry). The BIFAP program is a non-profit initiative funded by the Spanish Agency of Medicines and Medical Devices, a government agency that belongs to the Ministry of Health, in collaboration with regional health authorities. The main use of BIFAP is for research purposes to evaluate the adverse and beneficial effects of drugs and drug utilisation patterns in the general population under real conditions of use.

Institut Municipal Assistència Sanitària Information System (IMASIS)

The Institut Municipal Assistència Sanitària Information System (IMASIS) is the Electronic Health Record (EHR) system of Parc de Salut Mar Barcelona (PSMar) which is a complete healthcare services organisation. Currently, this information system includes and shares the clinical information of two general hospitals (Hospital del Mar and Hospital de l'Esperança), one mental health care centre (Centre Dr. Emili Mira) and one social-healthcare centre (Centre Fòrum) including emergency room settings, which are offering specific and different services in the Barcelona city area (Spain). Currently, IMASIS contains clinical information for over 1 million patients with at least one diagnosis who have utilised the services of this healthcare system since 1990, from various settings including admissions, outpatient clinics, the emergency room, and major ambulatory surgery. The diagnoses are coded using The International Classification of Diseases ICD-9-CM and ICD-10-CM. The average follow-up period per patient, in Years, is 6.37 (SD $\pm$ 6.82). IMASIS-2 is the anonymised relational database of IMASIS, which is used for mapping to OMOP, including additional sources of information such as the Tumours Registry.

The Information System for Research in Primary Care (SIDIAP)

SIDIAP is collected from EHR records of individuals receiving primary care delivered through Primary Care teams consisting of GPs, nurses, and non-clinical staff. It is a regional database covering the region of Catalonia. The Catalan Health Institute manages 286 out of 370 such PCTs, covering 5.6 million individuals out of 8 million people residing in the Catalan region (75%), and is mapped to the Observational Medical Outcomes Partnership (OMOP) common data model (CDM). (Recalde et al., 2022; Raventós et al., 2023)

The database began collecting data in 2006 and is updated every six months. The mean follow-up is 15.5 years. The observation period for an individual can be the start of the database (2006) or when a person is assigned to a Catalan Health Institute primary care centre. The exit date can be when a person is transferred to a primary care centre that does not belong to the Catalan Health Institute, the date of death, or the date of end of follow-up in the database. The demographic composition within SIDIAP closely mirrors that of the broader Catalan population, encompassing a representative spectrum of geographic distribution, age, and sex proportions. SIDIAP obtains data from electronic health records. The dataset encompasses demographics, all-cause mortality, disease diagnoses, prescription and dispensation records of drugs, laboratory test results, socio-economic indicators, vaccination records, lifestyle information, parent-child linkage, and various clinical parameters. Diseases are classified under the International Classification of Diseases 10th revision (ICD-10). In this data source, prescription data provides more information than dispensation (such as dose, posology, and quantity).

Quality checks have been implemented, including central identification of duplicate patient IDs and visual inspection for temporal patterns in the registry of specific variables. Furthermore, the data undergoes assessment for availability (longitudinally and reliability), plausibility (range checks and unusual values), and visualisation tools. Specifically, for biochemistry data, consistency of measurements taken in different laboratories is assessed, and unit conversions are undertaken when necessary. SIDIAP is linked with numerous other databases. It integrates data from external sources, including laboratory biomarker data, drug prescription and dispensation records, hospital discharge records, mental health centres, and other specific disease registries. Drugs not prescribed in the GP setting might be underreported, and disease diagnoses made in specialist care settings are not included. Vital status (death date and cause) is collected through linkage with the civil registry. The main limitation is that SIDIAP covers only primary health care records. However, it is combined with data from various other sources through effective linkages. This study will use primary care data linked to hospital discharge data.

SIDIAP includes detailed information related to the perinatal period for approximately 40,000 deliveries per year. The quality of a wide number of data captured in SIDIAP has been demonstrated in validation studies, which include several diseases in the general population.(Recalde et al., 2022)

Valencia Health System Integrated Dataset (VID)

The Valencia Health System Integrated Dataset (VID) is a set of multiple, public, population-wide electronic databases for the Valencia Region, the fourth most populated Spanish region, with about 5 million inhabitants and an annual birth cohort of 48,000 new-borns, representing 10.7% of the Spanish population and around 1% of the European population. The VID provides exhaustive longitudinal information, including sociodemographic and administrative data (sex, age, nationality, etc.), clinical (diagnoses, procedures, diagnostic tests, imaging, etc.), pharmaceutical (prescription, dispensing), and healthcare utilization data from hospital care, emergency departments, specialized care (including mental and obstetrics care), primary care, and other public health services. It also includes a set of associated population databases and registries of significant care areas, such as cancer, rare diseases, vaccines, congenital anomalies, microbiology (including COVID-19 test results registry), and others, as well as public health databases from the population screening programmes. All the information in the VID databases can be linked at the individual level through a single personal identification code. The databases were initiated at different moments in time, but all in all the VID provides comprehensive individual-level data fed by all the databases since 2008 to date. IMPORTANT: The OMOP instance has been created using CONSIGN project data, where 1.96 million of females in fertile age are studied from the start of 2018 to the end of 2021.

Clinical Practice Research Datalink GOLD (CPRD GOLD)

The Clinical Practice Research Datalink (CPRD) GOLD is a database of anonymised electronic health records (EHR) from General Practitioner (GP) clinics in the UK that use the Vision® software system for their

management.(Herret et al., 2015) The source population encompasses 98% of the UK, registered with GPs responsible for non-emergency care and referrals. Participating GPs provide CPRD EHR for all registered patients who did not specifically request to opt out of data sharing. Covering 4.6% of the current UK population, GOLD includes 4.9% of contributing GP practices, providing comprehensive information within its defined source population. GOLD contains data from all four UK constituent countries, and the current regional distribution of its GP practices is 5.7% in England, 55.6% in Scotland, 28.4% in Wales, and 10.2% in Northern Ireland (May 2022).

GOLD data include patients' demographics, biological measurements, clinical symptoms and diagnoses, referrals to specialists/hospital and their outcome, laboratory tests/results, and prescribed medications. GOLD has been assessed and found broadly representative of the UK general population in terms of age, gender, and ethnicity. (Herret et al., 2015) GOLD has been widely used internationally for observational research, resulting in nearly 3,000 peer-reviewed publications, making GOLD the most influential UK clinical database to date. (Carey et al., 2023; Wigglesworth et al., 2023; Fahmi et al., 2023)

In 2019, CPRD launched AURUM and, since then, has encouraged practices from England to move from the software that feeds GOLD (Vision) to the one that feeds AURUM (Emis). GOLD data from 2019, therefore, mainly represents Wales/Scotland/Ni, and AURUM represents England. However, GOLD data collected before 2019 fully represent the UK. CPRD provides an updated list of practices that have moved from GOLD to AURUM for each build release. An overlap between GOLD and AURUM can occur because historical data for these practices have been transferred from Vision/GOLD to Emis/AURUM. When DARWIN EU® uses both databases, the safest and easiest solution would be to disregard these practices in GOLD. The licence also covers HES/ONS data, which can be requested on a study-by-study basis as linked data. This data only covers England and is planned to be mapped to OMOP in the future.

In terms of quality checks, the integrity, structure, and format of the data is reviewed. Collection-level validation ensures integrity by verifying that data received from practices contains only the expected data files and that all data elements are of the correct type, length, and format. Duplicate records are identified and removed. (Herret et al., 2015) Transformation-level validation checks for referential integrity between records ensure that no orphan records are included in the database (for example, that all event records link to a patient), while research-quality-level validation covers the actual content of the data. CPRD provides a patient-level data quality metric in the form of a binary 'acceptability' flag. (Herret et al., 2015) This is based on the recording and internal consistency of key variables, including date of birth, practice registration date, and transfer-out date.

UK Biobank (UKBB)

UK Biobank is a powerful biomedical database that can be accessed globally to enable new discoveries to improve public health. UK Biobank contains in-depth genetic, biomarker, imaging, and health information from over half a million volunteers living in the UK aged 40–69 years at the time of recruitment (2006–2010). UK Biobank has collected an unprecedented amount of biological and medical data as part of a large-scale long term prospective study. With their consent, they regularly provide blood, urine, and saliva samples, as well as detailed information about their lifestyle, which is then linked to their health-related records (e.g., primary care data, hospital data, cancer registry) to provide a deeper understanding of how individuals experience diseases. Since 2012 UK Biobank database, the largest and richest of its kind, is opened to applications from researchers. The resource is available in a strictly anonymised format to scientists from the UK and around the world, subject to verification that the research is health-related and in the public interest. Researchers are required to publish their results in an open-source publication site or in an academic journal and return their findings to the UK Biobank. At the time of writing nearly 4,056 research applications have been approved for the usage of UK Biobank data and 8,553 peer-reviewed articles based on them have been published.

Health Impact - Swedish Population Evidence Enabling Data-linkage (HI-SPEED)

The Health Impact - Swedish Population Evidence Enabling Data-linkage (HI-SPEED) study is a nationwide linked multi-register, regularly updated, observational study for timely response over time to scientific questions around effectiveness and safety of approved drugs that can arise suddenly, requiring rapid evidence for timely regulatory action - to protect patients' health and lives. The available data cover the entire Swedish population (approximately 10 million), including information on specialist care (National Patient Register), drug use (Prescribed Drug Register), cause of death (Cause-of-Death Register), sociodemographic data, and selected clinical data. Primary care visit diagnoses and procedures are available for 40% of the population (two largest Swedish regions). Most data start from 2015; prescription drug data on all prescriptions filled nationally are available from 2018. Data on hospital-administered drugs is not available if the prescription was not individually filled by the patient. The study population and all data are updated quarter yearly. HI-SPEED builds on the predecessor project SCIFI-PEARL (Swedish COVID-19 Investigation for Future Insights - a Population Epidemiology Approach using Register Linkage) that was initiated in 2020 to conduct research on COVID-19 and pandemic-related issues (<https://www.gu.se/en/research/scifi-pearl>).

ANNEX II. Questionnaires for Data Partners on pregnancy data and Mother-Child Linkage details

Questionnaire 1

Recently, there has been a growing interest from EMA on studies focused on the population of pregnant women and their children. Because of this, we are contacting you to explore what information is available in your source data regarding pregnancy, if this information has been mapped to the OMOP CDM, and if not, if you could map it for potential new studies.

1. Pregnancy start date
 - a. Is it available in the source data (yes/no)?
 - b. Has it been mapped to the OMOP CDM (yes/no)?
2. Pregnancy end date
 - a. Is it available in the source data (yes/no)?
 - b. Has it been mapped to the OMOP CDM (yes/no)?
3. Gestational length (weeks)
 - a. Is it available in the source data (yes/no)?
 - b. Has it been mapped to the OMOP CDM (yes/no)?
4. Pregnancy outcomes (livebirth, stillbirth, miscarriage, elective termination, etc)
 - a. Is it available in the source data (yes/no)?
 - b. Has it been mapped to the OMOP CDM (yes/no)?
5. Pregnancy mode of delivery (c-section, vaginal)
 - a. Is it available in the source data (yes/no)?
 - b. Has it been mapped to the OMOP CDM (yes/no)?
6. Number of foetuses in one pregnancy
 - a. Is it available in the source data (yes/no)?
 - b. Has it been mapped to the OMOP CDM (yes/no)?
7. Mother-child linkage
 - a. Is it available in the source data (yes/no)?
 - b. Has it been mapped to the OMOP CDM (yes/no)?
8. Birth weight
 - a. Is it available in the source data (yes/no)?
 - b. Has it been mapped to the OMOP CDM (yes/no)?
9. Apgar score
 - a. Is it available in the source data (yes/no)?
 - b. Has it been mapped to the OMOP CDM (yes/no)?
10. Other information that might be relevant to pregnancy? If yes, please specify

Questionnaire 2

Dear Data Partners,

Thank you for accepting to participate in the study P4-C4-001 - **DARWIN EU® - Enabling pregnancy and mother-child research in the DARWIN EU® Data network**. One of the objectives of the study is to obtain detailed information on availability, completeness, and quality of pregnancy-related and mother and child linkage in your data. For this, we would like to ask for your help by providing answers to the questions below.

Start in MS Forms: https://forms.cloud.microsoft/Pages/DesignPageV2.aspx?subpage=design&FormId=-30_TkhCrk-6KrUKlyv7M1XRc2V8nQFLvusYx954QlxURDlDMzlEWlpUQUUpCUkhTU1ZNWE43MIJXRyQIQCN0PWcu&Token=210889a3673d4d7bb683b0b1da6603cb

What is your name?

What is your e-mail?

Which database do you represent? (please note that if you are a data partner with access to multiple databases, we kindly ask you to provide a reply to this survey separately for each database).

1. What is the source of pregnancy information in your database?

Example: Pregnancy data is available through linkage with national birth registries [add link].

- a. Entry in EHR by clinician*
- b. Linkage with external registry/ies*
- c. Patient reported*
- d. Hospital records*
- e. Others (free text)*

2. If available, please share documentation or references on the source for pregnancy information.

3. If start of pregnancy date is available in your database, how is that date derived or recorded? (multiple answers possible)

- a. Date of Last menstrual period
- b. Date of Ultrasound results
- c. Date of recording of "pregnancy" within the patient's file (i.e., diagnosis)
- d. Date of positive Pregnancy test (urine or blood)
- e. Algorithm (please specify more details on the algorithm you use in question number 5 below)
- f. Others (please specify):

4. If the pregnancy end data is available in your database, how is it derived or recorded?

- a. Date of delivery
- b. Date of birth when mother-child linkage is available
- c. Other (please specify):

5. If you use an algorithm to derive information on pregnancy start- and end dates or pregnancy outcomes , please provide links to relevant publications, github repository, or any documentation with details on how these are implemented.
6. If the start- and end date of pregnancy have been already mapped to the OMOP-CDM, how and in which table has it been mapped? *Example: We used the PET to map pregnancy start and end dates.*
7. Are the following variables available in source?
 - g. Date of last Menstrual Period Yes No
 - h. Gestational age Yes No
 - i. Mode of delivery (vaginal spontaneous, vaginal assisted, c-section) Yes No
 - j. Pregnancy outcomes (live birth, stillbirth, abortion, ectopic, delivery) Yes No
 - k. Weight measurements during pregnancy Yes No
 - l. Alcohol consumption during pregnancy Yes No
 - m. Fertility procedures Yes No
 - n. Amniocentesis Yes No
 - o. Alpha-fetoprotein Yes No
 - p. Breastfeeding Yes No
8. If available, are the following variables already mapped to the OMOP CDM?
 - a. Date of last Menstrual Period Yes No
 - b. Gestational age Yes No
 - c. Mode of delivery (vaginal spontaneous, vaginal assisted, c-section) Yes No
 - d. Pregnancy outcomes (live birth, stillbirth, abortion, ectopic, delivery) Yes No
 - e. Weight measurements during pregnancy Yes No
 - f. Alcohol consumption during pregnancy Yes No
 - g. Fertility procedures Yes No
 - h. Amniocentesis Yes No
 - i. Alpha-fetoprotein Yes No
 - j. Breastfeeding Yes No
9. If the date of last menstrual period (LMP) is available, how is it determined?

Example: LMP is backdated from pregnancy outcome and outcome date. LMP is recorded at first pregnancy visit.
10. If the gestational age is available, how is it determined?

Example: Gestational age is recorded as a variable at the time of pregnancy outcome. Gestational age is determined based on outcome-specific average gestational age.

11. Are the following pregnancy outcomes recorded (live birth, stillbirth, abortion, ectopic, delivery-only)?

- a. Live birth Yes No
- b. Multiple births (twin, triplets, etc.) Yes No
- c. Ectopic pregnancy Yes No
- d. Stillbirth Yes No
- e. Induced abortion Yes No
- f. Spontaneous pregnancy loss Yes No
- g. Stillbirth (>20-28 weeks of gestation) Yes No
- h. Maternal death Yes No

12. If available, are the following variables already mapped to the OMOP CDM?

- a. Live birth Yes No
- b. Multiple births (twin, triplets, etc.) Yes No
- c. Ectopic pregnancy Yes No
- d. Stillbirth Yes No
- e. Induced abortion Yes No
- f. Spontaneous pregnancy loss Yes No
- g. Stillbirth (>20-28 weeks of gestation) Yes No
- h. Maternal death Yes No

13. Do you have information on ALL pregnancies occurring in patients included in your database or is there selective recording of pregnancies? Some examples where not all pregnancies are recorded (for good reasons) are described below

Examples: We only have information on pregnancies ending after the 12th week of gestation; We only have pregnancies recorded in the public health centers; We only have pregnancies for which a hospitalisation occurred.

14. Do you apply any rules to determine successive pregnancies?

Example: At least 182-day separation between livebirths is imposed.

15. Do you have information covering the entire pregnancy period?

a. Yes

b. No – If no, provide further clarification e.g., as an example

“ We only have data of pregnancies during the second and third trimester.”

16. Has the data source been used for research on pregnancy?

- a. Yes
- b. No

17. If so, please provide links to relevant publications.

18. Do you work with a third-party for your OMOP CDM mapping?

- a. Yes
- b. No

19. Are there any limitations and strengths that you would like to share about the available data that you have access to related to pregnancy (e.g., proportion of missingness, validity, changes over time etc.)?

For DP with mother-child linkage (available or under implementation) only:

20. What is the source for mother-child linkage? If possible, could you share details and/or references about the algorithm used and its reliability?

Example: LinkedFamily ID and birthdates are available.

21. Is this linkage available for all pregnancies and children in your database?

Example: It's only available for pregnancies occurring in specific regions.

22. If the mother-child linkage is currently being implemented in the source data, could you specify the expected timelines for implementation and the challenges you are currently experiencing?

23. Has the mother-child linkage been mapped to the OMOP-CDM?

24. If this linkage is mapped to the OMOP CDM, how have you mapped this information?

- a. Fact_relationship table
- b. Observation table
- c. Custom table
- d. Other (free text)

25. If this linkage is not yet mapped to the OMOP CDM, are you planning to do so in the next 3 months?

- a. Yes
- b. No
- c. If not, why and when could you do this? (free text)

26. Is father-child linkage available in your database?

- a. Yes
- b. No

27. If father-child linkage is available, has it been mapped to the OMOP-CDM?

- a. Yes
- b. No

28. Are there any limitations and strengths that you would like to share about the available data on mother-child linkage?
29. Are there any limitations and strengths that you would like to share about the available data on father-child linkage (if available)?
30. If father-child linkage is not available in your data, are you aware of any other databases in your country where this linkage is available?

If your database includes mother and child linkage and/or father and child linkage, can you share relevant papers/documents?

ANNEX III. Codes for pregnancy and infant information searches

Table S1. List of pregnancy related codes for search.

Concept ID	Name	Domain	Vocabulary
444094	Finding related to pregnancy	Condition	SNOMED
4299535	Pregnant	Condition	SNOMED
4169813	Wanted pregnancy	Condition	SNOMED
439658	Disorder of pregnancy	Condition	SNOMED
4024659	Gestational diabetes mellitus	Condition	SNOMED
4058439	Last menstrual period -1st day	Measurement	SNOMED
436747	Legal termination of pregnancy without complication	Condition	SNOMED
4014769	Urine pregnancy test, by visual color comparison methods	Measurement	SNOMED
434701	Anemia in mother complicating pregnancy, childbirth AND/OR puerperium	Condition	SNOMED
45757490	Pregnancy not yet confirmed	Condition	SNOMED
44792235	Pregnancy counselling	Procedure	SNOMED
4193062	Early stage of pregnancy	Condition	SNOMED
4128331	Pregnancy	Observation	SNOMED
440785	Vomiting of pregnancy	Condition	SNOMED
4060556	Uterine scar from previous surgery in pregnancy, childbirth and the puerperium	Condition	SNOMED
435875	Complication of pregnancy, childbirth and/or puerperium	Condition	SNOMED
44790261	Ultrasound monitoring of early pregnancy	Procedure	SNOMED
4025198	Hemorrhagic complication of pregnancy	Condition	SNOMED
4062791	Endocrine, nutritional and metabolic disease complicating pregnancy, childbirth and puerperium	Condition	SNOMED
3002209	Gestational age Estimated	Measurement	LOINC
4217975	Normal pregnancy	Condition	SNOMED
4297250	Termination of pregnancy	Procedure	SNOMED
40539858	Pregnancy with abortive outcome	Condition	SNOMED
4115153	Medical termination of pregnancy	Procedure	SNOMED
37018340	Induced termination of pregnancy	Procedure	SNOMED
4034097	Symptomatic disorders in pregnancy	Condition	SNOMED
4188598	High risk pregnancy	Condition	SNOMED
43021845	Psychosocial problems related to unwanted pregnancy	Condition	SNOMED
4167493	Pregnancy-induced hypertension	Condition	SNOMED
4240605	Termination of pregnancy	Condition	SNOMED
437611	Ectopic pregnancy	Condition	SNOMED
4063155	Disease of nervous system complicating pregnancy, childbirth and puerperium	Condition	SNOMED

Concept ID	Name	Domain	Vocabulary
4058243	Diabetes mellitus during pregnancy, childbirth and the puerperium	Condition	SNOMED
4145947	Small for gestational age fetus	Condition	SNOMED
4210126	Pregnancy-related examination	Procedure	SNOMED
4060296	Infections of the genital tract in pregnancy	Condition	SNOMED
43530950	Complication occurring during pregnancy	Condition	SNOMED
77340	Genitourinary tract infection in pregnancy	Condition	SNOMED
4336958	Term pregnancy	Condition	SNOMED
4081422	Elective termination of pregnancy	Condition	SNOMED
45765728	Supervision of high risk pregnancy	Observation	SNOMED
4062269	Cerebral venous thrombosis in pregnancy	Condition	SNOMED
4129024	Non-viable pregnancy	Condition	SNOMED
4134701	Patient advised to have pregnancy test	Observation	SNOMED
45757356	Pre-existing hypertensive chronic kidney disease in mother complicating pregnancy	Condition	SNOMED
44802473	Pregnancy advice for patients with epilepsy declined	Observation	SNOMED
4059982	Pregnant - blood test confirms	Condition	SNOMED
442594	Quadruplet pregnancy	Condition	SNOMED
37017051	Supervision of high risk pregnancy with history of previous intrauterine death	Observation	SNOMED
439894	Post-term pregnancy - delivered	Condition	SNOMED
199076	Tubal pregnancy	Condition	SNOMED
4212941	Aspiration curettage of uterus for termination of pregnancy	Procedure	SNOMED
4062790	Disease of the digestive system complicating pregnancy, childbirth and/or the puerperium	Condition	SNOMED
36713460	Supervision of pregnancy with history of infertility	Observation	SNOMED
37392365	Urine pregnancy test	Measurement	SNOMED
4142469	Pregnancy time period	Observation	SNOMED
42538560	Syndrome of infant of mother with gestational diabetes	Condition	SNOMED
43020670	Continuing pregnancy after intrauterine death of one twin with intrauterine retention of dead twin	Condition	SNOMED
45757815	Counselling for termination of pregnancy done	Observation	SNOMED
437062	Cracked nipple in pregnancy, the puerperium or lactation with antenatal complication	Condition	SNOMED
432396	Cracked nipple in pregnancy, the puerperium or lactation with postnatal complication	Condition	SNOMED
196475	Damage to pelvic organs AND/OR tissues following molar AND/OR ectopic pregnancy	Condition	SNOMED
74706	Failed attempted termination of pregnancy complicated by genital-pelvic infection	Condition	SNOMED
4118058	Fetal or neonatal effect of ectopic pregnancy	Condition	SNOMED

Concept ID	Name	Domain	Vocabulary
4051249	FH: Diabetes in pregnancy	Observation	SNOMED
3049700	In vitro fertilization pregnancy	Observation	LOINC
44807267	Management of gestational diabetes mellitus	Procedure	SNOMED
442081	Pelvic soft tissue abnormality in pregnancy, childbirth and the puerperium	Condition	SNOMED
4057978	Pre-existing hypertensive heart and renal disease complicating pregnancy, childbirth and the puerperium	Condition	SNOMED
442271	Shock following molar AND/OR ectopic pregnancy	Condition	SNOMED
443214	Gestational edema	Condition	SNOMED
4074427	Excision of ruptured ectopic tubal pregnancy	Procedure	SNOMED
4166442	FH: Twin pregnancy	Observation	SNOMED
4214186	Gestational trophoblastic disease	Condition	SNOMED
42535715	Hyperthyroidism in pregnancy	Condition	SNOMED
442772	Isoimmunization from non-ABO, non-Rh blood-group incompatibility affecting pregnancy	Condition	SNOMED
443054	Maternal gonorrhea during pregnancy, childbirth and the puerperium	Condition	SNOMED
44813931	Pertussis vaccination in pregnancy declined	Observation	SNOMED
44809992	Pertussis vaccination in pregnancy invitation second letter	Observation	SNOMED
4148890	Pregnancy benefits	Observation	SNOMED
4028479	Vaginal abnormality in pregnancy, childbirth and the puerperium	Condition	SNOMED
45757142	Viral hepatitis in mother complicating pregnancy	Condition	SNOMED
4135932	Uterine scar from previous surgery affecting pregnancy	Condition	SNOMED
432695	Post-term pregnancy	Condition	SNOMED
46284879	Pertussis vaccination in pregnancy	Procedure	SNOMED
4064787	Pre-pregnancy education	Procedure	SNOMED
45772076	Premature rupture of membranes in full term pregnancy with onset of labor more than 24 hours after rupture	Condition	SNOMED
4239301	Unwanted pregnancy	Condition	SNOMED
40483581	Combined tubal and intrauterine pregnancy	Condition	SNOMED
442915	Drug dependence during pregnancy - baby delivered	Condition	SNOMED
4113509	Failed medical termination of pregnancy, complicated by embolism	Condition	SNOMED
439329	Incomplete legal termination of pregnancy with embolism	Condition	SNOMED
4061672	IUD failure - pregnant	Observation	SNOMED
45757789	Postpartum gestational diabetes mellitus	Condition	SNOMED
4082401	Pregnancy prophylactic therapy	Procedure	SNOMED
4062574	Pregnancy-related glycosuria	Condition	SNOMED
197930	Renal hypertension complicating pregnancy, childbirth and the puerperium	Condition	SNOMED
4155848	Two previous induced terminations of pregnancy	Condition	SNOMED

Concept ID	Name	Domain	Vocabulary
2211752	Ultrasound, pregnant uterus, real time with image documentation, fetal and maternal evaluation plus detailed fetal anatomic examination, transabdominal approach; each additional gestation (List separately in addition to code for primary procedure)	Procedure	CPT4
197050	Abnormality of organs AND/OR soft tissues of pelvis affecting pregnancy	Condition	SNOMED
199891	Rhesus isoimmunization affecting pregnancy	Condition	SNOMED
4099889	Anemia of pregnancy	Condition	SNOMED
45757124	Gestational diabetes mellitus in childbirth	Condition	SNOMED
45769366	Assessment for termination of pregnancy	Procedure	SNOMED
4135209	Combined pregnancy	Condition	SNOMED
438196	Complete legal termination of pregnancy with delayed or excessive hemorrhage	Condition	SNOMED
4129020	Concealed pregnancy	Condition	SNOMED
4064792	Drugs in pregnancy education	Procedure	SNOMED
44807391	Gestational diabetes mellitus annual review	Observation	SNOMED
4316868	Gingival disease due to pregnancy	Condition	SNOMED
45767173	Partner pregnant	Observation	SNOMED
42539186	Pre-existing maternal disease complicating pregnancy	Condition	SNOMED
4174577	Secondary abdominal pregnancy	Condition	SNOMED
4034087	Single pregnancy	Condition	SNOMED
141639	Transient hypertension of pregnancy - delivered with postnatal complication	Condition	SNOMED
443242	Varicose veins of perineum and vulva in pregnancy and the puerperium	Condition	SNOMED
433823	Infectious disease in mother complicating pregnancy, childbirth AND/OR puerperium	Condition	SNOMED
4059748	Finding related to risk factor in pregnancy	Condition	SNOMED
45757123	Premature rupture of membranes in full term pregnancy with onset of labor unknown	Condition	SNOMED
4269150	Measure of pregnancy	Observation	SNOMED
4086797	Requests pregnancy termination	Observation	SNOMED
3003191	Choriogonadotropin (pregnancy test) [Presence] in Serum or Plasma	Measurement	LOINC
45769909	Low back pain in pregnancy	Condition	SNOMED
4218813	Third trimester pregnancy	Condition	SNOMED
4118411	Fetal or neonatal effect of multiple pregnancy	Condition	SNOMED
4072438	Date of last menstrual period	Measurement	SNOMED
40310686	Possible pregnancy	Observation	SNOMED
44791171	9 - 13 weeks gestational age	Condition	SNOMED
81086	Breast engorgement in pregnancy, the puerperium or lactation - delivered with postnatal complication	Condition	SNOMED
4210896	Fetal gestational age	Observation	SNOMED

Concept ID	Name	Domain	Vocabulary
4126257	Finding of measures of pregnancy	Condition	SNOMED
37018765	Gestational diabetes mellitus complicating pregnancy	Condition	SNOMED
43530908	Induced termination of pregnancy complicated by damage to pelvic organs and/or tissues	Condition	SNOMED
43531710	Induced termination of pregnancy complicated by embolism	Condition	SNOMED
4121775	Injection of amnion for termination of pregnancy	Procedure	SNOMED
442346	Legal termination of pregnancy complicated by embolism	Condition	SNOMED
4129022	Low risk pregnancy	Condition	SNOMED
45757107	Malignant neoplastic disease in pregnancy	Condition	SNOMED
438205	Maternal syphilis during pregnancy - baby not yet delivered	Condition	SNOMED
43530979	Miscarriage of tubal ectopic pregnancy	Condition	SNOMED
44809995	Pertussis vaccination in pregnancy invitation third letter	Observation	SNOMED
44784480	Pregnancy induced edema	Condition	SNOMED
4150401	Reason for termination of pregnancy	Observation	SNOMED
4299185	Removal of ectopic pregnancy from fallopian tube	Procedure	SNOMED
4041162	Serum pregnancy test negative	Condition	SNOMED
44791172	Under 9 weeks gestational age	Condition	SNOMED
4129689	Vulval abnormality in pregnancy, childbirth and the puerperium	Condition	SNOMED
4060429	Disease of the respiratory system complicating pregnancy, childbirth and/or the puerperium	Condition	SNOMED
4063159	Subluxation of symphysis pubis in pregnancy, childbirth and the puerperium	Condition	SNOMED
4307820	Unplanned pregnancy	Condition	SNOMED
3051741	Pregnancy associated plasma protein A [Mass/volume] in Serum or Plasma	Measurement	LOINC
321074	Pre-existing hypertension complicating pregnancy, childbirth and puerperium	Condition	SNOMED
4060558	Uterine scar from previous surgery in pregnancy, childbirth and the puerperium - delivered	Condition	SNOMED
4239938	First trimester pregnancy	Condition	SNOMED
438480	Abnormal glucose tolerance in mother complicating pregnancy, childbirth AND/OR puerperium	Condition	SNOMED
4230221	Fatigue during pregnancy	Condition	SNOMED
3009417	Choriogonadotropin.beta subunit (pregnancy test) [Presence] in Urine	Measurement	LOINC
4055971	Albuminuria in pregnancy without hypertension	Condition	SNOMED
45757093	Alcohol dependence in pregnancy	Condition	SNOMED
45757214	Anti-A sensitization in pregnancy	Condition	SNOMED
4062811	Benign essential hypertension complicating pregnancy, childbirth and the puerperium with postnatal complication	Condition	SNOMED
4028640	Bicornuate uterus in pregnancy, childbirth and the puerperium	Condition	SNOMED

Concept ID	Name	Domain	Vocabulary
200791	Congenital or acquired abnormality of vagina in pregnancy, childbirth and the puerperium	Condition	SNOMED
42537767	Damage to pelvic organs and tissues following ectopic pregnancy	Condition	SNOMED
4207588	Education about toxoplasmosis precautions during pregnancy	Procedure	SNOMED
436228	Fetal death due to termination of pregnancy	Observation	SNOMED
4060294	Genitourinary tract infection in pregnancy with postnatal complication	Condition	SNOMED
43530939	Induced termination of pregnancy complicated by sepsis	Condition	SNOMED
45757097	Inflammation of cervix in pregnancy	Condition	SNOMED
4063175	Multiple pregnancy with malpresentation with antenatal problem	Condition	SNOMED
195019	Pelvic soft tissue abnormality in pregnancy, childbirth and the puerperium with antenatal problem	Condition	SNOMED
197608	Peripheral neuritis in pregnancy	Condition	SNOMED
45757136	Physical abuse complicating pregnancy	Condition	SNOMED
4034085	Pregnancy with uncertain dates	Condition	SNOMED
4033204	Prurigo of pregnancy	Condition	SNOMED
4061423	Questionable if pregnancy was planned	Condition	SNOMED
4113503	Readmission for retained products of conception, legal termination of pregnancy	Condition	SNOMED
195014	Renal failure following molar AND/OR ectopic pregnancy	Condition	SNOMED
4042412	Serum pregnancy test positive	Condition	SNOMED
4129023	Teenage pregnancy	Condition	SNOMED
4239693	Termination of pregnancy on demand	Condition	SNOMED
4204840	Therapeutic termination of pregnancy	Condition	SNOMED
440462	Triplet pregnancy - delivered	Condition	SNOMED
36303380	Trisomy 18 risk based on maternal age + Pregnancy associated plasma protein A + Choriogonadotropin + Nuchal translucency [Likelihood] in Fetus	Measurement	LOINC
4145952	Tuberculosis in pregnancy	Condition	SNOMED
4041877	Urine pregnancy test equivocal	Condition	SNOMED
4060424	Mental disorders during pregnancy, childbirth and the puerperium	Condition	SNOMED
196758	Tumor of body of uterus affecting pregnancy	Condition	SNOMED
4015590	Counseling for termination of pregnancy	Procedure	SNOMED
4063038	Infections of bladder in pregnancy	Condition	SNOMED
4243056	Not pregnant	Condition	SNOMED
4091461	Standard pregnancy test	Measurement	SNOMED
4055286	Urine pregnancy test negative	Condition	SNOMED
4175637	Premature pregnancy delivered	Condition	SNOMED
4244438	Second trimester pregnancy	Condition	SNOMED
4098318	Fear of pregnancy	Condition	SNOMED

Concept ID	Name	Domain	Vocabulary
21493940	US for pregnancy in second or third trimester	Procedure	LOINC
40760833	Birth plurality of Pregnancy	Measurement	LOINC
40771967	Body weight Mother --pre current pregnancy	Measurement	LOINC
43021054	Dichorionic diamniotic twin pregnancy	Condition	SNOMED
44791170	14 - 20 weeks gestational age	Condition	SNOMED
42873164	Acute cystitis in pregnancy, antepartum	Condition	SNOMED
44810891	Advice on risks of harm to fetus from maternal medication during pregnancy	Procedure	SNOMED
4061789	Antenatal care: precious pregnancy	Observation	SNOMED
4152443	Antenatal scan unable to confirm pregnancy	Condition	SNOMED
4062567	Asymptomatic bacteriuria in pregnancy with postnatal complication	Condition	SNOMED
45757276	Bacterial vaginosis in pregnancy	Condition	SNOMED
77624	Breast engorgement in pregnancy, the puerperium or lactation - delivered	Condition	SNOMED
44783183	Combined intrauterine and ovarian pregnancy	Condition	SNOMED
315013	Congenital cardiovascular disorder during pregnancy - baby delivered	Condition	SNOMED
443249	Congenital cardiovascular disorders during pregnancy, childbirth and the puerperium	Condition	SNOMED
4113846	Delivery of viable fetus in abdominal pregnancy	Condition	SNOMED
37206226	Excessive hemorrhage due to and following induced termination of pregnancy	Condition	SNOMED
4074426	Excision of ectopic ovarian pregnancy	Procedure	SNOMED
40479820	Exposure to rubella in pregnancy	Observation	SNOMED
4118413	Fetal or neonatal effect of twin pregnancy	Condition	SNOMED
4050824	FH: Raised B.P. in pregnancy	Observation	SNOMED
44792185	Finding of gestational age	Condition	SNOMED
4126258	Finding of viability of pregnancy	Condition	SNOMED
44805600	Gestational diabetes information leaflet given	Observation	SNOMED
44802654	Health in Pregnancy grant claim form issued	Observation	SNOMED
435323	Hemorrhage in early pregnancy, delivered	Condition	SNOMED
435869	Illegal termination of pregnancy	Condition	SNOMED
43530921	Induced termination of pregnancy complicated by endometritis	Condition	SNOMED
43530911	Induced termination of pregnancy complicated by metabolic disorder	Condition	SNOMED
43530945	Induced termination of pregnancy complicated by urinary tract infection	Condition	SNOMED
441081	Maternal gonorrhea during pregnancy - baby not yet delivered	Condition	SNOMED
435606	Maternal syphilis during pregnancy - baby delivered	Condition	SNOMED
432371	Metabolic disorder following molar AND/OR ectopic pregnancy	Condition	SNOMED
606917	Multiple pregnancy following assisted fertilization	Condition	SNOMED

Concept ID	Name	Domain	Vocabulary
45757099	Neoplasm of uterus affecting pregnancy	Condition	SNOMED
4060193	No history of induced termination of pregnancy	Observation	SNOMED
4309719	One previous induced termination of pregnancy	Condition	SNOMED
44790233	Over 20 weeks gestational age	Condition	SNOMED
192971	Pelvic soft tissue abnormality in pregnancy, childbirth and the puerperium - delivered with postnatal complication	Condition	SNOMED
4239635	Polyp of cervix affecting pregnancy	Condition	SNOMED
4064793	Pregnancy dental education	Procedure	SNOMED
4077075	Pregnancy vitamin/iron prophylaxis	Procedure	SNOMED
4089558	Presentation of pregnancy	Condition	SNOMED
45757156	Pyemic and septic embolism in pregnancy	Condition	SNOMED
438481	Quadruplet pregnancy - delivered	Condition	SNOMED
4129702	Rectocele in pregnancy, childbirth and puerperium	Condition	SNOMED
45757120	Renal failure after ectopic pregnancy	Condition	SNOMED
437061	Retracted nipple in pregnancy, the puerperium or lactation	Condition	SNOMED
438823	Retracted nipple in pregnancy, the puerperium or lactation with postnatal complication	Condition	SNOMED
444251	Rubella in mother complicating pregnancy, childbirth AND/OR puerperium	Condition	SNOMED
4069106	Ruptured ectopic pregnancy	Condition	SNOMED
4113680	Ruptured tubal pregnancy	Condition	SNOMED
44804788	Self referral to termination of pregnancy service	Observation	SNOMED
45765733	Supervision of high risk pregnancy for multigravida age 15 years or younger	Observation	SNOMED
45773399	Supervision of high risk pregnancy with history of gestational diabetes mellitus	Observation	SNOMED
45765729	Supervision of high risk pregnancy with history of previous molar pregnancy	Observation	SNOMED
4012559	Surrogate pregnancy	Condition	SNOMED
4210192	Three previous induced terminations of pregnancy	Condition	SNOMED
4323204	Trichomonal vaginitis in pregnancy	Condition	SNOMED
4242401	Urinary tract infection following molar AND/OR ectopic pregnancy	Condition	SNOMED
4055285	Urine pregnancy test	Procedure	SNOMED
4326497	Vaginitis in pregnancy	Condition	SNOMED
4129034	Varicose veins of vulva in pregnancy	Condition	SNOMED
4079752	Vitamin supplement - pregnancy	Procedure	SNOMED
136755	Thyroid dysfunction during pregnancy - baby delivered	Condition	SNOMED
4142037	Gestational proteinuria	Condition	SNOMED
4084832	Referral for counseling for termination of pregnancy	Observation	SNOMED

Concept ID	Name	Domain	Vocabulary
3021584	Pregnancy associated plasma protein A [Units/volume] in Serum or Plasma	Measurement	LOINC
444077	Legal termination of pregnancy	Condition	SNOMED
36713030	Bleeding from female genital tract during pregnancy	Condition	SNOMED
37018878	High risk pregnancy due to recurrent pregnancy loss	Condition	SNOMED
44783943	Anti-D isoimmunization affecting pregnancy	Condition	SNOMED
35815052	Number of babies delivered from single pregnancy	Observation	UK Biobank
4168396	Dilation of cervix uteri and curettage for termination of pregnancy	Procedure	SNOMED
4058375	Disease of skin and subcutaneous tissue complicating pregnancy, childbirth and/or puerperium	Condition	SNOMED
36717636	Supervision of pregnancy with history of insufficient antenatal care	Observation	SNOMED
45765732	Supervision of high risk pregnancy for social problem	Observation	SNOMED
440453	Incomplete legal termination of pregnancy with delayed or excessive hemorrhage	Condition	SNOMED
37206229	Secondary hemorrhage due to and following legally induced termination of pregnancy	Condition	SNOMED
3045823	Multiple pregnancy	Observation	LOINC
4190767	Referral for termination of pregnancy	Observation	SNOMED
46287029	Pertussis vaccination in pregnancy given by other healthcare provider	Observation	SNOMED
4060295	Infections of kidney in pregnancy	Condition	SNOMED
80778	Genitourinary tract infection in pregnancy - delivered with postnatal complication	Condition	SNOMED
435604	Viral disease in mother complicating pregnancy, childbirth AND/OR puerperium	Condition	SNOMED
3014064	Pregnancy associated plasma protein A [Multiple of the median] in Serum or Plasma	Measurement	LOINC
3012266	Gestational age	Measurement	LOINC
44802464	Pregnancy advice for patients with epilepsy	Procedure	SNOMED
3004147	Gestational age estimated from Crown rump length on US	Measurement	LOINC
45757122	Premature rupture of membranes in full term pregnancy with onset of labor within 24 hours of rupture	Condition	SNOMED
4164219	Cardiac disease in pregnancy	Condition	SNOMED
4143296	Hemorrhoids in pregnancy	Condition	SNOMED
4197324	Pre-pregnancy folic acid education	Procedure	SNOMED
42536434	Seen in early pregnancy assessment unit	Observation	SNOMED
37017058	Supervision of high risk pregnancy with poor reproductive history	Observation	SNOMED
80156	Asymptomatic bacteriuria in pregnancy	Condition	SNOMED
439403	Abnormal glucose tolerance test during pregnancy - baby not yet delivered	Condition	SNOMED
4007285	Abnormal pregnancy	Condition	SNOMED

Concept ID	Name	Domain	Vocabulary
37396531	Acute fatty liver of pregnancy	Condition	SNOMED
3035266	Age at first pregnancy	Observation	LOINC
4063458	Alopecia of pregnancy	Condition	SNOMED
4061432	Antenatal care: 10 years plus since last pregnancy	Observation	SNOMED
4152444	Antenatal ultrasound confirms ectopic pregnancy	Condition	SNOMED
42535716	Asthma in pregnancy	Condition	SNOMED
76761	Asymptomatic bacteriuria in pregnancy - delivered	Condition	SNOMED
73526	Asymptomatic bacteriuria in pregnancy - delivered with postnatal complication	Condition	SNOMED
72381	Asymptomatic bacteriuria in pregnancy - not delivered	Condition	SNOMED
44809647	At risk of unwanted pregnancy	Condition	SNOMED
320456	Benign essential hypertension complicating pregnancy, childbirth and the puerperium - delivered with postnatal complication	Condition	SNOMED
4133981	Benign gestational thrombocytopenia	Condition	SNOMED
45757104	Blood coagulation disorder complicating pregnancy	Condition	SNOMED
444052	Bone AND/OR joint disorder of back in mother complicating pregnancy, childbirth AND/OR puerperium	Condition	SNOMED
196172	Bone AND/OR joint disorder of lower extremities in mother complicating pregnancy, childbirth AND/OR puerperium	Condition	SNOMED
443903	Bone AND/OR joint disorder of pelvis in mother complicating pregnancy, childbirth AND/OR puerperium	Condition	SNOMED
73282	Breast engorgement in pregnancy, the puerperium or lactation with antenatal complication	Condition	SNOMED
42534937	Candidiasis of vagina in pregnancy	Condition	SNOMED
45770288	Centile age adjustment for gestational age	Observation	SNOMED
4194693	Cervical pregnancy	Condition	SNOMED
21492990	Choriogonadotropin (pregnancy test) [Presence] in Urine by Rapid immunoassay	Measurement	LOINC
4321446	Cicatrix of cervix affecting pregnancy	Condition	SNOMED
441911	Complete legal termination of pregnancy with damage to pelvic organs or tissues	Condition	SNOMED
196754	Complete legal termination of pregnancy with genital tract or pelvic infection	Condition	SNOMED
435017	Complete legal termination of pregnancy with metabolic disorder	Condition	SNOMED
37110279	Concern about body image related to pregnancy	Condition	SNOMED
314426	Congenital cardiovascular disorder during pregnancy - baby not yet delivered	Condition	SNOMED
4227267	Cornual pregnancy	Condition	SNOMED
434715	Cracked nipple in pregnancy, the puerperium or lactation - delivered	Condition	SNOMED
45757115	Cyst of ovary in pregnancy	Condition	SNOMED
4028636	Cystocele in pregnancy, childbirth and the puerperium	Condition	SNOMED

Concept ID	Name	Domain	Vocabulary
4088928	Date symptom of pregnancy first noted	Observation	SNOMED
192691	Diabetes mellitus during pregnancy - baby not yet delivered	Condition	SNOMED
44791005	Diabetes: shared care in pregnancy - diabetologist and obstetrician	Observation	SNOMED
44810000	Did not attend pertussis vaccination in pregnancy	Observation	SNOMED
44806783	Discussion about ectopic pregnancy risk	Procedure	SNOMED
42537770	Disorder of vein following ectopic pregnancy	Condition	SNOMED
37017322	Dizygotic twin pregnancy	Condition	SNOMED
442913	Drug dependence during pregnancy - baby not yet delivered	Condition	SNOMED
443250	Drug dependence during pregnancy, childbirth and the puerperium	Condition	SNOMED
4095946	Duration of pregnancy at time of previous miscarriage	Observation	SNOMED
37209653	Ectopic pregnancy of left ovary	Condition	SNOMED
37209654	Ectopic pregnancy of right ovary	Condition	SNOMED
45770351	Education about thyroid disease in pregnancy	Procedure	SNOMED
42539701	Embolism due to and following ectopic pregnancy	Condition	SNOMED
4176268	Failed attempted termination of pregnancy with pelvic peritonitis	Condition	SNOMED
4266703	Fibrosis of perineum affecting pregnancy	Condition	SNOMED
4104311	Finding of gestational sac	Condition	SNOMED
442981	Finding of shape of pregnant abdomen	Condition	SNOMED
4311613	Five previous induced terminations of pregnancy	Condition	SNOMED
4295741	Four previous induced terminations of pregnancy	Condition	SNOMED
443328	Galactorrhea in pregnancy and the puerperium with postnatal complication	Condition	SNOMED
42535063	Gastroesophageal reflux disease in pregnancy	Condition	SNOMED
4080941	Generalized pustular psoriasis of pregnancy	Condition	SNOMED
4114526	Gestational	Observation	SNOMED
4104312	Gestational sac absent	Condition	SNOMED
4115863	Gestational sac present	Condition	SNOMED
4091535	Girth of pregnant abdomen	Measurement	SNOMED
43021968	History of gestational diabetes mellitus	Observation	SNOMED
2721459	Home management of gestational hypertension, includes administrative services, professional pharmacy services, care coordination and all necessary supplies and equipment (drugs and nursing visits coded separately); per diem (do not use this code with a...	Visit	HCPCS
441915	Illegal termination of pregnancy without complication	Condition	SNOMED
439325	Illegal termination of pregnancy, incomplete	Condition	SNOMED
4030415	Illegitimate pregnancy	Observation	SNOMED
434705	Incomplete legal termination of pregnancy with damage to pelvic organs or tissues	Condition	SNOMED

Concept ID	Name	Domain	Vocabulary
43530928	Induced termination of pregnancy complicated by acute renal failure with oliguria	Condition	SNOMED
43531711	Induced termination of pregnancy complicated by shock	Condition	SNOMED
4066151	Intraperitoneal pregnancy	Condition	SNOMED
432690	Maternal malaria during pregnancy - baby delivered	Condition	SNOMED
435324	Maternal malaria during pregnancy - baby not yet delivered	Condition	SNOMED
443016	Maternal malaria during pregnancy, childbirth and the puerperium	Condition	SNOMED
40487510	Medical termination of pregnancy using prostaglandin	Procedure	SNOMED
4344627	Multifetal pregnancy reduction	Procedure	SNOMED
4239091	Necrosis of liver of pregnancy	Condition	SNOMED
4060250	No antenatal care: not known pregnant	Observation	SNOMED
45763683	No history of ectopic pregnancy	Observation	SNOMED
4258684	No previous induced termination of pregnancy	Condition	SNOMED
4091280	Number of induced terminations of pregnancy	Observation	SNOMED
4090055	Number of miscarriages or induced terminations of pregnancy	Observation	SNOMED
2784554	Obstetrics, Pregnancy, Abortion	Procedure	ICD10PCS
4058253	Orthopedic disorders during pregnancy, childbirth and the puerperium	Condition	SNOMED
4028635	Pendulous abdomen in pregnancy, childbirth and/or the puerperium	Condition	SNOMED
197042	Peripheral neuritis in pregnancy - delivered	Condition	SNOMED
4062566	Peripheral neuritis in pregnancy with postnatal complication	Condition	SNOMED
46284892	Pertussis vaccination in pregnancy contraindicated	Observation	SNOMED
606771	Phlebitis in pregnancy	Condition	SNOMED
606774	Phlegmasia alba dolens in pregnancy	Condition	SNOMED
4028624	Polyp of cervix in pregnancy, childbirth and the puerperium	Condition	SNOMED
3043524	Pregnancy associated plasma protein A multiple of the median [Percentile]	Measurement	LOINC
4033654	Pregnancy eruption	Condition	SNOMED
37209655	Pregnancy of left fallopian tube	Condition	SNOMED
37209658	Pregnancy of right fallopian tube	Condition	SNOMED
4064794	Pregnancy prescription exemption education	Procedure	SNOMED
4302270	Pregnancy termination care	Observation	SNOMED
44788745	Pregnancy test kit given	Observation	SNOMED
4061784	Pregnant - on abdominal palpation	Condition	SNOMED
4059983	Pregnant - vaginal examination confirms	Condition	SNOMED
4088581	Pregnant abdomen finding	Condition	SNOMED
4061676	Pregnant, diaphragm failure	Observation	SNOMED
4082505	Pruritic folliculitis of pregnancy	Condition	SNOMED
45757171	Psychological abuse complicating pregnancy	Condition	SNOMED

Concept ID	Name	Domain	Vocabulary
42535233	Pyelonephritis in pregnancy	Condition	SNOMED
37019032	Pyogenic granuloma of gingiva co-occurrent and due to pregnancy	Condition	SNOMED
4097378	Rectocele affecting pregnancy	Condition	SNOMED
2004788	Removal of extratubal ectopic pregnancy	Procedure	ICD9Proc
433276	Retracted nipple in pregnancy, the puerperium or lactation - delivered with postnatal complication	Condition	SNOMED
441365	Retracted nipple in pregnancy, the puerperium or lactation with antenatal complication	Condition	SNOMED
45757160	Retraction of nipple in pregnancy	Condition	SNOMED
46269744	Rhinitis of pregnancy	Condition	SNOMED
4127693	Rounded pregnant abdomen	Condition	SNOMED
46269806	Salpingitis due to ectopic pregnancy	Condition	SNOMED
4113995	Salpingo-oophoritis following abortive pregnancy	Condition	SNOMED
45757144	Salpingo-oophoritis in pregnancy	Condition	SNOMED
4150823	Self-induced termination of pregnancy	Condition	SNOMED
4028620	Septate vagina in pregnancy, childbirth and the puerperium	Condition	SNOMED
45765727	Supervision of high risk pregnancy with history of previous cesarean section	Observation	SNOMED
4127705	Term pregnancy delivered	Condition	SNOMED
43530946	Termination of pregnancy without complication	Condition	SNOMED
4062906	Transient hypertension of pregnancy with postnatal complication	Condition	SNOMED
437931	Triplet pregnancy with antenatal problem	Condition	SNOMED
433812	Tuberculosis in pregnancy, childbirth and the puerperium - delivered	Condition	SNOMED
4145940	Tumor of uterine body in pregnancy, childbirth and the puerperium	Condition	SNOMED
4028913	Ultrasound procedure on female genital system AND/OR pregnancy related structure	Procedure	SNOMED
4028642	Undiagnosed pregnancy	Condition	SNOMED
4129016	Uterine fibroids in pregnancy, childbirth and the puerperium	Condition	SNOMED
37109810	Vaginal bleeding complicating early pregnancy	Condition	SNOMED
602838	Varicose veins of perineum in pregnancy	Condition	SNOMED
197339	Congenital abnormality of uterus, affecting pregnancy	Condition	SNOMED
441919	Twin pregnancy	Condition	SNOMED
4214930	Refer to early pregnancy unit	Observation	SNOMED
313543	Venous complication of pregnancy and/or puerperium	Condition	SNOMED
3018954	Choriogonadotropin (pregnancy test) [Presence] in Urine	Measurement	LOINC
45765734	Supervision of high risk pregnancy for primigravida age 15 years or younger	Observation	SNOMED
435018	Twin pregnancy - delivered	Condition	SNOMED
4250175	Pregnancy detection examination	Procedure	SNOMED

Concept ID	Name	Domain	Vocabulary
432968	Anemia during pregnancy - baby delivered	Condition	SNOMED
4066111	Cesarean section - pregnancy at term	Condition	SNOMED
35625971	Vomiting during third trimester of pregnancy	Condition	SNOMED
44789950	Gestational age	Observation	SNOMED
4088927	Pregnancy, childbirth and puerperium finding	Condition	SNOMED
321080	Hypertension complicating pregnancy, childbirth and the puerperium	Condition	SNOMED
4058246	Iron deficiency anemia of pregnancy	Condition	SNOMED
45768717	Incomplete induced termination of pregnancy	Condition	SNOMED
193525	Abdominal pregnancy	Condition	SNOMED
74415	Genitourinary tract infection in pregnancy - not delivered	Condition	SNOMED
4146482	Urinary tract infection in pregnancy	Condition	SNOMED
139895	Thyroid dysfunction during pregnancy, childbirth and the puerperium	Condition	SNOMED
43021951	Monochorionic diamniotic twin pregnancy	Condition	SNOMED
37110290	Disease of circulatory system complicating pregnancy childbirth and puerperium	Condition	SNOMED
441073	Complete legal termination of pregnancy	Condition	SNOMED
42872398	Maternal obesity complicating pregnancy, childbirth and the puerperium, antepartum	Condition	SNOMED
4121646	Varicose veins of genitalia in pregnancy	Condition	SNOMED
4034086	Viable pregnancy	Condition	SNOMED
4170118	Cystitis of pregnancy	Condition	SNOMED
4058113	Gestational edema with proteinuria	Condition	SNOMED
193532	Incomplete legal termination of pregnancy with genital tract or pelvic infection	Condition	SNOMED
432969	Multiple pregnancy	Condition	SNOMED
44802468	Pregnancy advice for patients with epilepsy not indicated	Observation	SNOMED
440158	Venereal disease in mother complicating pregnancy, childbirth AND/OR puerperium	Condition	SNOMED
440464	Complication following molar AND/OR ectopic pregnancy	Condition	SNOMED
607211	Excessive weight gain during pregnancy	Observation	SNOMED
192380	Failed attempted termination of pregnancy without complication	Condition	SNOMED
43531007	Pre-existing diabetes mellitus in pregnancy	Condition	SNOMED
4063037	Viral hepatitis complicating pregnancy, childbirth and the puerperium	Condition	SNOMED
4162865	Abnormal glucose tolerance test during pregnancy, childbirth and the puerperium	Condition	SNOMED
435868	Incomplete legal termination of pregnancy	Condition	SNOMED
4270154	Pregnancy observable	Observation	SNOMED
4061785	Pregnancy unplanned and unwanted	Condition	SNOMED
4061471	Hemorrhoids in pregnancy and puerperium	Condition	SNOMED

Concept ID	Name	Domain	Vocabulary
37392366	High sensitivity urine pregnancy test	Measurement	SNOMED
4113508	Failed medical termination of pregnancy, complicated by delayed or excessive hemorrhage	Condition	SNOMED
42538946	Hypertension complicating pregnancy	Condition	SNOMED
4064725	Malnutrition in pregnancy	Condition	SNOMED
198212	Spotting per vagina in pregnancy	Condition	SNOMED
43530907	Termination of pregnancy with complication	Condition	SNOMED
197031	Intrauterine pregnancy	Condition	SNOMED
4146460	Superficial thrombophlebitis in pregnancy	Condition	SNOMED
432967	Anemia during pregnancy - baby not yet delivered	Condition	SNOMED
432373	Continuing pregnancy after intrauterine death of one or more fetuses	Condition	SNOMED
441349	Legal termination of pregnancy with complication	Condition	SNOMED
42536561	Genital tract infection due to and following ectopic pregnancy	Condition	SNOMED
81357	Genitourinary tract infection in pregnancy - delivered	Condition	SNOMED
43531419	Short cervical length in pregnancy	Condition	SNOMED
439883	Legal termination of pregnancy complicated by genital-pelvic infection	Condition	SNOMED
4060542	Multiple pregnancy with malpresentation	Condition	SNOMED
4118054	Fetal or neonatal effect of maternal complication of pregnancy	Condition	SNOMED
137613	Eclampsia in pregnancy	Condition	SNOMED
4113510	Failed medical termination of pregnancy, without complication	Condition	SNOMED
36713391	Fear of complications of pregnancy	Condition	SNOMED
4200256	Insufficient weight gain of pregnancy	Condition	SNOMED
44809910	Pertussis vaccination in pregnancy invitation	Observation	SNOMED
434097	Twin pregnancy with antenatal problem	Condition	SNOMED
4041876	Urine pregnancy test requested	Observation	SNOMED
194700	Diabetes mellitus in mother complicating pregnancy, childbirth AND/OR puerperium	Condition	SNOMED
40493226	Excision of fallopian tube and surgical removal of ectopic pregnancy	Procedure	SNOMED
36712669	IVF - in-vitro fertilization pregnancy	Condition	SNOMED
606330	Varicose veins of bilateral lower limbs in pregnancy	Condition	SNOMED
4302252	High risk pregnancy care	Observation	SNOMED
432375	Triplet pregnancy	Condition	SNOMED
314103	Benign essential hypertension complicating pregnancy, childbirth and the puerperium - delivered	Condition	SNOMED
4027371	Bone AND/OR joint disorder in mother complicating pregnancy, childbirth AND/OR puerperium	Condition	SNOMED
433826	Continuing pregnancy after abortion of one fetus or more	Condition	SNOMED
4066260	Galactorrhea in pregnancy and the puerperium	Condition	SNOMED

Concept ID	Name	Domain	Vocabulary
36715505	Sequela of complication of pregnancy childbirth and the puerperium	Condition	SNOMED
200455	Failed attempted termination of pregnancy complicated by delayed and/or excessive hemorrhage	Condition	SNOMED
4141779	FH: Multiple pregnancy	Observation	SNOMED
196177	Genital tract AND/OR pelvic infection following molar AND/OR ectopic pregnancy	Condition	SNOMED
4041878	Urine pregnancy test positive	Condition	SNOMED
44805020	Incomplete termination of pregnancy	Condition	SNOMED
4129834	Pregnancy with isoimmunization	Condition	SNOMED
4061848	Retained intrauterine contraceptive device in pregnancy	Condition	SNOMED
4028644	Biochemical pregnancy	Condition	SNOMED
440787	Drug dependence in mother complicating pregnancy, childbirth AND/OR puerperium	Condition	SNOMED
4171359	Failed attempted termination of pregnancy	Condition	SNOMED
441922	Transient hypertension of pregnancy	Condition	SNOMED
21494044	US for pregnancy in first trimester	Procedure	LOINC
4232703	Measurement of pregnancy associated plasma protein A concentration	Measurement	SNOMED
4095475	Pregnancy test finding	Condition	SNOMED
4061425	Antenatal care of second pregnancy	Observation	SNOMED
443330	Breast engorgement in pregnancy, the puerperium or lactation	Condition	SNOMED
4059981	Pregnant - urine test confirms	Condition	SNOMED
192679	Renal disease in pregnancy AND/OR puerperium without hypertension	Condition	SNOMED
137940	Transient hypertension of pregnancy - delivered	Condition	SNOMED
194694	Failed attempted termination of pregnancy with complication	Condition	SNOMED
3048230	Gestational age in weeks	Measurement	LOINC
4270068	Laparoscopic treatment of ectopic pregnancy with salpingectomy	Procedure	SNOMED
4059984	Pregnant - planned	Condition	SNOMED
40480278	High risk pregnancy due to history of preterm labor	Condition	SNOMED
442920	Mental disorder during pregnancy - baby delivered	Condition	SNOMED
4052190	Wife pregnant	Observation	SNOMED
4176015	Hypothyroidism in pregnancy	Condition	SNOMED
4079751	Iron supplement in pregnancy	Procedure	SNOMED
43531008	Pre-existing type 1 diabetes mellitus in pregnancy	Condition	SNOMED
4340944	Cholestasis of pregnancy	Condition	SNOMED
4215253	Diabetic pre-pregnancy education	Procedure	SNOMED
36715485	Injury complicating pregnancy	Condition	SNOMED
44809835	Pertussis vaccination programme for pregnant women enhanced service administration	Observation	SNOMED

Concept ID	Name	Domain	Vocabulary
43531010	Pre-existing type 2 diabetes mellitus in pregnancy	Condition	SNOMED
4062357	Pregnancy smoking education	Procedure	SNOMED
4061686	Pregnant - on history	Condition	SNOMED
4082504	Pruritus of pregnancy	Condition	SNOMED
37206227	Secondary hemorrhage due to and following induced termination of pregnancy	Condition	SNOMED
37016152	Group B streptococcus carrier complicating pregnancy	Condition	SNOMED
442082	Multiple pregnancy with malpresentation - delivered	Condition	SNOMED
4057979	Pre-existing secondary hypertension complicating pregnancy, childbirth and puerperium	Condition	SNOMED
4228043	Pruritic urticarial papules and plaques of pregnancy	Condition	SNOMED
4145939	Congenital abnormality of uterus in pregnancy, childbirth and the puerperium	Condition	SNOMED
443243	Cracked nipple in pregnancy, the puerperium or lactation	Condition	SNOMED
443012	Diabetes mellitus during pregnancy - baby delivered	Condition	SNOMED
196170	Liver disorder in pregnancy - delivered	Condition	SNOMED
37398200	Serum beta human chorionic gonadotrophin pregnancy test	Measurement	SNOMED
4152445	Antenatal ultrasound confirms intrauterine pregnancy	Condition	SNOMED
442565	Gestational edema without hypertension	Condition	SNOMED
4062552	Pre-existing hypertensive heart disease complicating pregnancy, childbirth and the puerperium	Condition	SNOMED
4082412	Hysterotomy and termination of pregnancy	Procedure	SNOMED
43530910	Induced termination of pregnancy complicated by genital-pelvic infection	Condition	SNOMED
4060237	Pregnancy unplanned but wanted	Condition	SNOMED
4061411	Pregnant, sheath failure	Observation	SNOMED
36715504	Protozoal disease complicating pregnancy childbirth and the puerperium	Condition	SNOMED
4125593	Finding of pattern of pregnancy	Condition	SNOMED
4144408	Past pregnancy outcome	Observation	SNOMED
4094910	Pregnancy test positive	Condition	SNOMED
4061351	Spinal and epidural anesthesia-induced headache during pregnancy	Condition	SNOMED
21493938	US.doppler for pregnancy	Procedure	LOINC
4061787	Antenatal care of third pregnancy	Observation	SNOMED
40482050	Chemical pregnancy	Condition	SNOMED
4180111	Third trimester pregnancy less than 36 weeks	Condition	SNOMED
37110282	Venous complication of pregnancy	Condition	SNOMED
197043	Cervical observation during pregnancy and labor	Condition	SNOMED
4113187	Failed medical termination of pregnancy, complicated by genital tract or pelvic infection	Condition	SNOMED

Concept ID	Name	Domain	Vocabulary
36715476	Human immunodeficiency virus complicating pregnancy childbirth and the puerperium	Condition	SNOMED
4060312	Infections of urethra in pregnancy	Condition	SNOMED
443291	Maternal syphilis during pregnancy, childbirth and the puerperium	Condition	SNOMED
37017056	Supervision of high risk pregnancy with poor obstetric history	Observation	SNOMED
3050401	Trisomy 21 risk based on maternal age+Pregnancy associated plasma protein A+Choriogonadotropin+Nuchal translucency [Likelihood] in Fetus	Measurement	LOINC
321638	Benign essential hypertension complicating pregnancy, childbirth and the puerperium	Condition	SNOMED
4206395	Dietary education for gestational diabetes	Procedure	SNOMED
4064791	Pregnancy exercise education	Procedure	SNOMED
440481	Cracked nipple in pregnancy, the puerperium or lactation - delivered with postnatal complication	Condition	SNOMED
4118802	Delivered by cesarean section - pregnancy at term	Condition	SNOMED
45757050	Epilepsy in mother complicating pregnancy	Condition	SNOMED
194699	Liver disorder in pregnancy	Condition	SNOMED
4047842	Pregnancy review	Observation	SNOMED
4094448	Pregnancy test negative	Condition	SNOMED
36715467	Congenital anomaly of mother complicating pregnancy	Condition	SNOMED
316478	Heart disease in mother complicating pregnancy, childbirth AND/OR puerperium	Condition	SNOMED
45757788	Postpartum pregnancy-induced hypertension	Condition	SNOMED
4063126	Pregnancy diet education	Procedure	SNOMED
4094911	Pregnancy test equivocal	Condition	SNOMED
4149783	Abdominal pain in pregnancy	Condition	SNOMED
4119975	Operative termination of pregnancy	Procedure	SNOMED
200153	Ovarian pregnancy	Condition	SNOMED
4189561	Therapeutic termination of pregnancy by aspiration curettage	Procedure	SNOMED
4129027	Viable fetus in abdominal pregnancy	Condition	SNOMED
4299609	Therapeutic termination of pregnancy procedure	Procedure	SNOMED
4060559	Uterine scar from previous surgery in pregnancy, childbirth and the puerperium with antenatal problem	Condition	SNOMED
44809922	Pertussis vaccination in pregnancy invitation first letter	Observation	SNOMED
4062671	Pregnancy-induced edema and proteinuria without hypertension	Condition	SNOMED
4041161	Serum pregnancy test (B-HCG)	Procedure	SNOMED
314423	Benign essential hypertension complicating pregnancy, childbirth and the puerperium - not delivered	Condition	SNOMED
4199274	Finding of position of pregnancy	Condition	SNOMED
434416	Maternal tuberculosis during pregnancy, childbirth and the puerperium	Condition	SNOMED
4034094	Moderate proteinuric hypertension of pregnancy	Condition	SNOMED

Concept ID	Name	Domain	Vocabulary
43021952	Monochorionic monoamniotic twin pregnancy	Condition	SNOMED
4063127	Pregnancy alcohol education	Procedure	SNOMED
4084768	Uncertain viability of pregnancy	Condition	SNOMED
4059985	Unplanned pregnancy unknown if child is wanted	Condition	SNOMED

*List of codes created based on a search of codes in ATLAS using key words related to pregnancy.

Table S2. List of infant related codes for data search.

Concept ID	Name	Domain	Vocabulary
3016704	1 minute Apgar Score	Measurement	LOINC
3016162	10 minute Apgar Score	Measurement	LOINC
3004221	5 minute Apgar Score	Measurement	LOINC
314478	Abnormal fetal heart beat first noted during labor AND/OR delivery in liveborn infant	Condition	SNOMED
137974	Abnormal fetal heart beat noted before labor in liveborn infant	Condition	SNOMED
314099	Abnormal fetal heart rate	Condition	SNOMED
4032626	Abnormal fetal presentation	Condition	SNOMED
43021739	Abnormality of fetal heart	Condition	SNOMED
4014293	Ambulance birth	Condition	SNOMED
4206007	Anemia following fetal blood loss	Condition	SNOMED
36713463	Antenatal screening for fetal growth retardation	Measurement	SNOMED
36713462	Antenatal screening for malformation	Measurement	SNOMED
4129846	Antepartum fetal death	Observation	SNOMED
4014304	Apgar at 1 minute	Observation	SNOMED
4014468	Apgar at 1 minute = 0	Condition	SNOMED
4015430	Apgar at 1 minute = 1	Condition	SNOMED
4014470	Apgar at 1 minute = 10	Condition	SNOMED
4015289	Apgar at 1 minute = 2	Condition	SNOMED
4014305	Apgar at 1 minute = 3	Condition	SNOMED
4015291	Apgar at 1 minute = 4	Condition	SNOMED
4014469	Apgar at 1 minute = 5	Condition	SNOMED
4014306	Apgar at 1 minute = 6	Condition	SNOMED
4015292	Apgar at 1 minute = 7	Condition	SNOMED
4016052	Apgar at 1 minute = 8	Condition	SNOMED
4015431	Apgar at 1 minute = 9	Condition	SNOMED
4014472	Apgar at 10 minutes	Observation	SNOMED
4014473	Apgar at 10 minutes = 0	Condition	SNOMED
4016471	Apgar at 10 minutes = 10	Condition	SNOMED
4014578	Apgar at 10 minutes = 5	Condition	SNOMED

Concept ID	Name	Domain	Vocabulary
4016058	Apgar at 10 minutes = 9	Condition	SNOMED
4016464	Apgar at 5 minutes	Observation	SNOMED
4016467	Apgar at 5 minutes = 10	Condition	SNOMED
4014310	Apgar at 5 minutes = 6	Condition	SNOMED
4016056	Apgar at 5 minutes = 7	Condition	SNOMED
4016465	Apgar at 5 minutes = 8	Condition	SNOMED
4016466	Apgar at 5 minutes = 9	Condition	SNOMED
4167905	Apgar normal	Condition	SNOMED
4274664	Apgar score 9	Condition	SNOMED
4074585	Application of fetal scalp electrode	Procedure	SNOMED
4119774	Arterial malformation	Condition	SNOMED
4075418	Arteriovenous malformation	Observation	SNOMED
4318411	Arteriovenous malformation of kidney	Condition	SNOMED
45766216	Arteriovenous malformation of limb	Condition	SNOMED
4005616	Arteriovenous malformation of precerebral vessels	Condition	SNOMED
4150397	Baby birth weight 1 to 1.5 kilogram	Condition	SNOMED
4147717	Baby birth weight 1.5 to 2 kilogram	Condition	SNOMED
4149455	Baby birth weight 2 to 2.5 kilogram	Condition	SNOMED
4015276	Baby birth weight 4400-4499g	Condition	SNOMED
438258	Baby birth weight above 2.5kg	Condition	SNOMED
4014460	Baby birth weight equal to 10 percent to 24 percent (2850-3149g)	Condition	SNOMED
4015427	Baby birth weight equal to 25 percent to 49 percent (3150-3449g)	Condition	SNOMED
4015426	Baby birth weight equal to 3 percent to 9 percent (2500-2849g)	Condition	SNOMED
4014461	Baby birth weight equal to 50 percent to 74 percent (3450-3749g)	Condition	SNOMED
4014301	Baby birth weight equal to 75 percent to 89 percent (3750-4049g)	Condition	SNOMED
4150396	Baby birth weight equal to 751g-1kg	Condition	SNOMED
4014302	Baby birth weight equal to 90 percent to 96 percent (4050-4399g)	Condition	SNOMED
4014303	Baby birth weight equal to or greater than 96 percent (over 4499g)	Condition	SNOMED
4151169	Baby birth weight less than 751gm	Condition	SNOMED
4015425	Baby birth weight under 3 percent (under 2500g)	Condition	SNOMED
4078138	Baby normal at birth	Condition	SNOMED
4216316	Birth	Condition	SNOMED
4184996	Birth asphyxia	Condition	SNOMED
4014291	Birth detail	Observation	SNOMED
4182460	Birth details - finding	Condition	SNOMED
4016064	Birth details not known	Condition	SNOMED
4071596	Birth dislocation of the shoulder	Condition	SNOMED

Concept ID	Name	Domain	Vocabulary
4014612	Birth examination abnormal - for observation	Condition	SNOMED
4016487	Birth examination abnormal - referred	Condition	SNOMED
4173795	Birth fracture of radius	Condition	SNOMED
4170447	Birth fracture of ulna	Condition	SNOMED
4015277	Birth head circumference	Measurement	SNOMED
4015279	Birth head circumference equal to 25th-49th centile	Observation	SNOMED
4015280	Birth head circumference equal to 50th-74th centile	Observation	SNOMED
4214587	Birth history	Observation	SNOMED
4048141	Birth injury to face	Condition	SNOMED
4047862	Birth injury to phrenic nerve	Condition	SNOMED
4290881	Birth injury to scalp	Condition	SNOMED
4014465	Birth length	Observation	SNOMED
4015287	Birth length equal to 50th-74th centile	Condition	SNOMED
4015270	Birth of child	Condition	SNOMED
4264823	Birth outcome	Observation	SNOMED
40760833	Birth plurality of Pregnancy	Measurement	LOINC
45770412	Birth to 25 education, health and care plan-based care started	Observation	SNOMED
435641	Birth trauma	Condition	SNOMED
4264825	Birth weight	Measurement	SNOMED
4149609	Birth weight 1000-2499 g	Condition	SNOMED
4149610	Birth weight 999 g or less	Condition	SNOMED
4187520	Birth weight finding	Condition	SNOMED
35810139	Birth weight known	Observation	UK Biobank
46271901	Birth weight less than 500g	Condition	SNOMED
35810314	Birth weight of first child	Observation	UK Biobank
4260748	Bottle fed at birth	Observation	SNOMED
4070527	Brachial plexus palsy due to birth trauma	Condition	SNOMED
4260749	Breast fed at birth	Observation	SNOMED
46235338	Cell-free DNA.fetal/Cell-free DNA.total in Plasma cell-free DNA by Dosage of chromosome-specific cfDNA	Measurement	LOINC
375016	Central nervous system malformation in fetus affecting obstetrical care	Condition	SNOMED
4047852	Cephalhematoma due to birth trauma	Condition	SNOMED
4121803	Cerebral arteriovenous malformation	Condition	SNOMED
4048139	Cerebral edema due to birth injury	Condition	SNOMED
4071589	Cerebral hemorrhage due to birth injury	Condition	SNOMED
4347416	Cerebral injury due to birth trauma	Condition	SNOMED
4099094	Chiari malformation	Condition	SNOMED

Concept ID	Name	Domain	Vocabulary
4100706	Chiari malformation type I	Condition	SNOMED
4160043	Chiari malformation type II	Condition	SNOMED
4099847	Chiari malformation type III	Condition	SNOMED
4014738	Child examination - birth	Procedure	SNOMED
4100563	Closed spina bifida with Arnold-Chiari malformation	Condition	SNOMED
4165529	Coloboma, heart malformation, choanal atresia, retardation of growth and development, genital abnormalities, and ear malformations association	Condition	SNOMED
4008239	Condition of fetal membrane	Condition	SNOMED
4119780	Congenital arteriovenous malformation	Condition	SNOMED
4316228	Congenital arteriovenous malformation of gastrointestinal tract	Condition	SNOMED
4006982	Congenital cystic adenomatoid malformation of lung	Condition	SNOMED
4079975	Congenital malformation	Condition	SNOMED
4113437	Congenital malformation of anterior abdominal wall	Condition	SNOMED
4109796	Congenital malformation of clitoris	Condition	SNOMED
4068969	Congenital malformation of corpus callosum	Condition	SNOMED
4166587	Congenital malformation of ear	Condition	SNOMED
76797	Congenital malformation of genital organs	Condition	SNOMED
4342875	Congenital malformation of pancreas	Condition	SNOMED
4109612	Congenital malformation of salivary glands and ducts	Condition	SNOMED
443220	Congenital malformation of sternum	Condition	SNOMED
4173657	Congenital malformation of the respiratory system	Condition	SNOMED
4108905	Congenital malformation of the urinary system	Condition	SNOMED
4191712	Congenital malformation of thumb	Condition	SNOMED
4108878	Congenital malformation of tongue, mouth and pharynx	Condition	SNOMED
4108877	Congenital malformation of trachea and bronchus	Condition	SNOMED
4169977	Congenital malformation of upper alimentary tract	Condition	SNOMED
4109625	Congenital malformation of uterus and cervix	Condition	SNOMED
4100723	Congenital malformation of vitreous humor	Condition	SNOMED
4266662	Congenital malformation syndrome	Condition	SNOMED
4005286	Congenital malformation syndromes affecting facial appearance	Condition	SNOMED
4007583	Congenital malformation syndromes associated with short stature	Condition	SNOMED
4007584	Congenital malformation syndromes involving limbs	Condition	SNOMED
4006101	Congenital scoliosis due to bony malformation	Condition	SNOMED
4267553	Congenital vascular malformation	Condition	SNOMED
4015159	Consultant unit birth	Condition	SNOMED
4123911	Correction of congenital anorectal malformation	Procedure	SNOMED
442529	Cranial nerve injury due to birth trauma	Condition	SNOMED

Concept ID	Name	Domain	Vocabulary
45766188	Cutaneous capillary malformation	Condition	SNOMED
4083587	Date of birth	Observation	SNOMED
198798	Dieulafoy's vascular malformation	Condition	SNOMED
4064698	Disorder relating to long gestation AND/OR high birthweight	Condition	SNOMED
4195545	Disorder relating to short gestation AND/OR low birthweight	Condition	SNOMED
73534	Disproportion between fetal head and pelvis	Condition	SNOMED
2211763	Doppler echocardiography, fetal, pulsed wave and/or continuous wave with spectral display; complete	Procedure	CPT4
46273020	Doppler fetal echocardiography	Procedure	SNOMED
2784328	Drainage of Fetal Blood from Products of Conception, Open Approach	Procedure	ICD10PCS
2784530	Drainage of Fetal Blood from Products of Conception, Percutaneous Approach	Procedure	ICD10PCS
2784536	Drainage of Fetal Blood from Products of Conception, Percutaneous Endoscopic Approach	Procedure	ICD10PCS
2784542	Drainage of Fetal Blood from Products of Conception, Via Natural or Artificial Opening	Procedure	ICD10PCS
2784532	Drainage of Other Fetal Fluid from Products of Conception, Percutaneous Approach	Procedure	ICD10PCS
2784544	Drainage of Other Fetal Fluid from Products of Conception, Via Natural or Artificial Opening	Procedure	ICD10PCS
4090493	Early fetal heart deceleration	Condition	SNOMED
4045955	Embolization of intracranial arteriovenous malformation	Procedure	SNOMED
4009644	Erb-Duchenne palsy as birth trauma	Condition	SNOMED
434758	Exceptionally large at birth	Condition	SNOMED
81636	Excessive fetal growth affecting management of mother	Condition	SNOMED
40486543	Excision of arteriovenous malformation	Procedure	SNOMED
4140123	Excision of congenital arteriovenous malformation	Procedure	SNOMED
4045952	Excision of intracranial arteriovenous malformation	Procedure	SNOMED
4045953	Excision of intracranial venous malformation	Procedure	SNOMED
4173323	Extremely low birth weight infant	Condition	SNOMED
4070533	Eye damage due to birth trauma	Condition	SNOMED
4083524	Face to pubes birth	Condition	SNOMED
377680	Facial nerve injury as birth trauma	Condition	SNOMED
4225278	Facial palsy as birth trauma	Condition	SNOMED
45757775	Failed attempted vaginal birth after previous cesarean section	Condition	SNOMED
46235335	Fetal 1p36 deletion risk [Interpretation] based on Plasma cell-free+WBC DNA by Dosage of chromosome-specific cfDNA Qualitative	Measurement	LOINC
36305055	Fetal 22q11.2 deletion risk [Interpretation] based on Plasma cell-free DNA by Dosage of chromosome-specific cfDNA Qualitative	Measurement	LOINC

Concept ID	Name	Domain	Vocabulary
46235276	Fetal 5p deletion risk [Interpretation] based on Plasma cell-free+WBC DNA by Dosage of chromosome-specific cfDNA Qualitative	Measurement	LOINC
42512758	Fetal adenocarcinoma of upper lobe, lung	Condition	ICDO3
4004785	Fetal alcohol syndrome	Condition	SNOMED
4150653	Fetal anatomy study	Procedure	SNOMED
201083	Fetal AND/OR placental disorder affecting management of mother	Condition	SNOMED
46235270	Fetal Angelman syndrome risk [Interpretation] based on Plasma cell-free+WBC DNA by Dosage of chromosome-specific cfDNA Qualitative	Measurement	LOINC
44513795	Fetal biometry	Procedure	OPCS4
4141511	Fetal biometry using ultrasound	Measurement	SNOMED
4070024	Fetal biophysical profile	Procedure	SNOMED
3050723	Fetal blood [Volume] by Kleihauer-Betke method	Measurement	LOINC
79146	Fetal blood loss	Condition	SNOMED
4215480	Fetal blood loss from cut end of co-twin's cord	Condition	SNOMED
4223239	Fetal blood loss from placenta	Condition	SNOMED
4201300	Fetal blood loss from ruptured cord	Condition	SNOMED
4196670	Fetal blood loss from vasa previa	Condition	SNOMED
4085919	Fetal blood sampling	Procedure	SNOMED
2004810	Fetal blood sampling and biopsy	Procedure	ICD9Proc
314749	Fetal bradycardia	Condition	SNOMED
4016801	Fetal cell screen	Measurement	SNOMED
4307826	Fetal cephalhematoma	Condition	SNOMED
4126076	Fetal cerebral hemorrhage	Condition	SNOMED
43020943	Fetal chromosomal abnormality screening	Measurement	SNOMED
37310907	Fetal cleft lip	Condition	SNOMED
37310908	Fetal clubfoot	Condition	SNOMED
74104	Fetal condition affecting obstetrical care of mother	Condition	SNOMED
3001951	Fetal Crown Rump length US	Measurement	LOINC
4171253	Fetal cytomegalovirus syndrome	Condition	SNOMED
43531705	Fetal damage from maternal infectious disorder	Condition	SNOMED
4079844	Fetal death	Observation	SNOMED
4071603	Fetal death due to prelabor anoxia	Observation	SNOMED
436228	Fetal death due to termination of pregnancy	Observation	SNOMED
75605	Fetal death, affecting management of mother	Condition	SNOMED
4321536	Fetal demise from miscarriage	Condition	SNOMED
4323285	Fetal disorder	Condition	SNOMED

Concept ID	Name	Domain	Vocabulary
600520	Fetal disorder due to and following operative procedure on mother	Condition	SNOMED
605035	Fetal disorder due to chronic maternal cardiovascular disease	Condition	SNOMED
43530894	Fetal disorder due to disease in mother	Condition	SNOMED
600527	Fetal disorder due to maternal disorder of urinary tract	Condition	SNOMED
605059	Fetal disorder due to maternal nutritional disorder	Condition	SNOMED
4178809	Fetal disorder secondary to chemicals	Condition	SNOMED
80165	Fetal disproportion	Condition	SNOMED
4063305	Fetal distress - delivered	Condition	SNOMED
4122059	Fetal distress - in labor	Condition	SNOMED
73268	Fetal distress affecting management of mother	Condition	SNOMED
435369	Fetal distress, in liveborn infant	Condition	SNOMED
4149999	Fetal distress-affecting care	Condition	SNOMED
4172863	Fetal dysrhythmia	Condition	SNOMED
4334808	Fetal echocardiography	Procedure	SNOMED
4327042	Fetal electrocardiogram	Procedure	SNOMED
4188605	Fetal engagement	Condition	SNOMED
46273759	Fetal exposure to alcohol	Observation	SNOMED
46272991	Fetal exposure to drug	Observation	SNOMED
4024705	Fetal finding	Condition	SNOMED
37208955	Fetal gastroschisis	Condition	SNOMED
42534975	Fetal genitourinary abnormality	Condition	SNOMED
4210896	Fetal gestational age	Observation	SNOMED
4171113	Fetal growth abnormality	Condition	SNOMED
80204	Fetal growth restriction	Condition	SNOMED
43022054	Fetal heart failure with myocardial hypertrophy	Condition	SNOMED
4269426	Fetal heart feature	Observation	SNOMED
4091642	Fetal heart finding	Condition	SNOMED
4084439	Fetal heart monitoring	Observation	SNOMED
4084440	Fetal heart monitoring in labor	Observation	SNOMED
4248953	Fetal heart monitoring using fetal scalp electrode	Observation	SNOMED
4091643	Fetal heart rate	Measurement	SNOMED
4089747	Fetal heart rate variability	Measurement	SNOMED
4124317	Fetal heart sounds absent	Observation	SNOMED
4096384	Fetal heart sounds present	Condition	SNOMED
4049246	Fetal hemoglobin determination	Measurement	SNOMED
4209304	Fetal hemorrhage into co-twin	Condition	SNOMED
4289756	Fetal hydantoin syndrome	Condition	SNOMED

Concept ID	Name	Domain	Vocabulary
4127418	Fetal hydrocephalus suspected	Observation	SNOMED
43020782	Fetal hypertrophic cardiomyopathy	Condition	SNOMED
43020784	Fetal hypertrophic cardiomyopathy associated with renal disease	Condition	SNOMED
45766963	Fetal hypertrophic cardiomyopathy due to maternal diabetes mellitus	Condition	SNOMED
43020783	Fetal hypertrophic cardiomyopathy due to twin to twin transfusion syndrome	Condition	SNOMED
4171120	Fetal hypoxia	Condition	SNOMED
437098	Fetal intrauterine distress first noted during labor AND/OR delivery in liveborn infant	Condition	SNOMED
433311	Fetal intrauterine distress noted before labor in liveborn infant	Condition	SNOMED
4146847	Fetal monitoring scalp injury	Condition	SNOMED
36305428	Fetal Monosomy X risk [Interpretation] based on Plasma cell-free DNA by Dosage of chromosome-specific cfDNA Qualitative	Measurement	LOINC
4269424	Fetal movement feature	Observation	SNOMED
4129009	Fetal movements absent	Observation	SNOMED
4145315	Fetal movements felt	Condition	SNOMED
4015303	Fetal movements seen	Condition	SNOMED
3050718	Fetal Neural tube defect risk [Likelihood] Based on maternal age	Measurement	LOINC
3048559	Fetal Nuchal fold [Multiple of the median] Thickness US	Measurement	LOINC
3005033	Fetal Nuchal fold Thickness US	Measurement	LOINC
4153443	Fetal or neonatal effect of abnormal uterine contractions	Condition	SNOMED
438541	Fetal or neonatal effect of abnormality of chorion	Condition	SNOMED
4148097	Fetal or neonatal effect of abruptio placentae	Condition	SNOMED
43530903	Fetal or neonatal effect of breech delivery	Condition	SNOMED
439129	Fetal or neonatal effect of breech delivery and extraction	Condition	SNOMED
4071485	Fetal or neonatal effect of cesarean section	Condition	SNOMED
4070415	Fetal or neonatal effect of chorioamnionitis	Condition	SNOMED
44784311	Fetal or neonatal effect of complication of delivery	Condition	SNOMED
435927	Fetal or neonatal effect of complication of labor and/or delivery	Condition	SNOMED
4225723	Fetal or neonatal effect of complication of placenta, cord and/or membranes	Condition	SNOMED
441682	Fetal or neonatal effect of compression of umbilical cord	Condition	SNOMED
440203	Fetal or neonatal effect of condition of umbilical cord	Condition	SNOMED
4118058	Fetal or neonatal effect of ectopic pregnancy	Condition	SNOMED
4047593	Fetal or neonatal effect of external version before labor	Condition	SNOMED
4047718	Fetal or neonatal effect of forceps delivery	Condition	SNOMED
436218	Fetal or neonatal effect of hallucinogenic agent transmitted via placenta and/or breast milk	Condition	SNOMED

Concept ID	Name	Domain	Vocabulary
435917	Fetal or neonatal effect of incompetent cervix	Condition	SNOMED
4047723	Fetal or neonatal effect of instrumental delivery	Condition	SNOMED
4146728	Fetal or neonatal effect of malposition or disproportion during labor or delivery	Condition	SNOMED
4152171	Fetal or neonatal effect of malpresentation before labor	Condition	SNOMED
76521	Fetal or neonatal effect of malpresentation, malposition and/or disproportion during labor and/or delivery	Condition	SNOMED
437976	Fetal or neonatal effect of maternal anesthesia and/or analgesia	Condition	SNOMED
4070421	Fetal or neonatal effect of maternal anesthetic and analgesic agents during labor and delivery	Condition	SNOMED
4229439	Fetal or neonatal effect of maternal blood loss	Condition	SNOMED
4118054	Fetal or neonatal effect of maternal complication of pregnancy	Condition	SNOMED
4048130	Fetal or neonatal effect of maternal exposure to environmental chemical substances	Condition	SNOMED
4118057	Fetal or neonatal effect of maternal oligohydramnios	Condition	SNOMED
4152170	Fetal or neonatal effect of maternal polyhydramnios	Condition	SNOMED
4118056	Fetal or neonatal effect of maternal premature rupture of membrane	Condition	SNOMED
4213100	Fetal or neonatal effect of maternal transmission of substance	Condition	SNOMED
4005284	Fetal or neonatal effect of maternal use of alcohol	Condition	SNOMED
4048129	Fetal or neonatal effect of maternal use of nutritional chemical substance	Condition	SNOMED
4070519	Fetal or neonatal effect of maternal use of tobacco	Condition	SNOMED
441127	Fetal or neonatal effect of medicinal agent transmitted via placenta and/or breast milk	Condition	SNOMED
440211	Fetal or neonatal effect of morphologic abnormality of placenta	Condition	SNOMED
4118411	Fetal or neonatal effect of multiple pregnancy	Condition	SNOMED
436517	Fetal or neonatal effect of narcotic transmitted via placenta and/or breast milk	Condition	SNOMED
440829	Fetal or neonatal effect of noxious influences transmitted via placenta or breast milk	Condition	SNOMED
4050022	Fetal or neonatal effect of noxious substance transmitted via placenta	Condition	SNOMED
4070410	Fetal or neonatal effect of placenta previa	Condition	SNOMED
4047595	Fetal or neonatal effect of placental damage caused by cesarean section	Condition	SNOMED
4178368	Fetal or neonatal effect of placental disorder	Condition	SNOMED
438562	Fetal or neonatal effect of placental or breast transfer of anticonvulsant	Condition	SNOMED
4047587	Fetal or neonatal effect of placental or breast transfer of narcotics	Condition	SNOMED
201681	Fetal or neonatal effect of placental separation and/or hemorrhage	Condition	SNOMED

Concept ID	Name	Domain	Vocabulary
4071480	Fetal or neonatal effect of placental transfusion syndrome	Condition	SNOMED
4070424	Fetal or neonatal effect of precipitate delivery	Condition	SNOMED
433018	Fetal or neonatal effect of prolapsed cord	Condition	SNOMED
4118413	Fetal or neonatal effect of twin pregnancy	Condition	SNOMED
4070411	Fetal or neonatal effect of twin-to-twin transplacental transfusion	Condition	SNOMED
4047719	Fetal or neonatal effect of vacuum extraction delivery	Condition	SNOMED
432734	Fetal OR neonatal hemorrhage	Condition	SNOMED
36716545	Fetal or neonatal intracerebral non-traumatic hemorrhage	Condition	SNOMED
4230351	Fetal OR neonatal jaundice from infection	Condition	SNOMED
4067525	Fetal OR neonatal jaundice from polycythemia	Condition	SNOMED
4096143	Fetal OR neonatal jaundice from swallowed maternal blood	Condition	SNOMED
36716544	Fetal or neonatal non-traumatic intraventricular hemorrhage	Condition	SNOMED
4189562	Fetal oxytocin stress test	Procedure	SNOMED
46235535	Fetal Prader-Willi syndrome risk [Interpretation] based on Plasma cell-free+WBC DNA by Dosage of chromosome-specific cfDNA Qualitative	Measurement	LOINC
2004825	Fetal pulse oximetry	Procedure	ICD9Proc
4344626	Fetal reduction or destruction procedure	Procedure	SNOMED
46235337	Fetal sex in Cell-free+Maternal WBC DNA by Dosage of chromosome specific circulating cell free (ccf) DNA	Measurement	LOINC
4015304	Fetal size accords with dates	Condition	SNOMED
4014437	Fetal size does not accord with dates	Condition	SNOMED
37204820	Fetal spina bifida	Condition	SNOMED
4129553	Fetal spina bifida suspected	Observation	SNOMED
4079857	Fetal stress	Condition	SNOMED
4207963	Fetal structure	Spec Anatomic Site	SNOMED
317100	Fetal tachycardia	Condition	SNOMED
46235256	Fetal Triploidy risk [Interpretation] based on Plasma cell-free+WBC DNA by Dosage of chromosome-specific cfDNA Qualitative	Measurement	LOINC
43533847	Fetal Trisomy 13 risk [Interpretation] based on Plasma cell-free DNA by Dosage of chromosome-specific cfDNA Qualitative	Measurement	LOINC
46236729	Fetal Trisomy 18 risk [Interpretation] based on Plasma cell-free DNA by Sequencing Qualitative	Measurement	LOINC
3052868	Fetal Trisomy 18 risk [Likelihood] Based on maternal age	Measurement	LOINC
46236728	Fetal Trisomy 21 risk [Interpretation] based on Plasma cell-free DNA by Sequencing Qualitative	Measurement	LOINC
3049207	Fetal Trisomy 21 risk [Likelihood] Based on maternal age	Measurement	LOINC
4066018	Fetal valproate syndrome	Condition	SNOMED
4244067	Fetal warfarin syndrome	Condition	SNOMED
201366	Fetal-maternal hemorrhage	Condition	SNOMED

Concept ID	Name	Domain	Vocabulary
193535	Fetal-maternal hemorrhage - delivered	Condition	SNOMED
196757	Fetal-maternal hemorrhage with antenatal problem	Condition	SNOMED
4074580	Fetoscopy sampling of fetal blood	Procedure	SNOMED
4200201	Finding of birth outcome	Condition	SNOMED
4129010	Finding of fetal heart rate	Condition	SNOMED
4126887	Finding of pattern of fetal movement	Condition	SNOMED
4324549	Forceps delivery with rotation of fetal head	Procedure	SNOMED
4014452	GP unit birth	Condition	SNOMED
44805655	Health visitor new birth visit	Observation	SNOMED
440532	Heavy-for-dates at birth regardless of gestation period	Condition	SNOMED
4098015	Hereditary persistence of fetal hemoglobin	Condition	SNOMED
4079859	High birth weight	Condition	SNOMED
4014451	Home birth	Condition	SNOMED
43530710	Hydrocephalus due to Arnold Chiari malformation type 2	Condition	SNOMED
4173169	Hydrops fetalis	Condition	SNOMED
77662	Hydrops fetalis due to isoimmunization	Condition	SNOMED
4049032	Idiopathic hydrops fetalis	Condition	SNOMED
4136449	Infant feeding at birth	Observation	SNOMED
36716538	Injury of central nervous system due to birth trauma	Condition	SNOMED
381952	Injury to brachial plexus as birth trauma	Condition	SNOMED
36716733	Injury to external genitalia due to birth trauma	Condition	SNOMED
4171121	Intrapartum fetal hypoxia	Condition	SNOMED
4121753	Intrauterine fetal defect correction	Procedure	SNOMED
4048279	Intraventricular hemorrhage due to birth injury	Condition	SNOMED
4347418	Kidney injury due to birth trauma	Condition	SNOMED
4305334	Klumpke-Déjerine paralysis as birth trauma	Condition	SNOMED
4063307	Labor and delivery complicated by biochemical evidence of fetal stress	Condition	SNOMED
443247	Labor and delivery complicated by fetal heart rate anomaly	Condition	SNOMED
4059900	Labor and delivery complicated by fetal heart rate anomaly with meconium in amniotic fluid	Condition	SNOMED
443254	Labor fetal anoxia	Condition	SNOMED
4092355	Late fetal heart deceleration	Condition	SNOMED
4260747	Length of gestation at birth	Observation	SNOMED
72726	Light-for-dates with signs of fetal malnutrition	Condition	SNOMED
74469	Light-for-dates without fetal malnutrition	Condition	SNOMED
4048144	Liveborn with labor fetal distress	Condition	SNOMED
4048143	Liveborn with prelabor fetal distress	Condition	SNOMED

Concept ID	Name	Domain	Vocabulary
4047865	Liver rupture due to birth trauma	Condition	SNOMED
4047731	Local subdural hematoma due to birth trauma	Condition	SNOMED
4171116	Long fetal gestation	Condition	SNOMED
4171115	Low birth weight infant	Condition	SNOMED
4124986	Lymphatic malformation	Condition	SNOMED
42539784	Malformation of central nervous system of fetus	Condition	SNOMED
201121	Malformation of urachus	Condition	SNOMED
4010271	Manual rotation of fetal head	Procedure	SNOMED
4070523	Massive subgaleal hemorrhage due to birth trauma	Condition	SNOMED
444114	Maternal AND/OR fetal condition affecting labor AND/OR delivery	Condition	SNOMED
35810246	Maternal smoking around birth	Observation	UK Biobank
434759	Mild to moderate birth asphyxia	Condition	SNOMED
4149395	Mild to moderate birth asphyxia - apgar score 4-7 at 1 minute	Condition	SNOMED
4344627	Multifetal pregnancy reduction	Procedure	SNOMED
4163851	Multiple birth	Condition	SNOMED
36716289	Multiple birth order	Observation	SNOMED
75015	Multiple gestation with one OR more fetal malpresentations	Condition	SNOMED
4174996	Multiple malformation syndrome with early overgrowth	Condition	SNOMED
4274970	Multiple malformation syndrome with facial defects as major feature	Condition	SNOMED
4190397	Multiple malformation syndrome with limb defect as major feature	Condition	SNOMED
4220676	Multiple system malformation syndrome	Condition	SNOMED
46284810	New birth examination completed by other healthcare provider	Observation	SNOMED
44790222	Non routine obstetric scan for fetal observations	Procedure	SNOMED
76533	Non-immune hydrops fetalis	Condition	SNOMED
46236732	Noninvasive prenatal fetal 13 and 18 and 21 aneuploidy panel - Plasma cell-free DNA by Sequencing	Measurement	LOINC
4305304	Nonreassuring fetal status	Condition	SNOMED
4014720	Normal birth	Condition	SNOMED
4080889	Normal birth weight	Observation	SNOMED
4111608	Normal fetal growth	Observation	SNOMED
4082923	Normal fetal heart baseline pattern	Condition	SNOMED
4193440	Notification of birth	Observation	SNOMED
37396169	Obstructed labor due to fetal abnormality	Condition	SNOMED
194100	Obstructed labor due to fetal malposition	Condition	SNOMED
194711	Obstructed labor due to fetal malposition - delivered	Condition	SNOMED
193831	Obstructed labor due to fetal malposition with antenatal problem	Condition	SNOMED
45772074	Obstructed labor due to incomplete rotation of fetal head	Condition	SNOMED

Concept ID	Name	Domain	Vocabulary
4050588	Operation for pulmonary arteriovenous malformation	Procedure	SNOMED
2004811	Other fetal monitoring	Procedure	ICD9Proc
44513799	Other specified non-routine obstetric scan for fetal observations	Procedure	OPCS4
4014444	P/N care started at birth	Observation	SNOMED
40404007	Parachute malformation of mitral valve	Condition	SNOMED
35810245	Part of a multiple birth	Observation	UK Biobank
44513691	Percutaneous sampling of fetal blood	Procedure	OPCS4
4177366	Percutaneous transluminal ablation of congenital heart malformation	Procedure	SNOMED
45769223	Percutaneous transluminal arterial and venous embolization of arteriovenous malformation	Procedure	SNOMED
44511560	Percutaneous transluminal embolisation of arteriovenous malformation NEC	Procedure	OPCS4
45769210	Percutaneous transluminal venous embolization of arteriovenous malformation	Procedure	SNOMED
4322190	Perinatal jaundice due to fetal OR neonatal hepatitis	Condition	SNOMED
4076794	Peripheral arteriovenous malformation	Condition	SNOMED
197089	Peripheral nerve injury due to birth trauma	Condition	SNOMED
4103235	Phrenic nerve paralysis as birth trauma	Condition	SNOMED
4014292	Place of birth	Observation	SNOMED
4181974	Place of birth - finding	Condition	SNOMED
4276503	Placenta of multiple birth higher than twin	Condition	SNOMED
72693	Poor fetal growth affecting management	Condition	SNOMED
75639	Poor fetal nutrition	Condition	SNOMED
4015415	Postnatal care greater than 48 hours after birth	Observation	SNOMED
4015152	Postnatal care less than 48 hours after birth	Observation	SNOMED
4097427	Premature birth of fraternal twins, both living	Condition	SNOMED
4069200	Premature birth of multiple newborns	Condition	SNOMED
4272248	Premature birth of newborn	Condition	SNOMED
4241228	Premature birth of newborn female	Condition	SNOMED
4107401	Premature birth of newborn triplets	Condition	SNOMED
4029786	Premature birth of newborn twins	Condition	SNOMED
4119302	Pulmonary arteriovenous malformation	Condition	SNOMED
4083719	Questionable if patient date of birth correct	Observation	SNOMED
77619	Reduced fetal movement	Condition	SNOMED
4010436	Registrar-birth/death/marriage	Observation	SNOMED
4089069	Relation of fetal size to dates	Observation	SNOMED
4141806	Repair of arteriovenous malformation of coronary artery	Procedure	SNOMED

Concept ID	Name	Domain	Vocabulary
4012653	Response to memory test for own date of birth	Observation	SNOMED
4069761	Retained fetal tissue	Condition	SNOMED
40487400	Sampling of fetal blood and biopsy of fetus	Procedure	SNOMED
4047853	Scalp bruising due to birth trauma	Condition	SNOMED
4260976	Seen in fetal medicine clinic	Observation	SNOMED
432427	Severe birth asphyxia	Condition	SNOMED
4153452	Severe birth asphyxia - apgar score less than 4 at 1 minute	Condition	SNOMED
4014295	Single live birth	Condition	SNOMED
36713467	Singleton liveborn unspecified as to place of birth	Observation	SNOMED
257370	Skeletal injury due to birth trauma	Condition	SNOMED
46284319	Skull injury due to birth trauma	Condition	SNOMED
4145947	Small for gestational age fetus	Condition	SNOMED
4129783	Spinal arteriovenous malformation	Condition	SNOMED
44784364	Spine injury due to birth trauma	Condition	SNOMED
40482937	Split cord malformation	Condition	SNOMED
4015143	Spontaneous rupture of fetal membranes	Condition	SNOMED
35815060	Status of baby at birth	Observation	UK Biobank
4045956	Stereotactic radiosurgery of intracranial arteriovenous malformation	Procedure	SNOMED
4047867	Sternomastoid injury due to birth injury	Condition	SNOMED
443463	Stillbirth - unknown if fetal death intrapartum or prior to labor	Condition	SNOMED
4173012	Strong fetal movements	Condition	SNOMED
4071066	Subarachnoid hemorrhage due to birth injury	Condition	SNOMED
4047868	Subcutaneous fat necrosis due to birth injury	Condition	SNOMED
378544	Subdural and cerebral hemorrhage due to birth trauma	Condition	SNOMED
4130539	Subdural hemorrhage due to birth trauma	Condition	SNOMED
42538556	Subgaleal epicranial subaponeurotic hemorrhage due to birth injury	Condition	SNOMED
36712695	Suspected fetal abnormality affecting management of mother	Condition	SNOMED
43530882	Suspected fetal anencephaly	Observation	SNOMED
43530886	Suspected fetal damage from disease in the mother	Observation	SNOMED
43530884	Suspected fetal damage from maternal alcohol	Observation	SNOMED
43530891	Suspected fetal damage from maternal drug use	Observation	SNOMED
43530888	Suspected fetal damage from radiation	Observation	SNOMED
43530893	Suspected fetal damage from viral disease in mother	Observation	SNOMED
43530881	Suspected fetal disorder	Observation	SNOMED
4071588	Tentorial tear due to birth trauma	Condition	SNOMED
4227728	Term birth of fraternal twins, both living	Condition	SNOMED
4054968	Term birth of newborn	Condition	SNOMED

Concept ID	Name	Domain	Vocabulary
4246654	Term birth of newborn female	Condition	SNOMED
4306814	Tooth malformation	Condition	SNOMED
4047866	Torticollis due to birth injury	Condition	SNOMED
4094046	Triplet birth	Condition	SNOMED
4101844	Twin birth	Condition	SNOMED
4197294	Ultrasonic doppler for fetal heart sounds	Procedure	SNOMED
2793345	Ultrasonography of Fetal Umbilical Cord	Procedure	ICD10PCS
4237338	Ultrasound scan for fetal anomaly	Procedure	SNOMED
4082672	Ultrasound scan for fetal growth	Procedure	SNOMED
4237339	Ultrasound scan for fetal nuchal translucency	Procedure	SNOMED
4060626	Ultrasound scan for fetal viability	Procedure	SNOMED
2211752	Ultrasound, pregnant uterus, real time with image documentation, fetal and maternal evaluation plus detailed fetal anatomic examination, transabdominal approach; each additional gestation (List separately in addition to code for primary procedure)	Procedure	CPT4
438259	Umbilical hemorrhage after birth	Condition	SNOMED
4111194	Unable to remember own date of birth	Condition	SNOMED
4059697	US scan - fetal maturity	Procedure	SNOMED
37204873	Vascular malformation	Observation	SNOMED
4215653	Vein of Galen malformation	Condition	SNOMED
4119775	Venous malformation	Condition	SNOMED
4079858	Very low birth weight infant	Condition	SNOMED
4071599	Vulval hematoma due to birth trauma	Condition	SNOMED
4071632	Water birth delivery	Procedure	SNOMED
4172833	Weak fetal movements	Condition	SNOMED
433027	Fetal or neonatal effect of maternal hypertensive disease	Condition	SNOMED
436219	Fetal or neonatal effect of maternal infection	Condition	SNOMED
434744	Fetal or neonatal effect of maternal medical problem	Condition	SNOMED
4039090	O/E - fetal heart 100-120	Condition	SNOMED
4038571	O/E - fetal heart 120-160	Condition	SNOMED
4038572	O/E - fetal heart 160-180	Condition	SNOMED
4038945	O/E - fetal heart heard	Condition	SNOMED
4039088	O/E - fetal heart not heard	Observation	SNOMED
4038943	O/E - fetal movement diminished	Condition	SNOMED
4038942	O/E - fetal movements	Condition	SNOMED
4038426	O/E - fetal movements felt	Condition	SNOMED
4039610	O/E - fetal movements seen	Condition	SNOMED
4150948	O/E - fetal presentation	Condition	SNOMED

Concept ID	Name	Domain	Vocabulary
4039609	O/E - no fetal movements	Observation	SNOMED
4038938	O/E -fetal presentation unsure	Condition	SNOMED

*List of codes created based on a search of codes in ATLAS using key words related to foetal and infant health.

ANNEX IV. Detailed results

Table S3. Detailed comments of availability of pregnancy related information in data sources (questionnaire 1).

Country	Data sources	Comments*
Finland	FinOMOP-ACI Varha	There is extensive information on pregnancies, births, and newborns available in the "Ipana" system at our institute. The Ipana database is accessible to us, but due to technical limitations it is not currently available to be used as source data for OMOP. We recently found out that at our site, the Ipana source data is not consistently available at the moment: the comprehensive data spans the time period of 2008-11-28 to 2020-08-25. To map the full Ipana data into OMOP, we would need the data to be brought and modelled into our data warehouse first, and then do the OMOP mapping. It is not clear how long it will take before we have the full data in the data warehouse. Alternatively, we could use the current dataset (years ~2008-2020) as such and map it into OMOP. However, this still seems to be a task that would require plenty of ETL coding resources, and at this point it would have to queue for those resources.
Spain	BIFAP	Own algorithm for mother and child linkage has been developed and could be implemented in some regions.
Spain	IMASIS	Mother and child has been only partially implemented; DP is currently working on it.
Spain	VID	With no information on sex of the child (not able to create a person table)
The Netherlands	IPCI	Please note that the requested information can be derived through algorithms.
The United Kingdom	UKBB	Pregnancy end date is considered sensitive. However, pregnancy outcome and mode of delivery seem to be available. This is available in only 1/3 of pregnancies.

BIFAP=Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público; FinOMOP-ACI Varha= FinOMOP Auria Clinical Informatics; IMASIS= Institut Municipal Assistència Sanitària Information System; VID=Valencia Health System Integrated Dataset; IPCI= Integrated Primary Care Information; UKBB=UK BioBank.

*Detailed comments shared by some data sources.

Table S4. Results from Questionnaire 1.

See excel file ("Table S4-Questionnaire 1").

Table S5. Record counts (RC) and Person Counts (PC) of pregnancy-related concepts explored in the DARWIN EU® Portal.

See excel file ("Table S5-Preg counts").

Table S6. Record counts (RC) and Person Counts (PC) of pregnancy-related concepts explored in the DARWIN EU® Portal, by sex.

See excel file ("Table S6-Preg counts by sex").

Table S7. Record counts (RC) and Person Counts (PC) of pregnancy-related concepts explored in the DARWIN EU® Portal, by age.

See excel file ("Table S7-Preg counts by age")

Table S8. Record counts (RC) and Person Counts (PC) of infant-related concepts explored in the DARWIN EU® Portal.

See excel file ("Table S8-Inf counts")

Table S9. Record counts (RC) and Person Counts (PC) of infant-related concepts explored in the DARWIN EU® Portal, by sex.

See excel file ("Table S8-Inf counts by sex")

Table S10. Record counts (RC) and Person Counts (PC) of infant-related concepts explored in the DARWIN EU® Portal, by age.

See excel file ("Table S8-Inf counts by age")

Table S11. Results from Questionnaire 2.

See excel file ("Table S11-Questionnaire 2").

ANNEX V. Glossary

Additional definitions are available in the EMA Glossary of terms <https://www.ema.europa.eu/en/about-us/glossaries>.

Aggregated Data

Data collected and combined from multiple sources to generate summary information, typically anonymised.

Benefit-Risk Assessment

Evaluation of the positive therapeutic effects of a medicine compared to its risks (e.g. side effects).

Common Data Model (CDM)

A standardised data structure that enables data from multiple sources to be harmonised, making analysis consistent and reproducible. DARWIN EU® utilises the OMOP CDM maintained by the OHDSI community.

Complex Studies (C3)

Studies requiring the development or customisation of specific study designs, protocols, and Statistical Analysis Plans (SAPs), with extensive collection or extraction of data. Examples include etiological studies measuring the strength and determinants of an association between an exposure and the occurrence of a health outcome in a defined population considering sources of bias, potential confounding factors, and effect modifiers.

Coordination Centre (CC)

The central hub responsible for managing and overseeing the activities within DARWIN EU®. It is based at Erasmus University Medical Center in Rotterdam, the Netherlands.

Data Access

The process of obtaining permission to use specific datasets for regulatory or scientific studies.

Data Quality Framework

A set of standards and procedures to ensure accuracy, completeness, timeliness, and consistency of data used in DARWIN EU®.

Data Source

A database or repository of structured health-related data, such as electronic health records (EHRs), insurance claims, or registries.

DARWIN EU®

The European Medicines Agency's (EMA) federated network of real-world data sources designed to generate evidence to support regulatory decision-making.

EMA (European Medicines Agency)

The regulatory body responsible for the evaluation and supervision of medicinal products in the EU, overseeing DARWIN EU®.

Evidence Generation

The process of analysing real-world data to produce scientific information that can inform healthcare or regulatory decisions.

Federated Network

A data infrastructure where data remain at their original location but can be analysed in a harmonised way across multiple partners using a common model and tools.

GDPR (General Data Protection Regulation)

The EU regulation governing the protection of personal data and privacy, crucial to how DARWIN EU® handles health data.

Health Technology Assessment (HTA)

A systematic evaluation of properties and impacts of health technology, often using DARWIN EU® data to support assessments.

Metadata

Descriptive information about a data source (e.g. its content, quality, and structure), essential for identifying relevant databases in DARWIN EU® studies.

Off-the-Shelf Studies (OTS)

Studies for which a standard protocol per study/analysis type and standardised analytics may be developed and applied or adapted, typically relating to a descriptive research question. This includes studies on disease epidemiology, for example, the estimation of the prevalence or incidence of health outcomes in defined time periods and population groups, or drug utilisation studies at the population or patient level.

OHDSI (Observational Health Data Sciences and Informatics)

An open-science collaborative community that develops tools and standards (including the OMOP CDM) to enable large-scale analytics of observational health data. OHDSI provides the technical and scientific foundation for DARWIN EU®'s analytical ecosystem.

Patient-Level Data

Data related to individuals, de-identified, used for longitudinal or detailed analyses.

OMOP (Observational Medical Outcomes Partnership)

A common data model (CDM) that standardises the structure and content of observational healthcare data, enabling systematic analysis across disparate datasets. DARWIN EU® uses the OMOP CDM to ensure interoperability and consistency in real-world evidence generation.

Real-World Data (RWD)

Data relating to individual health status or healthcare delivery that is collected from routine clinical practice rather than from randomised controlled trials.

Real-World Evidence (RWE)

Clinical evidence derived from the analysis of RWD, used to inform decisions by regulators, payers, or clinicians.

Regulatory Decision-Making

The process by which authorities like EMA assess data to authorise, monitor, or modify the use of medicines in the EU.

Routine Repeated Studies (RR)

Studies that are either Off-the-Shelf or Complex studies repeated on a regular basis, following the same protocol and study code, but with updated data and/or different data partners.

Study Protocol

A detailed plan describing how a specific real-world study will be conducted, including objectives, design, data sources, and analyses.

Very Complex Studies (C4)

Studies which cannot rely only on electronic health care databases, or which would require complex methodological work, for example, due to the occurrence of events that cannot be defined by existing diagnosis codes, including events that do not yet have a diagnosis code, where it may be necessary to combine a diagnosis code with other data such as results of laboratory investigations. These studies might require the collection of data prospectively, or the inclusion of new (not previously onboarded) data sources.



Study Report

P4-C4-001

DARWIN EU[®] - PeriNet

OBJECTIVE 2

**Optimising and validating OMOP
pregnancy algorithms for data sources
with pregnancy-related data within the
DARWIN EU[®] Data Network**

15/07/2026

Version 4.0

Authors: Julieta Politi, Talita Duarte-Salles, Aniek Markus, Berta Raventós, Anum Zahra, Maarten Van Kessel, Adam Black, Natasha Yefimenko

Public

Study title	DARWIN EU® - PeriNet Objective 2: Optimising and validating OMOP pregnancy algorithms for data sources with pregnancy-related data within the DARWIN EU® Data Network
Study report version	V4.0
Date	15/07/2026
EUPAS number	EUPAS1000000812
Active substance	NA
Medicinal product	NA
Research question and objectives	<u>Research question</u> Can algorithm-based inference of pregnancy episodes and outcomes be reliably applied across different populations and healthcare settings within the DARWIN EU® Data Network? <u>Objectives</u> This study evaluated the reliability of inferring pregnancy episodes and their outcomes using algorithms applied within the DARWIN EU® Data Network. The specific objectives of this study were: 2a. To describe existing published pregnancy algorithms that have been applied in data sources mapped to the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM). 2b. To assess the availability of the required input variables for the identified pregnancy algorithms across the selected DARWIN EU® data sources. 2c. To implement the identified pregnancy algorithm(s) in the selected DARWIN EU® data sources and adapt for initial execution, accounting for data availability. 2d. To iteratively evaluate the internal consistency and plausibility of the performance of the selected pregnancy algorithms. 2e. To optimise the pregnancy algorithms in the selected DARWIN EU® data sources (based on results from Objective 2d). 2f. To validate the algorithm's performance at the episode-level and population-level within the DARWIN EU® Data Network.
Countries of study	Croatia, Denmark, Finland, France, Germany, Hungary, the Netherlands, Norway, Portugal, Spain, Sweden, and the United Kingdom
Authors	Julietta Politi (j.politi@darwin-eu.org), Talita Duarte-Salles (t.duarte@darwin-eu.org), Aniek Markus (a.markus@darwin-eu.org), Berta Raventos (b.raventos@darwin-eu.org), Anum Zahra (a.zahra@darwin-eu.org), Maarten van Kessel (m.vankessel@darwin-eu.org), Ger Inberg (g.inberg@darwin-eu.org), Adam Black (a.black@darwin-eu.org), Natasha Yefimenko (n.yefimenko@darwin-eu.org)

LIST OF ABBREVIATIONS

Acronyms/term	Description
AB	Abortion
AEMPS	Spanish Agency for Medicines and Medical Devices
ATC	Anatomical Therapeutic Chemical
BIFAP	Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público
CC	Coordination centre
CDM	Common Data Model
CDW Bordeaux	Clinical Data Warehouse of Bordeaux University Hospital
CPRD GOLD	Clinical Practice Research Datalink GOLD
DARWIN EU®	Data Analysis and Real-World Interrogation Network
DELIV	Delivery
DK-DHR	Danish - Data Health Registries
DQD	Data Quality Dashboard
ECT	Ectopic pregnancy
EHR	Electronic Health Records
EMA	European Medicines Agency
Erasmus MC	Erasmus Medical Center
EMDB-ULSGE	Egas Moniz Health Alliance database - Gaia E Espinho
ENCePP	European Network of Centres for Pharmacoepidemiology and Pharmacovigilance
EU	European Union
EUPAS	EU Post-Authorisation Studies Register
FinOMOP-TaUH Pirha	Tampere University Hospital patient cohort
FinOMOP-THL	Finland: Finnish Care Register for Health Care
GDPR	General Data Protection Regulation
HIP	Hierarchy-based Inference of Pregnancy
HI-SPEED	Health Impact - Swedish Population Evidence Enabling Data-linkage
ICD	International Classification of Diseases
IDIAPJGol	L'Institut d'Investigació en Atenció Primària de Salut Jordi Gol
IMASIS	Institut Municipal Assistència Sanitària Information System
InGef RDB	InGef Research Database
IPCI	Integrated Primary Care Information
IQR	Interquartile range
LB	Livebirth
LMP	Last menstrual period
NA	Not applicable
NAJS	Croatian National Public Health Information System

Acronyms/term	Description
NLHR	Norwegian Linked Health Registry data
OHDSI	Observational Health Data Sciences and Informatics
OMOP	Observational Medical Outcomes Partnership
PET	Perinatal Expansion Table
PPS	Pregnancy Progression Signature
PPV	Positive Predictive Value
PREG	Pregnancy
PSMar	Consorci Mar Parc de Salut Barcelona
QC	Quality check
RxNorm	Medical prescription normalised
SA	Spontaneous abortion
SB	Stillbirth
SIDIAP	The Information System for Research in Primary Care
SMPA-GU	Swedish Medical Products Agency - Gothenburg University
SNOMED	Systematized Nomenclature of Medicine
SUCD	Semmelweis University Clinical Data
WHO	World Health Organisation

1. TITLE

DARWIN EU® - PeriNet Objective 2: Optimising and validating OMOP pregnancy algorithms for data sources with pregnancy-related data within the DARWIN EU® Data Network

2. DESCRIPTION OF THE STUDY TEAM

Study team role	Names	Organisation
Principal Investigator	Julieta Politi ¹	Erasmus MC
Co-Principal Investigators	Talita Duarte-Salles Aniek Markus	Erasmus MC
Data Scientist	Aniek Markus Maarten van Kessel Ger Inberg Adam Black	Erasmus MC
Study Manager	Natasha Yefimenko	Erasmus MC
Epidemiologist	Julieta Politi Talita Duarte-Salles Berta Raventós Anum Zahra	Erasmus MC
Data source	Names	Data Partner Organisation*
NAJS	Zeljka Drausnik Irena Barišić Ivan Pristaš Jakov Vuković Anamaria Jurčević Marko Čavlina Antea Jezidžić Karlo Pintarić	Croatian Institute of Public Health
DK-DHR	Elvira Braüner Susanne Bruun	Danish Data Health Registries
FinOMOP-TaUH Pirha	Sampo Kukkurainen Leena Hakkarainen Harri Rantala	Pirkanmaa Welfare Services County, Tampere University Hospital
FinOMOP-THL	Gustav Klingstedt Tiina Wahlfors Toni Lehtonen	Finnish Institute for Health and Welfare
CDW Bordeaux	Romain Griffier Guillaume Verdy	CHU DE BORDEAUX - DIRECTION GENERALE
InGef RDB	Raeleesha Norris Annika Vivirito	Institute for Applied Health Research Berlin GmbH

	Alexander Harms Josephine Jacob	
SUCD	Zsolt Bagyura Loretta Kiss Tibor Héja Ágota Mészáros	Semmelweis University
IPCI	Katia Verhamme	Erasmus University Medical Center
NLHR	Saeed Hayati Petrica Ioan Olteanu Hedvig Nordeng	University of Oslo
EMDB-ULSGE	Firmino Machado Inês Nunes Teresa Monjardino	Clinical Academic Center Egas Moniz Health Alliance
BIFAP	Miguel Ángel Maciá Martínez Ana Llorente-Garcia Hermenegildo Carlos Martínez-Alcalá García Mar Martín-Perez Elisa Martín-Merino	AEMPS
IMASIS	Miguel Angel Mayer Angela Leis Juan Manuel Ramírez Anguita	PSMar
SIDIAP	Agustina Giuliadori Elena Roel Laura Granés	IDIAPJGol
HI-SPEED	Rickard Ljung Huiqi Li Fredrik Nyberg	SMPA-GU
CPRD GOLD	Antonella Delmestri	University of Oxford

*Data partners do not have an investigator role. Data partners execute code at their data source, review, and approve their results.¹Involved in the study until 30/04/2026.

3. ABSTRACT

Title

DARWIN EU® - PeriNet Objective 2: Optimising and validating OMOP pregnancy algorithms for data sources with pregnancy-related data within the DARWIN EU® Data Network.

Rationale and background

Pregnancy episodes and outcomes are often missing or inconsistently derived across individuals and data sources in routinely collected healthcare data. When structured pregnancy information is missing, pregnancy algorithms are required to infer structured pregnancy information using available input variables. Specifically, the extent to which Observational Medical Outcomes Partnership (OMOP)-based pregnancy algorithms perform across a wide range of data types and settings is uncertain, including whether further development is needed.

Research question and objectives

Research questions

Can algorithm-based inference of pregnancy episodes and outcomes be reliably applied across different populations and healthcare settings within the DARWIN EU® Data Network?

Objectives

This study evaluated the reliability of inferring pregnancy episodes and their outcomes using algorithms applied within the DARWIN EU® Data Network.

The specific objectives of this study were:

- 2a. To describe existing published pregnancy algorithms that have been applied in data sources mapped to the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM).
- 2b. To assess the availability of the required input variables for the identified pregnancy algorithms across the selected DARWIN EU® data sources.
- 2c. To implement the identified pregnancy algorithm(s) in the selected DARWIN EU® data sources and adapt for initial execution, accounting for data availability.
- 2d. To iteratively evaluate the internal consistency and plausibility of the performance of the selected pregnancy algorithms.
- 2e. To optimise the pregnancy algorithms in the selected DARWIN EU® data sources (based on results from Objective 2d).
- 2f. To validate the algorithm's performance at the episode-level and population-level within the DARWIN EU® Data Network.

Methods

Study design

This multi-phase methodological study aimed to implement, optimise, and validate a published pregnancy episode inference algorithm in real-world data mapped to the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM). The work included: (1) a scoping review to identify candidate OMOP-compatible pregnancy algorithms with selection of one algorithm for implementation; (2) assessment of the required inputs and its implementation across participating data sources; (3) refinement of concept sets and algorithm logic through iterative runs and quality control using internal consistency and plausibility checks; and (4) validation and benchmarking of the algorithm's output using episode and population-level comparisons to prespecified reference sources.

Population

The study population included all individuals registered as female sex (at data source entry) present in the data source during the study period, from 01/01/2010 (or data source start, if later) to the end of available data. Individuals were required to have at least 365 days of continuous observation time. To allow capture of complete full-term pregnancies, individuals required at least one pregnancy-related record by 31/12/2023 or by one year before the end of data availability, whichever came first.

Variables

Outcome:

The primary outcomes were the occurrence and timing of the algorithm-derived variables:

- Pregnancy Start Date
- Pregnancy End Date
- Pregnancy Outcome (live birth, stillbirth, ectopic pregnancy, spontaneous abortion, induced abortion, and delivery)

Relevant covariates:

Age, calendar year.

Data sources

1. Croatia: Croatian National Public Health Information System (NAJS)
2. Denmark: Danish Data Health Registries (DK-DHR)
3. Finland: Tampere University Hospital patient cohort (FinOMOP-TaUH Pirha)
4. Finland: Finnish Care Register for Health Care (FinOMOP-THL)
5. France: Clinical Data Warehouse of Bordeaux University Hospital (CDW Bordeaux)
6. Germany: InGef Research Database (InGef RDB)
7. Hungary: Semmelweis University Clinical Data (SUCD)
8. Netherlands: Integrated Primary Care Information (IPCI)
9. Norway: Norwegian Linked Health Registry data (NLHR)
10. Portugal: Egas Moniz Health Alliance database - Gaia E Espinho (EMDB-ULSGE)
11. Spain: Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público (BIFAP)
12. Spain: Institut Municipal Assistència Sanitària Information System (IMASIS)
13. Spain: The Information System for Research in Primary Care (SIDIAP)
14. Sweden: Health Impact - Swedish Population Evidence Enabling Data-linkage (HI-SPEED)
15. United Kingdom: Clinical Practice Research Datalink GOLD (CPRD GOLD)

Study size

No sample size was calculated, as this was an exploratory study that did not test a specific hypothesis.

Statistical analysis

We performed a scoping review to identify and summarise existing pregnancy algorithms and to identify those developed and adapted for use in the OMOP CDM. We selected the Jones et al. pregnancy episode algorithm (using the implementation by Smith et al.) based on prespecified criteria. The algorithm

combines a Hierarchy-based Inference of Pregnancy (HIP) component and a Pregnancy Progression Signature (PPS) component and applies 4 sequential steps to infer pregnancy episodes. Episodes were classified into outcome categories, including live birth (LB), stillbirth (SB), ectopic pregnancy (ECT), abortion (AB), spontaneous abortion (SA), pregnancies with a recorded delivery (DELIV), and pregnancies without a recorded outcome (PREG). We assessed the availability and completeness of required input variables across data sources. We evaluated the algorithm output using internal consistency and plausibility metrics (intermediate quality checks (QC)), including the proportion of inferred episodes with gestational length in a prespecified plausible range (0–308 days), the proportion of overlapping episodes within individuals, and outcome specificity based on the distribution of assigned pregnancy outcome categories. We iteratively optimised the algorithm over 3 runs based on QC findings. We then validated Run 3 outputs at the episode level (by comparing results with data in the Perinatal Expansion Table [PET], a standardised OMOP-based table designed to store pregnancy episodes and related perinatal information, where available) and population level (benchmarking to national statistics). We reported validation results per data source using standardised data source profiles (“fingerprints”).

Results

We executed the Jones et al. pregnancy episode algorithm (using the implementation by Smith et al.) across 15 OMOP-mapped data sources using a standardised QC. Run 3 identified 8,159,921 pregnancy episodes (range 33,251–1,682,637 per source; results were not received from EMDB-ULSGE for this Run), among 5,566,270 individuals. Median maternal age at first pregnancy start ranged 29.0–32.9 years across sources. The proportion of episodes with plausible gestational length (0–308 days) after Run 3 ranged between 73.0–99.8%, with 12/14 >90%. Overlap of pregnancy episode dates ranged between 0.4–6.7%; outcome specificity varied substantially (PREG range 0–46.4%). Run 3 outputs were summarised as data source “fingerprints” (standardised profiles of episode yield, plausibility, overlap, and outcome specificity) to support cross-database interpretation. Episode-level validation against information used by the PET was performed in five data sources (range sensitivity 67.3–91.9%, PPV 58.7–97.3%). Population-level benchmarking against national statistics showed that birth rates based on a combination of LB+DELIV were generally more comparable to national references than rates based on LB only, maternal age was broadly similar across sources, and C-section rates were highly variable across data sources. Using a cleanup step excluding implausible gestational length and overlapping episodes with Run 3 would result in a median reduction of 6.2% of pregnancy episodes (range 2.3–36.6%).

Discussion

Network implementation of the Jones et al. pregnancy episode algorithm (using the Smith et al. OMOP implementation) was executed using the *PregnancyIdentifier* R package, developed for this purpose. Execution was done across the 15 participating data sources of the DARWIN EU® Data Network over three runs, which enabled QC-driven optimisation over each iteration. Run 3 outputs showed that gestational length plausibility was high, episode overlap generally low, but outcome specificity remained heterogeneous across settings. Episode validation benchmarking against national statistics birth rates were similar or smaller, depending on definitional and denominator comparability. Run 3 data source “fingerprints” provide a preliminary basis for fit-for-purpose interpretation of cross-data source performance, which can help identify where additional restrictions or refinements are needed for timing-sensitive or outcome-specific analyses.

Further optimisation of the pregnancy algorithm may be possible, particularly targeting episode start-date estimation and back-calculation logic, though gains will vary depending on data availability and source-level recording quality. Use of an exclusionary “cleanup” step (already quantified in the fingerprints) could be an option to deal with implausibility in gestational length and overlapping episodes.

4. AMENDMENTS AND UPDATES

None.

5. MILESTONES

Study deliverable	Timelines (planned)	Timelines (actual)
Final Study Protocol	May 2025	September 2025
Creation of Analytical code	September/October 2025	August 2025
Execution of Analytical Code on the data	October/November 2025	March 2026
Draft Study Report	November/December 2025	March 2026
Final Study Report	April 2026	June 2026

6. RATIONALE AND BACKGROUND

Pregnancy episodes in electronic health records (EHR) data are often inconsistently coded within and across data sources. Although pregnancy represents a critical period for capturing exposures and clinical outcomes, structured variables, such as pregnancy start and end dates, gestational age at delivery, and pregnancy outcomes, are not systematically available. These limitations complicate the use of routinely collected healthcare data for pregnancy-related research.

Reliable identification of both pregnancy episodes and their outcomes is critical. Accurate exposure timing is essential for evaluating the effects of medications, infections, acute illnesses, chronic conditions, and other exposures during specific gestational windows. However, inconsistencies in the recording of pregnancy start dates and gestational age complicate accurate exposure classification and increase the risk of misclassification. Moreover, accurately capturing key outcomes, such as miscarriage, intrauterine foetal demise, stillbirth, or preterm birth, is important to prevent bias in study results. Pregnancy episodes are rarely directly recorded in structured fields, and, depending on the data source, must often be inferred indirectly. Validated algorithms that integrate multiple data elements—such as diagnostic and procedure codes, laboratory results, among others—alongside standardised coding conventions, can support more consistent and reproducible episode identification across sources.[1-3] Reliable identification of pregnancy episodes and outcomes is a prerequisite for meaningful analysis of exposures and risks in reproductive and perinatal pharmacoepidemiology. This report corresponds to the second objective of the “DARWIN EU® - Enabling pregnancy and mother-child research in the DARWIN EU® Data network - DARWIN-PeriNet” project. In this study we aimed to optimise and validate OMOP pregnancy algorithms for DARWIN EU® data sources with pregnancy-related data.

7. RESEARCH QUESTION AND OBJECTIVES

Research question

Can algorithm-based inference of pregnancy episodes and outcomes be reliably applied across different populations and healthcare settings within the DARWIN EU® Data Network?

Objectives

This study evaluated the reliability of inferring pregnancy episodes and their outcomes using algorithms applied within the DARWIN EU® Data Network.

The specific objectives of this study were:

2a. To describe existing published pregnancy algorithms that have been applied for use in data sources mapped to the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM).

2b. To assess the availability of the required input variables for the identified pregnancy algorithms across the selected DARWIN EU® data sources.

2c. To implement the identified pregnancy algorithm(s) in the selected DARWIN EU® data sources and adapt for initial execution, accounting for data availability.

2d. To iteratively evaluate the internal consistency and plausibility of the performance of the selected pregnancy algorithms.

2e. To optimise the pregnancy algorithms in the selected DARWIN EU® data sources (based on results from Objective 2d).

2f. To validate the algorithm's performance at the episode-level and population-level within the DARWIN EU® Data Network.

8. RESEARCH METHOD

8.1. Study design

A retrospective study was conducted using routinely collected healthcare data from 15 data sources across 12 European countries and in 10 EU member states. The study comprised:

- A scoping literature review conducted to address objective 2a, which assessed published pregnancy algorithms that have been applied in OMOP CDM data sources. One algorithm was selected based on pre-specified criteria.
- An assessment of the availability of the required input variables by data source that addressed objective 2b and informed objective 2c.
- A descriptive characterisation of the algorithms' output to address objective 2d, assessing elements such as the number and characteristics of the pregnancies identified (plausibility/consistency of findings, such as median gestational age, % of pregnancies with length >308 days, % of pregnancies with length of at least 1-day, maternal age distribution, % of overlapping pregnancy episodes, etc.). This informed objective 2e.
- A validation of the final algorithm's performance at the episode level by comparing the constructed pregnancy episodes to locally pre-processed reference data available in some sources (e.g. perinatal expansion tables). At the population level, validation involved evaluating the plausibility of aggregated metrics (e.g. pregnancy rates, age distributions, and outcome frequencies) against national statistics.

8.2. Follow-up

For this descriptive algorithm validation study, follow-up referred to the period during which pregnancy episodes could be inferred within an individual's valid observation time in each data source. Follow-up began on the latest of:

- The study start date (January 1, 2010), or
- The date on which the individual had accumulated at least 365 days of prior observation.

End of follow-up was defined as the earliest of:

- End of observation period

- Death
- Loss to follow-up (if applicable)

Pregnancy episodes were constructed within each individual's observed time in the data source, using algorithm-specific criteria to infer episode start and end dates (e.g. gestational timing back-calculated from delivery, or anchored on prenatal care-related interactions, including visits, procedures, measurements, or observations).

8.3. Study population with inclusion and exclusion criteria

The study population included all individuals registered as female sex (at data source entry) present in the data source during the study period.

Inclusion criteria

- All individuals who were present in the period from 01/01/2010 to 31/12/2024 (or the latest date of data availability)
- Participants were female sex (at data source entry)
- Aged 15 to 55 years
- Minimum 365 days of continuous data source history were available
- At least one record indicative of a pregnancy episode occurred up to 31/12/2023 (or one year before the data source end date), to ensure full-term pregnancies could be identified.

Exclusion criteria

None.

8.4. Study setting and data sources

This study was conducted using 15 routinely collected healthcare data from primary, secondary care, and registry-based data sources in the DARWIN EU® Data Network of data partners from 12 European countries in 10 EU member states. All data were a priori mapped to the OMOP CDM.

Table 1. Data sources.

Country	Name of Data source	Health Care setting	Type of Data	Number of active individuals	Calendar period covered by each data source	Study period	Contributing to
Croatia	NAJS	Primary, secondary, and hospital care	Registry	4.3M	01-01-2014 to 08-02-2025	2017*–2025	All objectives
Denmark	DK-DHR	Secondary and hospital care	Registry	5.98M	01-01-1995 to 18-01-2025	2010–2025	All objectives
Finland	FinOMOP-TaUH Pirha	Secondary and hospital care	EHR	263k	01-01-2007 to 08-03-2025*	2010–2025	All objectives

Country	Name of Data source	Health Care setting	Type of Data	Number of active individuals	Calendar period covered by each data source	Study period	Contributing to
Finland	FinOMOP-THL	Primary, secondary, and hospital care	Registry	5.7M	01-01-2011 to 01-10-2024	2011–2024	All objectives
France	CDW Bordeaux	Hospital care	EHR and claims data	247k	01-01-2005 to 22-06-2025	2010–2025	All objectives
Germany	InGef RDB	Primary, and secondary, and hospital care	Administrative claims data	7.43M	01-01-2015 to 01-04-2024	2015–2024	All objectives
Hungary	SUCD	Secondary and hospital care	EHR and claims data	227k	01-01-2005 to 30-11-2024	2010–2024	All objectives
Netherlands	IPCI	Primary care	EHR	1.32M	01-01-2006 to 15-04-2025	2010–2015	All objectives
Norway	NLHR@UiO	Primary and hospital care	Registry	5.5M	01-01-2008 to 29-10-2024	2018*–2024	All objectives. Additionally, algorithm results were compared with the PET.
Portugal	EMDB-ULSGE	Secondary care	EHR	130k	01-01-2008 to 01-11-2023	2010–2023	All objectives
Spain	BIFAP	Primary care and diagnosis at hospital discharge	EHR	23.2M	01-01-2005 to 31-12-2024	2010–2024	All objectives
Spain	IMASIS	Secondary and hospital care	EHR	253k	01-01-1990 to 10-01-2025	2010–2025	All objectives. Additionally, algorithm results were compared with the PET.
Spain	SIDIAP	Primary care	EHR	6.07M	01-01-2006 to 30-06-2023	2010–2023	All objectives. Additionally, algorithm results were

Country	Name of Data source	Health Care setting	Type of Data	Number of active individuals	Calendar period covered by each data source	Study period	Contributing to
							compared with the PET.
Sweden	HI-SPEED	Primary, secondary, and hospital care	Registry	10.6M	01-01-2015 to 10-09-2024	2015–2024	All objectives. Additionally, algorithm results were compared with the PET.
United Kingdom	CPRD GOLD	Primary care	EHR	2.63M	01-01-1987 to 01-01-2025	2010–2025	All objectives. Additionally, algorithm results were compared with the PET.

NAJS: Croatian National Public Health Information System, DK-DHR: Danish Data Health Registries, FinOMOP-TaUH Pirha: Tampere University Hospital patient cohort, FinOMOP-THL: Finnish Care Register for Health Care, CDW Bordeaux: Clinical Data Warehouse of Bordeaux University Hospital, InGef RDB: InGef Research Database, SUCD: Semmelweis University Clinical Data, IPCI: Integrated Primary Care Information, NLHR: Norwegian Linked Health Registry data, EMDB-ULSGE: Egas Moniz Health Alliance database - Gaia E Espinho, BIFAP: Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público, IMASIS: Institut Municipal Assistència Sanitària Information System, SIDIAP: The Information System for Research in Primary Care, HI-SPEED: Health Impact - Swedish Population Evidence Enabling Data-linkage, CPRD GOLD: Clinical Practice Research Datalink GOLD, PET: Perinatal Expansion Table, M: Million, EHR: Electronic health records.

*Final CDM version used covers up to 10/04/2025.

Data sources

1. Croatia: Croatian National Public Health Information System (NAJS)
2. Denmark: Danish Data Health Registries (DK-DHR)
3. Finland: Tampere University Hospital patient cohort (FinOMOP-TaUH Pirha)
4. Finland: Finnish Care Register for Health Care (FinOMOP-THL)
5. France: Clinical Data Warehouse of Bordeaux University Hospital (CDW Bordeaux)
6. Germany: InGef Research Database (InGef RDB)
7. Hungary: Semmelweis University Clinical Data (SUCD)
8. Netherlands: Integrated Primary Care Information (IPCI)
9. Norway: Norwegian Linked Health Registry data (NLHR)
10. Portugal: Egas Moniz Health Alliance database - Gaia E Espinho (EMDB-ULSGE)
11. Spain: Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público (BIFAP)
12. Spain: Institut Municipal Assistència Sanitària Information System (IMASIS)
13. Spain: The Information System for Research in Primary Care (SIDIAP)
14. Sweden: Health Impact - Swedish Population Evidence Enabling Data-linkage (HI-SPEED)
15. United Kingdom: Clinical Practice Research Datalink GOLD (CPRD GOLD)

Data source justification and key characteristics

These data sources fulfilled the criteria required in terms of having been onboarded into the DARWIN EU[®] network (at the time the study started), and the availability of pregnancy-related data, such as pregnancy start and end dates and mother-child linkage. This information was obtained through a questionnaire to data partners, along with information available in the DARWIN EU[®] Portal. The Portal contains metadata for each of the DARWIN EU[®] data sources, including aggregated counts for data already mapped to the OMOP CDM, which was used internally to inform feasibility assessments performed in Objective 1.

In order to assess the reliability of data sources, the data partners were asked to describe their internal data quality process on the source data as part of the DARWIN EU[®] onboarding procedure. To further ensure data quality, we utilised the *Achilles* tool,[4] which systematically characterises the data and generates data characteristics, such as age distribution, condition prevalence per year, and data density.

Data density measures included:

- 1) monthly record counts by data domain (which offers insights into data collection patterns and the start date of each data source)
- 2) measurement value distribution (i.e. min, max, quartiles for numeric values per measurement concept and per unit and counts for discrete measurement-value pairs).

The latter can be compared against expectations for the data based on predefined standards, historical trends, or known epidemiological patterns to identify potential anomalies or inconsistencies. Additionally, the data quality dashboard (DQD) provided more objective checks on plausibility of data completeness, consistency, and conformity across the data sources.

The DARWIN EU[®] Portal, as well as information from the onboarding documents, was used to assess whether data sources have information on pregnancy-related variables and pregnancy outcomes. Data within the DARWIN EU[®] portal is maintained up to date by extracting the release dates for each dataset in the network and monitoring when data are out-of-date with the expected refresh cycle (typically quarterly or half-yearly). In addition, it was important to understand the time covered by each released data source, as this can vary across different domains. To facilitate this, the *CDMONboarding* (and *Achilles*) packages [4] contain a 'data density' plot. This plot displays the number of records per OMOP domain monthly, allowing insights into when data collection started, when new sources of data were added, and until when data was included.

Finally, when the data sources were invited to participate, they were informed about the study objectives and asked to confirm their ability to participate in the study.

More details on the selection of data sources can be found in the study report of objective 1 of the DARWIN-PeriNet project.

Detailed information on the data sources participating in this study is provided in [Annex I](#) and [Annex II](#).

8.5. Study period

The study period covered from 01/01/2010 to 31/12/2024, or the most recent data available for each contributing data source. Not all data sources contributed data for the entire study period. Specific availability per data source is shown in [Table 1](#).

8.6. Variables

8.6.1. Outcome

Consistent with prior work,[5] a pregnancy episode was defined as the interval of time from the estimated last menstrual period date (pregnancy start) and the date of the pregnancy outcome (pregnancy end). Each episode included person ID, pregnancy start date, pregnancy end date, and pregnancy outcome.

The primary outcome was the occurrence of the algorithm-derived variables:

- Pregnancy Start: Start of the pregnancy episode.
- Pregnancy End: End of the pregnancy episode.
- Pregnancy Outcomes: live birth, stillbirth, ectopic pregnancy, spontaneous or induced abortion, and delivery (delivery-only codes are non-specific to the outcome of the pregnancy).

We derived the outcome concept sets from the final selected algorithm identified in Objective 2a. The codes were refined during the study execution following the DARWIN EU® phenotyping standard processes, which involved the review of code lists by clinical experts and the review of phenotypes after their execution in the participating data sources.[6] The final concept sets used for study implementation are included in the *PregnancyIdentifier* (<https://github.com/darwin-eu/pregnancyidentifier>) R package.

8.6.2. Covariates

The covariates for this objective were as follows:

- Age groups (in years) defined at the pregnancy start date, namely:
 - 15–19
 - 20–29
 - 30–39
 - 40–49
 - ≥50
- Calendar year

8.7. Study size

No sample size was calculated, as this was an exploratory study that did not test a specific hypothesis.

8.8. Data transformation

Analyses were conducted separately for each data source. Before study initiation, test runs of the analyses were performed on a subset of the data sources and quality control checks were performed. Once all the tests passed (see [Annex III](#)), the final study code package was released in the version-controlled GitHub Study Repository for execution against all the participating data sources.

The data partners locally executed the analytics against the OMOP CDM in R Studio and reviewed and approved the, by default, aggregated results.

The study results of all data sources were checked, after which they were made available to the team, and the dissemination phase started. All results were locked and timestamped for reproducibility and transparency.

8.9. Statistical methods

The study workflow and relationship between objectives are shown in **Figure 1**. All analyses were conducted separately for each data source. Results were summarised qualitatively in a table (objective 2a) and quantitatively using counts and percentages (objectives 2b to 2f).

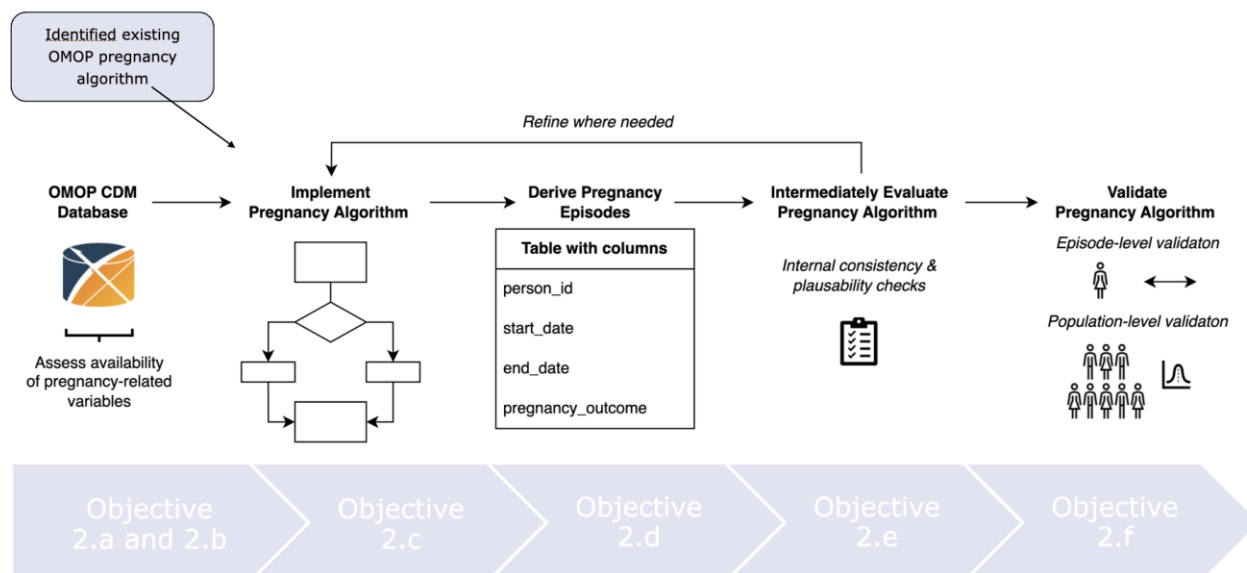


Figure 1. Study process and relationship with each study objective.

For Objective 2a, the scoping literature review focused on identifying existing published algorithms that had been developed or adapted for use in the OMOP CDM. The algorithm chosen for implementation was selected based on the following criteria: its ability to infer pregnancy start and end dates and outcomes, prior application to comparable data sources, and the availability of reproducible and well-documented code. When related algorithms existed, preference was given to refinements of earlier logic.

Applying these criteria, we selected the pregnancy episode inference algorithm described by Jones et al.[7]. This algorithm was subsequently reused by Smith et al., who applied this algorithm in comparable OMOP-mapped data, providing additional implementation precedent.[8] The algorithm builds on the published pregnancy episode logic of Matcho et al. and infers pregnancy episodes with estimated start and end dates, derives gestational length, and assigns episode end-type categories based on recorded pregnancy outcome evidence.[5, 7]

For network implementation, we used the pregnancy episode inference algorithm described by Jones et al., using the publicly available OMOP implementation provided by Smith et al. Published concept lists were harmonised to OMOP standard concepts and iteratively curated (HIP/PPS) to support consistent execution across participating data sources. The selected pregnancy episode algorithm produces person-level pregnancy episodes characterised by an estimated start date, estimated end date, and episode duration (gestational length, in days). For each episode, the algorithm assigns an episode end-type category based on available pregnancy outcome evidence. End-type categories are summarised using the high-level categories: Pregnancy (**PREG**): indicating that a pregnancy episode is identified, although the outcome of the episode is unknown; Delivery (**DELIV**): indicating that a record of delivery is observed, but whether the pregnancy ended in a live birth or stillbirth is not recorded. More specific outcome categories include live birth (**LB**), stillbirth (**SB**), spontaneous abortion (**SA**), induced abortion (**AB**), and ectopic pregnancy (**ECT**).

For Objective 2b, the published code lists proposed in the selected algorithm included over 1,000 codes and contained non-standard concepts. Non-standard concepts were converted to OMOP standard concepts

(and included relevant descendants where appropriate) before *CohortDiagnostics* and algorithm execution. Subsequently, we assessed the availability of the input variables required to implement the final selected algorithm across the selected data sources. The input concept sets were grouped into broad categories:

1. Hierarchy-based Inference of Pregnancy (HIP): used to identify possible pregnancy-related evidence, including outcomes and pregnancy-related procedures/encounters such as ultrasounds.
2. Pregnancy Progression Signature (PPS): used to support gestational timing signals and more specific pregnancy outcome evidence.

Descriptive analyses were conducted using the DARWIN EU® Portal and *CohortDiagnostics* to evaluate the presence and completeness of these input categories. Results were summarised using counts and percentages in each data source.

For Objective 2c, minimal adaptations were made to the Smith algorithm (a publicly available OMOP implementation of the pregnancy episode inference algorithm described by Jones et al.) to enable implementation across data sources.[7, 8] This included updating input variables by replacing them with equivalents in the standard vocabulary and supporting integration in the DARWIN EU® workflow (e.g. using *CDMConnector*). These adjustments allowed for the initial execution of the algorithm across all selected data sources. Concept sets and logic were subsequently iteratively curated as part of Quality-Check (QC) driven optimisation (Objective 2e). The algorithm was implemented in a new analytical R package, *PregnancyIdentifier*, which was developed by the DARWIN EU® CC to support both this, as well as future, studies. The study was conducted using OMOP CDM mapped data and utilising existing R packages, developed by DARWIN EU®, such as *CohortCharacteristics* and *IncidencePrevalence*, as well as other packages, where necessary.

For Objective 2d, we conducted a series of internal consistency and plausibility checks to assess the quality and coherence of the pregnancy episodes derived in Objective 2c. These metrics (summarised in [Table 2](#)), which we refer to as intermediate QC checks, included four quality aspects: episode frequency (plausibility and consistency of counted pregnancies within and across data sources), episode duration (accuracy of gestational length and alignment with known clinical timelines per pregnancy outcome), episode construction (completeness and correctness of the algorithmic output), and episode outcomes (capture of key maternal and foetal outcomes linked to each pregnancy episode). Additionally, we computed the rate of identified pregnancies per 1,000 individuals of female sex (at birth) aged 15–55 years using the *IncidencePrevalence* package. The output was computed for each data source, and irregularities were flagged and summarised.

Table 2. Episode metrics to assess internal consistency and plausibility of derived pregnancy episodes.

Metric	Purpose	Expected Output
Episode Frequency		
Number of pregnancy episodes and pregnant individuals	Evaluated pregnancy frequency overall and per individual	Total episodes, total individuals, distribution of pregnancies per person (min, max, median, IQR), total number of first pregnancy episodes, total number of subsequent pregnancy episodes (pregnancies with ≥1 preceding episode for the same individual)
Maternal age distribution	Assessed pregnancy distribution by age	Distribution of maternal age at pregnancy start (min, max, median, IQR)
Pregnancy incidence rate	Evaluated number of pregnancies relative to data source population	Pregnancies per 1,000 individuals of female sex (at birth) aged 15–55
Episode Duration		

Metric	Purpose	Expected Output
Gestational duration distribution	Assessed plausibility of pregnancy length	Min, max, median, IQR; count and % of episodes <1 day, >308 days
Distribution of gestational age at anchor event (e.g. delivery, first antenatal visit, ultrasound after first trimester, gestational diabetes mellitus)*	Assessed whether anchor events occur at plausible gestational ages	Histogram/density plots of gestational age at event; % preterm, term, post-term
Gestational week bins	Summarised gestational age into clinically relevant categories	% in bins: <12 12–27+6 (extremely preterm) 28–31+6 (very preterm) 32–36+6 (moderate to late preterm) 37–40 -41+6 (term) >42+ (post-term)
Seasonal and temporal patterns	Assessed expected timing patterns, detect artifacts	Monthly/annual trends of pregnancy start and end
Episode Construction		
Overlapping pregnancy episodes	Identified episodes that occur simultaneously (overlapping in time) within individuals with more than one inferred pregnancy episode	Count and % of overlapping episodes; individual-level review
Missing or inconsistent dates	Identified technical errors in episode construction	Count and % of episodes with missing or reversed dates
Control variables not used in the algorithm (e.g. breech presentation, preterm labour, polyhydramnios)*	Validated episode timing using external events not used in construction	Number of occurrences, number or occurrences overlapping with pregnancy episodes, percent overlap
Episode Outcomes		
Mode of delivery*	Evaluated the proportion of delivery modes	% vaginal and c-section deliveries
Pregnancy outcome events	Evaluated the proportion of delivery outcomes	Proportion of pregnancy outcome events and distribution of duration by pregnancy outcome type (Section 8.6.2.)

IQR: Interquartile Range.

* = Metrics were assessed only in Run 3.

In Objective 2e, we used the findings from Objective 2d to iteratively optimise the algorithm to improve its performance across various types of data sources. A total of three iterations of Objectives 2d and 2e were done (see [Figure 1](#)). Some examples of how these metrics guided the optimisation process are explained below:

- *CohortDiagnostics* was used to evaluate the concept sets used in the algorithm, identifying orphan codes, and reviewing the frequency and source contribution of input concepts. This helped to ensure that all relevant codes were included.
- Temporal patterns from *IncidencePrevalence* (e.g. sudden drops in pregnancy rates) were used to detect potential issues related to algorithm logic, changes in coding practices, or incomplete source data.

- Episode metrics ([Table 3](#)) guided the review of episode-building logic, including the alignment of date windows and the selection of anchor events.
- Data partner feedback was requested in cases where anomalies appeared specific to a data source or data source type, to assess whether adjustments to input hierarchies or logic rules were necessary.

HIP and PPS concept sets were iteratively curated between runs to increase outcome-specific capture and timing evidence, incorporating published Matcho concepts and concepts used in existing local pregnancy algorithms in participating data sources.[2, 5, 9, 10] The final Run 3 concept sets comprised approximately 1,500 HIP and 120 PPS standard concepts.

The algorithm was optimised within the *PregnancyIdentifier* R package. Following each optimisation step, the updated package was rerun across all selected data sources to reassess evaluation metrics and ensure improvements were achieved. Optimisation was conducted over three iterative runs (version 1–version 3) of the *PregnancyIdentifier* package. Some metrics were refined across runs. In Run 1, overlap percentages were calculated using all inferred pregnancy episodes as the denominator rather than individuals with more than one pregnancy episode. In addition, study-period restrictions affecting episode counts were not fully aligned to the intended analysis window in an early run and were corrected in the subsequent iteration. We corrected concept issues introduced during conversion from non-standard to standard OMOP concepts in Version 2. Accordingly, comparisons of QC metrics across runs should be interpreted cautiously, and later runs provide the most reliable basis for cross-source evaluation. The final validation was performed using the Run 3 output.

For Objective 2f, in addition to internal consistency and plausibility checks as performed for Objective 2d, further validation was done at the individual episode-level and aggregate population-level (see details below). Last, we quantified the impact of a hypothetical cleanup excluding overlapping episodes and episodes with implausible gestational length (>308 days).

Episode-level validation

This involved comparing results from the selected algorithm to those derived from locally pre-processed pregnancy episodes included in PET, a standardised OMOP-based table designed to store pregnancy episodes and related perinatal information, which was available in some data sources (NLHR@UiO, SIDIAP, HI-SPEED, IMASIS, and CPRD GOLD).[3] Although the local extraction might not represent a definitive gold standard nor is expected to produce identical results (e.g. prioritisation of birth registry data not mapped to OMOP CDM), it served as a high-quality reference.

We compared the two sets of derived pregnancy episodes. Each set consisted of identified pregnancy episodes that were characterised by a person ID, pregnancy start date, pregnancy end date, and pregnancy outcome. To enable comparison (see definitions including an * below for more details) we linked pregnancy episodes between sets based on person ID and timing by: 1) pairing each PET episode with every algorithm-identified episode with an overlap of at least 1 day and 2) repeatedly matching episodes with the largest overlap first. This resulted in a one-to-one matching between the two sets of derived pregnancy episodes. We evaluated consistencies and discrepancies using the metrics outlined below:

- Pregnant Individuals & Pregnancy Episodes: counted the number of identified pregnant individuals and pregnancies in both sets.
- Pregnancy Start & End Date*: calculated the difference between the estimated dates for the same episode, where feasible. We computed the distribution of differences (histogram, mean/median) and showed the overlap/differences in episodes derived between both sets in a Venn diagram (i.e. identified by both, identified by PET only, identified by the selected algorithm only). From this Venn

diagram, we computed the PPV and sensitivity of the selected algorithm compared to PET (taking the latter as the reference).

- Pregnancy Duration*: compared the distribution of episode durations in both sets (histogram, mean/median).
- Pregnancy Outcome*: linked pregnancy episodes between sets and created a confusion matrix with the number of identified outcomes per type (e.g. live birth, stillbirth, etc.) in both sets, where possible. Using PET as the reference, we assessed the accuracy of the algorithm-derived outcome classification output.

Population-level validation

External validity of our results for derived pregnancy episodes was compared with national statistics obtained from Euro-Peristat and EUROSTAT ([Table 3](#), [Table S1](#)).[11, 12] National statistics for mode of delivery in Norway were obtained from the National Institute for Health and Welfare.[13] Pregnancy outcome rates for the UK were obtained from the Office for National Statistics.[14]

Table 3. Metrics and sources for reference data to assess the external validity of the algorithm's output at the national level.

Metric	Source for reference data
Birth rate	EUROSTAT (e.g. 2014, 2023): Crude birth rate: expressed as the ratio of the number of live births during the year to the average population in that year
Maternal age at birth	EUROSTAT (e.g. 2014, 2023): Mean age of woman at birth of first child
Pregnancy outcome rates	EUROSTAT (e.g. 2014, 2023): Number of live births per year Euro-Peristat (e.g. 2015, 2019): Foetal mortality rate: foetal deaths at or after 24-weeks threshold* (per 1000 total births)
Gestational duration distribution	Euro-Peristat (e.g. 2015, 2019): Number of live births at each completed gestational age bin (<32, 32–36, 37–38, 39–41, and ≥42 weeks)
Mode of delivery	Euro-Peristat (e.g. 2015, 2019): Mode of delivery rate (% vaginal and c-section deliveries)

*For the purpose of this validation, we only considered foetal losses at or after 24 weeks.

8.10. Deviations from the protocol

Deviation number	Protocol version	Date	Section of study protocol	Deviation	Reason
1	4.1	10/10/2025	8.5. Study population with inclusion and exclusion criteria	An additional criterion of age of inclusion between 15 to 55 years of age was added.	The inclusion of reproductive age group was found to be more in line with the implemented pregnancy algorithm.
2	4.1	10/10/2025	8.6.2. Other covariates, including confounders, effect modifiers, and other variables	Age groups for stratification were adjusted according to the inclusion criterion for age mentioned in deviation 1. Age groups for stratification included: 10–19, 20–29, 30–39, 40–49, >=50 years.	These age groups were more in accordance with the age ranges of individuals included in the study.
3	4.1	10/10/2025	8.2. Study setting and data sources	At protocol stage, PET was available for three data sources (NLHR, SIDIAP, HI-SPEED) which has currently increased to five data sources (IMASIS, CPRD GOLD).	This allows for more extensive comparison of pregnancy episodes across data sources.
4	4.1	10/10/2025	8.8.3 Statistical model specification and assumptions of the analytical approach considered	Detailed QC checks based on anchor-event timing, external control variables, and delivery-mode completeness were only performed in Run 3, as earlier runs prioritised stabilising HIP/PPS concept sets and core episode-construction logic.	In practice, these specific QC checks were applied only in the final iteration (Run 3). This deviation was implemented because early iterations focused on stabilising the core pregnancy signal inputs (HIP/PPS concept sets) and resolving major episode-construction issues (e.g. implausible durations and overlap) identified in aggregate metrics. Applying the more detailed, downstream QC checks before concept-set and logic stabilisation would

Deviation number	Protocol version	Date	Section of study protocol	Deviation	Reason
					have produced difficult to interpret results. Therefore, these QC checks were deferred to Run 3 once the core concept sets and episode-construction logic were sufficiently stable and were then used to support the final internal consistency assessment and validation.
5	4.1	10/10/2025	8.8.3 Statistical model specification and assumptions of the analytical approach considered	The protocol specified rerunning the <i>PregnancyIdentifier</i> package after each optimisation step until no further site-generic improvements could be achieved. In practice, optimisation was limited to three network runs.	Due to the study timeline constraints, it was agreed to limit iterations to three runs. The prespecified final validation was performed using Run 3 outputs, and data sources were assessed for sufficiency based on the available QC and benchmarking results.
6	4.1	10/10/2025	8.8.3 Statistical model specification and assumptions of the analytical approach considered	A hypothetical cleanup step was introduced, which quantified the impact of excluding overlapping episodes and episodes with implausible gestational length (>308 days).	This was not prespecified in the protocol and was added, as it was deemed an important metric to support fit-for-purpose assessment of the algorithm outputs for studies.

9. RESULTS

The full set of results for this study is available through an interactive web application, Shiny app, at [EUPAS1000000812](https://eupas1000000812).

9.1. Main results

9.1.1. Algorithm scoping review (objective 2a)

A total of 20 studies published between 2004 and 2024 were included in this scoping review. Algorithms were assessed in terms of the data source type, healthcare setting, use of OMOP CDM data, and study period. Additionally, studies were also examined for methodology utilisation for pregnancy episode

identification, algorithm validation, and limitations. The number of pregnancies identified in the included studies ranged from 18,970 to 9,647,843 pregnancies (See [Annex Table S2](#)). The included studies originated from a range of data sources, including administrative or claims databases, electronic health records, and primary care data, spanning multiple geographic regions, such as the United States, Canada, the United Kingdom, and European countries (including Germany, Spain, and France).

Most studies identified pregnancies based on diagnosis and procedural codes handled in a hierarchical or rule-based manner (15 out of 20 studies), commonly estimating pregnancy start dates using LMP (12/20), while fewer used mother-infant linkage (5/20 studies). Most studies (15/20) had some type of validation (internal or external) performed. Amongst these, only 5 conducted gold-standard chart review, whereas others relied on external comparisons with national statistics or registries (9/20), internal validation using linked datasets (6/20), or indirect consistency checks (2/20). Only two studies [5, 8] conducted external validation of algorithms in new data sources, thereby assessing transportability across different health care settings, countries, and populations. Most studies assessed pregnancy-related outcomes such as live births, stillbirths, abortions (spontaneous or elective), ectopic pregnancies, while nearly half assessed gestational age and pregnancy start and end dates as outcomes. Four of the studies used OMOP CDM data sources.

Across studies, several limitations were identified, including i) underestimation of early pregnancy losses (particularly spontaneous abortions and pregnancies not resulting in healthcare visits), ii) incomplete or inaccurate recording of gestational age and LMP which introduced uncertainty in estimating pregnancy start dates, iii) misclassification of outcomes (e.g. identification of preterm vs term births). Finally, reliance on only one data source from one country limited generalisability and cross-settings comparability.

Details summarising all algorithms can be found in [Annex Table S2](#).

Applying the prespecified criteria, we selected the pregnancy episode inference algorithm described by Jones et al.; for implementation, we used the version by Smith et al.[7],[8] The algorithm builds on the published pregnancy episode logic of Matcho et al. and infers pregnancy episodes with estimated start and end dates, derives gestational length, and assigns episode end-type categories based on recorded pregnancy outcome evidence.[5, 7]

9.1.2. *PregnancyIdentifier* R package

We developed a new analytical R package *PregnancyIdentifier* that can be used to infer pregnancy episodes from OMOP CDM sources across the DARWIN EU® Data Network. This package implements our adapted and optimised version of the pregnancy episode inference algorithm described by Jones et al. [7], building from the publicly available OMOP implementation provided by Smith et al.[8] The algorithm follows the following four-step pipeline:

1. Hierarchy-based Inference of Pregnancy (HIP): This step detects pregnancy episodes by identifying high-specificity pregnancy outcomes (e.g. concepts indicating delivery or abortion) and gestational age evidence (e.g. number of weeks pregnant), serving as outcome anchors for episodes.
2. Pregnancy Progression Signature (PPS): This step constructs episodes from a variety of pregnancy-related concepts and their expected timing windows, utilising temporal information to define plausible periods for pregnancy-related events to occur.
3. Merge: The HIP and PPS episodes are integrated for each person by aligning episodes based on temporal overlap and prioritisation of high-specificity pregnancy outcomes to remove overlapping or duplicate episodes.
4. Episode Start Date (ESD) refinement: This final step refines the estimated pregnancy start dates by using gestational timing evidence and/or expected ranges per pregnancy outcome.

More information on the algorithm and its use can be found on GitHub pages: <https://darwin-eu-dev.github.io/PregnancyIdentifier/>. Key inputs to the algorithm are concept sets to identify pregnancy-related events, pregnancy progression signals, and outcome-specific min/max term durations.

9.1.3. Input variables (objective 2b)

Across input concepts, 360 non-standard concepts were identified and converted to OMOP standard concepts. Using *CohortDiagnostics* across the 15 participating data sources, HIP input concepts were observed in a total of 10,046,693 individuals (range 30,255–3,706,577 across data sources), whereas PPS input concepts were observed in 464,953 individuals (range 773–573,591) (**Table 4**).

Table 4. Number of individuals with availability of pregnancy algorithm input concepts across 15 data sources at Run 1 (*CohortDiagnostics*).

Concept set	Concepts (n)	BIFAP	CDW Bordeaux	DK-DHR	EMDB-ULSGE	FinOMOP-THL	HI-SPEED	IMAS IS	IPCI	InGef RDB	NAJS	NLHR@UIO	SIDIAP	SUCD	FinOMOP-TaUH Pirha	CPRD GOLD
HIP input concepts	770	3,455,073	81,757	3,706,577	30,255	1,052,235	2,790,794	180,398	231,553	1,986,973	1,059,604	1,557,723	1,189,837	448,477	134,652	10,046,693
PPS input concepts	75	573,591	4,380	82,212	773	66,009	73,758	33,407	9,636	67,785	308,555	87,313	330,249	28,338	75,295	464,953

Counts reflect individuals with ≥ 1 record mapped to each concept set.

NAJS: Croatian National Public Health Information System, DK-DHR: Danish Data Health Registries, FinOMOP-TaUH Pirha: Tampere University Hospital patient cohort, FinOMOP-THL: Finnish Care Register for Health Care, CDW Bordeaux: Clinical Data Warehouse of Bordeaux University Hospital, InGef RDB: InGef Research Database, SUCD: Semmelweis University Clinical Data, IPCI: Integrated Primary Care Information, NLHR@UIO: Norwegian Linked Health Registry data, EMDB-ULSGE: Egas Moniz Health Alliance database - Gaia E Espinho, BIFAP: Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público, IMASIS: Institut Municipal Assistència Sanitària Information System, SIDIAP: The Information System for Research in Primary Care, HI-SPEED: Health Impact - Swedish Population Evidence Enabling Data-linkage, CPRD GOLD: Clinical Practice Research Datalink GOLD, PET: Perinatal Expansion Table, HIP: Hierarchy-based Inference of Pregnancy, PPS: Pregnancy Progression Signature.

9.1.4. Multi-data source implementation and iterative improvements (objective 2c)

Across 15 data sources, the pregnancy episode algorithm was executed iteratively (Run 1–Run 3) to evaluate episode yield, gestational-length plausibility, overlap, and outcome typing. **Table 5** summarises episode-level QC metrics by data source and run, and **Table 6** summarises the number of data sources showing improvement for key QC endpoints between runs (Run 1 to Run 2 and Run 2 to Run 3).

Improvement showed as lower proportions of implausible gestational lengths, fewer overlapping episodes, and greater outcome specificity, reflected by fewer episodes assigned to PREG, where a pregnancy episode is identified but, without a recorded outcome, its gestational length is likely underestimated and DELIV (delivery confirmed but live/stillbirth status unresolved). Noting that DELIV, as an intermediate category (delivery confirmed but live birth or stillbirth status unresolved), may be the most specific outcome available in sources where live birth is not separately recorded. We interpreted episode counts descriptively: higher counts do not necessarily indicate better performance, as they may reflect over-fragmentation, duplicate episodes, or inclusion of non-pregnancy-specific concepts rather than genuine improvement in pregnancy episode identification.

The most consistent improvements in gestational-length plausibility were observed from Run 1 to Run 2, whereas changes from Run 2 to Run 3 were more heterogeneous across data sources. Changes across QC endpoints should be interpreted jointly. Refinements to episode construction and outcome assignment can redistribute episodes across end-type categories (DELIV, PREG, DELIV+PREG) and affect inferred duration and overlap patterns. Cross-run comparisons also need careful interpretation, as we refined selected QC components iteratively throughout the study, including concept definition corrections in Run 2, alignment of pregnancy episode counts to the intended study period, and correction of the pregnancy episode overlap denominator definition in later runs.

Pregnancy episode counts showed marked variability across data sources by run (Run 1: 29,213–3,762,235; Run 2: 25,067–2,064,880; Run 3: 33,251–1,682,637). In Run 3, the pregnancy algorithm identified a total of 8,159,921 pregnancy episodes among 5,566,270 individuals in the available data sources (results for EMDB-ULSGE are not received for this Run).

Gestational length plausibility: In Run 1, gestational length plausibility (0–308 days) was high across data sources (range 91.8–100%). In Run 2, gestational length plausibility (0–308 days) ranged between 79.6–100%, with 1/15 <90% (FinOMOP-TaUH Pirha). In Run 3, gestational length plausibility (0–308 days) ranged between 73.0–99.8%, with 2/14 <90% (FinOMOP-TaUH Pirha, SIDIAP).

From Run 1 to 2, gestational length plausibility (0–308 days) improved in 8/15 data sources (DK-DHR, EMDB-ULSGE, FinOMOP-THL, IMASIS, IPCI, InGef RDB, SIDIAP, and SUCD). Similarly, the proportion of episodes with extremely long gestational length (>52 weeks) decreased from Run 1 to Run 2 in 6/15 data sources (BIFAP, DK-DHR, IMASIS, IPCI, SIDIAP, and SUCD).

From Run 2 to 3, gestational length plausibility (0–308 days) increased in NLHR, was unchanged in CDW Bordeaux and CPRD GOLD, and decreased in the remaining 11/14 data sources; where decreases occurred, they ranged from 0.4 to 8.2 percentage points (median 3.2 points). Over the same period, the proportion of extremely long gestational length episodes (>52 weeks) was unchanged in 4/14 data sources and increased in 10/14; most increases were small (0.1–0.2 percentage points), with the largest increases in SIDIAP (3.7) and FinOMOP-TaUH Pirha (8.1).

Overlapping pregnancy episodes (%) varied across data sources and runs (**Table 5**). In Run 1, overlap ranged from 0.7% to 30.1% across data sources. In Run 2, overlap ranged from 0.5% to 23.5%. Between Run 1 and 2, overlap decreased in CDW Bordeaux, DK-DHR, EMDB-ULSGE, FinOMOP-THL, IMASIS, InGef RDB, NAJS, and SIDIAP, and increased in the remaining data sources. Between Run 2 and 3, overlap decreased in all data sources (range 0.4%–6.7%).

Outcome type identification remained heterogeneous across data sources and runs (PREG ranges: Run 1 0–57.1%, Run 2 0–69.8%, Run 3 0–46.4%; DELIV ranges: Run 1 18.8–87.9%, Run 2 13.5–96.0%, Run 3 0.9–74.6%). From Run 1 to Run 2, the proportion of episodes ending as PREG (pregnancy identified, no recorded outcome) decreased in 3/15 data sources (BIFAP, FinOMOP-TaUH Pirha, and SIDIAP), while other sources showed no change or increased. From Run 2 to Run 3, the PREG category decreased in 11/14 sources and increased in 2/14 (HI-SPEED and SIDIAP).

For transparency on how iterative refinements were derived and applied, **Table 7** summarises the purpose of each run, the key issues identified, and the corresponding site-generic changes introduced before rerunning the algorithm across all data sources. Following the completion of Run 3, these outputs were used to create data source-specific summaries and evaluate the different output domains. Additionally, Run 3 outputs were used for population-level benchmarking against national statistics and for PET comparisons where available (reported in subsequent sections and in the Shiny app for full detail: [EUPAS1000000812](https://eupas1000000812)).

Table 5. Cross-database quality control metrics for inferred pregnancy episodes (15 data sources).

Version ²	Episodes	Percentage of Total Episodes (%)			Outcome (%) ¹		
		Plausible (0–308d) ³	>52 weeks ⁴	Overlapping ⁵	PREG	DELIV	PREG+DELIV
BIFAP							
Run 1	1,848,653	98.6	1.3	0.7	28.8	38.2	67
Run 2	2,064,880	98.3	0.1	7.1	3.1	42.2	45.3
Run 3	1,682,637	95.9	2.5	1.7	2.6	35.2	37.8
CDW Bordeaux							
Run 1	78,767	99.8	0	7.3	1.3	63.7	65
Run 2	108,599	98.5	0	3.2	2.2	70.4	72.6
Run 3	107,102	98.5	0	0.8	2.1	70.4	72.5
CPRD GOLD							
Run 1	3,762,235	100	0	4	0	87.9	87.9
Run 2	613,336	99.8	0	6.9	1.8	55.8	57.6
Run 3	691,169	99.8	0	1.9	1.1	48	49.1
DK-DHR							
Run 1	1,241,203	97.8	0.4	15	0.1	19.8	19.9
Run 2	1,325,143	99.2	0	3.9	6.1	50.7	56.8
Run 3	1,068,924	97.7	0.2	2.5	0.1	30.6	30.7
EMDB-ULSGE							
Run 1	29,213	99.9	0.1	4.6	0.8	48.9	49.7
Run 2	25,067	100	0	0.9	1.1	96.1	97.2
Run 3	†	†	†	†	†	†	†
FinOMOP-THL							
Run 1	566,804	99.2	0.1	5.2	0.2	50.9	51.1
Run 2	620,087	99.6	0	3.0	3.9	72.5	76.4
Run 3	856,955	96.4	0.1	2.4	0.1	49.6	49.7
FinOMOP-TaUH Pirha							

Version ²	Percentage of Total Episodes (%)				Outcome (%) ¹		
	Episodes	Plausible (0–308d) ³	>52 weeks ⁴	Overlapping ⁵	PREG	DELIV	PREG+DELIV
Run 1	91,725	98.5	0.1	2.9	19.2	59.5	78.7
Run 2	105,724	79.6	1.1	15.6	1.8	77.5	79.3
Run 3	80,247	73	8.1	4.3	1.8	74.6	76.4
HI-SPEED							
Run 1	1,004,560	99.7	0	6.7	0.1	70	70.1
Run 2	1,551,473	97.8	0	14.2	0.2	74.9	75.1
Run 3	1,135,090	96.6	0.1	2.3	0.3	60.2	60.5
IMASIS							
Run 1	79,623	93.5	2.6	12.1	38.3	22.3	60.6
Run 2	30,703	99.5	0	0.5	40.1	57.7	97.8
Run 3	33,251	95.5	0	0.4	27.4	1.3	28.7
IPCI							
Run 1	398,709	92.3	7.1	0.8	57.1	24.2	81.3
Run 2	412,555	99.4	0.2	7.4	69.8	21.9	91.7
Run 3	306,014	99	0.3	2.2	46.4	42.4	88.8
InGef RDB							
Run 1	695,105	99.8	0	30.1	0	23.5	23.5
Run 2	679,790	99.9	0	23.5	0.3	13.5	13.8
Run 3	640,749	91.7	0.2	6.7	0	2.4	2.4
NAJS							
Run 1	372,509	99.5	0	27.7	0.1	36.8	36.9
Run 2	287,852	99.5	0.1	5.6	21.5	33.1	54.6
Run 3	389,321	97.4	0.2	4.7	11.5	9.2	20.7
NLHR@Uio							
Run 1	820,061	99.3	0	7.5	0	18.8	18.8
Run 2	351,677	94.3	0	16.7	0.6	57.8	58.4
Run 3	400,865	94.9	0	3.4	0	0.9	0.9
SIDIAP							
Run 1	859,077	91.8	4.7	15.5	40.3	40.3	80.6
Run 2	791,748	92.5	0.2	6.9	10.9	35.6	46.5
Run 3	638,009	89.2	3.9	3	11.6	36.5	48.1
SUCD							
Run 1	163,002	99.1	0.3	2	13.5	30.7	44.2
Run 2	88,747	99.4	0	4.7	16.6	47.2	63.8
Run 3	129,588	93.8	0.1	1.5	5.8	5.4	11.2

NAJS: Croatian National Public Health Information System, DK-DHR: Danish Data Health Registries, FinOMOP-TaUH Pirha: Tampere University Hospital patient cohort, FinOMOP-THL: Finnish Care Register for Health Care, CDW Bordeaux: Clinical Data Warehouse of

Bordeaux University Hospital, InGef RDB: InGef Research Database, SUCD: Semmelweis University Clinical Data, IPCI: Integrated Primary Care Information, NLHR@UiO: Norwegian Linked Health Registry data, EMDB-ULSGE: Egas Moniz Health Alliance database - Gaia E Espinho, BIFAP: Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público, IMASIS: Institut Municipal Assistència Sanitària Information System, SIDIAP: The Information System for Research in Primary Care, HI-SPEED: Health Impact - Swedish Population Evidence Enabling Data-linkage, CPRD GOLD: Clinical Practice Research Datalink GOLD, PET: Perinatal Expansion Table.

Run 1– Run 3 denote the *PregnancyIdentifier* algorithm version corresponding to each iteration (Runs 1–3).

¹Outcomes include: PREG, which indicates a pregnancy episode was identified but, without a recorded outcome, gestational length is likely underestimated; DELIV, which confirms a delivery was recorded but the vital status of the birth (live birth or stillbirth) is unresolved; and PREG+DELIV represents the combined proportion of episodes not resolved to a specific pregnancy outcome (live birth, stillbirth, spontaneous abortion, induced abortion, or ectopic pregnancy). While a decrease in PREG is a primary quality signal, a persistently high DELIV proportion may reflect the limits of outcome coding granularity in a given source rather than poor algorithm performance.

²Version run.

³Plausible: episode duration ≥ 0 and ≤ 308 days.

⁴>52 weeks: episodes with gestational length >364 days.

⁵Overlap: episodes sharing any date range within the same person.

†Run 3 results do not include EMDB-ULSGE, as the CC lost communication with the DP, and no results were received.

Table 6. Summary of quality control (QC) metric improvements across runs.

QC metric	Run 1 → Run 2	Run 2 → Run 3*	Highlights
QC: Number of data sources where % of episodes >308 days decreased	8/15	1/14	Run 1 to Run 2: decreased in 7/15 data sources; 2/15 remained unchanged. Run 2 to Run 3: increased in 11/14 data sources, remained unchanged in 2/14, and decreased in 1/14.
QC: Number of data sources with a decrease in % of PREG (low-information outcome proxy)	3/15	11/14	Run 2 to 3: improved in 11/14 data sources, remaining unchanged in 1, and increasing in the remaining.
Context: Number of data sources where DELIV % increased	10/15	3/14	Direction not inherently better/worse.
Context: Number of data sources where DELIV+PREG % decreased	4/15	13/14	Context only.
QC: Number of data sources where % overlap decreased	8/15	14/14	Run 2 to 3: improved in all data sources.

Summary metrics are reported as the number of data sources for which a change in the metric was observed over the total participating data sources per run. DELIV and DELIV+PREG rows are provided for context because changes in episode construction and outcome assignment can redistribute episodes across end-type categories.

*Run 3 results do not include EMDB-ULSGE, so totals are over 14 data sources. PREG, indicates a pregnancy episode was identified, but without a recorded outcome, gestational length is likely underestimated. DELIV indicates a delivery was recorded, but the vital status of the birth (live birth or stillbirth) is unresolved. PREG+DELIV represents the combined proportion of episodes not resolved to a specific pregnancy outcome (live birth, stillbirth, spontaneous abortion, induced abortion, or ectopic pregnancy). While a decrease in PREG is a primary quality signal, a persistently high DELIV proportion may reflect the limits of outcome coding granularity in a given source rather than poor algorithm performance.

Table 7. Summary of the runs describing the main issues and corresponding changes to address them.

	Run 1	Run 2	Run 3
Purpose	Test execution of the algorithm across the data network and obtain baseline results.	Improved the algorithm with high-level solutions based on first aggregate-level results.	Further refinements in concept sets and logic based on more advanced test cases. Also incorporates the final validation with national statistics and PET comparisons (if available).
Main issues in current run	Extremely long gestational lengths	Implausible gestational lengths Many overlapping pregnancies Low granularity of outcome category (e.g. unknown/PREG)	Many overlapping pregnancies Low granularity of outcome category (e.g. PREG/DELIV)
Major changes of given run compared to previous run	–	Optimisation of concept sets (removal of overly generic concepts and addition of missing gestational indicators) Correction of gestational-age/date-difference logic ¹ (maximum 308 days) Alignment of episodes to the prespecified study period ²	Technical/QC issues and refinements ³ from earlier runs Addition of advanced test cases ⁴ (limited information, multiple episodes, late capture) Refinement of overlap resolution rules Final concept set refinement (see change log) Addition of attrition/cleanup visualisation Code refactoring to improve usability and maintenance

Summary of runs indicates the key changes and issues addressed between runs.

¹The original algorithm used 301 days as the maximum gestational-age threshold at specific steps in the algorithm logic. This was modified to 308 days in Run 2 to align with the prespecified plausibility window. This threshold is not applied as a global rule, episodes exceeding 308 days can still arise at certain merging steps. This is intentional: the aim is to flag implausible episodes as a QC signal rather than exclude or truncate them systematically.

²In Run 1, episode counts were not fully restricted to the prespecified study period, meaning episodes outside the intended analysis window were inadvertently included. This was corrected in Run 2 to ensure counts reflected only the intended study period and to support valid cross-source comparisons.

³Technical and QC issues and refinements from earlier runs included: gestational age value handling in the HIP stage; person-scoped overlap resolution in the merge stage; and handling of episode date validation filters. Full details are documented in the *PregnancyIdentifier* package and available in the Shiny pp under “Versions”.

⁴Advanced test cases refer to edge-case scenarios used to stress-test the algorithm beyond aggregate-level checks, including: episodes with limited pregnancy-related evidence; individuals with multiple closely spaced pregnancy episodes; and late-capture pregnancies where the first recorded evidence appeared late in gestation.

9.1.5. Internal consistency and plausibility checks in the data sources grouped by settings (primary care/secondary care/claims/registry)

This section summarises internal consistency and plausibility checks using Run 3 outputs, complemented by population-level benchmarking and PET comparisons, where available.

Across Run 3, the algorithm identified a total of 8,159,921 pregnancy episodes (range 33,251–1,682,637, excluding EMDB- ULSGE) among 5,566,270 individuals across 14 data sources. Summary metrics are shown in [Table 8](#).

Table 8. Run 3 summary metrics across participating data sources.

Domain	Summary across data sources
Episodes identified	8,159,921
Episodes identified per data source range	33,251–1,682,637
Individuals with ≥1 episode	5,566,270
Median maternal age at first pregnancy start (range, years)	29.0–32.9
Outcome distribution, range	
Livebirth (LB)	0.0–86.0%*
Delivery (DELIV)	0.9–74.6%
Spontaneous abortion (SA)	2.5–15.22%
Abortion (AB)	1.4–23.0%
Ectopic pregnancy (ECT)	0.7–2.2%
Stillbirth (SB)	0.2–0.65%
Pregnancy (PREG)	0.0–46.4%
Gestational length plausibility (0–308 days)	73.0–99.8%

Ranges reflect variation across data sources with Run 3 outputs available (Run 3 results do not include EMDB-ULSGE). Pregnancy (PREG): indicating that a pregnancy episode is identified, although the outcome of the episode is unknown; Delivery (DELIV): indicating evidence of a pregnancy episode ending in delivery, although the vital status of the delivery is unknown. More specific outcome categories include live birth (LB), stillbirth (SB), spontaneous abortion (SA), induced abortion (AB), and ectopic pregnancy (ECT).

*A range low of 0.0% for LB reflects that in some data sources, live birth-specific outcome coding was not available. In these sources, the combined LB+DELIV likely provides a more informative estimate of birth frequency.

To facilitate cross-database interpretation in a heterogeneous network, we present a standardised data source summary (“fingerprint”) for each data source, grouped by care setting (primary care, secondary care, claims, registry). Individual “fingerprints” were not explicitly prespecified in the protocol; they are used here as a structured way to present the protocol-defined evaluation domains consistently across data sources. Each fingerprint summarises: i) Episode metrics to assess internal consistency and plausibility of episodes, including episode outcomes Pregnancy (**PREG**): indicating that a pregnancy episode is identified, although the outcome of the episode is unknown; Delivery (**DELIV**): indicating evidence of a pregnancy episode ending in delivery, although the vital status of the delivery is unknown. More specific outcome categories include live birth (**LB**), stillbirth (**SB**), spontaneous abortion (**SA**), induced abortion (**AB**), and ectopic pregnancy (**ECT**); ii) Population-level validation, using national statistics; iii) Episode-level validation: PET comparison in sources with PET availability; and iv) hypothetical cleanup: to provide an approximation of higher quality episodes.

Population-level validation in this report is based on the most recent available reference values; additional years and supporting results are provided in the Shiny app ([EUPAS1000000812](https://eupas1000000812)).

Table 9 provides an overview of Run 3 across all data sources. **Table 10** summarises external benchmarking against national statistics (EUROSTAT, Euro-Peristat), where feasible. **Table 11** summarises PET comparisons for data sources with PET available.

9.1.6. Primary care

Netherlands: Integrated Primary Care Information (IPCI)

Episode-level internal consistency and plausibility. Run 3 identified 306,014 pregnancy episodes among 215,372 women (episodes per women with ≥1 episode: mean 1.42; median 1). Gestational length

plausibility was 99.0% (0–308 days), and 0.3% >52 weeks. Episode construction indicators showed 2.2% overlapping episodes. Specific outcomes were distributed as LB 0.11%, SB 0.65%, SA 6.4%, AB 3.4%, ECT 0.67%, DELIV 42.4%, and PREG 46.4%. Delivery mode was observed in episodes classified as DELIV (129,738) and was available in 7,552 episodes. Restricting to episodes with delivery mode recorded, 100% were C-section deliveries.

Population-level validation. The 2023 birth rate was 0 per 1,000 general population, compared with 9.2 in the reference statistics ([Table S1](#)); however, when LB+DELIV was used as the livebirth definition, the birth rate was 3.2. The maternal age at delivery (2023) was 30.6 years (reference 30.4 years, [Table S1](#)). Among births with delivery mode recorded, no episodes were classified as LB in this data source for 2019 (see [Table 3](#) for reference year rationale). Accordingly, the C-section rate was derived from the combined LB+DELIV category among births with recorded delivery mode; in 2019, this proportion was 100% (reference 16.4%, [Table S1](#)), which should be interpreted cautiously.

Hypothetical cleanup. Excluding overlaps and implausible durations (>308 days), 10,382 episodes (3.39% of the initially identified episodes) would be removed and 295,632 would remain.

Spain: Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público (BIFAP)

Episode-level internal consistency and plausibility. Run 3 identified 1,682,637 pregnancy episodes among 1,348,361 women (episodes per women with ≥ 1 episode: mean 1.24; median 1). Gestational length plausibility was 95.9% (0–308 days), and 2.5% >52 weeks. Episode construction indicators showed 1.7% overlapping episodes. Specific outcomes were distributed as LB 43.22%, SB 0.28%, SA 12.49%, AB 4.95%, ECT 1.27%, DELIV 35.19%, and PREG 2.6%. Delivery mode was available in episodes classified as LB (406,023/727,221) and DELIV (92,343/592,198). Restricting to episodes with delivery mode recorded, 10.63% and 92.90%, respectively, were C-section deliveries.

Population-level validation. The 2023 birth rate was 3.3 per 1,000 general population, compared with 6.6 in the reference statistics ([Table S1](#)); however, when LB+DELIV was used as the livebirth definition, the birth rate was 4.3. The maternal age at delivery (2023) was 32.1 years (reference 31.5 years, [Table S1](#)). Among births with delivery mode recorded, the C-section rate in 2019 was 9.1% based on LB episodes alone and 14.1% when calculated using the combined LB+DELIV category (reference 18.2%, [Table S1](#)).

Hypothetical cleanup. Excluding overlaps and implausible durations (>308 days), 86,531 episodes (5.1% of the initially identified episodes) would be removed and 1,596,106 would remain.

Spain: The Information System for Research in Primary Care (SIDIAP)

Episode-level internal consistency and plausibility. Run 3 identified 638,009 pregnancy episodes among 508,160 women (episodes per woman with ≥ 1 episode: mean 1.25; median 1). Gestational length plausibility was 89.2% (0–308 days), and 3.9% >52 weeks. Overlap occurred in 3.0% of episodes. Specific outcomes were distributed as LB 31.5%, SB 0.30%, SA 6.2%, AB 12.7%, ECT 1.2%, DELIV 36.5%, and PREG 11.6%. Delivery mode was available in episodes classified as LB (173,142/201,225) and DELIV (83,415/232,699). Restricting to episodes with delivery mode recorded, 19.95% and 52.50%, respectively, were C-section deliveries.

Population-level validation. The 2023 birth rate was 4.4 per 1,000 general population (reference 6.6, [Table S1](#)). However, when LB+DELIV was used as the livebirth definition, the birth rate was 4.5. The maternal age at delivery (2023) was 30.7 years (reference 31.5 years, [Table S1](#)). Among LB episodes with delivery mode recorded, the C-section rate (2019) was 19.2 % (reference 18.2%, [Table S1](#)).

Episode-level validation using PET. The algorithm identified 638,009 episodes, while the PET identified 600,805, with 404,471 episodes identified by both (233,538 algorithm-only; 196,334 PET-only), corresponding to a sensitivity 67.3% and PPV 63.4% (PET as reference).

Hypothetical cleanup. Excluding overlaps and implausible durations (>308 days), 80,677 episodes (12.65% of the initially identified episodes) would be removed and 557,332 episodes would remain.

United Kingdom: Clinical Practice Research Datalink GOLD (CPRD GOLD)

Episode-level internal consistency and plausibility. Run 3 identified 691,169 pregnancy episodes among 501,050 women with ≥ 1 episode (episodes per woman with ≥ 1 episode: mean 1.38; median 1). Gestational length plausibility was 99.8% (0–308 days), and 0.0% >52 weeks. Episode construction indicators showed 1.9% overlapping episodes. Specific outcomes were distributed as LB 20.4%, SB 0.25%, SA 15.2%, AB 13.6%, ECT 1.3%, DELIV 48.0%, and PREG 1.1%. Delivery mode was recorded for 146,927/332,016 (44.3%) episodes classified as DELIV and 102,958/141,192 (72.9%) episodes classified as LB. Restricting to episodes with delivery mode recorded, 38.67% (DELIV) and 4.82% (LB) were C-section deliveries, respectively.

Population-level validation. The 2023 birth rate was 1.7 per 1,000 general population (reference 10.0, [Table S1](#)). However, when LB+DELIV definition was used, the rate was 4.6. The maternal age at delivery (2023) was 29.9 years (reference 30.9, [Table S1](#)). Among births with delivery mode recorded, the C-section rate in 2019 was 4.7% based on LB episodes alone and 26.4% when calculated using the combined LB+DELIV category (reference 32.2%, [Table S1](#)).

Episode-level validation using PET. PET comparison showed that the algorithm identified 691,169 episodes, while PET identified 506,847, with 405,778 episodes identified by both (285,391 algorithm-only; 101,069 PET-only), corresponding to a sensitivity 80.1% and PPV 58.7% (PET as reference).

Hypothetical cleanup. excluding overlaps and implausible durations (>308 days), 16,222 episodes (2.35% of the initially identified episodes) would be removed and 674,947 would remain.

9.1.7. Secondary care

Finland: Tampere University Hospital patient cohort (FinOMOP-TaUH Pirha)

Episode-level internal consistency and plausibility. Run 3 identified 80,247 pregnancy episodes among 52,072 women with ≥ 1 episode (episodes per woman with ≥ 1 episode: mean 1.54; median 1). Gestational length plausibility was 73.0% (0–308 days), and 8.1% >52 weeks. Episode construction indicators showed 4.3% overlapping episodes. Specific outcomes were distributed as LB 0.29%, SB 0.25%, SA 9.9%, AB 11.4%, ECT 1.8%, DELIV 74.6%, and PREG 1.8%. Delivery mode was recorded for 22,735/59,872 (37.97%) episodes classified as DELIV and 56/235 (23.83%) episodes classified as LB. Restricting to episodes with delivery mode recorded, 43.07% (DELIV) and 37.50% (LB) were C-section deliveries, respectively.

Population-level validation. The 2023 birth rate was 0 per 1,000 general population (reference 7.8, [Table S1](#)). However, when LB+DELIV were used, the rate was 7.6. The maternal age at delivery (2023) was 29.6 years (reference 30.0, [Table S1](#)). Among births with delivery mode recorded, no episodes were classified as LB in this data source for 2019. Accordingly, the C-section rate was derived from the combined LB+DELIV category among births with recorded delivery mode; in 2019, this proportion was 97.3% (reference 20.9%, [Table S1](#)), which should be interpreted cautiously.

Hypothetical cleanup. Excluding overlaps and implausible durations (>308 days), 29,437 episodes (36.57% of the initially identified episodes) would be removed (remaining 50,810).

France: Clinical Data Warehouse of Bordeaux University Hospital (CDW Bordeaux)

Episode-level internal consistency and plausibility. Run 3 identified 107,102 pregnancy episodes among 78,871 women with ≥ 1 episode (episodes per woman with ≥ 1 episode: mean 1.35; median 1). Gestational length plausibility was 98.5% (0–308 days) and 0.0% >52 weeks. Episode construction indicators showed 0.8% overlapping episodes. Specific outcomes were distributed as LB 0.03% (see [Table 8](#) footnote), SB 0.31%, SA 2.5%, AB 23.0%, ECT 1.7%, DELIV 70.4%, and PREG 2.1%. Delivery mode was recorded for 65,384/75,359 (86.76%) episodes classified as DELIV and 19/32 (59.38%) episodes classified as LB.

Restricting to episodes with delivery mode recorded, 10.83% (DELIV) and 26.32% (LB) were C-section deliveries, respectively.

Population-level validation. The 2023 birth rate was 0 per 1,000 general population (reference 9.9, [Table S1](#)); however, when the livebirth definition combined LB+DELIV, the birth rate was 5.16. The maternal age at delivery (2023) was 29.5 years (reference 29.1, [Table S1](#)). Among births with delivery mode recorded, the C-section rate in 2019 was 100% based on LB episodes alone (reference 31.8%), but 13.3% when the denominator included LB+DELIV category.

Hypothetical cleanup. excluding overlaps and implausible durations (>308 days), 2,478 episodes (2.31% of the initially identified episodes) would be removed and 104,624 would remain.

[Hungary: Semmelweis University Clinical Data \(SUCD\)](#)

Episode-level internal consistency and plausibility. Run 3 identified 129,588 pregnancy episodes among 94,777 women with ≥ 1 episode (episodes per woman with ≥ 1 episode: mean 1.37; median 1). Gestational length plausibility was 93.8% (0–308 days), and 0.1% >52 weeks. Episode construction indicators showed 1.5% overlapping episodes. Specific outcomes were distributed as LB 59.1%, SB 0.29%, SA 10.7%, AB 17.5%, ECT 1.2%, DELIV 5.4%, PREG 5.8%. Delivery mode was recorded for 1,021/7,010 (14.56%) episodes classified as DELIV and 47,207/76,541 (61.68%) episodes classified as LB. Restricting to episodes with delivery mode recorded, 99.51% (DELIV) and 81.14% (LB) were C-section deliveries, respectively.

Population-level validation. The 2023 birth rate was 5.65 per 1,000 general population (reference 9.1, [Table S1](#)); however, when the livebirth definition combined LB+DELIV, the birth rate was 6.47. The maternal age at delivery (2023) was 31.5 years (reference 28.8, [Table S1](#)). Among LB episodes with delivery mode recorded, the C-section rate (2019) was 81.2% (reference 17.4, [Table S1](#)).

Hypothetical cleanup. Excluding overlaps and implausible durations (>308 days), 10,800 episodes (8.33% of the initially identified episodes) would be removed and 118,788 would remain.

[Spain: Institut Municipal Assistència Sanitària Information System \(IMASIS\)](#)

Episode-level internal consistency and plausibility. Run 3 identified 33,251 pregnancy episodes among 26,286 women with ≥ 1 episode (episodes per woman with ≥ 1 episode: mean 1.26; median 1). Gestational length plausibility was 95.5% (0–308 days), and 0.0% >52 weeks. Episode construction indicators showed 0.4% overlapping episodes. Specific outcomes were distributed as LB 54.0%, SB 0.23%, SA 12.6%, AB 2.8%, ECT 1.7%, DELIV 1.3%, and PREG 27.4%. Delivery mode was recorded for 21/428 (4.91%) episodes classified as DELIV and 11,816/17,938 (65.9%) episodes classified as LB. Restricting to episodes with delivery mode recorded, 90.5% (DELIV) and 33.2% (LB) were C-section deliveries, respectively.

Population-level validation. The 2023 birth rate was 1.96 per 1,000 general population (reference 6.6, [Table S1](#)); however, when the livebirth definition combined LB+DELIV, the rate was 2.0. Maternal age at delivery (2023) was 31.3 years (reference 31.5, [Table S1](#)). Among LB episodes with delivery mode recorded, the C-section rate (2019) was 19.9% (reference 18.2%, [Table S1](#)).

Episode-level validation using PET. PET comparison showed that the algorithm identified 33,251 episodes, while PET identified 20,429, with 18,782 episodes identified by both (14,469 algorithm-only; 1,647 PET-only), corresponding to a sensitivity 91.9% and a PPV 56.5% (PET as reference).

Hypothetical cleanup. Excluding overlaps and implausible durations (>308 days), 2,066 episodes (6.21% of the initially identified episodes) would be removed and 31,185 would remain.

9.1.8. Claims

Germany: InGef Research Database (InGef RDB)

Episode-level internal consistency and plausibility. Run 3 identified 640,749 pregnancy episodes among 441,969 women with ≥ 1 episode (episodes per woman with ≥ 1 episode: mean 1.45; median 1). Gestational length plausibility was 91.7% (0–308 days), and 0.2% >52 weeks. Episode construction indicators showed 6.7% overlapping episodes. Specific outcomes were distributed as LB 86.0%, SB 0.41%, SA 8.1%, AB 1.4%, ECT 1.7%, DELIV 2.4%, and PREG 0.0%. Delivery mode was recorded for 620/15,591 (3.98%) episodes classified as DELIV and 395,435/550,925 (71.8%) episodes classified as LB. Restricting to episodes with delivery mode recorded, 71.7% (DELIV) and 40.6% (LB) were C-section deliveries, respectively.

Population-level validation. The 2023 birth rate was 6.8 per 1,000 general population (reference 8.3, [Table S1](#)); however, when the livebirth definition combined LB+DELIV, the rate was 7.0. Maternal age at delivery (2023) was 31.0 years (reference 29.8, [Table S1](#)). Among LB episodes with delivery mode recorded, the C-section rate (2019) was 39.2% (reference 41.5, [Table S1](#)).

Hypothetical cleanup. Excluding overlaps and implausible durations (>308 days), 59,193 episodes (9.24% of the initially identified episodes) would be removed and 581,556 would remain.

9.1.9. Registry-based

Croatia: Croatian National Public Health Information System (NAJS)

Episode-level internal consistency and plausibility. Run 3 identified 389,321 pregnancy episodes among 255,849 women with ≥ 1 episode (episodes per woman with ≥ 1 episode: mean 1.52; median 1). Gestational length plausibility was 97.4% within 0–308 days, 2.6% >308 days, and 0.3% >52 weeks. Episode construction indicators showed 4.7% overlapping episodes. Specific outcomes were distributed as LB 60.4%, SB 0.22%, SA 11.0%, AB 5.5%, ECT 2.2%, DELIV 9.2%, PREG 11.5%. Delivery mode was recorded for 18,917/35,653 (53.06%) episodes classified as DELIV and 202,074/235,222 (85.91%) episodes classified as LB. Restricting to episodes with delivery mode recorded, 95.5% (DELIV) and 22.2% (LB) were C-section deliveries, respectively.

Population-level validation. The 2023 birth rate was 6.5 per 1,000 general population (reference 8.3, [Table S1](#)); however, when the livebirth definition combined LB+DELIV, the rate was 7.7. The maternal age at delivery (2023) was 32.6 years (reference 29.3, [Table S1](#)). Among births with delivery mode recorded, the C-section rate in 2019 was 20.0% when restricted to LB episodes and 25.6% when calculated using the combined LB+DELIV category (reference 26.2, [Table S1](#)).

Hypothetical cleanup. Excluding overlaps and implausible durations (>308 days), 26,403 episodes (6.8% of the initially identified episodes) would be removed and 362,918 would remain.

Denmark: Danish Data Health Registries (DK-DHR)

Episode-level internal consistency and plausibility. Run 3 identified 1,068,924 pregnancy episodes among 581,154 women with ≥ 1 episode (episodes per woman with ≥ 1 episode: mean 1.84; median 2). Gestational length plausibility was 97.7% (0–308 days), and 0.2% >52 weeks. Episode construction indicators showed 2.5% overlapping episodes. Specific outcomes were distributed as LB 44.5%, SB 0.27%, SA 9.1%, AB 14.0%, ECT 1.4, DELIV 9.2%, and PREG 4.7%. Delivery mode was recorded for 99,092/ 327,139 (30.29%) episodes classified as DELIV and 146,520/ 475,845 (30.8%) episodes classified as LB. Restricting to episodes with delivery mode recorded, 63.06% (DELIV) and 65.9% (LB) were C-section deliveries, respectively.

Population-level validation. Population-level benchmarking, the 2023 birth rate was 0.0 per 1,000 general population (reference 9.7); however, when the livebirth definition combined LB+DELIV, the rate was 9.0. Maternal age at birth (2023) was 29.1 years (reference 30.1). Among LB episodes with delivery mode recorded, the 2019 C-section rate was 65.4% (reference 17.9%, [Table S1](#)).

Hypothetical cleanup. Excluding overlaps and implausible durations (>308 days), 50,629 episodes (4.7% of the initially identified episodes) would be removed (remaining 1,018,295).

Finland: Finnish Care Register for Health Care (FinOMOP-THL)

Episode-level internal consistency and plausibility. Run 3 identified 856,955 pregnancy episodes among 470,190 women with ≥ 1 episode (episodes per woman with ≥ 1 episode: mean 1.82; median 2). Gestational length plausibility was 96.4% (0–308 days), and 0.1% >52 weeks. Episode construction indicators showed 2.4% overlapping episodes. Specific outcomes were distributed as LB 21.4%, SB 0.23%, SA 11.8%, AB 15.1%, ECT 1.9%, DELIV 49.6%, and PREG 0.1%. Delivery mode was recorded for 136,348/424,771 (32.10%) episodes classified as DELIV and 117,921/183,132 (64.4%) episodes classified as LB. Restricting to episodes with delivery mode recorded, 21.2% (DELIV) and 5.6% (LB) were C-section deliveries, respectively.

Population-level validation. The 2023 birth rate was 0.7 per 1,000 general population (reference 7.8, [Table S1](#)); however, when the livebirth definition combined LB+DELIV, the rate was 7.1. The maternal age at delivery (2023) was 29.2 years (reference 30.0, [Table S1](#)). Among births with delivery mode recorded, the C-section rate in 2019 was 8.3% when restricted to LB episodes and 13.3% when calculated using the combined LB+DELIV category (reference 20.7, [Table S1](#)).

Hypothetical cleanup. Excluding overlaps and implausible durations (>308 days), 52,034 episodes (6.1% of the initially identified episodes) would be removed (remaining 804,921).

Norway: Norwegian Linked Health Registry data (NLHR@UiO)

Episode-level internal consistency and plausibility. Run 3 identified 400,865 pregnancy episodes among 277,819 women with ≥ 1 episode (episodes per woman with ≥ 1 episode: mean 1.44; median 1). Gestational length plausibility was 94.9% (0–308 days), and 0.0% >52 weeks. Episode construction indicators showed 3.4% overlapping episodes. Specific outcomes were distributed as LB 67.8%, SB 0.48%, SA 14.6%, AB 14.2%, ECT 1.9%, DELIV 0.9%, and PREG 0.0%. Delivery mode was recorded for 69/3,518 (1.96%) episodes classified as DELIV and 25,494/271,945 (9.4%) episodes classified as LB. Restricting to episodes with delivery mode recorded, 100.0% (DELIV) and 100.0% (LB) were C-section deliveries, respectively.

Population-level validation. The 2023 birth rate was 8.6 per 1,000 general population (reference 9.4, [Table S1](#)). However, when the livebirth definition combined LB+DELIV, the rate was 8.8. Maternal age at birth (2023) was 29.9 years (reference 30.1, [Table S1](#)). Among births with delivery mode recorded, the C-section rate in 2019 was 100% when restricted to LB episodes and remained 100% when calculated using the combined LB+DELIV category (reference 16.0%, [Table S1](#)). However, delivery mode was recorded in only 2 % of episodes, representing a very small and likely non-representative subset of all deliveries in this source. This C-section rate should therefore not be interpreted as population-representative and is reported for completeness only.

Episode-level validation using PET. PET comparison showed that the algorithm identified 400,865 episodes, while PET identified 463,400 episodes, with 390,113 episodes identified by both (10,752 algorithm-only; 73,287 PET-only), corresponding to a sensitivity 84.2% and PPV 97.3% (PET as reference).

Hypothetical cleanup. Excluding overlaps and implausible durations (>308 days), 34,282 episodes (8.6% of the initially identified episodes) would be removed (remaining 366,583).

Sweden: Health Impact - Swedish Population Evidence Enabling Data-linkage (HI-SPEED)

Episode-level internal consistency and plausibility. Run 3 identified 1,135,090 pregnancy episodes among 714,640 women with ≥ 1 episode (episodes per woman with ≥ 1 episode: mean 1.59; median 1). Gestational length plausibility was 96.6% (0–308 days), and 0.1% >52 weeks. Episode construction indicators showed 2.3% overlapping episodes. Specific outcomes were distributed as LB 8.9%, SB 0.27%, SA 12.2, AB 16.6%, ECT 1.5%, DELIV 60.2%, and PREG 0.3%. Delivery mode was recorded for 107,100/682,805 (15.69%)

episodes classified as DELIV and 18,337/101,164 (18.1%) episodes classified as LB. Restricting to episodes with delivery mode recorded, 0.38% (DELIV) and 0.43% (LB) were C-section deliveries, respectively.

Population-level validation. In 2023, the inferred birth rate was 0.9 per 1,000 general population (reference 9.5, [Table S1](#)). However, when the livebirth definition combined LB+DELIV, the rate was 7.3. The maternal age at delivery (2023) was 29.9 years (reference 30.0, [Table S1](#)). Among births with delivery mode recorded, the C-section rate in 2019 was 0.4% when restricted to LB episodes and 0.5% when calculated using the combined LB+DELIV category, which is inconsistent with the national reference rate (reference 25.7%, [Table S1](#)).

Episode-level validation using PET. PET in this data source is restricted to episodes from 22 weeks of gestation onward. PET comparison showed that the algorithm identified 1,135,090 episodes, while PET identified 990,989, with 782,869 episodes identified by both (352,221 algorithm-only; 208,120 PET-only), corresponding to a sensitivity 79.0% and PPV 69.0% (PET as reference).

Hypothetical cleanup. Excluding overlaps and implausible durations (>308 days), 67,196 episodes (5.9% of the initially identified episodes) would be removed (remaining 1,067,894).

Table 9. Run 3 overview of episode-level QC metrics by data source (grouped by setting).

Data source ¹	Episodes (N)	Plausible 0–308d (%) ²	>52w (%)	Overlap (%)	PREG (%)	DELIV (%)	DELIV+PREG (%)	PET available (Y/N)
Primary care								
IPCI	306,014	99	0.3	2.2	46.4	42.4	88.8	N
SIDIAP	638,009	89.2	3.9	3	11.6	36.5	48.1	Y
BIFAP	1,682,637	95.9	2.5	1.7	2.6	35.2	37.8	N
CPRD GOLD	691,169	99.8	0	1.9	1.1	48	49.1	Y
Secondary care								
CDW Bordeaux	107,102	98.5	0	0.8	2.1	70.4	72.5	N
FinOMOP-TaUH Pirha	80,247	73	8.1	4.3	1.8	74.6	76.4	N
SUCD	129,588	93.8	0.1	1.5	5.8	5.4	11.2	N
EMDB-ULSGE	†	†	†	†	†	†	†	N
IMASIS	33,251	95.5	0	0.4	27.4	1.3	28.7	Y
Claims								
InGef RDB	640,749	91.7	0.2	6.7	0	2.4	2.4	N
Registry-based								
NAJS	389,321	97.4	0.2	4.7	11.5	9.2	20.7	N
DK-DHR	1,068,924	97.7	0.2	2.5	0.1	30.6	30.7	N
FinOMOP-THL	856,955	96.4	0.1	2.4	0.1	49.6	49.7	N
NLHR@UiO	400,865	94.9	0	3.4	0	0.9	0.9	Y
HI-SPEED	1,135,090	96.6	0.1	2.3	0.3	60.2	60.5	Y

NAJS: Croatian National Public Health Information System, DK-DHR: Danish Data Health Registries, FinOMOP-TaUH Pirha: Tampere University Hospital patient cohort, FinOMOP-THL: Finnish Care Register for Health Care, CDW Bordeaux: Clinical Data Warehouse of Bordeaux University Hospital, InGef RDB: InGef Research Database, SUCD: Semmelweis University Clinical Data, IPCI: Integrated Primary Care Information, NLHR: Norwegian Linked Health Registry data, EMD-ULSGE: Egas Moniz Health Alliance database - Gaia E Espinho, BIFAP: Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público, IMASIS: Institut Municipal Assistència Sanitària Information System, SIDIAP: The Information System for Research in Primary Care, HI-SPEED: Health Impact -

Swedish Population Evidence Enabling Data-linkage, CPRD GOLD: Clinical Practice Research Datalink GOLD, PET: Perinatal Expansion Table, LB: livebirth, DELIV: Delivery, PREG: Pregnancy.

¹Run 1–Run 3 denote the *PregnancyIdentifier* algorithm version corresponding to each iteration (Runs 1–3).

²Plausible gestational length was defined as >0 and ≤308 days. '>52w' was defined as >364 days. Overlap denotes the percentage of episodes overlapping in time for the same person. PREG indicates a pregnancy episode was identified, but without a recorded outcome, gestational length is likely underestimated. DELIV indicates a delivery was recorded, but the vital status of the birth (live birth or stillbirth) is unresolved. PREG+DELIV represents the combined proportion of episodes not resolved to a specific pregnancy outcome (live birth, stillbirth, spontaneous abortion, induced abortion, or ectopic pregnancy). While a decrease in PREG is a primary quality signal, a persistently high DELIV proportion may reflect the limits of outcome coding granularity in a given source rather than poor algorithm performance. PET availability indicates whether a Perinatal Expansion Table was available for comparison.

†Run 3 results do not include EMDB-ULSGE, as the CC lost communication with the DP, and no results were received.

Table 10. Population-level benchmarking by data source (grouped by setting).

Data source	Birth rate (per 1,000)	Birth rate (LB+DELIV) (per 1,000)	External Reference birth rate (per 1,000)	Maternal age (mean)	External Reference age	C-section* (%)	C-section * (LB+DELIV) (%)	External Reference C-section rate (%)
Primary care								
IPCI	0	3.2	9.2	30.6	30.4	–	100	16.4
BIFAP	3.3	4.3	6.6	32.1	31.5	9.1	14.1	18.2
SIDIAP	4.4	4.5	6.6	30.7	31.5	19.2	21.0	18.2
CPRD GOLD	1.7	4.6	10	29.9	30.9	4.7	26.4	32.2
Secondary care								
FinOMOP-TaUH Pirha	0	7.6	7.8	29.6	30	–	97.3	20.9
CDW Bordeaux	0	5.2	9.9	29.5	29.1	100.0	13.3	31.8
SUCD	5.7	6.5	9.1	31.5	28.8	81.2	81.5	17.4
EMDB-ULSGE-	†	†	†	†	†	†	†	†
IMASIS	2.0	2.0	6.6	31.3	31.5	19.9	20.0	18.2
Claims								
InGef RDB	6.8	7.0	8.3	31.0	29.8	39.2	39.3	41.5
Registry								
NAJS	6.5	7.7	8.3	32.6	29.3	20.0	25.6	26.2
DK-DHR	0	9.0	9.7	29.1	30.1	65.4	62.4	17.9
FinOMOP-THL	0.7	7.1	7.8	29.2	30	8.3	13.3	20.7
NLHR@UiO	8.6	8.8	9.4	29.9	30.1	100	100	16
HI-SPEED	0.9	7.3	9.5	29.9	30	0.4	0.5	25.7

External references used in this table are populated in accordance with [Table 3](#) and specified in [Table S1](#).

NAJS: Croatian National Public Health Information System, DK-DHR: Danish Data Health Registries, FinOMOP-TaUH Pirha: Tampere University Hospital patient cohort, FinOMOP-THL: Finnish Care Register for Health Care, CDW Bordeaux: Clinical Data Warehouse of Bordeaux University Hospital, InGef RDB: InGef Research Database, SUCD: Semmelweis University Clinical Data, IPCI: Integrated Primary Care Information, NLHR@UiO: Norwegian Linked Health Registry data, EMDB-ULSGE: Egas Moniz Health Alliance database - Gaia E Espinho, BIFAP: Base de Datos para la Investigación Farmacoe epidemiológica en el Ámbito Público, IMASIS: Institut

Municipal Assistència Sanitària Information System, SIDIAP: The Information System for Research in Primary Care , HI-SPEED: Health Impact - Swedish Population Evidence Enabling Data-linkage, CPRD GOLD: Clinical Practice Research Datalink GOLD, PET: Perinatal Expansion Table.

LB: livebirth. DELIV: delivery.

*The C-section rate was calculated among births with delivery mode recorded, using the relevant birth definition (LB only, or LB+DELIV where applicable) whilst births without recorded delivery mode were excluded from the denominator.

“—” indicates C-section rate among LB was not calculated for data sources in which no LB episodes were identified.

†Run 3 results do not include EMDB-ULSGE, as the CC lost communication with the DP, and no results were received.

Table 11. PET comparison summary by data source (grouped by setting).

Data source	Algorithm episodes	PET episodes	Matched	Algorithm-only	PET-only	Sensitivity (%)	PPV (%)	Start date difference (days) median (IQR)	Start = 0d (%)	End date difference (days) median (IQR)	End = 0d (%)	Outcome agreement (accuracy) * (0–1)
Primary care												
SIDIAP	638,009	600,805	404,471	233,538	196,334	67.3	63.4	-1 (-4 to 31)	8.4	0 (0 to 1)	50.2	0.9
CPRD GOLD	691,169	506,847	405,778	285,391	101,069	80.1	58.7	28 (24 to 57)	0.1	0 (0 to 0)	89.6	0.9
Secondary care												
IMASIS	33,251	20,429	18,782	14,469	1,647	91.9	56.5	13 (2 to 28)	9.8	0 (0 to 1)	61	1
Registry												
NLHR@UiO	400,865	463,400	390,113	10,752	73,287	84.2	97.3	36 (26 to 67)	0.1	0 (0 to 1)	67.2	1
HI-SPEED**	1,135,090	990,989	782,869	352,221	208,120	79.0	69.0	28 (23 to 35)	0	0 (0 to 1)	57.9	1

NAJS: Croatian National Public Health Information System, DK-DHR: Danish Data Health Registries, FinOMOP-TaUH Pirha: Tampere University Hospital patient cohort, FinOMOP-THL: Finnish Care Register for Health Care, CDW Bordeaux: Clinical Data Warehouse of Bordeaux University Hospital, InGef RDB: InGef Research Database, SUCD: Semmelweis University Clinical Data, IPCI: Integrated Primary Care Information, NLHR@UiO: Norwegian Linked Health Registry data, EMDB-ULSGE: Egas Moniz Health Alliance database - Gaia E Espinho, BIFAP: Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público, IMASIS: Institut Municipal Assistència Sanitària Information System, SIDIAP: The Information System for Research in Primary Care, HI-SPEED: Health Impact - Swedish Population Evidence Enabling Data-linkage, CPRD GOLD: Clinical Practice Research Datalink GOLD, PET: Perinatal Expansion Table.

LB: livebirth. DELIV: delivery. PREG: Pregnancy.

PREG indicates a pregnancy episode was identified, but without a recorded outcome, gestational length is likely underestimated. DELIV indicates a delivery was recorded, but the vital status of the birth (live birth or stillbirth) is unresolved. PREG+DELIV represents the combined proportion of episodes not resolved to a specific pregnancy outcome (live birth, stillbirth, spontaneous abortion, induced abortion, or ectopic pregnancy). DELIV proportion may reflect the limits of outcome coding granularity in a given source rather than poor algorithm performance.

PPV: Positive predictive value.

Sensitivity and PPV were estimated using PET as the reference.

*Outcome agreement (accuracy) was assessed only for outcome categories directly represented in PET. Episodes classified as ECT or Unknown were excluded as no corresponding PET categories were available. As PET represents live birth using the combined LB+DELIV category, LB and DELIV episodes were assessed jointly. Because excluded categories are those where algorithm-PET disagreement would be most likely, accuracy estimates apply only to the subset of directly comparable outcomes and should be interpreted as an upper bound of true agreement across all identified episodes.

**In HI-SPEED, PET was restricted to episodes from >22 weeks of gestation onward. Because comparison was conducted generically at the PET level, agreement estimates should be interpreted with caution.

10. DISCUSSION

10.1. Key results

We executed and optimised the Jones et al. algorithm (Smith et al. OMOP implementation), which we incorporated into the developed *PregnancyIdentifier* R package, by executing it across 15 data sources of the DARWIN EU® Data Network over three iterative runs. Input concepts were available across data sources for both HIP concepts and PPS concepts, supporting the feasibility of network implementation, although their extent was variable.

Pregnancy episode counts varied across data sources, with Run 3 identifying 8,159,921 pregnancy episodes among 5,566,270 individuals from 14 data sources (Run 3 range 33,251–1,682,637 episodes per data source).

Run 3 gestational length plausibility (0–308 days) was generally high, exceeding 90% in 12 of the 14 data sources (range 73.0–99.8%) whilst pregnancy episode overlap was generally low ranging from 0.4–6.7%. Pregnancy outcome identification specificity varied, with episodes having a PREG outcome (indicates that a pregnancy episode was identified, but without a recorded outcome) ranging from 0–46.4% across 14 sources.

Population-level benchmarking against national statistics showed that birth rates based on a combined measure using LB+DELIV were generally more comparable to national reference values than using only LB-based rates. Maternal age at pregnancy start was also broadly consistent with reference values. C-section rates were the most heterogeneous metric across data sources. Episode-level validation against the PET, performed in 5 data sources, showed sensitivity ranging from 67.3–91.9% and PPV ranging from 58.7–97.3%. Among these, SIDAP had the lowest concordance (sensitivity 67.3%, PPV 63.4%).

10.2. Strengths and limitations of the research methods

This study has several key strengths. First, we developed a standard analytical package, *PregnancyIdentifier*, which, when used together with the proposed evaluation framework, enabled a systematic evaluation across heterogeneous OMOP-mapped data sources. This supported consistent reporting of pregnancy episode metrics and outcome categories. The harmonised framework also enabled consistent QC, internal validation against PET comparisons, and external validation through population-level benchmarking. Second, we optimised an existing algorithm with incremental performance improvements across most participating sites over three runs. Third, we performed multi-site validation to evaluate algorithm transportability, addressing a key gap in the literature and informing readiness of the DARWIN EU® Data Network for pregnancy studies.

Several limitations apply. First, while run-level QC diagnostics identified the main issues driving suboptimal algorithm performance across iterations (see [Table 7](#)), we did not investigate site-specific factors underlying performance differences across sites. Because data were collected for care and administration rather than research, completeness and accuracy may vary across data sources. This is especially relevant for the inputs used to construct pregnancy episodes, including gestational age signals, pregnancy outcomes, and prenatal care. Capture depends on healthcare contact, coding practices, and the setting of each data source, which affects episode construction and comparability. Outcome granularity also varied across data sources. In some data sources, episodes where a pregnancy was identified but no specific outcome could be assigned (PREG) were more common. This may limit live birth versus pregnancy-loss analyses unless analyses are restricted to episodes with known outcomes (e.g. excluding PREG episodes or requiring a recorded delivery or loss code). In some data sources, sparse gestational dating concepts and fixed-value clustering of inferred durations suggest reliance on back-calculation assumptions. PET comparisons, where available, showed that start date agreement varied, with some sources performing well, while others showed non-trivial differences, consistent with this constraint and highlighting where

timing refinement remains limited. Differences in care pathways and care settings (e.g. primary care versus hospital capture; inpatient versus outpatient recording; public versus private service use) likely influenced which pregnancy-related events were observed and how completely episodes could be constructed, contributing to residual heterogeneity despite site-generic optimisation. In addition, data sources differed in scope and structure (national versus regional coverage; varying contributing providers), and some records reflected administrative interactions with limited clinical detail, which can lead to incomplete pregnancy episode construction. Addressing these limitations will require different approaches depending on their underlying cause: some may be mitigated through further algorithm development (e.g. refining back-calculation logic or overlap resolution); others through mother-child linkage or linkage to complementary sources (e.g. birth registries or maternity records); and others only through improvements in source data recording quality, which lies outside of the algorithm's scope.

Second, due to the study timelines, iterations were limited to three runs rather than continuing until no further site-generic improvements could be achieved, as originally planned. Run 3 supports the prespecified final validation and enables assessment of data source sufficiency using available QC and benchmarking results. Where implausible durations, overlap, or low-information outcome categories persist, further refinement may still be beneficial, but only where issues reflect addressable algorithm behaviour rather than underlying data quality limitations. The hypothetical cleanup step, excluding overlapping episodes and episodes with implausible length, is already quantified in the data source fingerprints and represents a possible approach when designing studies. Additionally, direct comparability of some QC metrics across iterative runs was limited because selected implementation issues were identified and corrected during the QC process, including concept-set specification issues, study-period alignment for episode counts, and the denominator used for overlap estimation.

Third, the algorithm was designed to apply site-generic rules across all sources. No site-specific hierarchies were implemented. NLHR illustrates what this means in practice: PET pre-processing is done using the UiO algorithm, which applies source-specific prioritisation rules that prioritise registry data and enrich episodes with primary and secondary care records depending on gestational week and available evidence, a level of local specificity the generic algorithm cannot replicate. Whether this could be addressed within an OMOP-based implementation is uncertain, as coding may not preserve sufficient record-origin metadata to distinguish registry-derived from administratively derived evidence at the concept level. This remains an open methodological question for future algorithm development.

Fourth, not all validation strategies were feasible across all data sources. The internal episode-level validation against PET was only possible in five data sources where the PET had been implemented. While Abellan et al. defined the required output standard, the preprocessing pathway each data partner uses to populate PET is determined locally and is not centrally overseen.[3] This means that differences in how episodes were derived prior to PET entry may have influenced comparisons but cannot be fully characterised at this stage. Guidance on consistent PET implementation will be addressed in Objective 3, and systematic characterisation of PET episode content, including outcomes, timing, and other characteristics across sources is the focus of Objective 4, which will indirectly inform understanding of cross-source variation in PET implementation. The PET therefore should not be treated as a uniform reference standard in the current analysis, and comparisons across the five sources should be interpreted accordingly. For example, in one source (HI-SPEED), the PET included only episodes above 22 weeks of gestation onward, but matching was performed generically against PET rather than adapted to this source-specific constraint. In addition, source data used to populate the PET may differ across sources; for example, some sources may refine pregnancy start date using ultrasound-based information (SIDIAP), whereas others may rely on recorded gestational age, LMP date, or back-calculation. Importantly, episodes identified by the algorithm only or by PET only are not necessarily invalid, they likely reflect different capture mechanisms. Algorithm-only episodes may include early pregnancies or losses not captured in the PET, while PET-only episodes may reflect pregnancies the algorithm did not detect. Sensitivity and PPV

estimates should therefore be interpreted as measures of concordance between two complementary episode sets rather than as absolute measures of algorithm accuracy. In the remaining sources, validation relied mainly on benchmarking against national statistics. These comparisons help assess aggregate plausibility and provide contextual external validity, but they depend on definitional alignment and denominator/catchment comparability and cannot detect individual-level misclassification or episode-construction errors. Furthermore, we used overlapping episodes as a flag for potential fragmentation, but overlaps may not always indicate errors. Overlaps most likely arise from data fragmentation or from administrative record duplication across encounters. Multiple gestations may also contribute, though they represent a relatively small proportion of pregnancies overall. Overlap-resolution and the hypothetical cleanup (provided in the data source fingerprint) steps may remove a small number of valid pregnancies. Note that reviewing site-specific performance at study start will remain important at this stage to ensure reliable usage in future studies.

Finally, as this study used routinely collected healthcare data, we need to account for common data quality limitations. For example, pregnancy episodes in routine healthcare data mostly reflect clinically recognised and recorded pregnancies. As a result, early pregnancies that did not lead to a documented clinical encounter (including very early losses) may be under-ascertained, and timing and outcome detail can be incomplete or fragmented across care settings—particularly in early gestation (e.g. <8 weeks). Under-ascertainment likely varies by source. For example, NLHR (Norway) has mandatory reporting from gestational week 12 onward and is nationally representative, which may reduce under-recording beyond that point. In contrast, regionally scoped EHR systems and other sources with partial pathway coverage may miss pregnancy-related care occurring outside contributing providers or settings, which can reduce early pregnancy capture and outcome specificity.

10.3. Interpretation

Consistent with ENCePP guidance, we evaluated the algorithm quality using a fit-for-purpose framework covering pregnancy ascertainment, timing validity, outcome specificity, and episode construction consistency.[16] Clinical detail varies by source type and country, so we treated between-source differences in outcome specificity and timing signals as expected heterogeneity, not uniform “performance” differences.[17] Accurate start dates are especially relevant for timing-sensitive analyses and misclassified pregnancy timing can bias prenatal exposure estimates; PET timing comparisons therefore provide key evidence on timing uncertainty where available.[15]

QC metric patterns differed by setting. Primary care data sources generally supported pregnancy episode identification with high plausibility, but outcome specificity was variable and for some data sources dominated by outcome categories where a pregnancy was identified but no specific outcome could be assigned (PREG), consistent with delivery and outcome details being recorded outside primary care. Secondary care sources showed wider heterogeneity in gestational length plausibility and overlap, including outliers with elevated proportions of episodes exceeding the 308-day plausibility threshold and >52-week episodes, noting that while episodes beyond 308 days are potentially implausible, they are not always erroneous, indicating that start and end date inference depends heavily on the availability of gestational dating signals in hospital data. The claims source showed a low proportion of pregnancy episodes with unknown outcome (PREG) and high live birth capture (LB), but higher episode overlap, consistent with administrative fragmentation. Registry-based sources generally showed high plausibility and more structured outcome capture, but completeness of delivery-mode recording and agreement with national statistics remained source-dependent. These patterns should not be interpreted as certain source types not being suitable for use; rather, they highlight that fit-for-purpose assessment is essential. The suitability of any source depends on the research question: sources with higher proportions of unknown-outcome episodes may still support pregnancy episode identification but are less suited to outcome-specific analyses without additional restrictions, while sources with higher overlap or timing uncertainty may require cleanup steps before use in timing-sensitive analyses.

Feasibility assessments for future studies should integrate these algorithm performance patterns alongside broader data source characteristics.

Run 3 helped identify data sources where the algorithm was able to infer more reliable pregnancy episodes characterised by high gestational length plausibility, low episode overlap, and more outcome-specific capture. Less favourable QC patterns should not be interpreted as poor data quality overall. They reflect differences in coding granularity, timing-signal availability, and care-setting capture, and they indicate where additional refinement, restrictions (e.g. excluding implausible durations), or alternative validation strategies may be needed for specific objectives.

Episode-level QC metrics should be interpreted together: episode-construction, gestational length plausibility, and outcome-coding granularity.[18]

Several data sources showed reasonable-quality episode output, with PREG near 0%, high plausibility, and modest overlap, suggesting suitability for outcome-specific and timing-sensitive analyses (CPRD GOLD, InGef RDB, DK-DHR, FinOMOP-THL, NLHR, and HI-SPEED). In contrast, other sources showed good gestational length plausibility but more limited pregnancy outcome specificity (e.g. IPCI and IMASIS with high PREG) or high pregnancy episode overlap and/or reduced plausibility (e.g. NAJS; and FinOMOP-TaUH Pirha as an outlier with elevated >308-day and >52-week episodes). Among these, FinOMOP-TaUH Pirha showed the weakest overall profile: the lowest gestational length plausibility (73.0%), highest proportion of episodes >52 weeks (8.1%), and the largest hypothetical cleanup proportion in the network (36.57%). This source's output should be considered exploratory pending further site-specific optimisation. However, remaining pregnancy episodes following an exclusionary cleanup step could still be of value (n=50,810) although they may be less representative.

Episode-level validation using the PET was available in five data sources (SIDIAP, CPRD GOLD, IMASIS, NLHR, HI-SPEED) and showed variable concordance, with PPV ranging from 58.7% (CPRD GOLD) to 97.3% (NLHR) and sensitivity ranging from 67.3% (SIDIAP) to 91.9% (IMASIS). Results should be interpreted as measures of concordance between two complementary episode sets rather than as absolute measures of algorithm accuracy though (see [Section 10.2](#) for further details).

Importantly, the purpose of this work was to evaluate how the algorithm performed when implemented across a heterogeneous network, rather than to support a single downstream study design. As such, the Run 3 outputs provide a structured basis for assessing data source fitness for purpose by answering: a) which sources are better able to support pregnancy episode identification; b) which are more likely to support outcome-stratified analyses; and c) where residual uncertainty remains because of timing limitations or episode-construction artefacts (e.g. overlap and extreme durations). These findings can inform the feasibility assessment for future studies and help define prespecified restrictions to ensure that timing and outcome assumptions fit the intended application. They also provide a basis for further refinement of the current algorithm and for development of future pregnancy-identification approaches within the network.

In several sources, incomplete capture of the live birth outcome limited direct comparison with crude birth rates when defined using the live birth outcome alone. To support interpretation in data sources with lower outcome specificity, we therefore also examined an alternative operational definition combining LB and DELIV (LB+DELIV). Sources in which LB counts were 0 or implausibly low, despite non-zero LB+DELIV rates, were interpreted as more likely reflecting differences in outcome-coding granularity and capture than a true absence of live births (for example only those complicated deliveries may be specifically coded with an alternative pregnancy outcome whilst DELIV is assumed to represent live births). Under this alternative definition, live birth-related estimates were more consistent with expected ranges.[11, 12, 19]

Benchmarking for other outcomes (e.g. spontaneous abortion, induced abortion, ectopic pregnancy) was not consistently feasible because external reference definitions and denominators are not harmonised across countries and are sensitive to care-seeking and recording pathways; we therefore interpreted these

outcomes as descriptive within-network distributions rather than population-representative rates. By way of context, early pregnancy loss affects an estimated 15–20% of clinically recognised pregnancies, but ascertainment in routine data depends strongly on care-seeking and recording pathways, particularly in early gestation.[20, 21] Model-based estimates suggest an induced abortion rate of approximately 17 (15–20) per 1,000 women aged 15–49 in Europe and Northern America (2015–2019), but comparability to routine healthcare data remains limited by differences in reporting systems and capture pathways.[22]

Maternal age at birth (2023) was generally within expected ranges and varied across sources, consistent with differences in catchment and care-setting capture. As delivery-mode benchmarking was strongly source-dependent, extreme C-section proportions (near 100% or near 0%) were observed in some data sources and interpreted as indicating variable completeness and/or mapping of delivery-mode fields rather than population-representative obstetric practice. National-statistics benchmarking was interpreted as a contextual check on capture and definitional alignment and used to inform fit-for-purpose assessments, particularly for analyses requiring live-birth specificity and delivery mode.

We iteratively enriched the HIP and PPS concept sets to increase capture of outcome-specific concepts and timing signals, drawing on published Matcho concepts and algorithms already implemented in participating data sources.[5, 7] These updates aimed to improve gestational length plausibility and outcome specificity; remaining heterogeneity likely reflects differences in concept availability and local coding practices across sources. Review of pregnancy algorithms used in study-specific data sources (NLHR, SIDIAP, BIFAP) informed additional concepts and supported interpretation of source-specific coding granularity.[2, 9, 10]

Across runs, the most consistent improvement in gestational length plausibility was observed from Run 1 to 2, with reductions in the proportion of episodes >308-days in 7/15 sources, whereas changes from Run 2 to 3 were smaller and more heterogeneous, likely reflecting source-specific differences in coding structure and algorithm behaviour. By contrast, Run 3 showed clearer reductions in pregnancy episode overlap and, in many data sources, outcome specificity. QC metrics are interdependent and should be interpreted jointly: refinements to episode construction (including overlap handling) and outcome assignment can shift episodes across end-type categories (DELIV, PREG, DELIV+PREG) and affect inferred gestational lengths. For example, tightening the gestational length window may cause episodes that previously anchored to a delivery-related code (DELIV) to no longer meet the timing criteria, reclassifying them as pregnancies with unknown outcome (PREG), and vice versa when rules are relaxed. Similarly, changes to overlap-resolution rules can affect which episodes are retained and how their outcomes are classified, further shifting the distribution across DELIV, PREG, and DELIV+PREG categories. A change that appears to worsen one metric may therefore reflect redistribution across categories rather than a true decline in performance. Persistent variability in pregnancies where no specific outcome could be assigned (PREG), alongside extreme gestational lengths and overlap in some sources, reflects how sensitivity inferred timing and outcome classification depends on local database coding practices and available timing evidence.

Accuracy of gestational length start-date remains hard to verify in routine data sources. PET comparisons (where available) showed non-trivial start-date differences among matched episodes. Inaccurate pregnancy start dates creates potential timing misclassification for analyses that depend on gestational-week-specific exposure windows and aligns with recent methodological work showing that pregnancy identification and timing approaches in real-world data can introduce selection and timing-related bias in prenatal exposure studies.[15]

Future algorithm developments could be prioritised based on study type requests (e.g. whether there is a requirement for time-sensitive, outcome-specific or both). Improving episode start-date estimations will be important for timing-sensitive analyses (e.g. trimester-specific risk windows) but is hard to validate without an external reference (e.g. PET fields, linked maternity records, or validated gestational-age signals). When granular start date signals are unavailable, start dates often rely on rule-based back-calculation from events recorded later in pregnancy, which can produce discrete-value heaping and residual timing uncertainty.

This does not mean the algorithm cannot be used for timing-sensitive analyses; rather, some degree of timing misclassification should be anticipated and accounted for. Users should assess start-date uncertainty as part of feasibility, apply restrictions where possible (e.g. limiting analyses to episodes with specific gestational dating evidence), and consider sensitivity analyses that vary the exposure window to assess the impact of timing uncertainty on study results. Expanding episode-level validation for start dates (e.g. using PET where available, linked obstetric datasets, or enhanced gestational timing concepts) would enable more targeted refinement of start-date logic in future iterations. Finally, the gestational length boundaries applied per outcome used in the algorithm, which define the plausible duration windows used to classify episodes as live birth, spontaneous abortion, stillbirth, and other outcomes, could be further explored and refined to improve interpretability for timing-sensitive and outcome-specific analyses. Additionally, future work could help to establish acceptable fit-for-purpose performance thresholds for pregnancy algorithms that determine their adequacy for specific research applications. Accordingly, interpretation remains context-dependent and should consider the intended use and the data source.[5, 7, 17] Collaboration with data partners is essential to interpret source-specific findings, optimise use of locally available data to improve performance, guide methodological refinement, and support expansion across the network.

10.4. Generalisability

This study informs the transportability of a pregnancy episode algorithm across heterogeneous data sources within the DARWIN EU® Data Network, rather than the generalisability of pregnancy episodes to a single underlying population. Within each data source, inferred pregnancy episodes represent clinically recognised and recorded pregnancies captured in the contributing care pathways; very early pregnancies and losses without documented encounters may be under-ascertained, and care occurring outside the data source scope may not be captured.

11. CONCLUSION

Network-wide implementation of the Jones et al. pregnancy episode algorithm (using the Smith et al. OMOP implementation) was feasible across 15 OMOP-mapped data sources within the DARWIN EU® Data Network. The developed *PregnancyIdentifier* package and proposed four-dimensional evaluation framework, covering episode frequency, duration, construction, and outcomes, provide a reusable, standardised foundation for pregnancy episode inference and quality assessment across the network.

Run 3 outputs confirm that the network may support pregnancy episode identification at scale, but sources readiness varies. Several data sources showed high gestational length plausibility, low episode overlap, and specific outcome capture and may be ready for use as is or with post-processing restrictions. Others showed meaningful limitations in outcome specificity, episode timing, or both, and require site-specific refinement or further validation before use in analyses with precise gestational or outcome-level requirements.

The data source “fingerprints” provide a structured, transparent basis for fit-for-purpose assessment, on the suitability of included data sources for future analyses, restrictions needing to be prespecified, and where additional validation may be needed.

Future work could focus on improving episode start-date estimation for timing-sensitive analyses, by expanded gestational timing concepts with targeted PET validation. Second, establishing fit-for-purpose performance thresholds that define minimum acceptable algorithm output for specific study designs, enabling more systematic source selection decisions across the network. Continued collaboration with data partners will be essential to interpret source-specific findings, optimise local implementation, and support the expansion of pregnancy research capacity within DARWIN EU®.

12. REFERENCES

1. Huybrechts KF, Bateman BT, Hernández-Díaz S. Use of real-world evidence from healthcare utilization data to evaluate drug safety during pregnancy. *Pharmacoepidemiol Drug Saf.* 2019;28(7):906-22.
2. Nordeng H, Lupattelli A, Engjom HM, van Gelder M. Detecting and Dating Early Non-live Pregnancy Outcomes: Generation of a Novel Pregnancy Algorithm From Norwegian Linked Health Registries. *Pharmacoepidemiol Drug Saf.* 2024;33(9):e70002.
3. Abellan A, Burn E, Trinh NTH, Burkard T, Callahan A, Fernández-Bertolín S, et al. Expanding the OMOP Common Data Model to Support Perinatal Research in Network Studies. *Pharmacoepidemiol Drug Saf.* 2025;34(2):e70106.
4. DeFalco F, Ryan P, Schuemie M, Huser V, Knoll C, Londhe A, et al. Achilles: Achilles Data Source Characterization. R package version 1.7.2. 2023.
5. Matcho A, Ryan P, Fife D, Gifkins D, Knoll C, Friedman A. Inferring pregnancy episodes and outcomes within a network of observational databases. *PLOS ONE.* 2018;13(2):e0192033.
6. Dernie F, Corby G, Robinson A, Bezer J, Mercade-Besora N, Griffier R, et al. Standardised and Reproducible Phenotyping Using Distributed Analytics and Tools in the Data Analysis and Real World Interrogation Network (DARWIN EU). *Pharmacoepidemiol Drug Saf.* 2024;33(11):e70042.
7. Jones SE, Bradwell KR, Chan LE, McMurry JA, Olson-Chen C, Tarleton J, et al. Who is pregnant? Defining real-world data-based pregnancy episodes in the National COVID Cohort Collaborative (N3C). *JAMIA Open.* 2023;6(3):o0ad067.
8. Smith LH, Wang W, Keefe-Oates B. Pregnancy episodes in All of Us: harnessing multi-source data for pregnancy-related research. *J Am Med Inform Assoc.* 2024;31(12):2789-99.
9. Sanchez Ortiz S, Llorente García A, Astasio P, Huerta C, Cea Soriano L. An algorithm to identify pregnancies in BIFAP Primary Care database in Spain: Results from a cohort of 155 419 pregnancies. *Pharmacoepidemiology and Drug Safety.* 2020;29(1):57-68.
10. Lestón Vázquez M, Vilaplana-Carnerero C, Gomez-Lumbreras A, Prat-Vallverdu O, Marsal JR, Vedia Urgell C, et al. Drug exposure during pregnancy in primary care: an algorithm and observational study from SIDIAP database, Catalunya, Spain. *BMJ Open.* 2023;13(8):e071335.
11. Eurostat. Fertility statistics Luxembourg: European Commission [cited 2026 Apr 14]. Available from: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Fertility_statistics.
12. Euro-Peristat. European Perinatal Health Report: Core indicators of the health and care of pregnant women and babies in Europe from 2015 to 2019 Paris: Euro-Peristat Project 2022 [cited 2026 Apr 14]. Available from: https://www.europeristat.com/wp-content/uploads/2022/11/Euro-Peristat_Fact_sheets_2022_for_upload.pdf.
13. (THL) TNIfHaW. Parturients and delivery procedures: THL; 2026 [Available from: https://sampo.thl.fi/pivot/prod/en/synre/toimenpiteet/summary_timebar?sairaala_0=10822&aiemmatsynnytykset_0=10768&sikioisyys_0=10837&iika_0=10452&raskausviikot_0=10711&bmi_0=73019&mittaritit_0=10863].
14. (ONS) OfNS. statistical bulletin, Births in England and Wales: 2023 ONS website; 2024 [updated 28 October 2024. Available from: Births in England and Wales: 2023
15. Latour CD, Edwards JK, Funk MJ, Suarez EA, Boggess K, Wood ME. Pregnancy identification method as a source of bias in studies of prenatal exposures using real-world data. *Am J Epidemiol.* 2025;16. European Network of Centres for Pharmacoepidemiology and Pharmacovigilance E. Annex 2. Guidance on methods

for the evaluation of medicines in pregnancy and breastfeeding [cited 2026 Apr 14]. Available from: https://encepp.europa.eu/encepp-toolkit/methodological-guide/annex-2-guidance-methods-evaluation-medicines-pregnancy-and-breastfeeding_en.

17. van Rijt-Weetink YRJ, van Gendt J, Egberts TCG, van Hunsel F, Lewis DJ, Yates LM, et al. Quality of Clinical Information in Pregnancy Pharmacovigilance Data Sources-A Contribution of the ConcePTION Project. *Pharmacoepidemiol Drug Saf.* 2025;34(7):e70182.
18. Gatto NM, Campbell UB, Rubinstein E, Jaksa A, Mattox P, Mo J, et al. The Structured Process to Identify Fit-For-Purpose Data: A Data Feasibility Assessment Framework. *Clin Pharmacol Ther.* 2022;111(1):122-34.
19. Kniffka MS, Schöley J, Lee S, Bertens LCM, Been JV, Gunnarsdóttir J. Stillbirth rate trends across 25 European countries between 2010 and 2021: the contribution of maternal age and multiplicity. *Eur J Public Health.* 2025.
20. Rossen LM, Ahrens KA, Branum AM. Trends in Risk of Pregnancy Loss Among US Women, 1990-2011. *Paediatr Perinat Epidemiol.* 2018;32(1):19-29.
21. Quenby S, Gallos ID, Dhillon-Smith RK, Podsek M, Stephenson MD, Fisher J, et al. Miscarriage matters: the epidemiological, physical, psychological, and economic costs of early pregnancy loss. *Lancet.* 2021;397(10285):1658-67.
22. Bearak J, Popinchalk A, Ganatra B, Moller A-B, Tunçalp Ö, Beavin C, et al. Unintended pregnancy and abortion by income, region, and the legal status of abortion: estimates from a comprehensive model for 1990-2019. *The Lancet Global Health.* 2020;8(9):e1152-e61.

13. ANNEXES

ANNEX I. Description of data sources

Croatian National Public Health Information System (NAJS)

#	Section	Description
1	Data source identification and country	NAJS (Croatian National Public Health Information System) Croatia
2	Data partner information section	Croatian Institute of Public Health Department of Data Science and Analytics
3	Coverage and timespan	Data collection since: 1998 Extent: Nation-wide. Geographic coverage covers whole Croatia, with various levels of resolution for different registries. Current estimates for the population in Croatia will be available at: https://podaci.dzs.hr/hr/podaci/stanovnistvo/procjena-stanovnistva/ for each year. The total and active person count in the NAJS data is larger than the current population of Croatia. This explained by: a) the person table included deceased and all previously insured people and b) there is no information about insurance ending, c) healthcare is also used by people with dual citizenship from neighbouring countries It is known that a lot of people emigrated (300k-400k) and weren't included in the last population census but still are in the NAJS database. There is also an influx of immigrant workers that are insured and registered but weren't included in the census.
4	Healthcare setting / type of data	Primary care – General Practitioner, and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. Primary care – gps, and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. For both inpatient and outpatient setting diagnoses, medication, procedures, and measurements are captured. The year of availability of information depends on the setting • 2014-2025 for biochemical lab tests in primary care from EHR patients records (measurements with results) • 2015-2025 for primary care data from EHR patient records (conditions, procedures, and drug prescriptions) • 2015- 2024 for inpatient hospital data from EHR administrative records (conditions, procedures, measurements without results and drug administrations) • 2016-2025 for health risk assessment data entered by GPs (measurements with results - height, weight...) • 2016-2022 for secondary conciliatory care data from EHR administrative records (conditions, procedures, measurements without results and drug administrations) • 2016-2022 for emergency care data from EHR patient records (conditions) • 2017-2025 for hospital records from registry data (conditions and procedures) • 2020-2025 for vaccination data from EHR patient records
5	Data collection process	Inpatient hospital billing systems, and Other. Data is entered by clinicians at healthcare contact, then combined by CIPH into the NAJS database and integrated with registries for public health purposes.
6	General representativeness	The data is collected from the evidence of public health records collected for public health purposes, as the majority of health care in Croatia is public and under single health insurance provider. Personal details are collected to a better extent for insured individuals compared to uninsured patients, who are excluded in the ETL process.
7	Data content /source coding	Medication prescriptions are recorded with ATC codes with an additional 3 digit code denoting the package. Diagnoses with ICD10 codes (Australian modification). Procedures with local source codes. Lab results with local source codes.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. Records from 2017 include insured patients with reliable IDs. Uninsured patients do not have reliable IDs. For example, if a patient changed her status from insured to uninsured, or vice versa,

#	Section	Description
		she could be counted several times, as could tracking records from before 2017 and after. By using the unique personal identifier for Croatian citizens, it can be checked and verified.
9	Quality control (data source specific)	There is a network of registry personnel (leaders, administrators, coders, sources) working on data coverage and other quality dimensions. An analytical team routinely checks for erroneous entries in hospital records, removing double entries, false dates, and overlapping stays. Entries without enough data or with obviously erroneous dates from primary care analysis are being excluded.
10	Linkage	The national death registry is updated yearly, with one year lag, but the fact of someone's death (just the date) is updated daily, without the cause of death or any other additional details. Primary care is updated weekly and hospital level care monthly. Specific registries are included in NAJS (e.g. diabetes registry), where inclusion criteria vary across these registries.
11	Vital status	NAJS is linked to the national death registry.
12	Limitations	Hospital data is available from 2017 onwards. This is often used as start of data collection, while laboratory and GP data is captured before that (since 2014 and 2015 respectively). Drug duration is often not available, and set to 1 day for administration and 30 days for prescription. Hospital discharge summaries are currently not captured in NAJS. Hospital drug administration data is less reliable than prescription data from primary care, with some drugs (monoclonal antibodies / precision medicine drugs) that require additional approval not being recorded at all.
13	Main references	No main reference provided.
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111155 Website: https://www.hzjz.hr/nacionalni-javnozdravstveni-informacijski-sustav-najs/

Danish Data Health Registries (DK-DHR)

#	Section	Description
1	Data source identification and country	DK-DHR (Danish Data Health Registries) Denmark
2	Data partner information section	Danish Medicines Agency (DKMA) Data Analytics Centre (DAC)
3	Coverage and timespan	Data collection since: 1995 Extent: Nation-wide. The data is representative of the entire Danish population.
4	Healthcare setting / type of data	Community pharmacists, and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. The following data elements are collected: diagnosis (including rare diseases and pregnancy data), hospital admissions, discharge and ICU data, Cause of death, Drug prescriptions, dispensing, vaccination and contraception, Procedures (surgical and non-surgical hospital), and Sociodemographic information (sex and age only).
5	Data collection process	Outpatient electronic health records, and Inpatient hospital electronic health records, and Registries, and Other. All causes of deaths, all retrieved drug prescriptions, all records of vaccinations, all hospital inpatient and outpatients contacts including disease diagnoses and hospital surgical and non-surgical procedures, histologically confirmed incident cancers, laboratory test results for the entire Danish population from 1/1/1995 onwards.
6	General representativeness	The data is representative of the entire Danish population. Healthcare is free in Denmark, so we do not expect any bias in data collection based on socio-economic status.

#	Section	Description
7	Data content /source coding	Diagnoses and causes of death are collected using the ICD-10 vocabulary. ATC and RxNorm are used for Drugs. SNOMED codes are used for Procedures.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. Patients have unique identifiers used to link datasets.
9	Quality control (data source specific)	The data we have received relating to nationwide Danish Health Data registries offer an opportunity for large-scale, population-based studies with several advantages 1) Their large size improves the precision of estimates and enables the study of rare exposures and outcomes with long-term latency, 2) Inclusion of nearly all individuals in the target population ensures that the data reflect routine clinical care and all clinical segments of the source population, 3) Data are collected independently of each research study, thus minimising certain types of bias, e.g. non-response, and the influence from attention to the research question on the diagnostic process. Before the source data is sent to us, the Danish Health Data Authority runs and does comprehensive checks of the registry table data validity of the variables, breaks in data, changes in variable coding, missingness, etc. We perform checks of missingness/completeness in relation to requested variables. In essence, we are receiving a dump of a mirror of the data that is controlled by the SDS. The documentation performed by SDS is available online, in Danish primarily https://www.esundhed.dk/Dokumentation (all variables), but also in English https://sundhedsdatastyrelsen.dk/da/english/health_data_and_registers/national_health_registers
10	Linkage	There is no linkage in this data source.
11	Vital status	The Cause of Death registry (DAR) is used, the cause of death is collected using ICD-10 codes.
12	Limitations	DK-DHR has the following limitations, which may be relevant confounders for certain complex Darwin EU studies: - We lack information on key socio-economic status (SES) factors, such as occupation, education, and income. These variables may be important for analysis in some studies. - We only have complete data on lifestyle factors (such as smoking status and weight) for pregnant women. - We have no information on patient contacts in primary care (visits to the GP). Consequently, the incidence of chronic diseases like Type 2 Diabetes (T2D) and asthma must be determined using drug prescriptions as a proxy. Stillborn children will not have any records in our CDM. This means that e.g. birth length of stillborns is not recorded.
13	Main references	Schmidt M, Schmidt SAJ, Adelborg K, Sundbøll J, Laugesen K, Ehrenstein V, Sørensen HT "The Danish health care system and epidemiological research: from health care contacts to database records." Clinical epidemiology (2019): 31372058
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111217 Website: https://sundhedsdatastyrelsen.dk/da/english/health_data_and_registers/healthdatadenmark

Tampere University Hospital patient cohort (FinOMOP-TaUH Pirha)

#	Section	Description
1	Data source identification and country	FinOMOP-TaUH Pirha (Tampere University Hospital patient cohort) Pirkanmaa, Finland
2	Data partner information section	Pirkanmaa Welfare Services County, Tampere University Hospital Department of Research, Development and Education
3	Coverage and timespan	Data collection since: 2007 Extent: Regional. TaUH Research Database includes all specialities/all patient groups treated in the Tampere University Hospital,

#	Section	Description
4	Healthcare setting / type of data	Secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. Secondary, and tertiary care given in the region, including given clinical and pathology diagnoses, diagnostic and therapeutic procedures, laboratory findings, radiology and pathology reports, medication given in the hospital and electronic prescriptions, and continuous medical records (free text), including discharge letters since 2007.
5	Data collection process	Outpatient electronic health records, and Inpatient hospital electronic health records. The data is entered by the clinician at point of contact and processed in a data warehouse for secondary use.
6	General representativeness	The data covers patients visiting the hospital and therefore will not reflect the general population.
7	Data content /source coding	ICD10fi: Finnish modification of ICD10; SNOMED2-TMP: a local SNOMED2-based vocabulary of organ-diagnosis pairs in pathology; NCSPfi: Finnish national version of NCSP/Nomesco vocabulary for procedures; LABfi: Finnish national vocabulary (Kuntaliitto) for laboratory measurements; LABfi_TMP: Local additions to the national vocabulary for laboratory measurements; UNIT_FIN: measurement units; MICROBefi_TMP: Microbe names in measurement data; ATC: all drugs except anticancer drugs; active ingredient (in Finnish): for anticancer drugs; VNR: Nordic medicine product identifiers (https://pharmaca.fi/en/health/pharma/vnr/); Hilmo_eala: Finnish national medical specialty vocabulary.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. In some cases, patients can be re-registered with a new ID. Patients are identified by their social security number, and each patient gets a patient ID in the EHR, associated with the social security number. However, there are cases when the social security number has changed, and the associated patient ID also has changed. Approximately 4% of all patient IDs in the database are such duplicates. Majority of such duplicates are newborns who have first been assigned a temporary social security number, which has later been changed to a permanent social security number. An unidentified person can have been given a temporary social security number, and sometimes a person can change their social security number, for example, because of threat (stalking) or gender reassignment. It is technically possible to link all the instances of a person in the database however this is not currently done.
9	Quality control (data source specific)	The source database tables are checked for completeness on a general level: whether there are null values in the fields, what the date ranges are for each table, are there gaps in the data, and how many unique patients can be found in each data table. Before mapping to the OMOP CDM, we have identified the data tables and the fields in the tables that are required to build the OMOP CDM. The White Rabbit and Rabbit-in-a-Hat tools were used in this process.
10	Linkage	Data from many nation-wide registries can be combined, which is subject to obtaining special data permits.
11	Vital status	The source of vital status (date of death) is the Digital and Population Data Services Agency.
12	Limitations	No database-specific limitations documented. General limitations for the data type applicable.
13	Main references	No main reference provided.
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111135 Website: https://www.pirha.fi/en/web/english

Finnish Care Register for Health Care (FinOMOP-THL)

#	Section	Description
1	Data source identification and country	FinOMOP-THL (Finnish Care Register for Health Care) Finland
2	Data partner information section	Finnish Institute for Health and Welfare (THL) The Department of Data and Analytics
3	Coverage and timespan	Data collection since: 1998 Extent: Nation-wide. The current CDM population comprises all persons having been alive and residing in Finland since the beginning of 2011.
4	Healthcare setting / type of data	Primary care – General Practitioner, and primary care specialists (e.g. paediatricians), and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care, and other (specify). THL maintains health registers that cover both public and private, primary, and specialised inpatient, urgent and outpatient health care encounters in Finland, starting from 2011. The entire public sector and private inpatient encounters have been included since 2011, while private outpatient encounters, including occupational care, are included since 2020. Since 1998, the Care Register for Health Care (TerveysHilmo) has covered both public outpatient and inpatient specialized care and private inpatient care. Since 2009, the Finnish National Vaccination Register has covered all vaccinations from the public sector and from a large part of private vaccination providers, with the data coverage from both sections being very good to complete from 2020 onwards. Since 2011, the Register of Primary Health Care Visits (AvoHilmo) has covered public primary care. Since 2020, the register has also covered private outpatient care and occupational care. In addition, the CDM also contains positive COVID-19 test results from the Finnish National Infectious Diseases Register. The register itself covers all laboratory confirmations for around 70 specific microbes from 1995 onwards, but only COVID-19 has currently been mapped to the CDM. The CDM also includes prescription records from multiple different sources. Both Care Register for Health Care and Register of Primary Health Care Visits contain very basic prescription data recorded during health care encounters that include just the ATC-code and trade name of medication. More comprehensive prescription data from Kanta Prescription Centre, maintained by Social Insurance Institution of Finland (Kela), has been integrated into the CDM since its 2024 release. The Kanta Prescription records are based electronic prescriptions, which were adopted by most public health care providers in 2010 and by most private providers by 2017.
5	Data collection process	Outpatient electronic health records, and Inpatient hospital electronic health records, and Registries. Data is entered by clinicians upon healthcare contact and processed by THL (Kela in the case of Kanta Prescription Centre).
6	General representativeness	The THL data has national coverage and is therefore well representative of the Finnish population. Using the complete population as a basis for the person table also serves to facilitate calculations on a population level, e.g. incidence rates.
7	Data content /source coding	The following coding systems have been OMOP-mapped, typically to a good level of completeness: ICD10fi Finnish Extension, ICPC-2, ATC, Toimenpideluokitus (procedure classification adapted from the Nordic Classification of Surgical Procedures (NCSP)), Terveystieteiden erikoissalat (Hilmo specific provider speciality), Rokotustapa (AR/YDIN National classification for vaccine administration), Tupakointistatus (AR/YDIN National classification for smoking status). Vaccinations are identified on product level based on batch number, trade name, vaccine title, and ATC-code. This is mapped on brand and type in the OMOP CDM.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. Each patient in THL has a unique identifier.

#	Section	Description
9	Quality control (data source specific)	The source data collection undergoes a structural and semantic validation before entry into the source database. Additionally, some coded variables undergo quality assessment against the respective code systems post entry into the database. The source registers are also assessed for completeness and coverage, with the aim of improving future collection in the areas where data is lacking.
10	Linkage	THL is already a linkage of multiple Finnish registries (see above).
11	Vital status	The National Population registry data forms the basis for forming the patient population. This ensures an up-to-date location (municipality of residence) of patients, as well as complete death occurrences (although not the cause of death).
12	Limitations	All drug records in the CDM are currently based on prescriptions. Kanta Prescription Centre also includes information on drug dispensings, but these have not currently been converted into OMOP CDM. Depending on the type of the medication, a single prescription can be for up to two years dosage of the drug. This means a patient can have up to two-year breaks between observations while actively using the drug. Observation of a prescription also does not mean that the patient necessarily bought or used the medication. The CDM does not currently have meaningful end dates or days of supply for drug exposure. This information is not available for Care Register for Health Care or Register of Primary Health Care Visits, and for Kanta Prescription Centre it is only available in unstructured, free-form format that has not been converted into OMOP CDM. The source system for prescription drug data is only available to the DP until Dec-2023, therefore drug_exposure records from this source are only available in that time range. Not all private health care records are covered for the CDM's entire follow-up time from 2011 onwards. For Register of Primary Health Care Visits and Finnish National Vaccination Register, records from private health care have been available from 2020 onwards. For Kanta Prescription Centre, the coverage of private health care records has been good from 2017 onwards. The inclusion of private health care mainly presents itself as an increase in the number of observations, meaning that it has to be accounted for when interpreting any time series data from the CDM.
13	Main references	Häkkinen, Pirjo; Mölläri, Kaisa; Saukkonen, Sanna-Mari; Väyrynen, Riikka; Mielikäinen, Lasse; Järvelin, Jutta "Hilmo - Sosiaali- ja terveydenhuollon hoitoilmoitus 2020 : Määrittelyt ja ohjeistus : Voimassa 1.1.2020 alkaen" Terveyden ja hyvinvoinnin laitos (2019):
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111187 ; https://catalogues.ema.europa.eu/data-source/1111191 Website: https://thl.fi/en/statistics-and-data/data-and-services/register-descriptions ; https://www.kanta.fi/en/research-and-knowledge-management

Clinical Data Warehouse of Bordeaux University Hospital (CDW Bordeaux)

#	Section	Description
1	Data source identification and country	CDW Bordeaux (Clinical Data Warehouse of Bordeaux University Hospital) Nouvelle-Aquitaine, France
2	Data partner information section	CHU DE BORDEAUX - DIRECTION GENERALE Gironde / Nouvelle-Aquitaine
3	Coverage and timespan	Data collection since: 2005 Extent: Regional. It covers the population of Bordeaux metropolitan area, and possibly beyond, as the health care centre for referrals and expertise for the Nouvelle Aquitaine region. The database contains data from 2005 onwards.
4	Healthcare setting / type of data	Secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care, and claims data. The database currently captures information about patient demographics, visit details, conditions, procedures, drugs, measurements, and mortality.

#	Section	Description
5	Data collection process	Outpatient electronic health records, and Inpatient hospital electronic health records, and Inpatient hospital billing systems, and Biobank. The integrated data is extracted from the hospital production information system via a real-time research database. The data is then processed and quality controlled by a team dedicated to maintaining the database. Internal evaluations were carried out to ensure consistency between the research database and the patient bedside software.
6	General representativeness	This is the 6th largest metropolitan area in France, and CHUBX is the largest hospital in the region. More than 75% of the patients administered to Bordeaux university hospital reside in the Gironde departments, with almost 50% coming directly from the Bordeaux metropolitan area. The hospital also captures additional cases from Nouvelle-Aquitaine region.
7	Data content /source coding	Diagnosis source data is coded in ICD-10 terminology. Procedures are coded in CCAM (French terminology). Laboratory measurements are coded in local terminology and partially mapped to LOINC. Drugs are coded through a local terminology and then mapped to UCD (French terminology), as well as ATC codes.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. We use the hospital's unique identifier to generate the patient identifier in OMOP. If two identities are merged at the hospital, the merge is taken into account in the CDW. An automatic (hourly) detection of suspected duplicated identities has been implemented at the hospital since 2020, with merging of duplicated identities by a specialized team. Identities since 2015 were processed retrospectively. Thus, the rate of identity duplication in the database is low, especially since 2015.
9	Quality control (data source specific)	- The integrated data comes from the hospital production information system through a real-time replicated database. Consistency evaluations between the replicated database and the production system are performed by the technical team in charge of maintaining the replicated database. - In the same way, consistency checks are performed between the replicated database and the data integrated into the i2b2 CDW. In addition, dashboards enable monitoring the data integrated into the i2b2 CDW, in particular by controlling the amount of data available over time, and its evolution, according to the various data sources. - An internal evaluation was carried out to ensure the consistency between the data integrated into i2b2 and the data available in the software used at the patient's bedside. In addition, many use cases were performed on the i2b2 CDW, with return to the patient chart and comparison of the data integrated into i2b2 and the data available in the care file.
10	Linkage	Death certificates (without the cause of death).
11	Vital status	The database is linked to the French death registry.
12	Limitations	CDW Bordeaux is limited to events captured in the hospital setting and thus does not include patient events not treated by the hospital (e.g. rare cancers). Patient events that are not included in CDW Bordeaux are rare disease treatments or specialist events that occur outside of CHUBX.
13	Main references	Cossin S, Diouf S,Griffier R,Le Barrois d'Orgeval P,Diallo G,Jouhet V "Linkage of Hospital Records and Death Certificates by a Search Engine and Machine Learning." JAMIA open (2021): 33709061
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111112 Website: https://www.chu-bordeaux.fr/

InGef Research Database (InGef RDB)

#	Section	Description
1	Database Identification and country	InGef Research Database (InGef RDB) Germany (DE)
2	Data partner information section	InGef – Institute for Applied Health Research Berlin GmbH InGef is a research service provider within the German statutory health insurance system.
3	Coverage and timespan	The Research Database contains approximately 10.5 million insured persons from 50 of the 94 statutory health insurances in Germany (as of Oct 2025). All individuals in the research database are included in the OMOP CDM. Due to data protection laws for personal data in Germany, only the last 10 years of data are available in the RDB and thus in the OMOP CDM.
4	Healthcare setting / type of data	In general, the data in the RDB reflects most of the health services paid for by statutory health insurances. In the OMOP CDM, the health services are reduced to primary and secondary care. This means that treatments by GPs and specialists (e.g. paediatricians), prescription medicines dispensed by pharmacies, hospital inpatient and outpatient care, as well as information about all insured individuals are included. The following data elements are presented in the OMOP CDM: demographic information, diagnoses, procedures, dispensing drugs and advanced therapy medicinal products, vaccinations, pregnancy data (via diagnoses and procedures) and contraception.
5	Data collection process	In Germany, data exchange between medical service providers and statutory health insurances is organized via central data collection centers. In addition to receiving and forwarding data, these centers also store data and provide access to third parties under strict regulations in data warehouses. The RDB contains selected, condensed and anonymized information from one of these data warehouses. Data sovereignty remains with the individual health insurance companies.
6	General representativeness	The RDB covers about 11% of the German population and is comparable to the German population in terms of the distribution of age and sex. Most health insurances that contribute to the RDB have nationwide coverage, meaning that the database covers all regions of Germany. Since almost all services covered by statutory health insurances are specified in national legislation, healthcare provision all over Germany is well represented in the RDB. Additionally, in Germany it is very common to stay with the same health insurance throughout life, which results in a good longitudinal coverage over the entire period of 10 years.
7	Data content /source coding	The coding in the research database complies with national classification and coding rules in Germany. Diagnoses are coded according to ICD-10-GM. Inpatient and outpatient surgeries or procedures are recorded as OPS codes (German classification of Operations and Procedures). The dispensing of drugs in pharmacies is recorded using the PZN (pharmaceutical registration number). For drugs that miss a PZN-to-RxNorm mapping, the ATC code is used instead. In some cases, dispensed drugs can be coded using OPS codes (e.g. in hospitals) or EBM codes (fee schedule for outpatient treatments).
8	Data Harmonization	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network.
9	Quality control (database specific)	The data transmitted by healthcare providers complies with the standardized requirements and formats of the Association of Statutory Health Insurances (GKV-SV). Before being imported into the research database, the data elements are checked for data format, completeness and plausibility. After each update of the research database, various counts are compared with the previous update to verify completeness. Due to the anonymity of the database, direct validation of the data (e.g. using medical records as the gold standard) is not possible.

#	Section	Description
10	Linkage	Due to the anonymization of the source data, linkage is not possible.
11	Vital status	The date of death is recorded as the last day of the quarter in which the death occurred (i.e. 30/31st of Mar/Jun/Sept/Dec) as reported to the health insurance (no linkage to death registry). The cause of death is not available.
12	Limitations	Ambulatory diagnoses and procedures are summarised in the source on a quarterly basis. Both are mapped to the observation table with the date set to the last day of the respective quarter (i.e. 30/31st of Mar/Jun/Sept/Dec) and the concept "History of event within 3 months" (observation_concept_id 1340222), with the actual diagnosis or procedure concept_id recorded in the field "value_as_concept_id". There is no vocabulary for the German pharmaceutical product codes (PZN). A direct source-to-standard-mapping has been done manually by InGef but is incomplete. The drug exposure duration is unknown. Following OMOP conventions, the end date is always set to dispensing date + 29. Outpatient and inpatient procedures are recorded as OPS codes (German Procedure Classification), for which the vocabulary is incomplete.
13	Main references	Andersohn F, Walker J "Characteristics and external validity of the German Health Risk Institute (HRI) Database." Pharmacoepidemiology and drug safety (2016): PMID 26530279 doi: 10.1002/pds.3895 Ludwig M, Enders D, Basedow F, Walker J, Jacob J: Sampling strategy, characteristics and representativeness of the InGef research database. Public Health 2022. doi: 10.1016/j.puhe.2022.02.013
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111207 Website: https://www.ingef.de/en/

Semmelweis University Clinical Data (SUCD)

#	Section	Description
1	Data source identification and country	SUCD (Semmelweis University Clinical Data) Budapest, Hungary
2	Data partner information section	Semmelweis University -
3	Coverage and timespan	Data collection since: 2010 Extent: Regional. The general catchment area of SU is the central region of the country, Budapest city and Pest county, although patients can be referred from anywhere in Hungary. The total population of Budapest and Pest county is approximately 4,200,000 people. The total population of Hungary is around 9,500,000.
4	Healthcare setting / type of data	Secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care, and claims data, and other (specify). diagnostic data (laboratory tests, radiology, pathology)
5	Data collection process	Insurance/administrative claims, and Outpatient electronic health records, and Inpatient hospital electronic health records, and Inpatient hospital billing systems, and Registries. Data is extracted directly from the source database. From there, the data entry in the system is heavily controlled and validated on the user interface before being made available for further research.
6	General representativeness	SU captures information on patients who are covered by the public health insurance system. This covers all Hungarian citizens, and therefore the database should mirror the source population well. Although, besides Semmelweis University Clinics, there are multiple hospitals in

#	Section	Description
		the region, and data on visits in other hospitals is not represented in the database. Therefore, the patient population is not directly representative of the general population.
7	Data content /source coding	Regarding SU's source data, procedures and diagnoses are coded in SNOMED, measurements are coded in LOINC, and drugs are stored in RxNorm and ATC.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. Patients have a unique identifier (SSN).
9	Quality control (data source specific)	The clinical database is the source database and therefore it has to be treated as a trusted database. Data entry in the systems is heavily controlled by validation on the user interface, and there are large number of rules that controls the data on the insurer's side that has to be corrected in the system by the users to be able to close the encounters. OMOP mapping is done in the framework by EHDEN recognized partners under quality check by the EHDEN society.
10	Linkage	No known linkages.
11	Vital status	Source for vital status unknown.
12	Limitations	Medication prescribed in secondary care is fully present in our database, but medication given in the hospital is rarely documented. General limitations for the data type applicable. General practitioner data is not present in our database; therefore, this part of the patient journey is not represented.
13	Main references	No main reference provided.
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1000000184 Website: https://www.semmelweis.hu

Integrated Primary Care Information (IPCI)

#	Section	Description
1	Data source identification and country	IPCI (Integrated Primary Care Information) Netherlands
2	Data partner information section	Erasmus University Medical Center Department of Medical Informatics
3	Coverage and timespan	Data collection since: 2006 Extent: Nation-wide. IPCI is a Dutch database that contains patient records from 2006 onwards. However, it mainly covers the central part of the country, including the most densely populated area (the 'Randstad') and non-urban areas. IPCI contains information on all patients registered with GPs responsible for non-emergency care and referrals. A patient is registered at birth or at first encounter with the GP.
4	Healthcare setting / type of data	Primary care – General Practitioner. Data is collected from primary care EHR. This includes demographic information, complaints and symptoms, diagnoses, laboratory test results, lifestyle factors (in limited amount), and correspondence with secondary care, such as referral and discharge letters.
5	Data collection process	Outpatient electronic health records. Data is entered into the EHR system by the GPs, during or after the visit. The patient dossiers are collected by Erasmus MC data managers and combined in one harmonized database. Several checks are done on this database to ensure correct data processing. Persons can have dossiers at multiple GPs.

#	Section	Description
6	General representativeness	More than 99% of the Dutch population has health insurance, and almost all citizens are registered with a general practitioner. Over 12 months, around 78% of the population has at least one contact with their GP. IPCI included around 350 GP practices out of around 5000 in the country (~ 7%). The demographic composition of the IPCI population mirrors that of the general Dutch population in terms of age and sex.
7	Data content /source coding	Dutch GPs use mainly Dutch standard codes, like ICPC-1 and Diagnostische Bepalingen maintained by NHG. And for therapy the G-Standard is used, maintained by ZIndex.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. Patients can be registered under different IDs, but since a patient can only be registered at one GP at a time, the observations periods will not overlap.
9	Quality control (data source specific)	Prior to each data release, extensive quality control steps are performed, e.g. comparison of patient characteristics between practices, and checks to identify abnormal temporal data patterns in practices. For each practice, around 200 quality indicators are obtained. Of these indicators, a quarter refer to population characteristics, e.g. number of birth and mortalities relative to practice size, temporal consistency. The other indicators are based on medical data, e.g. distribution of measurement values, frequencies of diagnoses and procedures relative to age, completeness of data. The indicators are combined in a couple of quality scores for each practice. For these scores, cut-off values for acceptable quality have been defined. Practices with a score below a cut-off are excluded for research. This approach has shown to be very important, for example to check if data from practices that just joined the database are at an acceptable level of quality. The details of the approach, like the cut-off values for acceptance, are based on years of experience. In addition, trends are compared with the previous database release. Extensive quality control steps are performed before each data release. These include comparing patient characteristics between practices and checks to identify abnormal temporal data patterns in practices. Additional checks include over 200 indicators related to population characteristics (e.g. reliability of birth and mortality rates) and medical data (e.g. availability of durations of prescriptions and completeness of laboratory results). Records of low quality are excluded from the database.
10	Linkage	Linkage requires additional approval steps and needs to be assessed on a case-by-case basis. IPCI is not routinely linked with other databases.
11	Vital status	Vital status (death date and cause) is collected based on GP records.
12	Limitations	The main limitation comes with the fact that IPCI is limited to GP records, and although it contains information on referrals and discharge letters, it may not fully capture specific hospital information. IPCI does not include coded/detailed data about medications/procedures/test results from the hospital or other care-providers.
13	Main references	de Ridder MAJ, de Wilde M, de Ben C, Leyba AR, Mosseveld BMT, Verhamme KMC, van der Lei J, Rijnbeek PR "Data Resource Profile: The Integrated Primary Care Information (IPCI) database, The Netherlands." International journal of epidemiology (2022): 35182143
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/42618 Website: http://www.ipci.nl

Norwegian Linked Health Registry data (NLHR)

#	Section	Description
1	Data source identification and country	NLHR (Norwegian Linked Health Registry data) Norway
2	Data partner information section	University of Oslo Faculty of Mathematics and Natural Science – Department of Pharmacy
3	Coverage and timespan	Data collection since: 2008 Extent: Nation-wide. Norway has a universal public health care system, consisting of primary and specialist health care services covering a population of approximately 5.4 million inhabitants.
4	Healthcare setting / type of data	Primary care – General Practitioner, and primary care specialists (e.g. paediatricians), and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. The following registries are included: the Medical Birth Registry of Norway (MBRN), the Norwegian Prescription Registry (NorPD), the Norwegian Patient Registry (NPR), Norway Control and Payment of Health Reimbursement (KUHR), the Norwegian Surveillance System for Communicable Diseases (MSIS), the Norwegian Immunisation Registry (SYSVAK), the National Death Registry, and the National Registry (NR).
5	Data collection process	Registries. Many population-based health registries were established in the 1960s, with use of unique personal identifiers facilitating linkage between registries. Data in these health registries are used for health analysis, health statistics, improving the quality of healthcare, research, administration, and emergency preparedness.
6	General representativeness	The NLHR data covers the full Norwegian population.
7	Data content /source coding	NPR: ICD-10 for diagnosis, ATC and some special codes for drug use, Norwegian codes for clinical procedures (surgery (NCSP), medicine (NCMP) and diagnostic imaging, image-guided intervention, and nuclear medicine (NCRP)). KUHR: ICD-10 and ICPC-2 and ICPC-2B for diagnosis/procedure. NorPD: ATC. SYSVAK and MSIS: national classifications. MBRN: custom classifications by questionnaires (incl. check box variables in Maternity health care card)
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. Linkage between the registries was facilitated using project-specific person IDs generated from unique personal identification assigned at birth or immigration for all legal residents in Norway.
9	Quality control (data source specific)	In-house data quality checks of rates of common conditions, drug exposures, and outcomes. We compare obtained rates with official national statistics (e.g. birth statistics, yearly rates of drug dispensing, and diagnosis by age and gender). We also review missing data and outliers and inform registry holders of any unusual patterns.
10	Linkage	The NLHR is, by definition, a linkage of datasets. Helsedata.no is one central portal to apply for 11 national health registries, including all the registries that have been mapped to the OMOP CDM.
11	Vital status	The national death registry is linked.
12	Limitations	No database-specific limitations documented. General limitations for the data type applicable.
13	Main references	Mitter VR, Lupattelli A, Bjørk MH, Nordeng HME "Identification and characterization of migraine in pregnancy: A Norwegian registry-based cohort study." Cephalalgia : an international journal of headache (2024): 38663979
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1000000409 Website: https://www.mn.uio.no/farmasi/english/research/groups/pharma-safe/

Egas Moniz Health Alliance database - Gaia E Espinho (EMDB-ULSGE)

#	Section	Description
1	Data source identification and country	EMDB-ULSGE (Egas Moniz Health Alliance database - Unidade Local de Saúde de Gaia e Espinho) Distrito de Aveiro, Portugal
2	Data partner information section	Clinical Academic Center Egas Moniz Health Alliance
3	Coverage and timespan	Data collection since: 2004 Extent: Regional. The ULSGE includes 32 primary care centres, assisted by three hospitals (Hospital Eduardo Santos Silva, Hospital Distrital Vila Nova de Gaia, Hospital Nossa Senhora da Ajuda) and one specialized rehabilitation centre (Centro de Reabilitação do Norte). It serves the population of the municipalities of Gaia and Espinho, with a total population of 350,000 patients (100% coverage) and indirect coverage of 1,2 million inhabitants considering the high differentiation of clinical care.
4	Healthcare setting / type of data	Primary care – General Practitioner, and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care, and other (specify). The integrated database captures information about demographics, visit occurrences, diagnoses, medications, and procedures.
5	Data collection process	Outpatient electronic health records, and Inpatient hospital electronic health records. The data is extracted directly from the source EHR system.
6	General representativeness	The healthcare facilities collecting the data are part of the national health service and treat patients of all socioeconomic levels. Therefore, the patient sample should be representative of the population.
7	Data content /source coding	Source data terminologies include ATC, RxNorm, ICD-9, ICD-10 codes, and ICPC-2 codes.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. The same patients cannot be registered under a different id in the OMOP CDM data.
9	Quality control (data source specific)	We have a two-step approach for data quality: A set of unit tests are implemented as data quality scenarios. These scenarios are tested at the ETL level and ran every time the ETL is executed. Those tests ensure that known issues in data quality are addressed and check automatically on every run. In addition, data quality is assessed through the OHDSI DataQualityDashboard project and reported as well. There is room for implementation of specific data quality checks that may be relevant at the study level, if required.
10	Linkage	No known linkages.
11	Vital status	Vital status captured through two mechanisms: 1) recording of deaths in hospital admissions (secondary and tertiary care); and 2) verification against the national patient registry, which is triggered by follow-up appointments in primary care.
12	Limitations	No database-specific limitations documented. General limitations for the data type applicable.
13	Main references	No main reference provided.
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111133 Website: https://www.emha.pt/

Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público (BIFAP)

#	Section	Description
1	Data source identification and country	BIFAP (Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público (Pharmacoepidemiological Research Database for Public Health Systems)) Spain
2	Data partner information section	AEMPS Pharmacoepidemiology and Pharmacovigilance Division - Medicines for human use Department
3	Coverage and timespan	Data collection since: 2001 Extent: Regional. Spanish National Health Service (SNS) from 9 of the 17 regions in Spain. The population currently included represents 36% of the total Spanish population.
4	Healthcare setting / type of data	Primary care – General Practitioner, and community pharmacists, and primary care specialists (e.g. paediatricians), and hospital inpatient care. BIFAP includes a collection of databases linked at individual patient level. The main one is the Primary care Database, given the central role of PCPs in the SNS. Linked, there are additional important structural databases, like the medicines dispensed at community pharmacies and the patients' hospital diagnosis at discharge. 7 out of the 9 regions have linkage to hospital data. However, hospital data is available for different time periods for each region. From 2014 onwards, linkage to hospital data is available for >68% of patients.
5	Data collection process	Insurance/administrative claims, and Outpatient electronic health records, and Inpatient hospital electronic health records, and Registries. Data in BIFAP is collected from Primary Care and Hospital EHR.
6	General representativeness	Spain has a SNS that provides universal access to health services through the Regional Healthcare Services. Primary care physicians (PCPs), both general practitioners and paediatricians, have a central role. They act as gatekeepers of the system and exchange information with other levels of care to ensure the continuity of care. Most of the population (98.9%) is registered with a PCP and, in addition, most drug prescriptions are written at the primary care level.
7	Data content /source coding	The BIFAP source data is coded in SNOMED, ICD, ICPC-2 (diagnoses), AEMPS (drugs), and local lab codes.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. Pseudonymized ID numbers are generated at regional level. The Personal Identification Code for the Autonomous Community (CIPA) is used to perform the pseudonymisation procedure. Therefore, upon changing practice or de- and re-registration within the same region, (Autonomous Community) the patient in BIFAP is correctly identified as the same person with the same ID number. However, the same patient would obtain different ID numbers if the patient moves to a different region and is registered in a primary care practice in the new region. The percentage of people who are de-registered due to moving to other region in relation to the BIFAP population is, for example, 5% in Madrid and 4% Castilla y Leon. This situation would have a very limited impact on the data analysis due to the following: - The proportion is low (less than 5%) in relation to the overall population in BIFAP. - In BIFAP, only stable residents are included. This means that patients living in another region for a foreseen short time period and are provisionally assigned to a primary care practice are not included in BIFAP. - Medical events of those patients who have more than one ID do not overlap in time, since dates of events correspond to different periods. This means that counts of these events are never duplicated. - A number of study designs allows the same patient to be part of different cohorts or to be selected both as case and as control, provided that their person-time experience correspond to a different period of time. In all these cases, the impact in study analysis of duplicated IDs would be negligible.

#	Section	Description
9	Quality control (data source specific)	Patients who meet any of the following disability criteria are discarded: - Non-owners of the individual health card - Date of birth before 01/01/1801 - Active patients over 115 years of age - Patients without clinical records (only contains administrative information) - Patients marked as "fictitious" in the clinical history - Badly coded sex - Inactive without termination date - Start date = End date - Clinical records prior to date of birth
10	Linkage	The following data are also linked at individual patient level and available. For a subset of the BIFAP population (regions and/or periods of time): • Information on dispensation of medicines at hospital pharmacies from outpatients and inpatients. • Registration of Causes of Death by the National Institute for Statistics. From the start of the COVID-19 pandemic: • Vaccines COVID-19 Administration Registry linked to patients included in BIFAP. • Diagnosis Tests of COVID-19 linked to patients included in BIFAP, for some regions.
11	Vital status	Source for vital status unknown.
12	Limitations	Primary care is available from 2001 but is considered complete since 2005. Hospital discharge has different coverage periods per region Spain, with most starting between 2014–2016. This means that for different regions and different time periods there is a different coverage of healthcare events. In the release of July 2025, the laboratory results are not covered. These will be added again at the next release, expected at the end of 2025.
13	Main references	Maciá-Martínez MA, Gil M,Huerta C,Martín-Merino E,Álvarez A,Bryant V,Montero D "Base de Datos para la Investigación Farmacoepidemiológica en Atención Primaria (BIFAP): A data resource for pharmacoepidemiology in Spain." Pharmacoepidemiology and drug safety (2020): 32337840
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/21501 Website: http://www.bifap.org/index_EN.html

Institut Municipal Assistència Sanitària Information System (IMASIS)

#	Section	Description
1	Data source identification and country	IMASIS (Institut Municipal Assistència Sanitària Information System) Catalunya, Spain
2	Data partner information section	Consorci Mar Parc de Salut Barcelona (PSMar) Direcció de Control de Gestió (Management and Control Department)
3	Coverage and timespan	Data collection since: 1990 Extent: Regional. The primary catchment area covers the Barcelona city area, in Catalonia (Spain). However, all recorded visits are included, which may also include patients residing outside this area.
4	Healthcare setting / type of data	Secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. The available information includes diagnoses, procedures, drug administrations, dispensations and prescriptions, laboratory tests, visit types, cancer data, and vaccinations. These data are linked to primary care centers, which currently contribute information on diagnoses, and treatments, as well as to mortality data.
5	Data collection process	Outpatient electronic health records, and Inpatient hospital electronic health records, and Other. The data collection process begins at the point of care, when a patient is attended in any facility

#	Section	Description
		within our hospital network or associated primary care centers. Clinical information is recorded directly into the electronic health records (EHRs) and subsequently complemented with data from external sources, including the regional vaccine registry, the cancer registry, the national mortality registry, and outpatient treatment records. All information is consolidated into a relational database using standardized formats and subjected to multiple quality control procedures to ensure completeness and consistency. The resulting datasets are then extracted, transformed, and loaded (ETL) into an OMOP Common Data Model (CDM) database on a quarterly to semi-annual basis.
6	General representativeness	The data source includes patients across all socioeconomic levels, ages, and genders, suggesting that the dataset is representative of the population of Barcelona requiring secondary (hospital-based) healthcare at any given time.
7	Data content /source coding	Health outcomes, diagnoses, and procedures are recorded using ICD-9-CM and ICD-10-CM classification. Laboratory tests are recorded in the source data using LOINC terminology and drugs using the national codes classification. Cancer registry is based on ICD-O-3 classification. Additionally, a set of demographics data are available such as age and sex.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. A patient cannot be registered under different ID numbers. They are registered under the same ID, even when changing practice of re-registration or admission.
9	Quality control (data source specific)	Several quality control procedures have been implemented during the creation and update of IMASIS. The process includes a series of filters that exclude any data whose format or content do not match the specifications of the target variables. Some examples of these actions include the following. - Records containing codes that do not belong to the standardized vocabularies used for encoding are reviewed within their respective domains and, if necessary, discarded. - Records with implausible values, such as dates preceding a patient's birthdate, are identified and corrected or removed. - Before each IMASIS update, the number of records in the source files is verified. Record counts from the new IMASIS instance are compared with both the source files and previous IMASIS versions to detect potential inconsistencies or anomalies in the source data. - In addition, if the Data Quality Dashboard (DQD) package identifies quality issues in the OMOP instance, for example diagnoses recorded for patients with an implausible gender, these discrepancies are reviewed and corrected in the source data.
10	Linkage	Information of diagnosis and treatments received in outpatient settings and in primary care as well as national mortality registry, cancer registry and regional vaccination registry.
11	Vital status	The hospital's death registry records the age at death for each patient when the event occurs within the hospital. In addition, deaths occurring outside the hospital are captured through linkage with the national mortality registry; however, in these cases, the specific cause of death is not currently available.
12	Limitations	No database-specific limitations documented. General limitations for the data type applicable.
13	Main references	Mayer MA, Furlong LI, Torre P, Planas I, Cots F, Izquierdo E, Portabella J, Rovira J, Gutierrez-Sacristan A, Sanz F "Reuse of EHRs to Support Clinical Research in a Hospital of Reference." Studies in health technology and informatics (2015): 25991136
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/104316 Website: https://www.parcdesalutmar.cat/en/

The Information System for Research in Primary Care (SIDIAP)

#	Section	Description
1	Data source identification and country	SIDIAP (The Information System for Research in Primary Care) Catalunya, Spain
2	Data partner information section	IDIAPJGol
3	Coverage and timespan	Data collection since: 2006 Extent: Regional. SIDIAP is a database of primary care electronic health records of the population of Catalonia, North-East Spain. It contains pseudo-anonymised records of more than 8 million people, of which 6.1 million are active as of 2022, representing around 76% of the Catalan population.
4	Healthcare setting / type of data	Primary care – General Practitioner, and hospital inpatient care. SIDIAP captured data includes routine visits, socio-demographics, diagnoses, laboratory tests, drugs (prescribed and dispensed), referrals, sick leaves and lifestyle information.
5	Data collection process	Outpatient electronic health records, and Inpatient hospital electronic health records. Data is entered by primary care physicians upon healthcare contact, supplemented with hospital discharge records. The Institut Catala de la Salut (Catalan Health Institute) is the data controller.
6	General representativeness	It was previously shown that the captured SIDIAP population is highly representative of the entire Catalan region in terms of geographic, age, and sex distributions.
7	Data content /source coding	SIDIAP data covers all services that occur at the Primary Care Centres, as well as support services, such as sexual and reproductive health or home end-of-life care. Drugs are coded in ATC-WHO terminology in the source data. Health outcomes are captured in ICD-10CM codes. The SIDIAP contains all laboratory tests and results performed in primary health centres. Demographics, geographical, as well as socio-economic factors are recorded for each patient.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. A patient has a unique id, also upon changing practice or re-registration.
9	Quality control (data source specific)	Internal and external validation processes are carried out to determine the data quality of the SIDIAP information at each data update. These include stratifying the data by geographical regions and year in order to identify differences in data collection that need to be harmonized (e.g. recording of specific information under different codes). The measurement units of variables measuring one characteristic are also homogenized (e.g. transformation of the data from every laboratory that measures haemoglobin to grams per decilitre). Visual inspection of all data included in the database by week is also conducted, allowing one to see temporal patterns in the registry of a certain variable. With this information, the SIDIAP team can issue recommendations to researchers about the most common variable(s) where certain information is recorded (e.g. there are several variables with information concerning the women's menopausal status and with these visual inspection tools the SIDIAP team can inform the researchers about which related variables have the largest number of records and could be more helpful to capture menopause). Data availability (longitudinally and reliability), plausibility (range checks and unusual values), and consistency are inspected through visualisation tools. In addition, before accessing the data for a requested project, research teams have access to a quality-control report. This document contains counts, years, percentiles, maximums and minimums, incidences, and prevalence of the data requested for the project, allowing detection of inconsistencies in the data extraction prior to data delivery. External validation processes of the SIDIAP database mainly include assessing the data recorded in SIDIAP through linkage to external gold standard data sources, by analysing free text, or by sending questionnaires to health professionals.

#	Section	Description
10	Linkage	SIDIAP is linked to a hospital discharge database, pharmacy dispensation, and primary care laboratories. It can also be linked to other registries in Catalonia on a project by project basis.
11	Vital status	Mortality is fully captured in SIDIAP. The cause of death is not available, but can be linked to the Spanish death registry on a project by project basis.
12	Limitations	The SIDIAP data is not representative of individuals not using public primary care, and conditions that are usually followed by specialist care might not be properly captured. In addition, there is limited information on lifestyle variables. Patients are followed until Death or when transferring to another primary health care centre that does not contribute to SIDIAP.
13	Main references	Recalde M, Rodríguez C, Burn E, Far M, García D, Carrere-Molina J, Benítez M, Moleras A, Pistillo A, Bolívar B, Aragón M, Duarte-Salles T "Data Resource Profile: The Information System for Research in Primary Care (SIDIAP)." International journal of epidemiology (2022): 35415748
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/50190 Website: https://www.sidiap.org/index.php/en

Health Impact - Swedish Population Evidence Enabling Data-linkage (HI-SPEED)

#	Section	Description
1	Data source identification and country	HI-SPEED (Health Impact - Swedish Population Evidence Enabling Data-linkage) Sweden
2	Data partner information section	SMPA-GU, Läkemedelsverket, Box 26, 751 03 Uppsala, Sweden - Box 469, 405 30 Gothenburg, Sweden Pharmacoepidemiology and Analysis Department (FeA) - School of Public Health and Community Medicine, Institute of Medicine
3	Coverage and timespan	Data collection since: 2015 Extent: Nation-wide. The catchment area includes the whole of Sweden, covering the full population of approximately 11.7 million.
4	Healthcare setting / type of data	Primary care – General Practitioner, and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. Primary care (GPs) is available only for the 2 largest regions (~40% of national population) The following data elements are collected: Socio-demographics, dispensed drug prescriptions, cause of death, diagnoses and procedures from secondary (specialist) care and inpatient visits or clinical events, as well as from primary care visits (40%pop only).
5	Data collection process	Registries. The data is acquired from the relevant Swedish national and regional registries, only once all legislative, GDPR and ethical approvals have been granted. Therefore only relevant data is passed on, which will then be entered and processed by the study team. The data are updated several times annually.
6	General representativeness	The coverage includes all patients of all sociodemographic characteristics. Therefore it should mirror the source population to a very good extent.
7	Data content /source coding	Medicines are coded with ATC and NPLID (National Product ID), ICD10-SE is used for diagnoses and the Swedish procedure coding system (KVA) is used for clinical procedures.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. Patients have a uniquely id across datasets.
9	Quality control (data source specific)	The source data are obtained from the relevant Swedish National and Regional Registers. The registers perform some regular quality controls on their data. After receiving the data, we

#	Section	Description
		perform additional checks and cleaning. We also run regular quality checks on the data we manage.
10	Linkage	Data on specialist care is acquired from the National Patient Register, mortality information is provided by the Cause-Of-Death Registry. Drug data is provided by the Patient Drug Register. And similarly for other registers. Data are linked very accurately using the national personal ID number, and pseudonymized before delivery to HI-SPEED. All data are updated 2-4 times per year.
11	Vital status	Data on death and underlying + contributing causes-of-death are extracted from the Cause-of-Death registry (i.e. based on death certificates).
12	Limitations	General limitations for the data type applicable. This is a research project where all studies require ethics approval. Data collection since: 2015 for most data, except prescribed drug register (from 2018), and some COVID-related data (tests, vaccination) from 2020 Primary care is only available for a subset.
13	Main references	Nyberg F, Franzén S, Lindh M, Vanfleteren L, Hammar N, Wettermark B, Sundström J, Santosa A, Björck S, Gisslén M "Swedish Covid-19 Investigation for Future Insights - A Population Epidemiology Approach Using Register Linkage (SCIFI-PEARL)." Clinical epidemiology (2021): 34354377
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/node/4463/ Website: https://www.gu.se/en/research/scifi-pearl

Clinical Practice Research Datalink GOLD (CPRD GOLD)

#	Section	Description
1	Data source identification and country	CPRD GOLD (Clinical Practice Research Datalink GOLD) United Kingdom
2	Data partner information section	University of Oxford NDORMS
3	Coverage and timespan	Data collection since: 1987 Extent: Nation-wide. CPRD GOLD consists of patients in contributing practices using Vision software. Historically this covered the whole of the UK, but the number of contributing practices in the England is dropping. In January 2025 only 3 practices from England were a part of CPRD GOLD, while historical patient data were from the whole of the UK, and will continue to be so. In the future, no practices from England will be present, only practices from Scotland, Wales, and Northern Ireland.
4	Healthcare setting / type of data	Primary care – General Practitioner, and primary care specialists (e.g. paediatricians), and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. CPRD GOLD data include patient demographics, biological measurements, clinical symptoms and diagnoses, referrals to specialist/hospital and their outcome, laboratory tests/results, and prescribed medications.
5	Data collection process	Outpatient electronic health records. Data are entered by clinicians into the EHR. Data is processed by CPRD that provides data releases for research.
6	General representativeness	In the last 10 years, the CPRD GOLD regional distribution of currently contributing general practitioner (GP) practices has significantly shifted, resulting in many new practices joining from Scotland, Wales, and Northern Ireland, and fewer participating from England. These changes have affected the CPRD GOLD population size, regional coverage, and eligibility for data linkages. CPRD GOLD January 2024 contains >21.3 million historical and current patients (12.9 in England,

#	Section	Description
		3.1 in Wales, 4.7 in Scotland, 0.7 in Northern Ireland). Of these, nearly 3 million are currently registered in a GP practice and represent ~4.3% of the estimated current UK population (0.1% in England, 32.3% in Wales, 28.6% in Scotland, 16.2% in Northern Ireland). Patients currently registered in CPRD GOLD January 2024 are broadly representative of the UK population with respect to age and sex. Reference: https://doi.org/10.1093/ije/dyaf077
7	Data content /source coding	Gemsript, Read, dm+d
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. In GOLD, a patient can be registered under different ID numbers upon changing practice or re-registration. Researchers are not able to identify these patients, as the data are anonymised. However, GOLD covers less than 5% of the current UK GP practices and it is unlikely that an individual who does change GP practice ends up in another GP practice which uses the Vision software and accepts the CPRD data collection agreement. The very small number of duplicated IDs will have different observation periods and should not have an impact on the data analyses.
9	Quality control (data source specific)	CPRD GOLD only includes practices whose data quality is assessed to be up-to-standard (UTS). Each practice is associated to an UTS date set when the data quality standards become satisfactory, and CPRD recommend using only longitudinal data starting from this UTS date. Every time CPRD collect the EHR from a practice, checks are run for the data quality standards, and if they are not adequate, the EHR is not accepted. When the data quality becomes acceptable again, CPRD updates the practice UTS date. CPRD also checks data quality standards at the patient level, and associates each patient with a flag, reporting if its data are acceptable for clinical research. Only patients with acceptable data quality are included in the population to be mapped to CDM.
10	Linkage	CPRD GOLD can be linked to several sources, however our Oxford OMOP CDM is only linked to the CPRD GOLD Ethnicity Record and to the CPRD Townsend Deprivation Index at the Practice Level.
11	Vital status	The date of death in CPRD GOLD has been validated against the Population registry (ONS) mortality data. Reference: https://doi.org/10.1002/pds.4747
12	Limitations	The main limitation is due to the fact that CPRD GOLD is limited to GP records, and although it contains information on referrals and discharge letters, it may not fully capture specific hospital information. Events from hospital and specialist care are not covered.
13	Main references	Sanchez-Santos MT, Axson EL, Dedman D, Delmestri A "Data Resource Profile Update: CPRD GOLD." International journal of epidemiology (2025): 40499193
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111113 Website: https://www.cprd.com/data/primary-care-data/cprd-gold

ANNEX II. Fitness for use assessment

The information provided below was supplemented by a specific pregnancy questionnaire (carried out in Objective 1), which data partners used to provide additional detail on the availability and structure of relevant variables in each data source. These insights helped contextualise and inform analytic decisions.

Croatian National Public Health Information System (NAJS)

NAJS was included in this study because it is a primary, secondary, and hospital data source that has data on pregnancy and mother-child linkage, although this data has not been mapped.

Data availability in NAJS began in 2014; however, data owners recommended its use starting in 2017, when data is considered of sufficient quality for research purposes. The date of the most recent data extraction is 2025. The median length of the observation period was 3640 days (IQR: 3110–3740).

Danish Data Health Registries (DK-DHR)

DK-DHR was included in this study because it is a registry-based data source that includes secondary and hospital care (including emergency room visits) health data, with partially mapped pregnancy data and mother-child linkage. Births, procedures (including caesarean procedures), and diagnoses were already mapped from the patient registry.

Data availability began in 1995, and the date of the most recent data extraction is 2025. The median length of the observation period was 7920 days (IQR: 2610–10,900).

Tampere University Hospital patient cohort (FinOMOP-TaUH Pirha)

FinOMOP-TaUH Pirha was included in this study because it is a secondary care and hospital data source with unmapped pregnancy and mother-child linkage data.

Moreover, data availability and follow-up were considered sufficient for this study. Data availability began in 2007, and the date of most recent data extraction is 2024. The median length of the observation period (i.e. the continuous time a person was observable in the data source) was 4230 days (IQR: 374–7980).

Finnish Care Register for Health Care (FinOMOP-THL)

FinOMOP-THL was included in this study because it is a registry-based data source including private and public, primary, secondary, and hospital health data, with unmapped pregnancy data and mother-child linkage.

Moreover, data availability and follow-up were considered sufficient for this study. Data availability began in 2011, and the date of most recent data extraction is 2024. The median length of the observation period was 4020 days (IQR: 4020–5030).

Clinical Data Warehouse of Bordeaux University Hospital (CDW Bordeaux)

CDW Bordeaux was included in this study because it is a hospital data source with mapped information on pregnancy.

Moreover, data availability and follow-up in CDW Bordeaux were considered sufficient for this study. Data availability began in 2005, and the date of the most recent data extraction was 2024. The median length of the observation period was 384 days (60–2450).

InGef Research Database (InGef RDB)

InGef RDB was included in this study because it is a claims data source including primary, secondary, and hospital care data, with unmapped pregnancy-related information.

Moreover, data availability and follow-up in InGef RDB were considered sufficient for this study. Data availability for InGef RDB includes the last 10 years (by law), and the date of the most recent data extraction was in 2024. The median length of the observation period was 3560 days (IQR: 1400–3650).

Semmelweis University Clinical Data (SUCD)

SUCD was included in this study because it is a secondary care and hospital data source that has data on pregnancy and mother-child linkage, although this data has not been mapped.

Moreover, data availability and follow-up in SUCD was considered sufficient for this study. Data availability began in 2005, and the date of most recent data extraction is 2025. The median length of the observation period was 266 days (IQR: 0–2170).

Integrated Primary Care Information (IPCI)

IPCI was included in this study because it is a primary care data source with mapped pregnancy data, with mother-child linkage currently under development.

Moreover, data availability and follow-up were considered sufficient for this study. Data availability began in 2006, and the date of the most recent data extraction is 2024. The median length of the observation period was 1730 days (IQR: 791–3070).

The Norwegian Linked Health Registry data (NLHR), Norway

NLHR was included in this study because it is a registry data source that has information on pregnancy. Additionally, NLHR had pre-processed pregnancy data using the PET (3) and had previously conducted pregnancy-related studies using source data and OMOP-mapped data and had been recognised for its value in providing reliable and comprehensive data for pregnancy research.

There were study-specific limitations present in NLHR, namely that while notifications of pregnancies and pregnancy outcomes were mandatory in Norway for all pregnancies past gestational week 12, those dated before could be missed. However, the UiO Pregnancy algorithm improved the detection and dating of early non-live pregnancy outcomes that would have gone unnoticed if relying solely on the medical birth registry information by linking the Medical Birth Registry of Norway (MBRN) records with hospital and primary care records.

This database has 2 available CDMs, which are: The Medical Birth Registry of Norway (MBRN) (NLHR@University of Oslo (UiO):PERINATAL CDM) and NLHR (NLHR@UiO CDM dataset). Both CDMs use the UiO pregnancy algorithm and cover the entire Norwegian population. However, they have different source populations, based on the data source that provides the linkage IDs (MBRN vs. national registry) and the data period covered. Specifically (see [Table](#) below):

- **NLHR@UiO:PERINATAL CDM**, IDs come from the medical birth registry, covering all individuals with a pregnancy event >12 gestational weeks, and covers the period 2008–2019.
- **NLHR@UiO CDM**, IDs come from the National Registry covering the entire Norwegian population and covers the period 2018–2023.

Overview of the two Norwegian CDM datasets available.

Pregnancy algorithm)	Data source	Information	Years	Source study population	N pregnancies	Prevalence pregnancy outcomes					
						LB	SAB	T	SB	ECT	MO
UiO algorithm	NLHR@UiO, P1380	MBRN, NPR KUHR	2018–2023	All females in Norway*	463 400	68.9	15.3	14.4	0.2	1.1	0.1

Pregnancy algorithm)	Data source	Information	Years	Source study population	N pregnancies	Prevalence pregnancy outcomes					
						LB	SAB	T	SB	ECT	MO
UiO algorithm	NLHR@UiO:PERINATAL, P704	MBRN, NPR KUHR	2008–2018	Pregnancy episodes in MBRN*	859 449	74.8	13.1	11.0	0.3	0.7	0.1

*All females with a recorded pregnancy in MBRN, their partners and offsprings.

LB: Live births, SAB: spontaneous abortions, T: elective terminations, SB: stillbirths. ECT: ectopic pregnancies, MO: molar pregnancies. MBRN: medical birth registry, NPR: Patient registry, KUHR: Primary care registry.

Overall, both could identify similar pregnancy events. However, the NLHR@UiO CDM has a greater ability to detect early pregnancies (<12 gestational weeks), as its source population includes the entire Norwegian population. In contrast, the source population in the NLHR@UiO:PERINATAL CDM is based on the MBRN, which only include individuals with a pregnancy episode lasting beyond gestational week 12. Although the UiO pregnancy algorithm improved the identification of pregnancy events <12 gestational weeks, in this CDM it can only improve detection for individuals present in the MBRN. Consequently, differences between the two CDMs are likely driven by the NLHR@UiO:PERINATAL CDM's has lower ability to detect elective terminations, particularly among younger individuals, who were less likely to have had a prior pregnancy extending beyond 12 gestational weeks, and therefore were also less likely to appear in the MBRN.

Of the two available CDMs, we selected the NLHR@UiO CDM for this study because it had a greater ability to detect early pregnancies and its more granular information on drug exposures, making it more suitable for use in future studies.

The median length of observation period was 5840 days (5230–5840) for NLHR@UiO: CDM.

Egas Moniz Health Alliance database - Gaia E Espinho (EMDB-ULSGE)

The EMDB-ULSGE was included in this study because it is a secondary care data source, with unmapped pregnancy and mother-child linkage data.

Moreover, data availability and follow-up in EMDB-ULSGE were considered sufficient for this study. Data availability began in 2008, and the date of most recent data extraction is 2023. The median length of the observation period was 1590 days (IQR: 58–4770).

Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público (BIFAP)

BIFAP was included in this study because it is a primary care data source with ongoing development of mother-child linkage. It had previously implemented pregnancy identification algorithms on the source data and has been used in prior pregnancy-related studies.

Moreover, data availability and follow-up in BIFAP was considered sufficient for this study. Data availability began in 2005, and the date of the most recent data extraction was in 2024. The median length of the observation period was 4020 days (IQR: 1810–6260).

This database covers 12 regions in Spain. Since 2014, 10 regions have been linked to hospital discharge data.

Institut Municipal Assistència Sanitària Information System (IMASIS)

IMASIS was included in this study because it is a secondary care and hospital data source, with unmapped pregnancy-related information and is currently developing mother-child linkage.

Moreover, data availability and follow-up in IMASIS were considered sufficient for this study. Data availability in IMASIS began in 1990, and the date of the most recent data extraction in 2024. The median length of the observation period was 66 days (IQR: 0–2070).

The Information System for Research in Primary Care (SIDIAP), Spain

SIDIAP was included in this study because it is a primary care data source, which was linked to hospital discharge data in this study. This data source provided information on pregnancy, which has been pre-processed using the PET and had previously conducted pregnancy-related studies using OMOP-mapped data.

Moreover, data availability in SIDIAP was considered sufficient for this study. Data collection began in 2006, and the most recent data extraction date in June 2023. The median length of the observation period was 5670 days (IQR: 2220–6390).

In SIDIAP, voluntary pregnancy losses could be underreported. Additionally, data did not include data from private care settings.

Health Impact - Swedish Population Evidence Enabling Data-linkage (HI-SPEED)

HI-SPEED was included in this study because it is a Registry, primary, secondary, and hospital care data source that provides information on pregnancy in the general population. Additionally, this data source had pre-processed pregnancy data using the PET and had previously conducted pregnancy-related studies using OMOP-mapped data.

Moreover, data availability in HI-SPEED was considered sufficient for this study. Data availability in starts in 2015, and the date of the most recent data extraction in September 2024. The median length of the observation period was 3530 days (IQR: 3420–3530).

Clinical Practice Research Datalink GOLD (CPRD GOLD)

CPRD GOLD was included in this study because it is a primary care data source with available pregnancy and mother-child linkage at source. Several studies on source data had been conducted to inform the extraction of pregnancy data and provide descriptive analysis of its outputs. Additionally, this data source had previously conducted pregnancy-related studies using OMOP-mapped data, where pregnancy episodes were inferred from primary care data using a validated algorithm.

Moreover, data availability and follow-up in CPRD GOLD were considered sufficient for this study. Data availability began in 1987, and the date of the most recent data extraction in 2024. The median length of the observation period was 2150 days (IQR: 727–4930).

ANNEX III. Operational and reporting considerations

DATA MANAGEMENT

Data management

All data sources have previously mapped their data to the OMOP common data model. This enabled the use of standardised analytics and using DARWIN EU® tools across the network since the structure of the data and the terminology system was harmonised. The OMOP CDM was developed and maintained by the Observational Health Data Sciences and Informatics (OHDSI) initiative and is described in detail on the wiki page of the CDM: <https://ohdsi.github.io/CommonDataModel> and in The Book of OHDSI: <http://book.ohdsi.org>

The analytic code for this study was written in R and used standardised analytics wherever possible. Each data partner executed the study code against their data source containing individual data and then returned the results (csv files) which only contained aggregated data. The results from each of the contributing data sites were then combined in tables and figures for the study report.

Data storage and protection

For this study, personal data from individuals in various EU member states were processed, using information collected from national/regional electronic health record data sources. Due to the sensitive nature of this personal medical data, it is important to be fully aware of ethical and regulatory aspects and to strive to take all reasonable measures to ensure compliance with ethical and regulatory issues on privacy.

All data sources used in this study were already used for pharmaco-epidemiological research and have a well-developed mechanism to ensure that European and local regulations dealing with ethical use of the data and adequate privacy control were adhered to. In agreement with these regulations, rather than combining person-level data and performing only a central analysis, local analyses were run, which generated non-identifiable aggregate summary results.

QUALITY CONTROL

Data source quality control

The *CohortDiagnostics* (<https://github.com/OHDSI/CohortDiagnostics>) R package was run to assess the use of different codes across the data sources contributing to the study and identify any codes potentially omitted in error.

The study code for characterisation was based on DARWIN EU® R packages: *IncidencePrevalence* to estimate Incidence and Prevalence, and *CohortCharacteristics* to characterise the cohort by indication. These packages included numerous automated unit tests to ensure the validity of the codes, alongside software peer review and user testing. The R package used is available via GitHub.

ANNEX IV. List of stand-alone documents

Table S1. National statistics table utilised as reference for examining external validity of pregnancy algorithm.

Indicators			Year	Croatia	Denmark	Finland	France	Germany	Hungary	Netherlands	Norway	Portugal	Spain	Sweden	UK
Birth rate	Crude birth rate (per 1000 person)		2014	9.4	10.1	10.5	12.4	8.8	9.5	10.4	11.5	7.9	9.2	11.9	12
			2023	8.3	9.7	7.8	9.9	8.3	9.1	9.2	9.4	8.1	6.6	9.5	10
Maternal age at birth	Mean age of woman at birth of first child		2014	28.1	29.2	28.6	28.3	29.4	27.7	29.5	28.7	29.2	30.6	29.2	28.6
			2023	29.3	30.1	30	29.1	29.8	28.8	30.4	30.1	30.2	31.5	30	30.9
Pregnancy outcome rates	Number of live births per year		2014	39,566	56,870	57,232	819,328	714,927	93,281	175,181	58,976	82,367	426,076	114,907	7759,08
			2023	32,170	57,469	43,383	678,333	692,989	87,671	164,487	51,983	85,699	319,830	100,051	591,072
	*Foetal mortality rate: foetal deaths at or after 24-weeks threshold (per 1000 total births)		2015	4.7	3.6	3.1	8.4	3.8	5.1	3	4	2.9	3.1	3.8	4.4
			2019	4.5	2.9	2.7	8.4	4.1	4.7	2.3	3.5	3.3	3.1	3.1	3.2
Gestational duration distribution	Number of live births at each completed gestational age range	<32	2015	328	43	7,701	9,635	1,296	1,834	528	886	1,236	1,068	3,583	8,327**
		32–36		2,101	2,764	46,033	51,907	6,687	9,657	3,140	5,990	6,910	5,536	23,345	44,619
		37–38		8,173	10,002	164,040	184,069	24,306	38,904	10,883	23,136	20,686	21,208	82,306	143,892
		39–41		25,910	40,112	533,420	475,741	59,202	113,305	42,672	55,464	76,021	80,218	241,253	477,667
		≥42		747	2,207	6,339	3,874	144	2,111	2,290	115	3,367	8,433	5,710	18,403
	<32	2019	317	329	7,525	9,598	1,148	1,739	493	896	556	963	3,374	7,676	
	32–36		2053	2,094	41,266	52,243	6,264	8,919	2,871	6,082	2,994	5,332	19,480	41,839	
	37–38		7676	8,434	153,025	193,314	24,332	38,119	10,525	22,751	11,107	22,096	72,411	153,224	
	39–41		25,278	33,696	503,141	501,968	57,215	112,474	38,757	57,045	44,512	80,384	222,987	422,193	
	≥42		1,147	1,113	5,945	3,917	123	2,250	2,395	106	1,404	6,926	3,515	12,696	
		Vaginal ¹	2015	78.4	83.5	79.3	67.7	61.2	82.7	83.4	74.8	78.5	82.1	73.4	UK-NI: 69.9

Indicators		Year	Croatia	Denmark	Finland	France	Germany	Hungary	Netherlands	Norway	Portugal	Spain	Sweden	UK
Mode of delivery	Mode of delivery rate (% vaginal and c-section deliveries)													UK-S: 67.6 UK-W: 74.1
	C-section ²		21.6	16.5	20.7	32.3	38.8	17.3	16.6	16.1	21.5	17.9	26.7	UK-NI: 30.2 UK-S: 32.4 UK-W: 25.9
	Vaginal ¹	2019	73.8	82.1	79.1	68.2	58.5	82.7	83.6	72.2	79.7	81.8	74.3	UK-NI: 67.3 UK-S: 64.5 UK-W: 71.7
	C-section ²		26.2	17.9	20.9	31.8	41.5	17.4	16.4	16	20.3	18.2	25.7	UK-NI: 32.7 UK-S: 35.5 UK-W: 28.3

*Late Foetal Mortality Rate as the number of stillbirths (foetal deaths at or after 28 weeks gestation) divided by the total number of births (live births + stillbirths), expressed per 1,000 births

**UK: England and Wales, UK-NI: Northern Ireland, UK-S: Scotland, UK-W: Wales

¹Includes (vaginal spontaneous, instrumental, and vaginal-unknown)

²Includes (Caesarean-no labour or elective, Caesarean-during labour or emergency, Caesarean-type unknown)

Table S2. A scoping literature review to assess published pregnancy algorithms.

Study	Year of publication	Country	Data Source	Setting	OMOP CDM used	Study period	How pregnancy identified	Pregnancies identified	Validation measures	Outcomes captured	Limitations
Hardy et al.	2004	UK	GPRD	General practice	No	1991 to 1999	Estimating a beginning and endpoint of each distinct pregnancy in the data involved mapping the earliest detected pregnancy marker (diagnoses and	297,082 pregnancies	Indirect checks (like data consistency checks, assessment of unmapped pregnancy markers/out	Delivery codes (live or stillborn), live births, elective termination, records of foetal death	- Indirect validation (no manual chart review available, costly review of GPRD records). - First trimester had to be estimated

Study	Year of publication	Country	Data Source	Setting	OMOP CDM used	Study period	How pregnancy identified	Pregnancies identified	Validation measures	Outcomes captured	Limitations
							procedures codes) linking to its respective pregnancy outcome		comes, comparison with external studies) were used		indirectly (no data on LMP) - Uncertainty remained in distinguishing preterm from term pregnancies
Hornbrook et al.	2007	US	Kaiser Permanente Northwest Integrated Health Care Delivery System	Administrative and clinical databases	No	1998 to 2001	Algorithm searches for pregnancy indicators. Hierarchy based determination of pregnancy episode, episode end date, and outcome type	24,680 pregnancies in 21,001 individuals	External validation using structured chart review	Pregnancy outcomes such as live births, ectopic pregnancies, abortions, morbid pregnancy outcomes	- Gap in LMP data for low-risk pregnancies
Margulis et al.	2012	Canada	British Columbia linked administrative data	Administrative claims database	No	1998 to 2007	Estimate the gestational age at birth based on LMP, dating scan, preterm-status indicator, maternal age, singleton/multiple status, and claims for routine prenatal care tests. GA at birth is a field in the database (including recorded clinical exam at birth or	286,432 births in 209,532 individuals	Validated using clinical gestational age at birth recorded in the delivery discharge record	Gestational age	- Study restricted to live births in administrative data. - Gestational age recorded in completed weeks only, therefore maximum precision is ± 1 week

Study	Year of publication	Country	Data Source	Setting	OMOP CDM used	Study period	How pregnancy identified	Pregnancies identified	Validation measures	Outcomes captured	Limitations
							delivery discharge record)				
Cea-Soriano et al.	2013	UK	The Health Improvement Network	UK primary care database	No	1996 to 2010	Identification of first pregnancy during the follow-up based on LMP, end of pregnancy specific codes (pregnancy tests, prenatal visits) each classified in separate group. Linkage of deliveries with infants born during the study period	191,000 pregnancies	Not performed	Completed pregnancies and pregnancies losses (ectopic pregnancies, miscarriages, terminations, and stillbirths)	- Only first pregnancy was assessed. - Incomplete or misclassified ascertainment of elective terminations and abortions may have occurred - Under recording of preterm and post-term deliveries and imperfect recording of LMP in electronic medical records databases
Mikolajczyk et al.	2013	Germany	German Pharmacoepidemiological Research Database (GePaRD)	Data from four statutory health insurances	No	2005	For the identification of births (inpatient date of delivery, date of an inpatient operation or procedure indicating birth, date of an inpatient diagnosis indicating birth	94,261 pregnancy outcomes identified	Comparison with national statistics	Outcomes included term births, preterm births, stillbirths, induced abortions, spontaneous abortions, and ectopic pregnancies.	-Underestimation of induced and spontaneous abortions -External validation of pregnancy outcomes by medical chart review was not possible.

Study	Year of publication	Country	Data Source	Setting	OMOP CDM used	Study period	How pregnancy identified	Pregnancies identified	Validation measures	Outcomes captured	Limitations
							and ambulatory services. Hierarchical algorithm starting with births, followed by induced abortions, ectopic pregnancies, and spontaneous abortions. indicating birth.				
Li et al.	2013	US	Medication Exposure in Pregnancy Risk Evaluation Program (MEP REP)	Administrative health plan data with additional data source	No	2001 to 2007	A gestational age algorithm based on the delivery date and diagnosis codes for preterm birth from electronic health plan data performed well in classifying prenatal medication exposure status.	225,384 live births	Validation using the “gold standard” gestational age obtained from the infant birth certificate/records.	Gestational age	- Underestimated the prevalence of preterm delivery
Matcho et al.	2018	US and UK	Truven MarketScan® Commercial Claims and Encounters (CCAE), Truven MarketScan® Multi-state Medicaid (MDCD), and	Administrative health claims data	Yes	1987 to 2013	Definition of pregnancy episodes by inferring pregnancy start dates (using a derived hierarchy of available pregnancy markers, including	885,608 (CPRD), 1,087,626 (Optum), 3,278,013 (CCAE), and 518,112 (MDCD) pregnancies	Electronic patient profiles validation and completion of a survey by personnel familiar with	Pregnancy outcomes were classified as live births, stillbirths, abortions, and ectopic pregnancies.	- Less accuracy was found for start date estimations in stillbirth and abortion outcomes - Valid outcomes with record dates too close in days

Study	Year of publication	Country	Data Source	Setting	OMOP CDM used	Study period	How pregnancy identified	Pregnancies identified	Validation measures	Outcomes captured	Limitations
			the Optum ClinFormatics® (Optum) database and one non-US database, the United Kingdom (UK) based Clinical Practice Research Datalink (CPRD).				records such as last menstrual period and nuchal ultrasound dates) and outcomes classification.		observational data.		would not be captured - Ectopic pregnancy episodes were missed by algorithm because of delayed treatment
Blotière et al.	2018	France	French national health data system	Claims data	No	2007 to 2015	The algorithm searched for discharge diagnoses and medical procedures indicative of completion of a pregnancy coded. all records of codes indicative of completion of a pregnancy were grouped into 2 categories: (1) births (end of pregnancy ≥ 22 weeks after the last menstrual period [LMP]); (2) any terminations of pregnancy < 22	9,647,843 pregnancies	External validation using national French statistics	Pregnancy start date and outcomes (live births, stillbirths, elective abortions, therapeutic abortions, spontaneous abortions, and ectopic pregnancies)	- ICD-10 diagnosis codes and medical procedure codes that may be subject to coding errors, and data from medical records could not be used to validate pregnancy outcomes. - Some pregnancies may not have been identified, particularly anonymized abortions, which

Study	Year of publication	Country	Data Source	Setting	OMOP CDM used	Study period	How pregnancy identified	Pregnancies identified	Validation measures	Outcomes captured	Limitations
							weeks after the LMP.				can be requested by minors -Spontaneous abortions which are not managed in hospital were missed - Exact pregnancy outcome dates were not available before 2009
Wentzell et al.	2018	Germany	GePaRD (German claims database)	Claims data	No	2006 to 2014	Data include multiple dated codes (ICD-10-GM, OPS, EBM) for each pregnancy. A hierarchical algorithm selects the most reliable information and assigns outcome dates using two steps: first, it prioritizes pregnancy outcomes (births, induced abortions, ectopic pregnancies, spontaneous abortions); second, it prioritizes data	1,235,261 pregnancy outcomes identified	Clinical review, 20 longitudinal records manually reviewed by 2 clinical experts using blinded concordance assessment	Pregnancy outcomes (live births, induced abortions, ectopic pregnancy, spontaneous abortions)	-Data do not include claims of midwives - Outcomes not included (as spontaneous abortions, induced abortions, and perinatal death shortly after birth) -Pregnancy outcomes for which no medical care is sought like unrecognized early losses are not well reported.

Study	Year of publication	Country	Data Source	Setting	OMOP CDM used	Study period	How pregnancy identified	Pregnancies identified	Validation measures	Outcomes captured	Limitations
							sources within each outcome (inpatient records first, followed by outpatient codes).				-Only late spontaneous abortions are recorded
Minassian et al.	2019	UK	CPRD GOLD	Primary care	No	1987 to 2018	The pregnancy (hierarchical classification) algorithm used all available pregnancy data (entity codes) to determine the timing of pregnancy (start, end, and trimester dates), the outcome (live birth, stillbirth, or early pregnancy loss), and additional details including whether a pregnancy was preterm, post-term, or multiple.	5.8 million pregnancies in 2.4 million individuals	Internal validation by comparison with linked Hospital Episode Statistics. External validation by national statistics and published data.	Timing of pregnancy (start, end, and trimester dates), the outcome (live birth, stillbirth, or early pregnancy loss), and additional details including whether a pregnancy was preterm, post term, or multiple.	-Underestimates terminations and livebirths - Incomplete capture of early losses -Relied on UK-specific coding hierarchies. -No explicit GA estimation
MacDonald et al.	2019	US	Truven Health MarketScan Commercial Claims and Encounters Database	Health insurance claims database.	No	2011 to 2015	Identify pregnancies in two-step algorithm using pregnancy end codes, infant linkage, and pregnancy marker codes with	2.76 million pregnancies	Indirect validation by comparison with national statistics	Livebirths, spontaneous abortions, elective terminations, stillbirths	- Underestimated rates due to low reporting - Incomplete coverage of terminations by

Study	Year of publication	Country	Data Source	Setting	OMOP CDM used	Study period	How pregnancy identified	Pregnancies identified	Validation measures	Outcomes captured	Limitations
							prioritization for specific timing codes				insurance companies
Sánchez Ortiz et al.	2019	Spain	BIFAP	Primary care	No	2002 to 2015	Pregnancy episodes identified using diagnosis codes. Pregnancy duration was estimated using LMP, GA, and pregnancy outcomes. A hierarchical algorithm searched sequentially for codes: (1) pregnancy confirmation, (2) end-of-pregnancy outcomes (e.g. delivery, abortion, ectopic pregnancy, stillbirth), and (3) other pregnancy-related codes.	155,419 pregnancies	Not performed	Completed pregnancies, ectopic pregnancies, stillbirths, abortions.	-Possibly missed pregnancies from private clinics or delayed recording of GA - Algorithm not validated
Canelón et al.	2021	US	Penn Medicine EHR	EHR	No	2010 to 2017	Codes and their frequency of usage coupled with temporal information regarding the timing of code assignments to	63,334 deliveries in 50,560 individuals	External validation using Birth log from Department of Obstetrics	Patient delivery dates and delivery-specific details	- Minor Delivery date discrepancies: Complicated pregnancies (e.g. assisted reproduction, conjoined twins)

Study	Year of publication	Country	Data Source	Setting	OMOP CDM used	Study period	How pregnancy identified	Pregnancies identified	Validation measures	Outcomes captured	Limitations
							infer each patient delivery dates.				involved complex billing records, challenged accurate delivery date detection. - Missed early losses (deliveries under 20 weeks gestation) - Incomplete capture of foetal losses
Moll et al.	2021	US	IBM MarketScan Explorys Claims-EMR Data Set	Claims data	No	2016 to 2018	Pregnancy episodes were constructed following a hierarchy, starting with the outcomes considered most reliably coded with respect to outcome identification and timing in claims data and based on consultation with physicians advising the study.	34,204 pregnancies	Clinical adjudication EMR chart review by OB-GYN.	Full-term births, preterm births, spontaneous abortions, stillbirths	- Using only on structured EHR components as a reference method hasn't been validated in other studies. - Not all pregnancies identified in claims have corresponding EMRs, leading to a convenience sample for validation. - Pregnancies with shorter gestational periods (e.g. spontaneous

Study	Year of publication	Country	Data Source	Setting	OMOP CDM used	Study period	How pregnancy identified	Pregnancies identified	Validation measures	Outcomes captured	Limitations
											abortions) may be underrepresented due to fewer healthcare encounters.
Lyu et al.	2022	US	National COVID Cohort Collaborative (N3C)	EHR	Yes	2018 to 2021	Pregnancy start dates were estimated from GA records, and end dates from DOD records. Since some records contained only one of these, separate GA and DOD cohorts were created and then combined to determine the most accurate pregnancy timelines.	296,194 pregnancies in 270,897 individuals	Random sample of 30 patients from the final cohort was manually reviewed. Accuracy and granularity as well as clinical validity for GA and DOD were assessed.	Estimation of gestational age	- Small validation sample (n=40) - No linkage to infant records - Missing information about time of some examinations or laboratory tests administration
Bertoia et al.	2022	US	Optum Research Database	Insurance claims database	No	2015 to 2020	Two approaches were used to identify unique pregnancies and to estimate the LMP date. The first (Z3A) approach identified active (ongoing) pregnancies using Z3A codes. The	1,030,874 pregnancies in 841,196 individuals	None reported (validation ongoing)	Pregnancy outcomes (livebirth, stillbirth, abortion)	- Algorithms, have not been validated against medical records - Possible imprecise LMP estimation that could lead to misclassification of trimester timing and

Study	Year of publication	Country	Data Source	Setting	OMOP CDM used	Study period	How pregnancy identified	Pregnancies identified	Validation measures	Outcomes captured	Limitations
							second (non-Z3A) approach identified completed pregnancies and backdated LMP according to the type of outcome and outcome date.				medication exposure. - Unvalidated mother–infant linkage
Jones et al.	2023	US	National COVID Cohort Collaborative (N3C)	Harmonised information from single- and multi-hospital health systems	Yes	2018 to 2022	A composite algorithm to identify pregnancy episodes using gestational age-specific signatures of a progressing pregnancy for further episode support, and pregnancy start date inference.	816,471 pregnancies in 628,165 individuals	Clinical validation using external clinician "annotation review"	Gestational age and pregnancy outcomes (live births, (stillbirth, ectopic pregnancy, abortions), and 23.3% had unknown outcomes	- No source validation - Less certainty in inferences related to less common outcomes such as ectopic pregnancies or underreported outcomes like spontaneous abortions
Lestón Vázquez et al.	2023	Spain	Information System for the Improvement of Research in Primary Care (SIDIAPI)	Primary care	No	2011 to 2020	Three- step sequential algorithm: 1) identification of potential pregnancy (LMP to identify a potential pregnancy, positive pregnancy test, gestational week, foetal death and ICD- 10 codes	327,865 pregnancies in 250,910 individuals	Comparisons with vital statistics and other published data for Spain	Pregnancy episodes identification and description of drug exposure.	- Lack of information on women whose pregnancies are followed up in the hospital or in private settings. - Algorithm not validated

Study	Year of publication	Country	Data Source	Setting	OMOP CDM used	Study period	How pregnancy identified	Pregnancies identified	Validation measures	Outcomes captured	Limitations
							indicating pregnancy or abortion; 2) length of pregnancy (using LMP, the pregnancy episode end date); 3) Outcome of pregnancy (reasons for pregnancy termination using ICD- 10)				
Zhu et al.	2023	US	Mass General Brigham (MGB) and Research Patient Data Registry (RPDR)	Healthcare claims data	No	2000 to 2014	Identification potential cases of stillbirth, spontaneous abortion, and termination, based on the presence of International Classification of Diseases, 9th Revision, Clinical Modification (ICD- 9-CM) diagnosis and procedure codes and Current Procedural Terminology (CPT) procedure codes	314 stillbirths, 4321 spontaneous abortions, and 5527 terminations	Review of random sample of medical charts. - Positive predictive value (PPV) and its 95% confidence interval (CI) for the primary definitions of stillbirth, spontaneous abortion, and termination	Non-live births; (stillbirth, spontaneous abortion, termination)	- Early spontaneous abortions that occurred before clinical recognition of the pregnancy were not captured - Included two major hospitals in Boston

Study	Year of publication	Country	Data Source	Setting	OMOP CDM used	Study period	How pregnancy identified	Pregnancies identified	Validation measures	Outcomes captured	Limitations
Smith et al.	2024	US	All of Us Research Program data	Adults voluntary link EHR since 2018 + survey data	Yes	1979 to 2022	Identification of pregnancy episodes in the most recent release of All of Us data using an algorithm previously published by Jones et al.	18,970 episodes in 14,234 individuals	Validation included survey-EHR cross-validation and external comparison with findings from Jones et al. It assessed the frequency of 25 clinician-curated pregnancy.	Determining gestational age and pregnancy outcomes (Live birth, abortion, ectopic, still birth).	- Patients often receive care from multiple healthcare systems, resulting in fragmented records that result in missing information. For instance, a patient receiving prenatal care at 1 hospital - Gestational age could not be specifically validated at outcome

GPRD: General Practice Research Database, UK: United Kingdom, LMP: last menstrual period, USA: United States of America, GA: gestational age, EHR: Electronic Health Records, N3C: National COVID Cohort Collaborative, DOD: Date of delivery, ICD-10-GM: German Modification of the International Classification of Diseases, OPS: Operations and Procedures Coding System, EBM: statutory health insurance scheme named Einheitlicher Bewertungsmaßstab, OB-GYN: obstetrician-gynaecologist, BIFAP: Database for Pharmacoepidemiological Research in Primary Care, CPRD GOLD: Clinical Practice Research Datalink General Practice Online Database.

Published pregnancy algorithm references:

1. Hardy JR, Holford TR, Hall GC, Bracken MB. Strategies for identifying pregnancies in the automated medical records of the General Practice Research Database. *Pharmacoepidemiol Drug Saf.* 2004;13(11):749-59.
2. Hornbrook MC, Whitlock EP, Berg CJ, Callaghan WM, Bachman DJ, Gold R, et al. Development of an algorithm to identify pregnancy episodes in an integrated health care delivery system. *Health Serv Res.* 2007;42(2):908-27.
3. Margulis AV, Setoguchi S, Mittleman MA, Glynn RJ, Dormuth CR, Hernandez-Diaz S. Algorithms to estimate the beginning of pregnancy in administrative databases. *Pharmacoepidemiol Drug Saf.* 2013;22(1):16-24.
4. Cea-Soriano L, Garcia Rodriguez LA, Fernandez Cantero O, Hernandez-Diaz S. Challenges of using primary care electronic medical records in the UK to study medications in pregnancy. *Pharmacoepidemiol Drug Saf.* 2013;22(9):977-85.
5. Mikolajczyk RT, Kraut AA, Garbe E. Evaluation of pregnancy outcome records in the German Pharmacoepidemiological Research Database (GePaRD). *Pharmacoepidemiol Drug Saf.* 2013;22(8):873-80.

6. Li Q, Andrade SE, Cooper WO, Davis RL, Dublin S, Hammad TA, et al. Validation of an algorithm to estimate gestational age in electronic health plan databases. *Pharmacoepidemiol Drug Saf.* 2013;22(5):524-32.
7. Matcho A, Ryan P, Fife D, Gifkins D, Knoll C, Friedman A. Inferring pregnancy episodes and outcomes within a network of observational databases. *PLOS ONE.* 2018;13(2):e0192033.
8. Blotiere PO, Weill A, Dalichampt M, Billionnet C, Mezzarobba M, Raguideau F, et al. Development of an algorithm to identify pregnancy episodes and related outcomes in health care claims databases: An application to antiepileptic drug use in 4.9 million pregnant women in France. *Pharmacoepidemiol Drug Saf.* 2018;27(7):763-70.
9. Wentzell N, Schink T, Haug U, Ulrich S, Niemeyer M, Mikolajczyk R. Optimizing an algorithm for the identification and classification of pregnancy outcomes in German claims data. *Pharmacoepidemiol Drug Saf.* 2018;27(9):1005-10.
10. Minassian C, Williams R, Meeraus WH, Smeeth L, Campbell OMR, Thomas SL. Methods to generate and validate a Pregnancy Register in the UK Clinical Practice Research Datalink primary care database. *Pharmacoepidemiol Drug Saf.* 2019;28(7):923-33.
11. MacDonald SC, Cohen JM, Panchaud A, McElrath TF, Huybrechts KF, Hernandez-Diaz S. Identifying pregnancies in insurance claims data: Methods and application to retinoid teratogenic surveillance. *Pharmacoepidemiol Drug Saf.* 2019;28(9):1211-21.
12. Sanchez Ortiz S, Llorente Garcia A, Astasio P, Huerta C, Cea Soriano L. An algorithm to identify pregnancies in BIFAP Primary Care database in Spain: Results from a cohort of 155 419 pregnancies. *Pharmacoepidemiol Drug Saf.* 2020;29(1):57-68.
13. Canelon SP, Burris HH, Levine LD, Boland MR. Development and evaluation of MADDIE: Method to Acquire Delivery Date Information from Electronic health records. *Int J Med Inform.* 2021;145:104339.
14. Moll K, Wong HL, Fingar K, Hobbi S, Sheng M, Burrell TA, et al. Validating Claims-Based Algorithms Determining Pregnancy Outcomes and Gestational Age Using a Linked Claims-Electronic Medical Record Database. *Drug Saf.* 2021;44(11):1151-64.
15. Lyu T, Liang C, Liu J, Campbell B, Hung P, Shih YW, et al. Temporal Events Detector for Pregnancy Care (TED-PC): A rule-based algorithm to infer gestational age and delivery date from electronic health records of pregnant women with and without COVID-19. *PLoS One.* 2022;17(10):e0276923.
16. Bertoia ML, Phiri K, Clifford CR, Doherty M, Zhou L, Wang LT, et al. Identification of pregnancies and infants within a US commercial healthcare administrative claims database. *Pharmacoepidemiol Drug Saf.* 2022;31(8):863-74.
17. Jones SE, Bradwell KR, Chan LE, McMurphy JA, Olson-Chen C, Tarleton J, et al. Who is pregnant? Defining real-world data-based pregnancy episodes in the National COVID Cohort Collaborative (N3C). *JAMIA Open.* 2023;6(3):o0ad067.
18. Leston Vazquez M, Vilaplana-Carnerero C, Gomez-Lumbreras A, Prat-Vallverdu O, Marsal JR, Vedia Urgell C, et al. Drug exposure during pregnancy in primary care: an algorithm and observational study from SIDIAP database, Catalunya, Spain. *BMJ Open.* 2023;13(8):e071335.
19. Zhu Y, Bateman BT, Hernandez-Diaz S, Gray KJ, Straub L, Reimers RM, et al. Validation of claims-based algorithms to identify non-live birth outcomes. *Pharmacoepidemiol Drug Saf.* 2023;32(4):468-74.
20. Louisa H. Smith WW, BK-O. Pregnancy episodes in All of Us: harnessing multi-source data for pregnancy-related research. *Journal of the American Medical Informatics Association.* 2024.

ANNEX V. Glossary

Additional definitions are available in the EMA Glossary of terms <https://www.ema.europa.eu/en/about-us/glossaries>.

Aggregated Data

Data collected and combined from multiple sources to generate summary information, typically anonymised.

Benefit-Risk Assessment

Evaluation of the positive therapeutic effects of a medicine compared to its risks (e.g. side effects).

Common Data Model (CDM)

A standardised data structure that enables data from multiple sources to be harmonised, making analysis consistent and reproducible. DARWIN EU® utilises the OMOP CDM maintained by the OHDSI community.

Complex Studies (C3)

Studies requiring the development or customisation of specific study designs, protocols, and Statistical Analysis Plans (SAPs), with extensive collection or extraction of data. Examples include etiological studies measuring the strength and determinants of an association between an exposure and the occurrence of a health outcome in a defined population considering sources of bias, potential confounding factors, and effect modifiers.

Coordination Centre (CC)

The central hub responsible for managing and overseeing the activities within DARWIN EU®. It is based at Erasmus University Medical Center in Rotterdam, the Netherlands.

Data Access

The process of obtaining permission to use specific datasets for regulatory or scientific studies.

Data Quality Framework

A set of standards and procedures to ensure accuracy, completeness, timeliness, and consistency of data used in DARWIN EU®.

Data Source

A database or repository of structured health-related data, such as electronic health records (EHRs), insurance claims, or registries.

DARWIN EU®

The European Medicines Agency's (EMA) federated network of real-world data sources designed to generate evidence to support regulatory decision-making.

EMA (European Medicines Agency)

The regulatory body responsible for the evaluation and supervision of medicinal products in the EU, overseeing DARWIN EU®.

Evidence Generation

The process of analysing real-world data to produce scientific information that can inform healthcare or regulatory decisions.

Federated Network

A data infrastructure where data remain at their original location but can be analysed in a harmonised way across multiple partners using a common model and tools.

GDPR (General Data Protection Regulation)

The EU regulation governing the protection of personal data and privacy, crucial to how DARWIN EU® handles health data.

Health Technology Assessment (HTA)

A systematic evaluation of properties and impacts of health technology, often using DARWIN EU® data to support assessments.

Metadata

Descriptive information about a data source (e.g. its content, quality, and structure), essential for identifying relevant databases in DARWIN EU® studies.

Off-the-Shelf Studies (OTS)

Studies for which a standard protocol per study/analysis type and standardised analytics may be developed and applied or adapted, typically relating to a descriptive research question. This includes studies on disease epidemiology, for example, the estimation of the prevalence or incidence of health outcomes in defined time periods and population groups, or drug utilisation studies at the population or patient level.

OHDSI (Observational Health Data Sciences and Informatics)

An open-science collaborative community that develops tools and standards (including the OMOP CDM) to enable large-scale analytics of observational health data. OHDSI provides the technical and scientific foundation for DARWIN EU®'s analytical ecosystem.

Patient-Level Data

Data related to individuals, de-identified, used for longitudinal or detailed analyses.

OMOP (Observational Medical Outcomes Partnership)

A common data model (CDM) that standardises the structure and content of observational healthcare data, enabling systematic analysis across disparate datasets. DARWIN EU® uses the OMOP CDM to ensure interoperability and consistency in real-world evidence generation.

Real-World Data (RWD)

Data relating to individual health status or healthcare delivery that is collected from routine clinical practice rather than from randomised controlled trials.

Real-World Evidence (RWE)

Clinical evidence derived from the analysis of RWD, used to inform decisions by regulators, payers, or clinicians.

Regulatory Decision-Making

The process by which authorities like EMA assess data to authorise, monitor, or modify the use of medicines in the EU.

Routine Repeated Studies (RR)

Studies that are either Off-the-Shelf or Complex studies repeated on a regular basis, following the same protocol and study code, but with updated data and/or different data partners.

Study Protocol

A detailed plan describing how a specific real-world study will be conducted, including objectives, design, data sources, and analyses.

Very Complex Studies (C4)

Studies which cannot rely only on electronic health care databases, or which would require complex methodological work, for example, due to the occurrence of events that cannot be defined by existing diagnosis codes, including events that do not yet have a diagnosis code, where it may be necessary to combine a diagnosis code with other data such as results of laboratory investigations. These studies might require the collection of data prospectively, or the inclusion of new (not previously onboarded) data sources.



Study Report

P4-C4-001

DARWIN EU[®] - PeriNet

OBJECTIVE 3

Establishing a convention for the mapping of pregnancy-related data and parent-child linkage data to the OMOP CDM

19/05/2026

Version 2.0

Authors: Maxim Moinat, Aniek Markus, Natasha Yefimenko, Berta Raventos, Katia Verhamme, Talita Duarte Salles

Public

Study title	DARWIN EU® - PeriNet Objective 3: Establishing a convention for the mapping of pregnancy-related data and parent-child linkage data to the OMOP CDM
Study report version	V2.0
Date	19/05/2026
EUPAS number	EUPAS1000000812
Research question and objectives	What is the most feasible mapping option to accurately and effectively capture pregnancy-related data, specifically pregnancy episodes, perinatal data, and parent-child linkage, in the OMOP CDM? The specific objectives were: 3a. To identify and describe different options for mapping pregnancy and parent-child data to the OMOP CDM. 3b. To evaluate the different options identified in objective 3a in terms of extensiveness, ETL (Extract, Transform, Load) implementation, and ease of analysis. 3c. To choose the overall best option for mapping based on 3b and to create a guidance for DARWIN EU® data partners on how to map pregnancy-related data to the OMOP CDM. 3d. To support DPs agreeing to map to the Perinatal Expansion Tables (PET) to facilitate participation in studies.
Countries of study	Norway, Spain, Sweden, The Netherlands
Authors	Maxim Moinat m.moinat@darwin-eu.org Natasha Yefimenko n.yefimenkonosova@darwin-eu.org Aniek Markus a.markus@darwin-eu.org Berta Raventos b.raventos@darwin-eu.org Katia Verhamme k.verhamme@darwin-eu.org Talita Duarte-Salles t.duarte@darwin-eu.org

LIST OF ABBREVIATIONS

Acronyms/term	Description
ATC	Anatomical Therapeutic Chemical classification
ATLAS	OHDSI Atlas cohort definition interface
BMI	Body Mass Index
CDM	Common Data Model
DP	Data Partner
EHR	Electronic Health Records
EMA	European Medicines Agency
ETL	Extract, Transform, Load
GP	General Practitioner
HI-SPEED	Health Impact - Swedish Population Evidence Enabling Data-linkage Database
IPCI	Integrated Primary Care Information Database
NLHR	Norwegian Linked Health Registry Database
PC	Primary Care
OHDSI	Observational Health Data Sciences and Informatics
OMOP	Observational Medical Outcomes Partnership
PET	Perinatal Expansion Tables
SIDIAP	Information System for Research in Primary Care Database
IDIAPJGol	Institut d'Investigació en Atenció Primària de Salut Jordi Gol
IMASIS	Institut Municipal Assistència Sanitària Information System

1. TITLE

DARWIN EU® - PeriNet Objective 3: Establishing a convention for the mapping of pregnancy-related data and parent-child linkage data to the OMOP CDM

2. DESCRIPTION OF THE STUDY TEAM

Study team role	Names	Organisation
Principal Investigator	Maxim Moinat	Erasmus MC
Data Scientist	Maarten van Kessel Aniek Markus Sofia Bazakou	Erasmus MC
Epidemiologist	Berta Raventós Katia Verhamme Talita Duarte-Salles Julieta Politi	Erasmus MC
Study Manager	Natasha Yefimenko	Erasmus MC
Data source	Names	Data Partner Organisation*
HI-SPEED	Fredrik Nyberg Huiqi Li	Gothenburg University
IPCI	Mees Mosseveld Katia Verhamme	Erasmus MC
NLHR	Hedvig Nordeng Nhung Trinh Saeed Hayati	University of Oslo
SIDIAP	Anna Palomar-Cros Laura Granés-González	IDIAPJGol

*Data partners do not have an investigator role. Data partners execute code at their data source, review, and approve their results.

3. ABSTRACT

Title

DARWIN EU® - PeriNet Objective 3: Establishing a convention for the mapping of pregnancy-related data and parent-child linkage data to the OMOP CDM.

Rationale and background

There is currently no widely recognised convention for representing pregnancy episodes, pregnancy-related variables, and parent-child linkage in the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM). This study objective aimed to establish a harmonised convention for the mapping of pregnancy-related data to the OMOP CDM, while also supporting data partners (DPs) that map to the Perinatal Expansion Tables (PET) for participation in ongoing studies. This is important to facilitate reproductive and perinatal pharmacoepidemiologic research across the DARWIN EU® network.

Research question

What is the most feasible mapping option to accurately and effectively capture pregnancy-related data, specifically pregnancy episodes, perinatal data, and parent-child linkage, in the OMOP CDM?

Objectives

This study objective aimed to establish a harmonised convention for the mapping of pregnancy-related data, including pregnancy episodes, perinatal variables, and parent-child linkage data, to the OMOP CDM to enable reproductive and perinatal pharmacoepidemiologic studies.

- 3a. To identify and describe different options for mapping pregnancy and parent-child data to the OMOP CDM.
- 3b. To evaluate the different options identified in objective 3a in terms of extensiveness, ETL (Extract, Transform, Load) implementation, and ease of analysis.
- 3c. To choose the overall most suitable option for mapping based on 3b and to create guidance for DARWIN EU® data partners on how to map pregnancy-related data to the OMOP CDM.
- 3d. To support DPs agreeing to map to the PET to facilitate participation in studies.

Methods

This study objective identified and evaluated multiple mapping options for representing pregnancy episodes, pregnancy-related variables, and parent-child linkage in the OMOP CDM in the context of DARWIN EU®. Mapping options were assessed based on extensiveness, ease of ETL implementation, and ease of analysis. The evaluation was performed by two reviewers: Reviewer 1 reflected the Data Partner perspective and Reviewer 2 reflected the OMOP CDM expert perspective. In addition, a survey was conducted among four participating DPs to assess the expected ease of ETL implementation of each mapping option. Based on the evaluation, recommendations were developed for future pregnancy and perinatal research in DARWIN EU®. In parallel, practical support was provided for DPs implementing the PET, including mapping guidance and dedicated data quality checks to assess conformance of PET implementations.

Data sources

1. Norway: Norwegian Linked Health Registry data (NLHR)
2. Spain: The Information System for Research in Primary Care (SIDIAP)
3. Sweden: Health Impact - Swedish Population Evidence Enabling Data-linkage (HI-SPEED)
4. The Netherlands: Integrated Primary Care Information (IPCI).

Results

In this DARWIN EU® PeriNet study objective, we described and evaluated mapping options across three areas of pregnancy-related information: pregnancy episodes, pregnancy-related variables, and parent-child linkage. Four mapping options for pregnancy episodes, two for pregnancy-related variables, and two mapping options for parent-child linkage were identified and assessed based on extensiveness, ease of ETL implementation, and ease of analysis. Two reviewers scored the mapping options based on extensiveness, ease of analysis, and ease of ETL implementation. Four DPs scored the mapping options on expected ease of ETL implementation. The evaluation supported the following long-term recommendation for mapping pregnancy-related data:

- Pregnancy episodes: Episode table
- Pregnancy-related variables: standard OMOP event tables
- Parent-child linkage: FACT_RELATIONSHIP table

At the same time, the PET was recognised as a valuable intermediate analytical dataset that remains useful in the short term for continuity in ongoing studies. The findings therefore support a two-tiered approach: continued short-term support for the PET, combined with a gradual long-term transition towards a standardised OMOP-based solution.

Discussion

The PET should be recognised as an important step in the development of pregnancy and perinatal research in OMOP, as it provided a pragmatic analytical structure at a time when a standard Episode-based solution was not yet available. However, where the same information can be represented using standard OMOP tables, dedicated expansion tables may introduce redundancy, increase ETL burden, and reduce interoperability with standard tooling. The availability of the EPISODE and EPISODE_EVENT tables in OMOP CDM v5.4 provides a strong standard alternative for representing pregnancy episodes and linking them to related events. Together with standard OMOP event tables for pregnancy-related variables, this supports a more scalable and interoperable long-term solution. To support both the short-term and long-term approaches, implementation guidance and data quality tools were developed for PET adoption and for transition towards the recommended OMOP-based convention.

4. AMENDMENTS AND UPDATES

None.

5. MILESTONES

Study milestones and deliverables	Planned dates*	Actual dates*
Final Study Protocol	December 2025	January 2026
Creation of PET mapping guideline	January 2026	January 2026
Draft Study Report	March 2026	April 2026
Final Study Report	March 2026	To be confirmed by EMA

*Planned dates are dependent on obtaining approvals from the internal review boards of the data sources.

6. RATIONALE AND BACKGROUND

The Observational Medical Outcomes Partnership Common Data Model (OMOP CDM) is a widely adopted standard for harmonising and analysing real-world data from diverse healthcare settings, including primary care, hospital care, registries, and administrative claims. The OMOP CDM continues to evolve in response to research needs identified by the user community. A recent example is the oncology domain, which led to the introduction of the EPISODE and EPISODE_EVENT tables in OMOP CDM v5.4, accompanied by structural conventions developed by the Themis workgroup and expansion of oncology-related vocabulary mappings [1,2]. These developments provide a standard mechanism to represent clinically meaningful longitudinal episodes and to link them to relevant clinical events recorded elsewhere in the OMOP CDM.

When considering changes to the OMOP CDM or new mapping conventions, it is useful to distinguish between two steps. The first is to define the analytical information required to address the research questions of interest. The second is to determine whether these analytical requirements can be met using existing OMOP CDM tables and fields, or whether additional structural elements or conventions are needed.

In this DARWIN EU® PeriNet study objective, we evaluated how reproductive and perinatal data can best be represented in the OMOP CDM. Pregnancy and perinatal research introduce specific modelling challenges, including the representation of pregnancy episodes, the linkage of related events to a pregnancy, the distinction between multiple births, and the support of parent-child linkage. In addition, differences in clinical practice and recording across countries and data sources result in heterogeneous data with varying levels of granularity.

The first step was already undertaken by members of the Observational Health Data Science and Informatics (OHDSI) community through the development of the Perinatal Expansion Tables (PET) [3,4]. The PET were developed collaboratively, with input from perinatal domain experts, epidemiologists, data analysts, and other stakeholders, and defines an analytical structure intended to capture the variables considered most relevant for pregnancy and perinatal research. It consists of a pregnancy table containing information per pregnancy episode and an infant table containing information on the newborn infant. These tables are described in [Annex II](#).

The objective 3 from the DARWIN EU® PeriNet study focused on the second step. Specifically, we assessed whether the analytical requirements reflected in the PET require structural extensions to the OMOP CDM, or whether they could be represented using existing standard OMOP tables and conventions. This includes evaluation of the EPISODE and EPISODE_EVENT tables as a potential standard mechanism for representing pregnancy episodes, despite their original development for oncology use cases. In addition, we assess how parent-child linkage can best be represented. Although the OMOP CDM already includes the

FACT_RELATIONSHIP table to capture relationships between persons or other records [5], its use for parent-child linkage has not yet been widely adopted, largely because supporting conventions and practical use cases have so far been limited.

7. RESEARCH QUESTION AND OBJECTIVES

Research question

What is the most feasible mapping option to accurately and effectively capture pregnancy-related data, specifically pregnancy episodes, perinatal data, and parent-child linkage, in the OMOP CDM?

Objectives

This study aimed to establish a harmonised convention for the mapping of pregnancy-related data, including pregnancy episodes, perinatal variables, and parent-child linkage data, to the OMOP CDM to enable reproductive and perinatal pharmacoepidemiologic studies.

The specific objectives were:

- 3a. To identify and describe different options for mapping pregnancy and parent-child data to the OMOP CDM.
- 3b. To evaluate the different options identified in objective 3a in terms of extensiveness, ETL (Extract, Transform, Load) implementation, and ease of analysis.
- 3c. To choose the overall most suitable option for mapping based on 3b and to create a guidance for DARWIN EU® data partners on how to map pregnancy-related data to the OMOP CDM.

In addition, the following objective was added, since the PET tables are used in PeriNet Objective 4, described in another report:

- 3d. To support DPs agreeing to map to the PET to facilitate participation in studies.

8. RESEARCH METHOD

8.1. Data partners

This study was conducted using routinely collected data from primary and secondary care data sources in the DARWIN EU® Data Network of data partners from four European countries ([Table 1](#)).

1. Norway: Norwegian Linked Health Registry data (NLHR)
2. Spain: The Information System for Research in Primary Care (SIDIAP)
3. Sweden: Health Impact - Swedish Population Evidence Enabling Data-linkage (HI-SPEED)
4. The Netherlands Integrated Primary Care Information (IPCI).

A detailed description of the included data sources is included in [Annex I](#).

Table 1. Description of the selected data sources.

Country	Name of Data Source	Justification for Inclusion	Health Care setting	Type of Data	Number of active subjects*	Data lock for the last update	Involvement in the objective
Norway	NLHR	PET implemented from pregnancy register, and knowledge on pregnancy data	Primary and secondary care	Registry	6,945,300	2024-11-28	Evaluation of mapping options
Spain	SIDIAP	PET implemented from EHR, and knowledge on pregnancy data	Primary care	EHR	5,951,200	2023-08-30	Evaluation of mapping options
Sweden	HI-SPEED	PET implemented, and knowledge on pregnancy data	Primary and secondary care	Registry	10,563,700	2025-01-29	Evaluation of mapping options
The Netherlands	IPCI	Pregnancy data directly from EHR	Primary care	EHR	1,332,500	2025-04-16	Evaluation of mapping options

EHR = Electronic Health Records, PC = Primary Care, PET= Perinatal Extension Tables, IPCI = Integrated Primary Care Information, NLHR = The Norwegian Linked Health Registry, SIDIAP = The Information System for Development of Research in Primary Care , HI-SPEED= Health Impact - Swedish Population Evidence Enabling Data-linkage.

*Number of individuals with an active observation period in 6 months before data lock.

8.1.1. Data Selection

The selected data sources were invited to participate based on the availability of data on pregnancy start and end dates, as well as mother-child linkage, along with information available in the DARWIN EU® Portal[6] (for data already mapped to the OMOP CDM) and PET implementation. At the time of the protocol approval for this study, three DARWIN EU® Data Partners (DPs) (HI-SPEED, NLHR, SIDIAP) had implemented the PET. IPCI has no mother-child linkage data and no PET implementation. The data sources cover different clinical settings and regions of Europe.

8.2. Description of mapping options

For this objective we identified mapping options through interaction with OMOP CDM experts and documentation of OMOP CDM v5.4.

8.3. Evaluation of mapping options

A proper assessment of changes to the OMOP CDM is needed because:

- 1) Unnecessary model complexity. Every new table increases the conceptual and technical complexity of the CDM. This causes the model to be difficult to learn, implement, govern, document, and maintain.
- 2) Risk of duplicating information. If the same perinatal concepts can already be stored using existing clinical tables, then new tables may create redundant representations of the same clinical facts. Variables are better derived analytically from existing events and dates than stored as separate persistent records. Storing derivable values can create discrepancies when source data are updated or derivation rules change.
- 3) Ambiguity about the source of truth. Once the same information can exist in both core tables and in expansion tables, it becomes unclear which representation should be used for analysis, quality control, and phenotype development.
- 4) Higher Extraction, Transform, and Load (ETL) burden. New tables require additional extraction, transformation, loading, testing, and maintenance work.
- 5) Tooling incompatibility. New expansion tables are not supported in ATLAS[7], *DataQualityDashboard*[8], cohort definitions, or standardized analytics without custom modifications.

For each mapping option we therefore described the following:

- Which information is stored? This covers a description of the field and how these can be used to store pregnancy related variables.
- At what phase is the mapping option implemented, at 'ETL time' or in the 'analysis phase'? From a pure modelling perspective, implementation at ETL time is preferred, as this only requires a one-time implementation instead of the need to repeat the mapping for each analysis. However, in some cases, variables are very study-specific and need to be defined at the analysis phase.
- Does the mapping option work on standard OMOP CDM tables, or does it require extensions?
- How to perform a certain task with the given options, specifically:
 - Characterisation of pregnancy episodes and outcomes.
 - Characterisation of pregnant patients with certain conditions (e.g. subgroup with T2DM) or exposed to certain drugs (e.g. use of antihypertensive drugs in the year prior).

In addition, we provided an example record to show how the data associated to a pregnancy episode, pregnancy-related variables, and parent-child linkage would be captured.

The different mapping options were assessed against a set of pre-defined evaluation criteria by two reviewers, a Data Partner (DP; Reviewer 1) and an OMOP Expert (Reviewer 2). All identified solutions were evaluated based on three criteria: Extensiveness, ETL implementation, and Ease of analysis, described in more detail in [Table 2](#). Findings were presented as a narrative in which advantages and disadvantages were described. In addition, a score between 0–3 was given, where higher scores indicate better perceived mapping options. Four DPs reviewed the impact on the ETL implementation using a scoring form.

Table 2. Evaluation Criteria.

Domain	Description
Extensiveness	This criterion measures if the solution enables capture of all the necessary variables and relations required for the studies. In addition, it assesses if additional variables, which may become of interest in future studies, could be added with minimal changes to the solution.
ETL Implementation	This criterion focuses on the requirements for implementing the mapping option at the data partner site. For this criterion, a questionnaire was shared among the four DPs. This questionnaire asked to score the ease of implementation of each mapping option, scoring between 0–3.
Ease of Analysis	This criterion evaluates if there is existing analytical code that supports the quality control of the ETL of pregnancy and perinatal data for the respective solution. Furthermore, it was assessed if new development of analytical code by DARWIN EU® CC is needed to use the data in studies, and if so, what would the complexity of this development be?

8.4. Creation of recommendations

Based on the output of the evaluation, the most promising solution was described, and recommendations were made on how to implement this solution into the ETL process. This resulted in an implementation guide which was discussed with the DPs.

8.5. Support PET mapping for DPs agreeing to map

To support DPs for ongoing studies using the PET, such as Objective 4 of this DARWIN EU® PeriNet project, we created a mapping guidance document for the necessary ETL steps. In addition, the *DataQualityDashboard* [8] was extended in a study-specific package to cover the PET for ETL quality assessment.

9. RESULTS

9.1. Description of mapping options (Objective 3a)

The current version of the OMOP CDM (v5.4) consists of 39 relational tables, grouped in six logical sets (**Figure 1**). The ‘Standardized clinical data’ contains the tables relevant for this objective. The PERSON table contains main demographic information, e.g., biological sex, and birth date. Each person has at least one observation period, describing the period in which health information is expected to be recorded for that person.

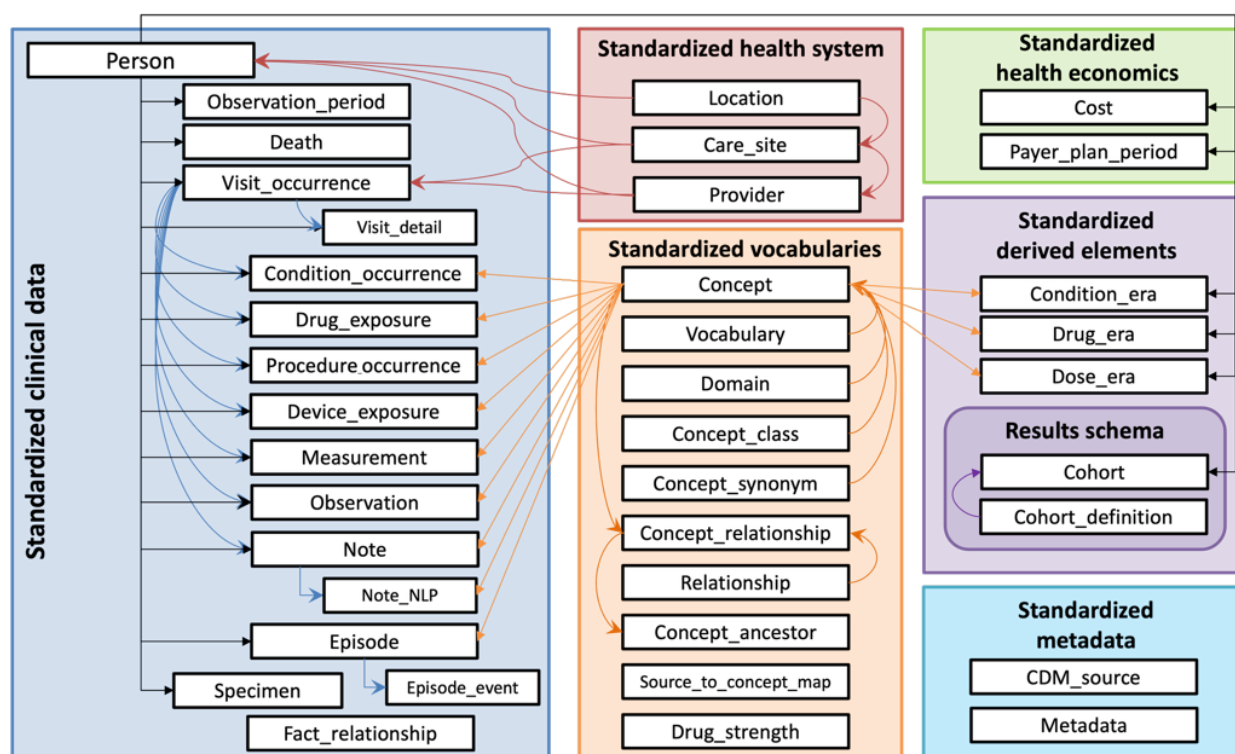


Figure 1. Overview of the OMOP CDM v5.4 with added episode table.

The most important tables are the ‘event tables’, each capturing a different type of longitudinal healthcare event. Most notably:

- Visit Occurrence – healthcare encounters, e.g. hospitalisation or General Practitioner (GP) visit
- Condition Occurrence – diagnoses and symptoms, e.g. diabetes gravidarum
- Drug Exposure – prescriptions or administrations, e.g. folic acid prescription
- Procedure Occurrence – diagnostic, surgery, or other therapy, e.g. c-section
- Measurement – vital signs and lab results, e.g. pre-pregnancy Body Mass Index (BMI)

For more information about the OMOP CDM tables we refer to:

<https://ohdsi.github.io/CommonDataModel/>

In the next sections we describe the mapping options for pregnancy episodes, as well as for pregnancy-related variables.

9.1.1. Mapping options for pregnancy episodes

We identified four options to capture pregnancy episodes. An overview of these options is provided in **Annex II**.

Option 1. PET

As proposed by Abellan et al. [2], capturing pregnancy episodes in a new table ‘PREGNANCY’ with start date, end date, and pregnancy outcome, as well as a pre-defined set of other variables.

Option 2. OMOP Condition Era

As described on <https://ohdsi.github.io/CommonDataModel/>:

“A Condition Era is defined as a span of time when the Person is assumed to have a given condition. Similar to Drug Eras, Condition Eras are chronological periods of Condition Occurrence, and every Condition Occurrence record should be part of a Condition Era. Combining individual Condition Occurrences into a single Condition Era serves two purposes:

- It allows aggregation of chronic conditions that require frequent ongoing care, instead of treating each Condition Occurrence as an independent event.
- It allows aggregation of multiple, closely timed doctor visits for the same Condition to avoid double counting the Condition Occurrences. For example, consider a Person who visits her Primary Care Physician and who is referred to a specialist. At a later time, the Person visits the specialist, who confirms the Primary Care Physician’s original diagnosis and provides the appropriate treatment to resolve the condition. These two independent doctor visits should be aggregated into one Condition Era.”

Pregnancy episodes can be treated as an ‘era’ of a condition, having a defined start and end date when the ‘condition’ is active.

Option 3. OMOP Episode/Episode Event

The EPISODE and EPISODE_EVENT tables [1,3] have been added in OMOP CDM v5.4 to capture detailed oncology information, e.g., disease episodes and treatment regimens. The EPISODE table aggregates lower-level clinical events (VISIT_OCCURRENCE, DRUG_EXPOSURE, PROCEDURE_OCCURRENCE, DEVICE_EXPOSURE) into a higher-level abstraction representing clinically and analytically relevant disease phases, outcomes, and treatments. The EPISODE_EVENT table connects qualifying clinical events (VISIT_OCCURRENCE, DRUG_EXPOSURE, PROCEDURE_OCCURRENCE, DEVICE_EXPOSURE) to the appropriate EPISODE entry. For example, cancers including their development over time, their treatment, and final resolution.

The EPISODE and EPISODE_EVENT Tables can be re-used for pregnancy episodes, where the episode table captures the period of being pregnant and the episode event link to healthcare events related to the pregnancy and birth.

For a full description of these tables see [Annex III](#).

Option 4. OMOP Cohort

It is an option to define pregnancy episodes as a phenotype or cohort definition based on inclusion/exclusion criteria at analysis time. A cohort is defined as a group of persons who satisfy a specific set of criteria over a given period. Building cohorts is a crucial part of any study performed on the OMOP CDM. A cohort can have a start and an end date which allow them to cover a period of time such as a pregnancy.

9.1.2. Mapping options for pregnancy-related variables

Option 1. PET

The PET contains two tables, ‘Pregnancy’ and ‘Infant’, containing 21 perinatal variables in total, such as gestational age (integer), pre-pregnancy BMI (numeric), and marital status. See [Annex II](#).

Option 2. OMOP event tables

The OMOP CDM consists of generic event tables, e.g., Condition Occurrence, Drug Exposure, Measurement, which can hold records of longitudinal healthcare events by referring to the OMOP Standardized

Vocabularies [10]. Examples of events are diagnosis of diabetes gravidarum, measurement of blood pressure or prescription of folic acid. These events can be linked directly to the pregnancy in the Episode table using the Episode Event table. See also Pregnancy Episode Option 3.

9.1.3. Mapping options for parent-child linkage

Option 1. PET

The PET contains two tables 'PREGNANCY' and 'INFANT' ([Annex II](#)), where the pregnant mother and infant are linked through the same 'pregnancy ID'. This is done by assigning each pregnancy a unique ID and then connecting both the mother and the infant(s) to that same pregnancy episode (see [Table 3](#)). Instead of directly linking two-person records, the PET links the mother and the infant through the shared pregnancy record, which also stores clinically important details like gestational age and delivery outcomes. This does not allow linking father to the pregnancy.

Table 3. Fields of the PET table relevant for linkage between mother and child.

Table	Field	Description	Example
pregnancy	pregnancy_id	Primary key of the pregnancy table	112233
pregnancy	person_id	Foreign key to person table	12345
infant	infant_id	Primary key of the infant table	998877
infant	pregnancy_id	Foreign key to pregnancy table	112233
infant	person_id	Foreign key to person table	98765

Option 2. FACT_RELATIONSHIP Table

The FACT_RELATIONSHIP table [4] can hold any relationship between two records within the OMOP CDM, both within the same table and between tables. It allows for connecting the parents with the child as long as both exist in the PERSON table (see example in [Table 4](#)). This means both mother-child and father-child can be represented in the same way and can be used to also include e.g. sibling relationships. The relations are bi-directional (i.e. separate relation for 'mother/father of' and 'child of') to enable easy association from both child to parent and parent to child.

Table 4. Fields of the OMOP FACT_RELATIONSHIP table with an example of a bi-directional linkage of mother and child.

Field	Description	Example Row 1	Row 2.
domain_concept_id_1	Specify domain that fact_id_1 refers to. For mother-child linkage this will always be the PERSON table	1147314 - person	1147314 - person
fact_id_1	Foreign key to event table	1	2
domain_concept_id_2	See above	1147314 - person	1147314 - person
fact_id_2	Foreign key to event table	2	1
relationship_concept_id	Child or Mother	4285883 - Child	4248584 - Mother

Row 1 links person with id '1' as the child of the person with id '2'. Row 2 links the person with id '2' as the mother of the person with id '1'.

Option 3. Custom Person Relationship

Another option is to introduce a new person relationship table in the OMOP CDM with three fields: person_id_1, person_id_2 and a relationship_concept_id. However, we decided not to evaluate this option because the same information can be placed in the FACT_RELATIONSHIP table by using fact_id_1 equal to person_id_1 and fact_id_2 to person_id_2.

9.2. Evaluation of described mapping options (Objective 3b)

9.2.1. Quantitative Assessment

Four DPs (A, B, C, and D) reviewed the expected ease of ETL implementation using a scoring form, scoring each mapping option between 0–3 (**Table 5**). Note that the Cohort Definition option for Pregnancy Episodes was not scored, as this is implemented in analysis and therefore requires no implementation in the ETL.

Two reviewers, a Data Partner (DP; Reviewer 1, different from DPs A, B, C, D) and an OMOP Expert (Reviewer 2), also assessed Extensiveness, Implementation, and Ease of Analysis, scoring each combination of criterion and mapping option between 0–3 (**Table 6**).

Table 5. Expected ease of Implementation for each option, scored by the four DPs.

Ease of implementation (0=very complex, 3=very easy)						
		Reviewer (DP)				
	Mapping option	A	B	C	D	Total
Pregnancy Episode	PET	2	2	3	0	7
	Condition Era	3	2	1	0	6
	Episode and Episode_Event	3	2	0	3	8
Pregnancy Variables	Cohort Definition	3	2	0	0	5
	PET	2	2	3	0	7
	OMOP Events	3	2	1	3	9
Parent-child linkage	PET	2	3	3	0	8
	Fact Relationship	3	2	2	3	10

Any maximum scores (3) corresponding to the best perceived mapping option (bolded in the table).

Table 6. Mapping options scored by Extensiveness, Implementation, and Ease of analysis.

			Extensiveness		Implementation		Ease of analysis		Total score
			Completeness	Maintenance	Required changes	Ease of implementation	Current support	Ease of development	
Pregnancy Episode	PET	Reviewer 1	2	3	1	2	1	3	12
		Reviewer 2	2	2	1	3	1	2	11
	Condition Era	Reviewer 1	0	3	3	3	3	0	12
		Reviewer 2	1	2	1	1	2	2	9
	EPISODE and EPISODE_EVENT	Reviewer 1	3	2	3	3	3	3	17
		Reviewer 2	3	3	2	2	2	3	15
	Cohort Definition	Reviewer 1	0	3	3	3	3	0	12
		Reviewer 2	1	2	3	3	1	1	11
Pregnancy-related variables	PET	Reviewer 1	2	3	1	2	1	3	12
		Reviewer 2	2	1	2	3	1	2	11
	OMOP Events	Reviewer 1	3	3	3	1.5	3	2	15.5
		Reviewer 2	3	3	3	2	3	3	17
Parent-child linkage	PET	Reviewer 1	3	3	1	1	1	3	12
		Reviewer 2	1	3	1	1	1	3	10
	Fact Relationship	Reviewer 1	1	3	3	3	2	1	13
		Reviewer 2	3	2	2	1	1	1	10

Rating goes from 0 (very complex) to 3 (very easy).

Any maximum scores (3) corresponding to the best perceived mapping option have been bolded. Reviewer 1: Data Partner Perspective. Reviewer 2: OMOP CDM Expert Perspective.

9.2.2. Qualitative Assessment

For an overview of all reflections of the individual reviewers, see [Annex IV](#).

Pregnancy Episodes

As shown in [Table 6](#), the mapping option 'EPISODE and EPISODE_EVENT' received the highest aggregate score (17 and 15 points, respectively) and was therefore identified as the preferred mapping option. From Reviewer 1 (R1), it received the highest scores across all criteria, except for maintenance. Similarly, Reviewer 2 (R2) consistently ranked it as the top choice. Note that for certain criteria, such as Required Changes & Complexity under ETL Implementation, multiple mapping options were among the top performers. See [Table 7](#) for the comments received by the reviewers.

Table 7. Assessment of Pregnancy Episodes solutions.

Criterion	Assessment
Extensiveness	Both reviewers state that the PET is the least flexible for adding new variables, and both gave a score of 1; R1 noted that it can only collect pre-defined variables, and R2 pointed out that adding new outcome variables requires updating the CDM and analytical packages. The Condition Era Table was seen as flexible by both reviewers but criticised for missing pregnancy-specific variables and not having the option to directly link the pregnancy outcome to the pregnancy, relying on linking by date of pregnancy. No major drawbacks were identified for EPISODE and EPISODE_EVENT Tables; it was recognised as a comprehensive option given that episodes can be linked to various event types (conditions, drugs, measurements, observations) without CDM changes. The mapping option Cohort Definitions raised concerns about what to do in case of unknown pregnancy outcomes (e.g. outcome information missing in the data) (R2) and overall unclarity on usage, being the least well-defined option (R1). Overall, EPISODE and EPISODE_EVENT Tables were referred to as the most complete option, while scoring good on maintenance.
ETL Implementation	The survey among four DPs shows that the EPISODE and EPISODE_EVENT is scored easiest overall (Total score = 8, Table 5), as it was given the maximum score of 3 by two DPs. The PET and Condition Era were overall scored with 7 and 6 points respectively. One DP scored the implementation of the PET as the easiest, and one DP scored all three evaluated implementations the same, with a score of 2. In contrast with the two reviewers, the DPs scored the Cohort definition as least easy to implement (Total score = 5, Table 5). In the full review (scores in Table 6), R2 states that PET's pregnancy-specific field names make it easy to understand and populate, but it lacks flexibility for other outcomes. It may cause double data recording for EHR sources, where data is mapped both as an OMOP CDM event (e.g. condition) and to the PET. The advantage of Condition Era is that it uses a familiar, existing table requiring minimal changes. However, using it for pregnancy episodes imposes the risk of using a standard OMOP table in an unintended way. EPISODE and EPISODE_EVENT Tables are seen as a favourable option, as it reuses existing OMOP tables, allowing gained experience with other OMOP tables to transfers to other future studies (e.g. for the Oncology domain, where these tables are also used). It is also seen as a good option that is flexible in ETL implementation by R1. Cohort Definition is viewed by R2 as easiest to map to, because it relies on pre-existing standard tables, but it still requires defining clear conventions for storing pregnancy data to be used in the cohort definition. Specifically, whether cohorts are populated as part of the study or part of the CDM ETL (R1). Overall, the Cohort Definition approach is scored as the easiest to implement; there is little work on the ETL side, as the complexity is shifted to the analysis phase.
Ease of Analysis	R2 notes that the PET is directly available for reporting (i.e. it is very easy to create a table 1 with summary characteristics of all pregnancies for a publication) but lacks current support of standard tooling, such as ATLAS and DARWIN EU® packages. R1 adds that there are no current packages to work with but thinks these may be adaptable and sees benefit in some variables being readily available. Condition Era works with ATLAS and DARWIN EU® packages, but requires linking outcomes by date, adding complexity. Standard tooling for EPISODE and EPISODE_EVENT Tables is under development but has benefits from shared use across domains (e.g. Oncology). Cohort Definitions are directly supported by ATLAS and DARWIN EU® packages but need to be run on study-per-study basis. This complicates new study feasibility (R2) and how to define cohorts that are usable across the network (R1). Overall, EPISODE and EPISODE_EVENT tables scored highest for ease of analysis.

Pregnancy-Related Variables

Both reviewers rated the mapping option ‘OMOP event tables’ highest with a clear margin (15.5 and 17 points, respectively). Hence, this option appears to be the most favourable mapping option overall. This option also aligns best with the episode table, which is the favoured pregnancy episode mapping option. See [Table 8](#) for the comments received by the reviewers.

Table 8. Assessment of Pregnancy-related Variables solutions.

Criterion	Assessment
Extensiveness	Both reviewers gave the OMOP event tables the maximum score (3 for completeness and 3 for maintenance) and highlighted the flexibility of this option, with the note that “ <i>we can add as many events as we want</i> ” (R1). In contrast, a disadvantage of the PET is that only pre-defined variables are available (R1), although these variables have been selected based on expert-opinion and considered the most useful for pregnancy research (R2).
ETL Implementation	R1 states that, for the PET, including new variables would take a lot of effort, as each new variable would be a new column, explaining the low score of 1. Both reviewers give the OMOP event tables a high score (i.e. very easy to implement) for the required changes, R2 commenting that ETL implementation will be very similar to other transformations to the OMOP event tables. The only foreseen downside is that the OMOP event tables do require at least a start date for each variable, which might not be available for data coming from dedicated registries. R2 gives the ease of implementation of OMOP event tables a low score (1.5), noting that implementation can be made easier with a predefined list of pregnancy-related events. The survey on ease of implementation for perinatal variables (or pregnancy-related variables) showed the same pattern as for pregnancy episode (Table 5). Two DPs scored OMOP Event tables as easiest, one DP favoured the PET, and one scored them both the same.
Ease of Analysis	There is agreement between reviewers that the OMOP event tables are better supported than the PET variables, giving that option a score of 2. R1 does expect that the OMOP event tables option is computationally more demanding as study code needs to search in multiple tables, especially if analysis cross more than one domain. For the PET option, R2 notes that data is already pre-processed in the context of pregnancy, making analysis easy. This makes it easy to find records meeting certain condition (R1). For the OMOP event option, this has to be done via other tables or fields, e.g. Episode Event or date (R2). There are no current specific packages to work with the PET, but it is expected to be easy to develop (R1)

Parent-Child Linkage

R1 scored the fact relationship mapping option slightly higher than the PET (13 vs 12), whereas R2 scored them equally (10 each). The overall low scores also reflect that both options have their downsides. See [Table 9](#) for the comments received by the reviewers.

Table 9. Assessment of Parent-Child Linkage solutions.

Criterion	Assessment
Extensiveness	The maintenance scores similarly for both mappings’ options, with R2 scoring the maintenance of the Fact Relationship option slightly lower (score 2). For ‘completeness’, the reviewers disagree and have opposite views. R1 notes that the PET is complete for linkage as child is linked to the pregnancy where the fact relationship is not as the link to a <i>pregnancy_id</i> is missing. R2 takes the opposite approach, noting that the PET is not complete as it is missing the possibility for father-child linkage. Overall, both mapping options are not ideal as both are missing a different core component (PET is missing father-child linkage and the Fact Relationship option missing direct linkage to pregnancy). This should be addressed in the mapping recommendation.
ETL Implementation	Both reviewers give a low score to the implementation, both maintenance and ease of implementation. The information needed is not always available in the source (R1) and needs to be linked to a pregnancy (R2). The <i>fact_relationship</i> table can be implemented independently from a pregnancy, however, as it can hold any relationship between OMOP records, it requires a more complex set up with fields indicating this is a person relationship. The <i>fact_relationship</i> table is an

Criterion	Assessment
	existing table making implementation easier than for a new table, although mapping guidance will be important (R2). The results from the ease of implementation survey (Table 5) are also split between the two options. Two DPs score the PET and two DPs score the FACT_RELATIONSHIP table as being easiest to implement.
Ease of Analysis	Neither mapping option is currently supported by the OHDSI or DARWIN EU[®] analytical tooling (R1), and both do not provide an easy lookup between mother and child. The PET requires linking through the pregnancy (R2) and the fact relationship does not have an explicit foreign key to the PERSON table requiring an additional filtering step (R2). R1 also notes that for the fact relationship option, no pregnancy_id is recorded, requiring an additional step to link perinatal variables of both mother and child. It is expected that both options can be easily integrated in the analytical code (R1,2).

9.3. Perinatal pregnancy recommendation (Objective 3c)

Based on the results of Objective 3b, we recommend a two-tiered approach consisting of a short-term and a long-term solution:

Short-term solution.

Given that the PET are already in use and have supported continuity in current DARWIN EU[®] pregnancy studies ([Section 9.4](#)), we recommend continuing to use the PET as an intermediate analytical table for ongoing studies. It has proven to be practical for this purpose and should therefore remain supported in the short term.

Long-term solution.

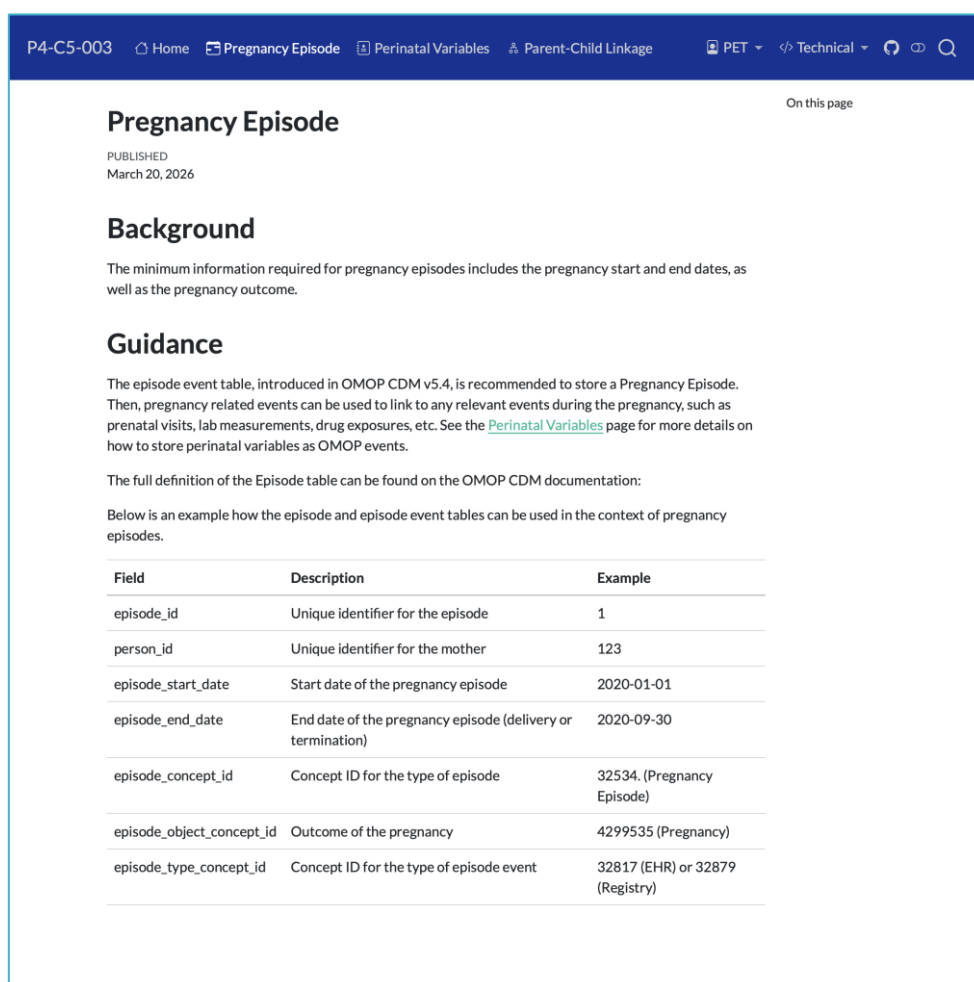
For future mapping, we recommend gradually phasing out the PET as a mapping target. Based on the evaluation, both reviewers preferred the OMOP EPISODE and EPISODE_EVENT tables for pregnancy episodes, combined with the standard OMOP event tables for perinatal variables. This approach builds on existing OMOP standards, integrates naturally with other OMOP tables, and avoids unnecessary duplication of data. To support this transition, mapping guidance should be developed describing how data from the PET, as well as from other source structures, can be transformed into the recommended OMOP-based approach. For parent-child linkage, the evaluation did not show a clear preference between the two mapping options. However, future work in DARWIN EU[®] would benefit from adopting one consistent convention. A key limitation is that the FACT_RELATIONSHIP table does not support a direct link between the child and the pregnancy episode. In previous DARWIN EU[®] studies, this was addressed by matching the child's birth date to the pregnancy end date, but this is not an ideal long-term solution. We therefore recommend establishing a mapping convention that creates an explicit link between the child and the pregnancy, for example by using the Episode Event table to link the child person record to the pregnancy episode. With this addition, the FACT_RELATIONSHIP table remains the preferred option, as it aligns best with the recommended use of the Episode table for pregnancy episodes.

We therefore recommend as long term solution the following for the three areas identified:

- Pregnancy episodes:
 - Record in the episode table, with at least an episode_start_date, episode_end_date and the episode concept for 'Pregnancy'. Pregnancy outcome should use a standard concept in episode_object_concept_id.
- Pregnancy-related variables:

- Following standard OMOP conventions, e.g. CONDITION_OCCURRENCE, DRUG_EXPOSURE, MEASUREMENT or OBSERVATION table.
- Parent-child linkage:
 - FACT_RELATIONSHIP table to capture any familial relationship between two persons.

An implementation guide for the recommended mapping options has been developed and is available at [PeriNetMapping](#) (See **Figure 2**). This includes guidance for both the short-term mapping option, describing the PET, and for the long-term solution, describing the Episode table.



Pregnancy Episode

PUBLISHED
March 20, 2026

Background

The minimum information required for pregnancy episodes includes the pregnancy start and end dates, as well as the pregnancy outcome.

Guidance

The episode event table, introduced in OMOP CDM v5.4, is recommended to store a Pregnancy Episode. Then, pregnancy related events can be used to link to any relevant events during the pregnancy, such as prenatal visits, lab measurements, drug exposures, etc. See the [Perinatal Variables](#) page for more details on how to store perinatal variables as OMOP events.

The full definition of the Episode table can be found on the OMOP CDM documentation:

Below is an example how the episode and episode event tables can be used in the context of pregnancy episodes.

Field	Description	Example
episode_id	Unique identifier for the episode	1
person_id	Unique identifier for the mother	123
episode_start_date	Start date of the pregnancy episode	2020-01-01
episode_end_date	End date of the pregnancy episode (delivery or termination)	2020-09-30
episode_concept_id	Concept ID for the type of episode	32534, (Pregnancy Episode)
episode_object_concept_id	Outcome of the pregnancy	4299535 (Pregnancy)
episode_type_concept_id	Concept ID for the type of episode event	32817 (EHR) or 32879 (Registry)

Figure 2. Screenshot of the Pregnancy Episode guidance page.

9.4. Supporting DPs in study participation (Objective 3d)

ETL Implementation

To support DPs in mapping to the PET to allow participation in current studies, we created the following materials: table definitions, an implementation guide, and example queries ([Figure 3](#)). This online resource has been shared with all DPs participating in DARWIN EU-PeriNet.

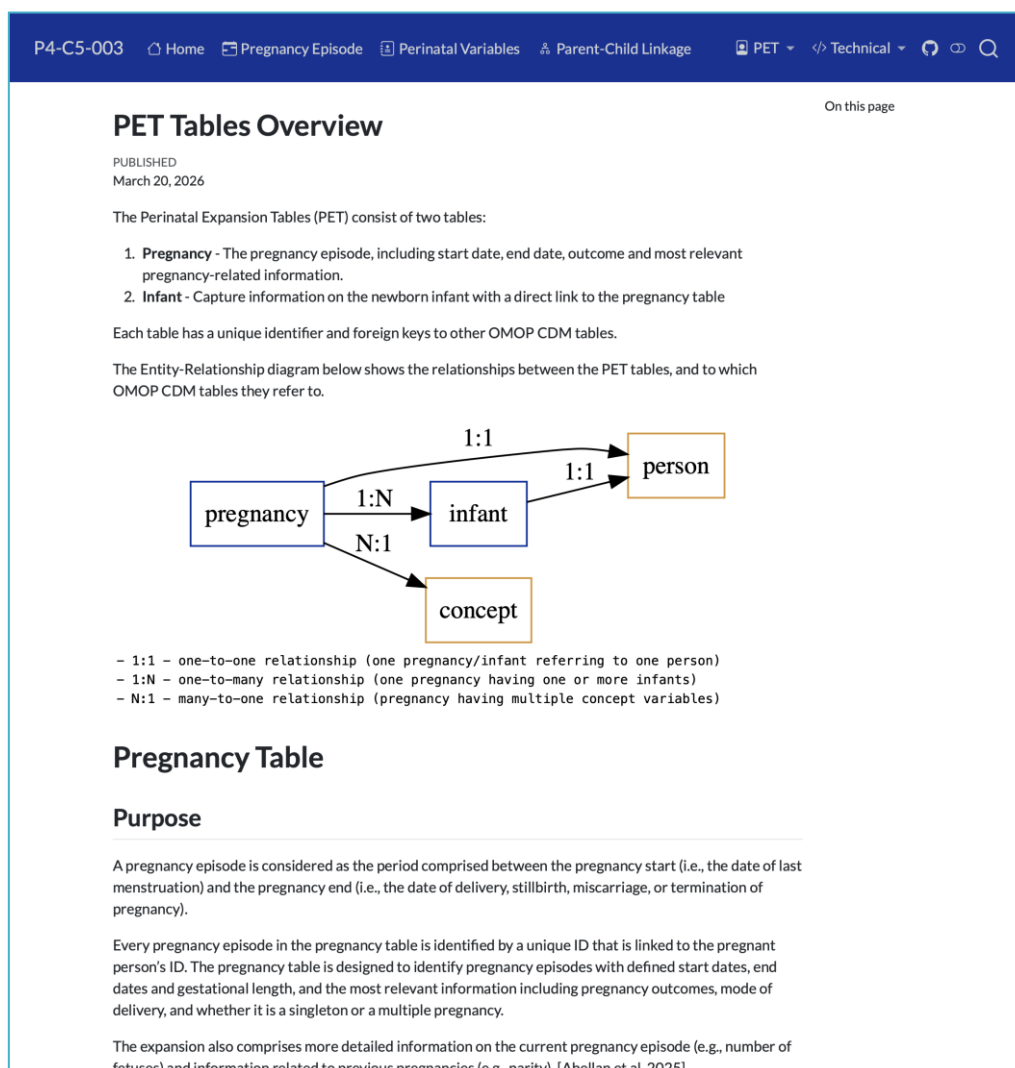


Figure 3. Screenshot of the PET Table Overview page.

In consultation with the study team, we made one adaptation to the definition of the PET, as described in the original publication. [3] In the original definition there was no reference to the person record of an infant, and some DPs used the `infant_id` for this purpose. We added the `person_id` field to the infant table to standardise the way an infant is linked, making it easier for analytics to find child demographics via the PERSON table.

ETL Quality Assessment

In order to evaluate the PET implementation, both new and existing, we developed a PET data quality assessment script. This implements a total of 147 checks on the pregnancy and infant tables ([Table 10](#)).

The following check types have been implemented:

- PET tables present and containing expected fields (cdmTable, cdmField)
- Required fields filled (isRequired)
- No end dates before start (plausibleStartBeforeEnd)
- No dates past the current date (future dates) (plausibleValueHigh)
- No dates before birth (plausibleAfterBirth)
- No dates after death (plausibleBeforeDeath)
- No pregnant males (plausibleGender)
- Standard concepts in concept_id fields (standardConceptCompleteness)
- No additional fields not in specification (noAdditionalFields)

Table 10. Number of checks implemented across PET by table and level.

	Pregnancy	Infant	Total
Table-level	3	2	5
Field-level	102	39	141
Concept-level	1	0	1
Total	106	41	147

The PET data quality was executed at the three DPs with PET implemented: HI-SPEED, NLHR, and SIDIAP. Running the PET data quality has identified multiple issues and respective DPs have been given feedback to solve these based on the implementation guide.

10. DISCUSSION

10.1. Key results

The PET should be recognised as an important and timely contribution to perinatal and pregnancy research in OMOP. At the time the PET was developed, the OMOP CDM did not yet provide a dedicated standard mechanism to represent clinically meaningful episodes such as pregnancy in a way that could be linked systematically to related events. In that context, the PET was a strong and pragmatic attempt to define an analytical structure containing the variables considered essential for pregnancy and perinatal studies. It provided a concrete way to organise pregnancy-level and infant-level information and has already demonstrated its practical value in ongoing DARWIN EU® studies, where it enabled continuity, supported study execution, and motivated the development of implementation guidance and dedicated data quality checks.

At the same time, the purpose of this objective was not only to assess what is useful analytically, but also to determine whether the same analytical goals can be achieved using standard OMOP structures without introducing unnecessary extensions. An analytical table such as the PET can be highly useful for study conduct, reporting, and harmonisation, while still not being the most appropriate long-term mapping target. As outlined in the evaluation framework, adding new tables where the information can already be represented in existing OMOP tables increases model complexity, creates additional ETL burden, and may introduce redundant representations of the same underlying clinical facts. It can also create ambiguity about the source of truth for analysis and quality control, particularly when information may exist both in standard event tables and in a dedicated expansion structure.

This is especially relevant for perinatal variables. Many of the variables needed for pregnancy research, such as diagnoses, drug exposures, procedures, measurements, and observations, are fundamentally longitudinal healthcare events and fit naturally within standard OMOP event tables. Storing such information again in pregnancy-specific fields may simplify some analyses in the short term, but it also risks duplication and reduced flexibility when new variables become relevant. In contrast, use of standard OMOP event tables allows the model to expand without structural change, benefits from established OMOP conventions, and better aligns with general-purpose tooling and future reuse across studies. The reviewers' assessment reflected this, with OMOP event tables being preferred over the PET for pregnancy-related variables.

A major development since the PET was introduced is the availability of the EPISODE and EPISODE_EVENT tables in OMOP CDM v5.4. Although these tables were initially introduced for oncology use cases, they are defined generically and provide an attractive standard mechanism for representing clinically meaningful episodes and linking them to relevant lower-level events. In the context of pregnancy research, this offers an important opportunity. Pregnancy is inherently episodic: it has a start and end, can be associated with an outcome, and is characterised by a range of related events across multiple OMOP domains. The Episode table provides a natural place to represent the pregnancy itself, while the EPISODE_EVENT table provides a structured way to connect the pregnancy to diagnoses, measurements, drug exposures, procedures, observations, and potentially additional conventions for linkage. This makes the EPISODE and EPISODE_EVENT mapping option conceptually strong, flexible, and well aligned with the broader OMOP model.

The evaluation performed in this study supports that position. Across the pregnancy episode options, the EPISODE and EPISODE_EVENT tables were consistently scored highest overall by both reviewers and were also rated favourably by the participating data partners in the ease-of-implementation survey. Reviewers described this option as complete, flexible, and capable of linking to any relevant event type without requiring changes to the CDM whenever new variables need to be captured. Potentially, the EPISODE table can be populated using the pregnancy algorithm developed in Objective 2 of the PeriNet study, which derives the start date, end date, and outcome of a pregnancy episode from data in the OMOP CDM. In addition, reuse of a standard OMOP structure has strategic value beyond this objective, because experience gained with the Episode framework can also be transferred to other domains where the same tables are used.

However, the findings also argue against an abrupt transition away from the PET. The PET is already implemented by several data partners, is being used in current DARWIN EU® studies, and has proven practical as an intermediate analytical dataset. For that reason, the two-tiered recommendation emerging from this work is important. In the short term, continued support for the PET is justified to maintain continuity for ongoing studies and to avoid disrupting implementations that are already functioning. In the long term, the preferred direction is to phase out the PET as a primary mapping target and move towards a standard OMOP-based approach using the Episode table for pregnancy episodes, standard OMOP event tables for pregnancy-related variables, and a consistent convention for parent-child linkage. This approach preserves the practical value of the PET where needed today, while guiding the network toward a more scalable, less redundant, and more interoperable solution for the future.

Overall, the PET should therefore be seen as an important transitional step in the maturation of pregnancy and perinatal modelling in OMOP. It helped define the analytical requirements of the domain and demonstrated what information needs to be captured to support meaningful research. The availability of the EPISODE and EPISODE_EVENT tables now makes it possible to address those same requirements in a way that is more consistent with standard OMOP design principles. The recommended two-tiered approach reflects this evolution: preserve the PET where it remains operationally valuable in the short term, while building toward a long-term convention that minimises redundancy, strengthens consistency, and integrates pregnancy research more fully into the standard OMOP ecosystem.

To support both elements of this two-tiered approach, we have developed practical implementation resources for current and future adoption. For the short-term use of the PET, we created mapping guidance, table definitions, example queries, and dedicated data quality scripts to support consistent implementation and assess conformance across data partners. For the long-term OMOP-based solution, we developed an implementation guide for the recommended mapping options and outlined the transition path from PET and other source structures towards the use of the EPISODE, EPISODE_EVENT, and standard OMOP event tables. Together, these materials are intended to facilitate adoption, improve implementation quality, and support a gradual transition towards a more standardised and scalable approach.

10.2. Strengths and limitations of the research methods

A strength of this study is that it addressed a concrete methodological question that is directly relevant for DARWIN EU® studies and for other data holders seeking to map pregnancy-related data to the OMOP CDM. The evaluation was based on clearly defined criteria, covering extensiveness, ease of ETL implementation, and ease of analysis, which ensured that both methodological and practical considerations were taken into account. This makes the findings relevant, not only from a modelling perspective but also from an operational perspective, for distributed research networks.

A further strength is that the assessment combined different perspectives. The review included both a Data Partner perspective and an OMOP expert perspective and was complemented by input from four data partners on expected ease of implementation. In addition, this objective produced practical outputs, including implementation guidance and data quality tools, which support immediate use in DARWIN EU® and future adoption of the recommended approach. The involvement of study team members in relevant OHDSI working groups further supports alignment with broader community developments and future tooling support.

This study also has limitations. First, the qualitative evaluation was performed by only two reviewers. Although both were selected based on their relevant expertise, the limited number of reviewers means that the findings may not fully reflect the range of views and implementation experiences across the wider DARWIN EU® network or OHDSI community.

Second, not all participating data partners had implemented all mapping options. As a result, the survey on ease of implementation was based on expected implementation effort rather than direct experience with each option. While all participating data partners had substantial experience with OMOP mapping, and most had experience with the PET, these findings should still be interpreted with this limitation in mind.

Third, the results for the ease of implementation for Pregnancy Episodes show a low inter-reviewer agreement. For instance, reviewer C scores the PET a score of 3 (very easy) and Episode/Episode Event a 0 (very complex) where reviewer D gives an opposite score: the opposite the PET a score of 0 (very complex) and Episode/Episode Event a 3 (very easy). It is likely that if a DP already has experience with e.g. implementing the PET or Episode table, they will perceive this as easy. Also given the total scores are close, (7, 6, and 8 for PET, Condition Era, and Episode respectively), we cannot conclude from these results that one option is expected to be easier to implement overall."

Fourth, the objective focused on evaluating and recommending mapping conventions rather than on large-scale validation of all options in fully implemented production settings. Some practical considerations may therefore only become clear during broader implementation in future studies. Continued experience within DARWIN EU® and the wider OHDSI community will help to further refine and support the proposed long-term approach.

11. CONCLUSION

In this report, we described and evaluated mapping options across three key areas of pregnancy-related information in the OMOP CDM: pregnancy episodes, pregnancy-related variables, and parent-child linkage. The evaluation, based on extensiveness, ease of ETL implementation, and ease of analysis, combined reviewer assessment with input from participating data partners and provided a structured basis for recommending a harmonised mapping approach for DARWIN EU®.

The findings support a clear long-term direction towards the use of standard OMOP structures wherever possible. For pregnancy episodes, the Episode table, supported by the EPISODE_EVENT table, was identified as the preferred option. For pregnancy-related variables, standard OMOP event tables were preferred, and for parent-child linkage, the FACT_RELATIONSHIP table was identified as the most suitable standard solution. Together, these options provide a more flexible, scalable, and interoperable approach for future pregnancy and perinatal research in OMOP.

At the same time, the PET remain valuable in the short term. The PET represented an important step in defining the analytical requirements for pregnancy and perinatal research and has already proven useful in ongoing DARWIN EU® studies. The overall recommendation is therefore a two-tiered approach: continue supporting the PET as an intermediate analytical dataset for current studies, while gradually transitioning towards a standardised OMOP-based solution for the longer term. Overall, objective 3 described in this report provides both a clear methodological recommendation and practical support for implementation. By combining evaluation, implementation guidance, and data quality tooling, it supports a more consistent and future-proof approach to representing pregnancy-related data within DARWIN EU® and beyond.

12. REFERENCES

1. CDM Working Group, Working group lead Blacketer C. OMOP Common Data Model, Available from: <https://ohdsi.github.io/CommonDataModel/>.
2. Belenkaya R, Gurley MJ, Golozar A, Dymshyts D, Miller RT, Williams AE, et al. Extending the OMOP Common Data Model and Standardized Vocabularies to Support Observational Cancer Research. *JCO Clin Cancer Inform*. 2021 Dec;(5):12–20.
3. Abellan A, Burn E, Trinh NTH, Burkard T, Callahan A, Fernández-Bertolín S, et al. Expanding the OMOP Common Data Model to Support Perinatal Research in Network Studies. *Pharmacoepidemiol Drug Saf*. 2025 Feb;34(2):e70106.
4. Mercade-Besora N, Raventós B, Trinh NT, Li H, Hayati S, Giuliadori A, et al. Effectiveness of COVID-19 mRNA primary and booster vaccination against infection and hospitalisation during pregnancy: a target trial emulation and meta-analysis of data from 4 European countries [Internet]. 2025 [cited 2025 Sep 9]. Available from: <http://medrxiv.org/lookup/doi/10.1101/2025.05.16.253275939>.
5. OMOP Common Data Model, FACT_RELATIONSHIP table. Available from: https://ohdsi.github.io/CommonDataModel/cdm54.html#fact_relationship.
6. DARWIN EU® Portal, Available from: <https://portal.darwin-eu.org/summary>.
7. OHDSI ATLAS. Available from: <https://github.com/ohdsi/atlas>.
8. OHDSI DataQualityDashboard. Available from: <https://github.com/ohdsi/dataqualitydashboard>.
9. OHDSI. The Book of OHDSI [Internet]. Available from: <https://ohdsi.github.io/TheBookOfOhdsi/>
10. OHDSI. The Book of OHDSI Chapter 5 Standardized Vocabularies [Internet]. Available from: <https://ohdsi.github.io/TheBookOfOhdsi/StandardizedVocabularies.html#StandardizedVocabularies>

13. ANNEXES

ANNEX I. Description of data sources

Norwegian Linked Health Registry data (NLHR)

#	Section	Description
1	Data source identification and country	NLHR (Norwegian Linked Health Registry data) Norway
2	Data partner information section	University of Oslo Faculty of Mathematics and Natural Science – Department of Pharmacy
3	Coverage and timespan	Data collection since: 2008 Extent: Nation-wide. Norway has a universal public health care system, consisting of primary and specialist health care services covering a population of approximately 5.4 million inhabitants.
4	Healthcare setting / type of data	Primary care – General Practitioner, and primary care specialists (e.g. paediatricians), and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. The following registries are included: the Medical Birth Registry of Norway (MBRN), the Norwegian Prescription Registry (NorPD), the Norwegian Patient Registry (NPR), Norway Control and Payment of Health Reimbursement (KUHR), the Norwegian Surveillance System for Communicable Diseases (MSIS), the Norwegian Immunisation Registry (SYSVAK), the National Death Registry, and the National Registry (NR).
5	Data collection process	Registries. Many population-based health registries were established in the 1960s, with use of unique personal identifiers facilitating linkage between registries. Data in these health registries are used for health analysis, health statistics, improving the quality of healthcare, research, administration, and emergency preparedness.
6	General representativeness	The NLHR data covers the full Norwegian population.
7	Data content /source coding	NPR: ICD-10 for diagnosis, ATC and some special codes for drug use, Norwegian codes for clinical procedures (surgery (NCSP), medicine (NCMP) and diagnostic imaging, image-guided intervention, and nuclear medicine (NCRP)). KUHR: ICD-10 and ICPC-2 and ICPC-2B for diagnosis/procedure. NorPD: ATC. SYSVAK and MSIS: national classifications. MBRN: custom classifications by questionnaires (incl. check box variables in Maternity health care card)
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5 and the OMOP standard vocabularies. The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. Linkage between the registries was facilitated using project-specific person IDs generated from unique personal identification assigned at birth or immigration for all legal residents in Norway.
9	Quality control (data source specific)	In-house data quality checks of rates of common conditions, drug exposures, and outcomes. We compare obtained rates with official national statistics (e.g., birth statistics, yearly rates of drug dispensing, and diagnosis by age and gender). We also review missing data and outliers and inform registry holders of any unusual patterns.
10	Linkage	The NLHR is, by definition, a linkage of datasets. Helsedata.no is one central portal to apply for 11 national health registries, including all the registries that have been mapped to the OMOP CDM.
11	Vital status	The national death registry is linked.
12	Limitations	No data source specific limitations documented. General limitations for the data type applicable.
13	Main references	Trinh NT, Houghtaling J,Bernal FL,Hayati S,Maglanoc LA,Lupattelli A,Halvorsen L,Nordeng HM "Harmonizing Norwegian registries onto OMOP common data model: Mapping challenges and opportunities for pregnancy and COVID-19 research." International journal of medical informatics (2024): 39153282

#	Section	Description
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1000000409 Website: https://www.mn.uio.no/farmasi/english/research/groups/pharma-safe/

The Information System for Research in Primary Care (SIDIAP)

#	Section	Description
1	Data source identification and country	SIDIAP (The Information System for Research in Primary Care) Catalunya, Spain
2	Data partner information section	IDIAPJGol
3	Coverage and timespan	Data collection since: 2006 Extent: Regional. SIDIAP is a database of primary care electronic health records of the population of Catalonia, North-East Spain. It contains pseudo-anonymised records of more than 8 million people, of which 6.1 million are active as of 2022, representing around 76% of the Catalan population.
4	Healthcare setting / type of data	Primary care – General Practitioner, and hospital inpatient care. SIDIAP captured data includes routine visits, socio-demographics, diagnoses, laboratory tests, drugs (prescribed and dispensed), referrals, sick leaves and lifestyle information.
5	Data collection process	Outpatient electronic health records, and Inpatient hospital electronic health records. Data is entered by primary care physicians upon healthcare contact, supplemented with hospital discharge records. The Institut Català de la Salut (Catalan Health Institute) is the data controller.
6	General representativeness	It was previously shown that the captured SIDIAP population is highly representative of the entire Catalan region in terms of geographic, age, and sex distributions.
7	Data content /source coding	SIDIAP data covers all services that occur at the Primary Care Centres, as well as support services, such as sexual and reproductive health or home end-of-life care. Drugs are coded in ATC-WHO terminology in the source data. Health outcomes are captured in ICD-10CM codes. The SIDIAP contains all laboratory tests and results performed in primary health centres. Demographics, geographical, as well as socio-economic factors are recorded for each patient.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5 and the OMOP standard vocabularies. The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. A patient has a unique id, also upon changing practice or re-registration.
9	Quality control (data source specific)	Internal and external validation processes are carried out to determine the data quality of the SIDIAP information at each data update. These include stratifying the data by geographical regions and year in order to identify differences in data collection that need to be harmonized (e.g. recording of specific information under different codes). The measurement units of variables measuring one characteristic are also homogenized (e.g. transformation of the data from every laboratory that measures haemoglobin to grams per decilitre). Visual inspection of all data included in the database by week is also conducted, allowing one to see temporal patterns in the registry of a certain variable. With this information, the SIDIAP team can issue recommendations to researchers about the most common variable(s) where certain information is recorded (e.g., there are several variables with information concerning the women's menopausal status and with these visual inspection tools the SIDIAP team can inform the researchers about which related variables have the largest number of records and could be more helpful to capture menopause). Data availability (longitudinally and reliability), plausibility (range checks and unusual values), and consistency are inspected through visualisation tools. In addition, before accessing the data for a requested project, research teams have access to a quality-control report. This document contains counts, years, percentiles, maximums and minimums, incidences, and prevalence of the data requested for the project, allowing detection

#	Section	Description
		of inconsistencies in the data extraction prior to data delivery. External validation processes of the SIDIAP database mainly include assessing the data recorded in SIDIAP through linkage to external gold standard data sources, by analysing free text, or by sending questionnaires to health professionals.
10	Linkage	SIDIAP is linked to a hospital discharge database, pharmacy dispensation, and primary care laboratories. It can also be linked to other registries in Catalonia on a project by project basis.
11	Vital status	Mortality is fully captured in SIDIAP. The cause of death is not available, but can be linked to the Spanish death registry on a project by project basis.
12	Limitations	The SIDIAP data is not representative of individuals not using public primary care, and conditions that are usually followed by specialist care might not be properly captured. In addition, there is limited information on lifestyle variables. Patients are followed until Death or when transferring to another primary health care centre that does not contribute to SIDIAP.
13	Main references	Recalde M, Rodríguez C, Burn E, Far M, García D, Carrere-Molina J, Benítez M, Moleras A, Pistillo A, Bolibar B, Aragón M, Duarte-Salles T "Data Resource Profile: The Information System for Research in Primary Care (SIDIAP)." International journal of epidemiology (2022): 35415748
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/50190 Website: https://www.sidiap.org/index.php/en

Health Impact - Swedish Population Evidence Enabling Data-linkage (HI-SPEED)

#	Section	Description
1	Data source identification and country	HI-SPEED (Health Impact - Swedish Population Evidence Enabling Data-linkage) Sweden
2	Data partner information section	SMPA-GU, Läkemedelsverket, Box 26, 751 03 Uppsala, Sweden - Box 469, 405 30 Gothenburg, Sweden Pharmacoepidemiology and Analysis Department (FeA) - School of Public Health and Community Medicine, Institute of Medicine
3	Coverage and timespan	Data collection since: 2015 Extent: Nation-wide. The catchment area includes the whole of Sweden, covering the full population of approximately 11.7 million.
4	Healthcare setting / type of data	Primary care – General Practitioner, and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. Primary care (GPs) is available only for the 2 largest regions (~40% of national population) The following data elements are collected: Socio-demographics, dispensed drug prescriptions, cause of death, diagnoses and procedures from secondary (specialist) care and inpatient visits or clinical events, as well as from primary care visits (40%pop only).
5	Data collection process	Registries. The data is acquired from the relevant Swedish national and regional registries, only once all legislative, GDPR and ethical approvals have been granted. Therefore only relevant data is passed on, which will then be entered and processed by the study team. The data are updated several times annually.
6	General representativeness	The coverage includes all patients of all sociodemographic characteristics. Therefore it should mirror the source population to a very good extent.
7	Data content /source coding	Medicines are coded with ATC and NPLID (National Product ID), ICD10-SE is used for diagnoses and the Swedish procedure coding system (KVA) is used for clinical procedures.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5 and the OMOP standard vocabularies. The format, structural and semantic conformance has been verified upon onboarding into the

#	Section	Description
		DARWIN EU® data network. Patients have a uniquely id across datasets.
9	Quality control (data source specific)	The source data are obtained from the relevant Swedish National and Regional Registers. The registers perform some regular quality controls on their data. After receiving the data, we perform additional checks and cleaning. We also run regular quality checks on the data we manage.
10	Linkage	Data on specialist care is acquired from the National Patient Register, mortality information is provided by the Cause-Of-Death Registry. Drug data is provided by the Patient Drug Register. And similarly for other registers. Data are linked very accurately using the national personal ID number, and pseudonymized before delivery to HI-SPEED. All data are updated 2-4 times per year.
11	Vital status	Data on death and underlying + contributing causes-of-death are extracted from the Cause-of-Death registry (i.e. based on death certificates).
12	Limitations	General limitations for the data type applicable. This is a research project where all studies require ethics approval. Data collection since: 2015 for most data, except prescribed drug register (from 2018), and some COVID-related data (tests, vaccination) from 2020 Primary care is only available for a subset.
13	Main references	Nyberg F, Franzén S, Lindh M, Vanfleteren L, Hammar N, Wettermark B, Sundström J, Santosa A, Björck S, Gisslén M "Swedish Covid-19 Investigation for Future Insights - A Population Epidemiology Approach Using Register Linkage (SCIFI-PEARL)." Clinical epidemiology (2021): 34354377
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/node/4463/ Website: https://www.gu.se/en/research/scifi-pearl

Integrated Primary Care Information (IPCI)

#	Section	Description
1	Data source identification and country	IPCI (Integrated Primary Care Information) Netherlands
2	Data partner information section	Erasmus University Medical Center Department of Medical Informatics
3	Coverage and timespan	Data collection since: 2006 Extent: Nation-wide. IPCI is a Dutch database that contains patient records from 2006 onwards. However, it mainly covers the central part of the country, including the most densely populated area (the 'Randstad') and non-urban areas. IPCI contains information on all patients registered with GPs responsible for non-emergency care and referrals. A patient is registered at birth or at first encounter with the GP.
4	Healthcare setting / type of data	Primary care – General Practitioner. Data is collected from primary care EHR. This includes demographic information, complaints and symptoms, diagnoses, laboratory test results, lifestyle factors (in limited amount), and correspondence with secondary care, such as referral and discharge letters.
5	Data collection process	Outpatient electronic health records. Data is entered into the EHR system by the GPs, during or after the visit. The patient dossiers are collected by Erasmus MC data managers and combined in one harmonized database. Several checks are done on this database to ensure correct data processing. Persons can have dossiers at multiple GPs.
6	General representativeness	More than 99% of the Dutch population has health insurance, and almost all citizens are registered with a general practitioner. Over 12 months, around 78% of the population has at least one contact with their GP. IPCI included around 350 GP practices out of around 5000 in the

#	Section	Description
		country (~ 7%). The demographic composition of the IPCI population mirrors that of the general Dutch population in terms of age and sex.
7	Data content /source coding	Dutch GPs use mainly Dutch standard codes, like ICPC-1 and Diagnostische Bepalingen maintained by NHG. And for therapy the G-Standard is used, maintained by ZIndex.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5 and the OMOP standard vocabularies. The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. Patients can be registered under different IDs, but since a patient can only be registered at one GP at a time, the observations periods will not overlap.
9	Quality control (data source specific)	Prior to each data release, extensive quality control steps are performed, e.g., comparison of patient characteristics between practices, and checks to identify abnormal temporal data patterns in practices. For each practice, around 200 quality indicators are obtained. Of these indicators, a quarter refer to population characteristics, e.g. number of birth and mortalities relative to practice size, temporal consistency. The other indicators are based on medical data, e.g. distribution of measurement values, frequencies of diagnoses and procedures relative to age, completeness of data. The indicators are combined in a couple of quality scores for each practice. For these scores, cut-off values for acceptable quality have been defined. Practices with a score below a cut-off are excluded for research. This approach has shown to be very important, for example to check if data from practices that just joined the database are at an acceptable level of quality. The details of the approach, like the cut-off values for acceptance, are based on years of experience. In addition, trends are compared with the previous database release. Extensive quality control steps are performed before each data release. These include comparing patient characteristics between practices and checks to identify abnormal temporal data patterns in practices. Additional checks include over 200 indicators related to population characteristics (e.g., reliability of birth and mortality rates) and medical data (e.g., availability of durations of prescriptions and completeness of laboratory results). Records of low quality are excluded from the database.
10	Linkage	Linkage requires additional approval steps and needs to be assessed on a case-by-case basis. IPCI is not routinely linked with other databases.
11	Vital status	Vital status (death date and cause) is collected based on GP records.
12	Limitations	The main limitation comes with the fact that IPCI is limited to GP records, and although it contains information on referrals and discharge letters, it may not fully capture specific hospital information. IPCI does not include coded/detailed data about medications/procedures/test results from the hospital or other care-providers.
13	Main references	de Ridder MAJ, de Wilde M, de Ben C, Leyba AR, Mosseveld BMT, Verhamme KMC, van der Lei J, Rijnbeek PR "Data Resource Profile: The Integrated Primary Care Information (IPCI) database, The Netherlands." International journal of epidemiology (2022): 35182144
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/42618 Website: http://www.ipci.nl

ANNEX II. Overview of mapping options

Table S1. Pregnancy Episodes.

Mapping Option	Information Stored	Implementation Phase	Standard OMOP or Extensions?	Simple Characterisation Study: Feasibility & Execution
Perinatal Expansion Tables (PET)	pregnancy_id, person_id, pregnancy_start_date, pregnancy_end_date, pregnancy_outcome*, pregnancy_single*, gestational_length_in_day*, pregnancy_mode_delivery*	ETL time	Requires extensions beyond standard OMOP CDM tables	Enables comprehensive characterisation of pregnancy episodes and outcomes. However, PET is not supported in ATLAS or standard DARWIN EU® analytical packages, limiting direct use for cohort definition and analysis.
Condition Era Table	condition_era_id, person_id, condition_concept_id (pregnancy), condition_era_start_date, condition_era_end_date, condition_occurrence_count, + record of pregnancy_outcome (condition_occurrence)	ETL time	Uses standard OMOP CDM tables	Supported within ATLAS and standard DARWIN EU® packages, allowing direct integration in cohort definition and analytical packages to perform characterization of pregnant patients with certain comorbidities or drug exposures.
Episode & Episode Event Tables	episode_id, person_id, episode_concept_id (pregnancy episode), episode_start_date, episode_end_date, episode_parent_id, episode_number, episode_object_concept_id (pregnancy_outcome & pregnancy_single, e.g. single livebirth, twin live birth, twin birth, one stillbirth one livebirth). + gestational_length_in_day (integer, computed from end-start) + any other variable (e.g. pregnancy_mode_delivery) could be linked via: episode_id	ETL time	Uses (new) standard OMOP CDM tables	Under active development in ATLAS/standard DARWIN EU® packages (also reusable in other domains, e.g. Oncology). Would allow direct integration in cohort definition and analytical packages to perform characterisation of pregnant patients with certain comorbidities or drug exposures.

Mapping Option	Information Stored	Implementation Phase	Standard OMOP or Extensions?	Simple Characterization Study: Feasibility & Execution
	event_id episode_event_field_concept_id			
Cohort Definitions	Cohorts defined per pregnancy outcome (livebirth, stillbirth, miscarriage, elective abortion) with cohort_definition_id, subject_id, cohort_start_date, cohort_end_date	Analysis phase	Uses standard OMOP CDM tables	Supported within ATLAS and standard DARWIN EU® packages, allowing direct integration in cohort definition and analytical packages to perform characterization of pregnant patients with certain comorbidities or drug exposures.

* pregnancy_outcomes: livebirth, miscarriage, stillbirth, termination, discordant; pregnancy_single: Yes, No; gestational_length_in_day: integer, computed from end-start or taken from source; pregnancy_mode_delivery: vaginal, C-section

Table S2. Pregnancy-Related Variables.

Mapping Option	Information Stored	Implementation Phase	Standard OMOP or Extensions?	Use Case: Study Feasibility & Execution
Perinatal Expansion Tables (PET)	The PET contains the following optional fields for pregnancy-related variables: pregnancy_marital_status, pregnancy_number_fetuses, pregnancy_number_fetuses, pregnancy_number_liveborn, pregnancy_number_liveborn prev_pregnancy_parity, prev_pregnancy_gravidity, prev_livebirth_number, pregnancy_number_fetuses, prev_top_number, pre_pregnancy_bmi, pregnancy_folic,	ETL Time	Requires extensions beyond standard OMOP CDM tables	

Mapping Option	Information Stored	Implementation Phase	Standard OMOP or Extensions?	Use Case: Study Feasibility & Execution
	prev_still_misc_number, pregnancy_art			
OMOP Event	Any event with a standard OMOP concept. e.g. condition occurrence, measurement, observation, procedure	ETL Time	Uses standard OMOP CDM tables	

* prev: previous, top: terminations of pregnancy, bmi: body mass index, folic: folic acid, still: stillbirth, misc: miscarriage, art: assisted reproductive technology

Parent-Child Linkage

Mapping Option	Information Stored	Implementation Phase	Standard OMOP or Extensions?	Use Case: Study Feasibility & Execution
Perinatal Expansion Tables (PET)	Linkage of infant to the mother via the pregnancy table.	ETL time	Requires extensions beyond standard OMOP CDM tables	
FACT_RELATIONSHIP table	Any familial relationship. In this context, mother to child and father to child. No linkage of infant to pregnancy.	ETL time	Uses standard OMOP CDM tables	

ANNEX III. EPISODE and EPISODE_EVENT Tables

The EPISODE table aggregates lower-level clinical events (VISIT_OCCURRENCE, DRUG_EXPOSURE, PROCEDURE_OCCURRENCE, DEVICE_EXPOSURE) into a higher-level abstraction, representing clinically and analytically relevant disease phases, outcomes, and treatments. The EPISODE_EVENT table connects qualifying clinical events (VISIT_OCCURRENCE, DRUG_EXPOSURE, PROCEDURE_OCCURRENCE, DEVICE_EXPOSURE) to the appropriate EPISODE entry. For example, cancers, including their development over time, their treatment, and final resolution.

Table S3. EPISODE Table.

CDM Field	User Guide	ETL Conventions	Datatype	Required	Primary Key	Foreign Key	FK Table	FK Domain
episode_id	A unique identifier for each Episode.		integer	Yes	Yes	No		
person_id	The PERSON_ID of the PERSON for whom the episode is recorded.		integer	Yes	No	Yes	PERSON	
episode_concept_id	The EPISODE_CONCEPT_ID represents the kind abstraction related to the disease phase, outcome, or treatment.	Choose a concept in the Episode domain that best represents the ongoing disease phase, outcome, or treatment. Please see [2] for cancers and [1] for non-cancers describing how these are defined. Accepted Concepts	integer	Yes	No	Yes	CONCEPT	Episode
episode_start_date	The date when the Episode begins.	Please see [1] for how to define an Episode start date.	date	Yes	No	No		
episode_start_datetime	The date and time when the Episode begins.		datetime	No	No	No		
episode_end_date	The date when the instance of the Episode is considered to have ended.	Please see [1] for how to define an Episode end date.	date	No	No	No		
episode_end_datetime	The date when the instance of the Episode is considered to have ended.		datetime	No	No	No		

CDM Field	User Guide	ETL Conventions	Datatype	Required	Primary Key	Foreign Key	FK Table	FK Domain
episode_parent_id	Use this field to find the Episode that subsumes the given Episode record. This is used in the case that an Episode are nested into each other.	If there are multiple nested levels to how Episodes are represented, the EPISODE_PARENT_ID can be used to record this relationship.	integer	No	No	No		
episode_number	For sequences of episodes, this is used to indicate the order the episodes occurred. For example, lines of treatment could be indicated here.	Please see [1] for the details of how to count episodes.	integer	No	No	No		
episode_object_concept_id	A Standard Concept representing the disease phase, outcome, or other abstraction of which the episode consists. For example, if the EPISODE_CONCEPT_ID is treatment regimen then the EPISODE_OBJECT_CONCEPT_ID should contain the chemotherapy regimen concept, like Afatinib monotherapy .	Episode entries from the 'Disease Episode' concept class should have an episode_object_concept_id that comes from the Condition domain. Episode entries from the 'Treatment Episode' concept class should have an episode_object_concept_id that come from the 'Procedure' domain or 'Regimen' concept class.	integer	Yes	No	Yes	CONCEPT	Procedure, Regimen
episode_type_concept_id	This field can be used to determine the provenance of the Episode record, as in whether the episode was from an EHR system, insurance claim, registry, or other sources.	Choose the EPISODE_TYPE_CONCEPT_ID that best represents the provenance of the record. Accepted Concepts . A more detailed explanation of each Type Concept can be found on the vocabulary wiki .	integer	Yes	No	Yes	CONCEPT	Type Concept
episode_source_value	The source code for the Episode as it appears in the source data. This code is mapped to a Standard Condition Concept in the Standardized Vocabularies and the		varchar(50)	No	No	No		

CDM Field	User Guide	ETL Conventions	Datatype	Required	Primary Key	Foreign Key	FK Table	FK Domain
	original code is stored here for reference.							
episode_source_concept_id	A foreign key to a Episode Concept that refers to the code used in the source.	Given that the Episodes are user-defined it is unlikely that there will be a Source Concept available. If that is the case then set this field to zero.	integer	No	No	Yes	CONCEPT	

EPISODE_EVENT Table

Table Description

The EPISODE_EVENT table connects qualifying clinical events (such as CONDITION_OCCURRENCE, DRUG_EXPOSURE, PROCEDURE_OCCURRENCE, MEASUREMENT) to the appropriate EPISODE entry. For example, linking the precise location of the metastasis (cancer modifier in MEASUREMENT) to the disease episode.

User Guide

This connecting table is used instead of the FACT_RELATIONSHIP table for linking low-level events to abstracted Episodes.

Table S4. EPISODE_EVENT Table.

CDM Field	User Guide	ETL Conventions	Datatype	Required	Primary Key	Foreign Key	FK Table	FK Domain
episode_id	A unique identifier for each Episode.		integer	Yes	Yes	No		
person_id	The PERSON_ID of the PERSON for whom the episode is recorded.		integer	Yes	No	Yes	PERSON	
episode_concept_id	The EPISODE_CONCEPT_ID represents the kind abstraction related to the disease phase, outcome or treatment.	Choose a concept in the Episode domain that best represents the ongoing disease phase, outcome, or treatment. Please see [2] for cancers and [1] for non-cancers describing how these are defined. Accepted Concepts	integer	Yes	No	Yes	CONCEPT	Episode
episode_start_date	The date when the Episode begins.	Please see [1] for how to define an Episode start date.	date	Yes	No	No		
episode_start_datetime	The date and time when the Episode begins.		datetime	No	No	No		
episode_end_date	The date when the instance of the Episode is considered to have ended.	Please see [1] for how to define an Episode end date.	date	No	No	No		
episode_end_datetime	The date when the instance of the Episode is considered to have ended.		datetime	No	No	No		
episode_parent_id	Use this field to find the Episode that subsumes the given Episode record. This	If there are multiple nested levels to how Episodes are represented, the	integer	No	No	No		

CDM Field	User Guide	ETL Conventions	Datatype	Required	Primary Key	Foreign Key	FK Table	FK Domain
	is used in the case that an Episode are nested into each other.	EPISODE_PARENT_ID can be used to record this relationship.						
episode_number	For sequences of episodes, this is used to indicate the order the episodes occurred. For example, lines of treatment could be indicated here.	Please see [1] for the details of how to count episodes.	integer	No	No	No		
episode_object_concept_id	A Standard Concept representing the disease phase, outcome, or other abstraction of which the episode consists. For example, if the EPISODE_CONCEPT_ID is treatment regimen then the EPISODE_OBJECT_CONCEPT_ID should contain the chemotherapy regimen concept, like Afatinib monotherapy .	Episode entries from the 'Disease Episode' concept class should have an episode_object_concept_id that comes from the Condition domain. Episode entries from the 'Treatment Episode' concept class should have an episode_object_concept_id that come from the 'Procedure' domain or 'Regimen' concept class.	integer	Yes	No	Yes	CONCEPT	Procedure, Regimen
episode_type_concept_id	This field can be used to determine the provenance of the Episode record, as in whether the episode was from an EHR system, insurance claim, registry, or other sources.	Choose the EPISODE_TYPE_CONCEPT_ID that best represents the provenance of the record. Accepted Concepts . A more detailed explanation of each Type Concept can be found on the vocabulary wiki .	integer	Yes	No	Yes	CONCEPT	Type Concept
episode_source_value	The source code for the Episode as it appears in the source data. This code is mapped to a Standard Condition Concept in the Standardized Vocabularies and the original code is stored here for reference.		varchar(50)	No	No	No		

CDM Field	User Guide	ETL Conventions	Datatype	Required	Primary Key	Foreign Key	FK Table	FK Domain
episode_source_concept_id	A foreign key to a Episode Concept that refers to the code used in the source.	Given that the Episodes are user-defined it is unlikely that there will be a Source Concept available. If that is the case then set this field to zero.	integer	No	No	Yes	CONCEPT	

ANNEX IV. Reviewer reflections

The evaluation was performed by two reviewers: Reviewer 1 (R1) was a participating Data Partner, reflecting the Data Partner perspective, while Reviewer 2 (R2) was an internal OHDSI ETL/Tooling experts, reflecting the OMOP CDM perspective.

Pregnancy Episodes

PET

Advantages:

- Only pre-defined variables can be collected (PET columns) (R1)
- Maintenance would be easy as we have predefined variables (R1)
- Some important variables are already there (e.g. pregnancy_outcome), which facilitates the analyses. (R1)
- all currently required elements captured; (R2)
- fields names clearly linked to pregnancy, easy to understand and populate from source registry (R2)
- analytical table, directly available for reporting (R2)

Downsides:

- Pregnancy_start_date and Pregnancy_end_date are mandatory but they are not always well recorded. Loss of follow-up is difficult to register there (R1)
- It is not flexible to changes (R1)
- No current specific packages to work with the PET but we think other packages may work/adapt easily (R1)
- Inaccurate end_dates recordings may lead to intersections between pregnancies (of same person) that affects the use of some packages (R1)
- Less flexible for changes, if a new outcome variable needs to be added, then the model needs to be updated, which also needs to be implemented in the analytical packages (R2)
- Potentially double recording of data, for non-registries this is an additional step to populate (R2)
- Not supported by ATLAS, not supported by standard DARWIN EU® packages (R2)

Condition Era Table

Advantages:

- We think required changes are NA as we cannot think of any potential changes. (R1)
- The table already exists and given a certain input, it doesn't need a specific implementation (R1)
- Model is flexible, new variables do not require a change in data model (R2)
- familiar structure, similar to drug_era and condition_occurrence (R2)
- can be used in ATLAS and standard DARWIN EU® packages (R2)
- We do not have experience with the condition era but as it is a CDM table packages may be adapted to their structure (R1)

Downsides:

- It is restricted to conditions. During pregnancy we have many measurements and observations that would not be captured in the “Era”. (R1)
- We are unsure about how to establish start and end based on the same condition (we don’t think in SIDIAP there is a condition systematically registered throughout pregnancy. (R1)
- Pregnancy related variables wouldn’t be captured with this option (R1)
- However, pregnancy related variables wouldn’t be capture with this option so analyses would be very complex (R1)
- outcome needs to be linked via date (R2)
- misuse of standard OMOP table (by convention era directly derived from condition occurrence, potentially clashing with existing condition era derivation script) (R2)

EPISODE and EPISODE_EVENT Tables**Advantages:**

- Complete (can be linked to any event: condition, drug, measurement, observation, etc). (R1)
- We could define hierarchical episodes if needed. (R1)
- Flexibility (R1)
- We think it would be easy, but we don’t have experience (R1)
- All elements captured. (R2)
- Model is flexible, new variables do not require a change in data model. OMOP vocabularies are extensive in terms of pregnancy outcomes. And if new concept is needed, then the analytical tooling does not need to be adapted. (R2)
- reuse of existing OMOP table, gained experience with table transfers to other fields such as Oncology (R2)
- support in ATLAS being developed, standard DARWIN EU® packages can be easily adopted. Same packages can be reused for Oncology studies. (R2)

Downsides:

- new table with new OMOP conventions that DPs are not yet familiar with (R2)
- analytical table needs to be derived (same as for other OMOP tables) - gives maximum flexibility (R2)

Cohort Definition**Advantages:**

- NA if we create the cohort for each study (R1)
- If cohort is part of the cdm, then it would be easy to implement. (R1)
- model is flexible, new variables do not require a change in data model. (R2)
- easy for DPs as mapping to standard OMOP tables (R2)
- can be used in ATLAS and standard DARWIN EU® packages (R2)

Downsides:

- We do not fully understand this option. Should we create a cohort for each study or would we have a pregnancy standard cohort already defined in our cdm? (R1)
- Pregnancy related variables wouldn't be captured with this option (R1)
- Complicated: we should work harder on definitions for each analysis? (R1)
- Not easy to find in which table we should search for specific variables (depends on how any variable is registered by each DP). (R1)
- potentially missed if outcome unknown. (R2)
- conventions needed where to store pregnancy data to make it usable for analysis (R2)
- pregnancy episodes defined on a study-per-study basis and result not stored for subsequent studies, so not usable for new study feasibility. (R2)
- Every outcome needs to be defined separately with a cohort definition; (R2)

Pregnancy-Related Variables**PET****Advantages:**

- Easy to find records that meet certain condition (e.g. pregnancies with pregnancy_outcome==xx) (R1)
- expert-based selection of variables (R2)
- easy implementation as subset of variables is already made (R2)
- explicit relation to pregnancy (R2)

Downsides:

- Only pre-defined variables (not flexible) (R1)
- Including new variables would take a lot of effort (R1)
- Not feasible to include a large list of variables (each of them would be included in a new column) (R1)
- No current specific packages to work with the PET but we think other packages may work/adapt easily (R1)
- additional variables require change of model (R2)
- some variables encoded as yes/no, some as numeric and some as concept (R2)
- outside of the predefined PET-variables, analysis needs to also query OMOP event tables. (R2)

OMOP Event Table**Advantages:**

- Flexibility. We can add as many events as we want. (R1)
- Complete (R1)
- Easy to maintain. (R1)
- Contents fully flexible (R2)

- Similar mapping as other OMOP mapping of variables (R2)
- standard OMOP analytics, data quality and characterisation can be applied (R2)

Downsides:

- Complexity of implementation would depend on having a predefined list of pregnancy-related events to be included (if we have a list would be easier) (R1)
- It would be computationally demanding as we need to search in all tables. (R1)
- If we have all conditions during pregnancy, we are not sure if this table facilitates the analyses. Instead, we could have an Event table populated with a pre-defined list of variables (pregnancy_outcome, delivery_mode, last menstrual period, eco, gestational_age, etc), and the user may adapt (add other variables) based on each study. (R1)
- some pregnancy variables might not have OMOP standard concept. These can still be as source values (R2)
- date field is required, might not be available for some data sources. Additional convention needed upon etl implementation. (R2).
- need to decide at analysis time if related to pregnancy, via episode event or based on date (R2)

Parent-Child Linkage

PET

Advantages:

- We understand that the proposed option here corresponds to the infant table. In that case we think it's a complete option (R1)
- It would facilitate analyses as we could filter specific variables already defined in the table (R1)
- explicit linkage of child to pregnancy, possible to link (R2)
- designed to make linkage from pregnancy registry (R2)

Downsides:

- We would need to link each child with the specific pregnancy, and we do not have a direct link in the source (R1)
- No current specific packages to work with the PET but we think other packages may work/adapt easily (R1)
- No father-child relationship (R2)
- For implementation, multiple steps to link mother to child (person->pregn->infant->person) (R2)

FACT_RELATIONSHIP table

Advantages:

- It doesn't require further efforts to implement this table (R1)
- Capture any familial relationship (R2)
- Exists in the OMOP CDM and has implementation examples from other domains (R2)
- With a clear mapping guidance, easy to implement for familial relationships (R2)

Downsides:

- Not complete: we do not have the pregnancy_id recorded. (R1)
- An option to improve this would be to register as a fact_relationship the child with the preg_id and the mother with the same preg_id, using relationships such as "birth of", "pregnancy of" (R1)
- Since pregnancy_id is not recorded, it is difficult to link. (R1)
- No explicit link to pregnancy (R2)
- No explicit foreign key to PERSON table, needing an additional step to perform the lookup (R2)



Study Report

P4-C4-001

DARWIN EU[®] - PeriNet

OBJECTIVE 4

Characterisation of pregnancies and mother-child linkage in data sources with pregnancy-related data

25/06/2026

Version 4.0

Authors: Berta Raventós, Julieta Politi, Maarten van Kessel, Ger Inberg, Ross Williams, Natasha Yefimenko, Katia Verhamme, Talita Duarte-Salles

Public

Study title	DARWIN EU® - PeriNet Objective 4: Characterisation of pregnancies and mother-child linkage in data sources with pregnancy-related data.
Study report version	V4.0
Date	25/06/2026
EUPAS number	EUPAS1000000812
Active substance	N/A
Medicinal product	N/A
Research question and objectives	The aim of this study was to characterise pregnancies and mother-child linkage over time for data sources with pregnancy-related data mapped to the OMOP CDM. The specific objectives of this study were: 4a. To describe pregnancy episodes within the participating data sources in terms of gestational length, pregnancy outcomes, mode of delivery, and the pregnancy episodes with mother-child linkage. 4b. To describe time trends in the number of pregnancy episodes, pregnancy outcomes, mode of delivery, and number of pregnancy episodes with mother-child linkage within the participating data sources. 4c. To characterise individuals with a pregnancy episode in terms of demographics, baseline comorbidities, health-related behaviours, medication use, pregnancy complications, and pregnancy outcomes. 4d. To characterise infants with mother-child linkage in terms of birth-related characteristics, early life factors, and follow-up time.
Countries of study	Norway, Spain, Sweden, the United Kingdom
Authors	Berta Raventós (b.raventos@darwin-eu.org) Julieta Politi (j.politi@darwin-eu.org) Maarten van Kessel (m.vankessel@darwin-eu.org) Ger Inberg (g.inberg@darwin-eu.org) Ross Williams (r.williams@darwin-eu.org) Natasha Yefimenko (n.yefimenkonosova@darwin-eu.org) Katia Verhamme (k.verhamme@darwin-eu.org) Talita Duarte-Salles (t.duarte@darwin-eu.org)

LIST OF ABBREVIATIONS

Acronyms/term	Description
APGAR	Appearance, Pulse, Grimace, Activity, and Respiration
ATC	Anatomical Therapeutic Chemical
BMI	Body Mass Index
CC	Coordination Centre
CDM	Common Data Model
CI	Confidence Interval
COVID-19	Coronavirus disease 2019
CPRD	Clinical Practice Research Datalink
DARWIN EU®	Data Analysis and Real-World Interrogation Network
DOI	Declaration of Interests
DQD	Data Quality Dashboard
DRE	Digital Research Environment
DTZ	Data Transfer Zone
ED	Emergency Department
EHR	Electronic Health Records
EMA	European Medicines Agency
ENCePP	European Network of Centres for Pharmacoepidemiology and Pharmacovigilance
ETL	Extract, Transform and Load
EU	European Union
EUPAS	EU Post-Authorisation Studies Register
GDPR	General Data Protection Regulation
GP	General Practitioner
HELLP	Hemolysis with a microangiopathic blood smear, elevated liver enzymes, and low platelets
HI-SPEED	Health Impact - Swedish Population Evidence Enabling Data-linkage
HIV	Human Immunodeficiency Virus
ICD	International Classification of Diseases
IMASIS	Institut Municipal Assistència Sanitària Information System
IRB	Institutional Review Board
LMP	Last Menstrual Period
NLHR	Norwegian Linked Health Registry data
OHDSI	Observational Health Data Sciences and Informatics
OMOP	Observational Medical Outcomes Partnership
PET	Perinatal Expansion Table
RxNorm	Medical prescription normalised
SNOMED	Systematized Nomenclature of Medicine

Acronyms/term	Description
SIDIAP	The Information System for Research in Primary Care
Tdap	Tetanus, Diphtheria & Pertussis
TORCH	Toxoplasmosis, others, rubella, cytomegalovirus, herpes simplex
UK	The United Kingdom
UiO	The University of Oslo

1. TITLE

DARWIN EU® - PeriNet Objective 4: Characterisation of pregnancies and mother-child linkage in data sources with pregnancy-related data.

2. DESCRIPTION OF THE STUDY TEAM

Study team role	Names	Organisation
Principal Investigator	Berta Raventós Talita Duarte-Salles	Erasmus MC
Data Scientist	Maarten van Kessel Ger Inberg Ross Williams ¹	Erasmus MC
Clinical Domain Expert	Julieta Politi ² Katia Verhamme	Erasmus MC
Study Manager	Natasha Yefimenko	Erasmus MC
Data source	Names	Data Partner Organisation*
NLHR	Saeed Hayati Petrica-Ioan Olteanu Hedvig Nordeng	University of Oslo
IMASIS	Miguel Angel Mayer Angela Leis Juan Manuel Ramírez Anguita	Parc de Salut Mar Barcelona
SIDIAP	Agustina Giuliadori Elena Roel Laura Granés	IDIAPJGol
HI-SPEED	Rickard Ljung Fredrik Nyberg Huiqi Li	Swedish Medical Products Agency - Gothenburg University (SMPA-GU)
CPRD GOLD	Antonella Delmestri	University of Oxford

*Data partners do not have an investigator role. Data partners execute code at their data source, review, and approve their results.

¹Added to the study team on 09/04/2026.

²Involved in the study until 30/04/2026.

3. ABSTRACT

Title

DARWIN EU® - PeriNet Objective 4: Characterisation of pregnancies and mother-child linkage in data sources with pregnancy-related data.

Rationale and background

The DARWIN EU® network offers a unique opportunity to inform research on pregnancy and perinatal outcomes. In this context, a comprehensive assessment of pregnancy-related data is crucial for supporting future feasibility studies on pregnancy and perinatal health within the network.

Research question and objectives

Research questions

What were the temporal trends in the capture of pregnancy-related data in selected data sources within the DARWIN EU® network, and what were the characteristics of individuals with recorded pregnancies or infants with mother-child linkage?

Objectives

The aim of this study was to characterise pregnancies and mother-child linkage over time for data sources with pregnancy-related data mapped to the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM).

The specific objectives of this study were:

- 4a. To describe pregnancy episodes within the participating data sources in terms of gestational length, pregnancy outcomes, mode of delivery, and the number of pregnancy episodes with mother-child linkage.
- 4b. To describe time trends in the number of pregnancy episodes, pregnancy outcomes, mode of delivery, and number of pregnancy episodes with mother-child linkage within the participating data sources
- 4c. To characterise individuals with a pregnancy episode in terms of demographics, baseline comorbidities, health-related behaviours, medication use, pregnancy complications, and pregnancy outcomes.
- 4d. To characterise infants with mother-child linkage in terms of birth-related characteristics, early life factors, and follow-up time.

Methods

Study design

A study cohort, including:

- A characterisation of pregnancy episodes, overall and over time (Objectives 4a and 4b).
- A characterisation of pregnancy episodes, focusing on characteristics of pregnant individuals (objective 4c).
- A characterisation of infants with mother-child linkage (Objective 4d).

Population

For objectives 4a to 4c, all pregnancy episodes recorded from 2010 to one year before the end of data availability in females aged 12 to 55 years at pregnancy start (i.e. index date) were included.

To describe infants (Objective 4d), all infants with mother-child linkage born from 2010 to one year before the end of data availability were included. Index date was defined as the recorded date of birth for infants.

Variables

Relevant covariates:

- Maternal characteristics (Objective 4c):
 - Demographics
 - Pregnancy outcome and duration
 - Health-related behaviours
 - Baseline comorbidities
 - Pregnancy complications
 - Medication use based on Anatomical Therapeutic Chemical (ATC) 3rd level classification system
 - Large-scale characterisation based on conditions and medications at ingredient-level
- Infant characteristics (Objective 4d):
 - Demographics
 - Gestational age
 - Birth weight
 - Appearance, Pulse, Grimace, Activity, and Respiration (APGAR) score (obtained 5 minutes after birth)
 - Small for gestational age
 - Breastfeeding
 - Congenital malformations
 - Mortality

Data sources

Data sources that have implemented the Perinatal Expansion Table (PET) or that had completed this process before the end of the implementation of this study objective were included:

1. Norway: Norwegian Linked Health Registry data (NLHR)
2. Spain: Institut Municipal Assistència Sanitària Information System (IMASIS)
3. Spain: The Information System for Research in Primary Care (SIDIAP)
4. Sweden: Health Impact - Swedish Population Evidence Enabling Data-linkage (HI-SPEED)
5. The United Kingdom: Clinical Practice Research Datalink GOLD (CPRD GOLD)

Study size

No sample size was calculated, as this was a descriptive study which did not test a specific hypothesis.

Statistical analysis

All pregnancy episodes were described in terms of pregnancy outcomes, mode of delivery, and number of pregnancies with mother-child linkage, overall and by calendar year. Gestational length, plurality, parity, and gravidity were also described. The analysis of Objectives 4c and 4d focused on characterising individuals with pregnancy episodes and infants with mother-child linkage based on relevant covariates. Medication use was assessed during pregnancy, in the 90 days prior to pregnancy start, and in the 90 days following both pregnancy start and pregnancy end. Pregnancy complications were assessed by trimester of gestation.

Characteristics were summarised (where available) using counts and percentages for categorical variables and medians with interquartile ranges for continuous variables. A minimum cell count of 5 was used when reporting results, with any smaller counts reported as "<5" and zero counts as "0".

Results

A total of 2,146,567 pregnancy episodes (n=367,803 in NLHR; n=15,330 in IMASIS; n=409,207 in SIDIAP; n=887,612 in HI-SPEED; n=466,615 in CPRD GOLD) and 1,482,391 infants with mother-child linkage were included (n=257,826 in NLHR; n=15,356 in IMASIS; n=312,291 in SIDIAP; n=896,918 in HI-SPEED; not available in CPRD GOLD).

New pregnancy episodes per year were generally stable over time but showed a marked decrease in SIDIAP and CPRD GOLD. Distribution of pregnancy outcomes differed between data sources. IMASIS primarily included live births (99.6%), whereas HI-SPEED mainly captured pregnancies that progressed beyond week 22. Stillbirths occurred in 0.2%–0.6% of pregnancies. Outcomes occurring in early gestation were not captured consistently, with elective terminations ranging from 0.5% to 14.3% and miscarriages from 8.3% to 27.8% in data sources where these outcomes were recorded in the PET (NLHR, SIDIAP, CPRD GOLD). C-sections ranged from 4.6% in NLHR to 22.6% in IMASIS.

Median age at pregnancy start was approximately 30 years. Current smoking ranged from 3.8% to 25.2%. During pregnancy, recorded folic acid use ranged from 3% to 28.2%, and vaccination varied from 1.7% to 18.3% for COVID-19, 8.5% to 29.7% for influenza, and 0.1% to 52.8% for Tdap. Pharmacological subgroups that were most frequently recorded during pregnancy included vaccines (J07A, J07B, J07C), "other analgesics and antipyretics" (N02B), "vitamin B12 and folic acid" (B03B), "iron preparations" (B03A), and "beta-lactam antibacterials penicillin" (J01C). The most common complications observed during pregnancy were antepartum haemorrhage (range: 3.8% to 12.0%), gestational diabetes (range: 2.8% to 8.0%), and preterm labour (range: 3.8% to 7.0%). Antepartum haemorrhage and ectopic pregnancies were more common in the first trimester, whereas other complications were generally more frequent in the third.

In infants with mother-child linkage, median age at the end of follow up ranged from 2 to 7 years. Birth weight ranged from 3,193g to 3,540g, and small for gestational age was observed in 8.2% of infants in NLHR, 10.1% in IMASIS, and 2.3% in HI-SPEED. Median APGAR score 5 minutes after birth was estimated at 10 across participating data sources. Breastfeeding in SIDIAP was observed in 88.5%, 59.4%, and 28.9%, in month 1, 6, and 12 after birth, respectively. Congenital malformations (including minor and major malformations, and structural and non-structural anomalies) were captured in 18.3% of infants in NLHR, 9.4% in SIDIAP, and 3.2% in HI-SPEED. Mortality was rare and was observed in <0.1% and <0.5% of cases within 30 days and one year after birth, respectively.

Discussion

This study provides a comprehensive overview of pregnancy-related health data in five data sources of the DARWIN EU® Data Network that have implemented the OMOP PET. The completeness and scope of the data varied across the included sources and depended upon the setting, data collection, data quality, and how the OMOP PET was populated. This suggests the need for operational rules aimed at ensuring the

validity of pregnancy episodes based upon plausibility, complemented by consistency in OMOP mapping and populating the PET, which has been addressed as part of DARWIN EU® PeriNet- Objective 3.

4. AMENDMENTS AND UPDATES

None.

5. MILESTONES

Study deliverable	Timelines (planned)	Timelines (actual)
Final Study Protocol	October 2025	November 2025
Creation of Analytical code	December 2025	January 2026
Execution of Analytical Code on the data	January 2026	February – April 2026
Draft Study Report	March 2026	April 2026
Final Study Report	May 2026	May 2026

6. RATIONALE AND BACKGROUND

Pregnant individuals are routinely excluded from clinical trials due to ethical considerations, resulting in a critical evidence gap on the safety and effectiveness of medicinal products at the time of marketing authorisation. Consequently, pregnancy is recognised as an important population for post-authorisation safety monitoring and is frequently included in risk management plans.[1, 2]

Observational healthcare data provide a key opportunity to address this evidence gap, as data sources may capture clinically relevant information on pregnancy, perinatal outcomes, and infant health. However, the availability, granularity, and research-readiness of such data vary considerably across data sources. A systematic assessment of pregnancy-related data, such as its structure, coding, and completeness, is essential to support robust pharmacoepidemiologic analyses in DARWIN EU® and to support alignment with the European Medicines Agency (EMA) Data Quality Framework, for example, in relation to extensiveness, relevance, and reliability.[3]

The adoption of the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) within the DARWIN EU® network provides a standardised approach for data harmonisation and analysis. Nevertheless, the use of OMOP CDM for perinatal research has been limited due to structural constraints in defining pregnancy episodes and key characteristics, such as gestational length, pregnancy outcomes, and mother-infant linkages. The recent development of the OMOP perinatal expansion addresses some of these limitations by introducing a standardised structure for perinatal data.[4] The expansion includes two main components: a pregnancy table and an infant table, which contain a set of variables related to each pregnancy episode and newborn infant, respectively. These tables are generated during the mapping process, also known as Extract, Transform, and Load (ETL) process, where source data is mapped into the standardised format required within the OMOP CDM. For simplicity, we refer to the pregnancy table as the Perinatal Expansion Table (PET) throughout this report. The implementation of the infant table was initially limited to one data source at the time when the perinatal expansion was published and has not yet been used in epidemiological studies.[4] At the time the protocol for this study objective was approved, four data sources within the DARWIN EU® Network had already implemented the PET and had either participated in or were currently involved in pregnancy-related studies.[5-7]

Successful utilisation of perinatal data, either through the perinatal expansion at ETL stage or through the creation of a temporary PET table through the use of pregnancy algorithms, depends on the underlying data capture from source data.[4, 8] This study aimed to characterise pregnancies and mother-child linkage over time in selected data sources that have implemented the PET within the DARWIN EU® Network. This report corresponds to the fourth objective of the “DARWIN EU® - Enabling pregnancy and mother-child research in the DARWIN EU® Data Network - DARWIN-PeriNet”. By assessing temporal trends in the

capture of pregnancy-related information and the characteristics of pregnant individuals and infants, this objective aimed to provide a broader understanding of pregnancy-related episodes and contribute to future feasibility assessments by offering insights into the populations and data involved.

7. RESEARCH QUESTION AND OBJECTIVES

Research question

What were the temporal trends in the capture of pregnancy-related data in data sources that have implemented the PET within the DARWIN EU® network, and what were the characteristics of individuals with recorded pregnancies or infants with mother-child linkage?

Objectives

The aim of this study was to characterise pregnancies and mother-child linkage over time for data sources with pregnancy-related data mapped to the OMOP CDM.

The specific objectives of this study were:

- 4a. To describe pregnancy episodes within the participating data sources in terms of gestational length, pregnancy outcomes, mode of delivery, and the number of pregnancy episodes with mother-child linkage.
- 4b. To describe time trends in the number of pregnancy episodes, pregnancy outcomes, mode of delivery, and number of pregnancy episodes with mother-child linkage within the participating data sources.
- 4c. To characterise individuals with a pregnancy episode in terms of demographics, baseline comorbidities, health-related behaviours, medication use, pregnancy complications, and pregnancy outcomes.
- 4d. To characterise infants with mother-child linkage in terms of birth-related characteristics, early life factors, and follow-up time.

8. RESEARCH METHODS

8.1. Study design

A cohort study was conducted using routinely collected health data from data sources included in the DARWIN EU® Network. The study comprised of:

The study comprised of:

- A characterisation of pregnancy episodes, overall (objective 4a) and over time (objective 4b).
- A characterisation of pregnancy episodes, focusing on individual characteristics of pregnant individuals (objective 4c).
- A characterisation of infants with mother-child linkage (objective 4d).

8.2. Follow-up

Objectives 4a, 4b, and 4c:

For Objectives 4a to 4c, analyses were conducted at the pregnancy episode level. Follow-up started from pregnancy start date. According to the PET specifications, this is based on ultrasound estimations or calculated from the last menstrual period (LMP) if ultrasound information is missing.[4] End of follow-up was defined as the earliest of end of pregnancy, loss to follow-up, death, or end of observation period (the latest available data), whichever occurred first. For Objective 4c, individuals were followed up to 90 days

after pregnancy end to assess certain postpartum covariates for characterisation, as described in [Section 8.6.3](#). Individuals were allowed to contribute with multiple entries, corresponding to each pregnancy recorded during the study period, provided they met the criteria specified in [Section 8.3](#).

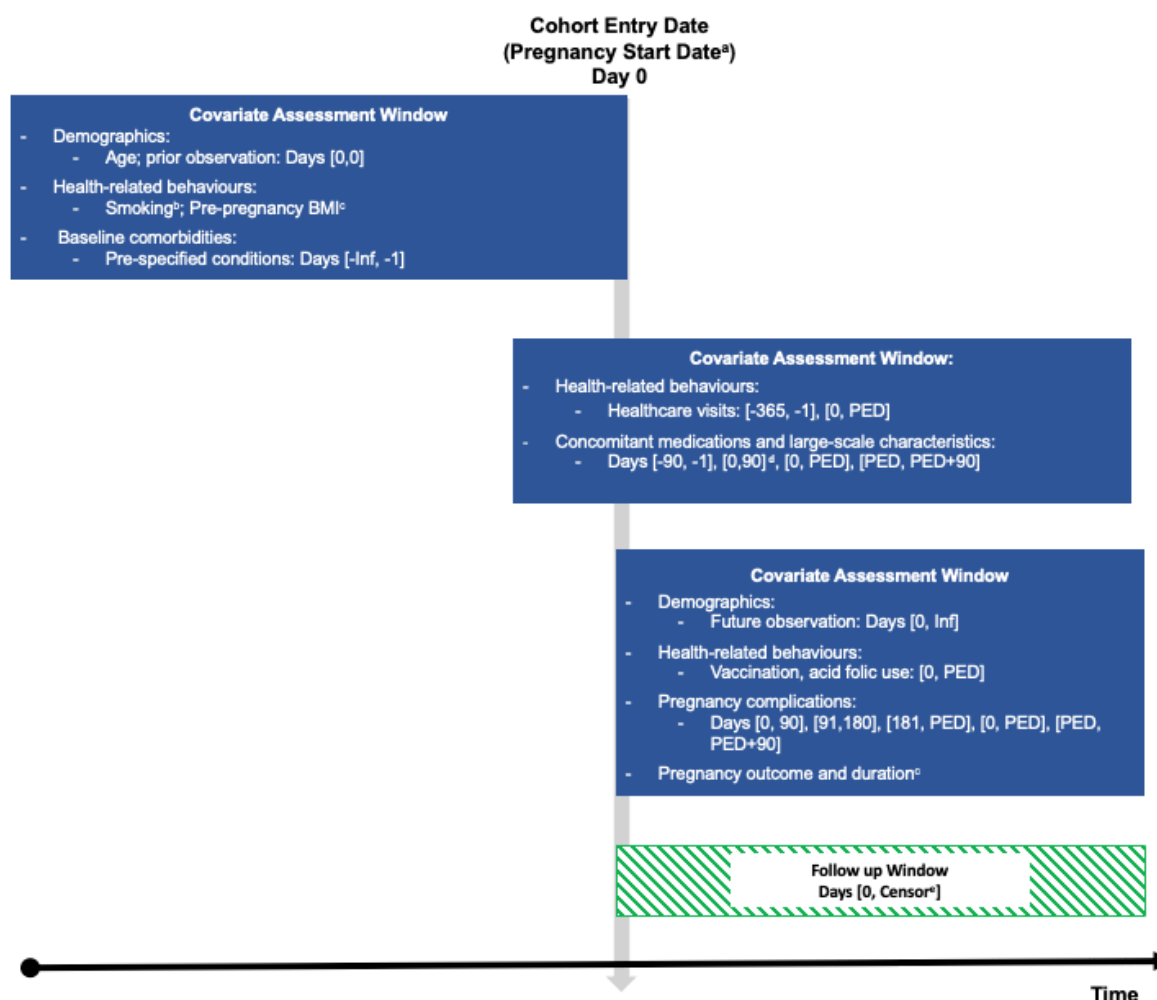


Figure 1. Graphical depiction of the study design for Objective 4c, assessing pregnant individuals.

- a. Cohort entry date was defined as the date of pregnancy start. The source population included pregnant females aged 12 to 55 years at pregnancy start.
- b. As recorded in the Perinatal Expansion Table. If not recorded there, it was ascertained within the time window of [-730, 90] days.
- c. As recorded in the Perinatal Expansion Table.
- d. The time window [0,90] days were applied to medications only.
- e. Defined as the earliest of 90 days after pregnancy end, loss to follow-up, death, or end of observation period (the latest available data), whichever occurred first.

BMI=Body Mass Index; COVID-19=coronavirus disease 2019; PED=Pregnancy end date; Tdap=Tetanus, Diphtheria, & Pertussis.

Objective 4d:

For infants with mother-child linkage, follow-up started on the recorded date of birth, as recorded in the maternal records (see [Section 8.10](#)). All covariates were assessed at index date, except for age and future observation (i.e. years of data availability), which were assessed at the end of follow-up. End of follow-up was defined as loss to follow-up, death, or end of observation period (the latest available data), whichever occurred first ([Figure S1](#)).

8.3. Study population with inclusion and exclusion criteria

Objectives 4a, 4b, and 4c:

For Objectives 4a to 4c, all pregnancy episodes recorded of females aged 12 to 55 years at pregnancy start (i.e. index date) were included. Therefore, participants could contribute multiple episodes if more than one pregnancy was recorded during the study period.

Inclusion criteria

- Pregnancy starting in the period of 01/01/2010 (or start of data availability, if later) to one year before end of data availability
- Females at index date
- Participants aged 12 to 55 years at index date

Exclusion criteria

- Pregnancies among individuals not observed in the data source at the pregnancy start or end date.
- Pregnancies with implausible duration (i.e. pregnancy end date < pregnancy start date, <1 day, or > 44 weeks of gestation)
- Pregnancies with conflicting dates (i.e. distinct pregnancies with overlapping dates)

Objective 4d:

For Objective 4d, infants with mother-child linkage were described. The infant population was restricted to those that could be linked to a recorded pregnancy episode in the PET in the previous objective (see [Section 8.10.](#)). Therefore, a documented parental relationship (i.e. a record indicating the infant is the child of another subject alone) was not sufficient unless the infant could be linked to a specific pregnancy episode.

Inclusion criteria

- All infants with mother-child linkage

Exclusion criteria

- Infants that could not be linked to a pregnancy episode included in Objectives 4a, 4b, and 4c.
- Infants born from a pregnancy episode that resulted in a live birth.

8.4. Study setting and data sources

This study was conducted using routinely collected data from 5 data sources in the DARWIN EU® network of data partners from 5 European countries, of which 2 are member states of the European Union (EU) ([Table 1](#)). All data were a priori mapped to the OMOP CDM.[9, 10]

Table 1. Data sources.

Country	Name of Data source	Health Care setting	Type of Data	Number of active individuals ¹	Calendar period covered by each data source ²	Contributing to
Norway	Norwegian Linked Health Registry data (NLHR)	Primary and secondary care	Registries	5.55M	2018–2023	All objectives
Spain	Institut Municipal Assistència Sanitària Information System (IMASIS)	Secondary care	EHR	1.77M	1990–2024	All objectives
Spain	The Information System for Research in Primary Care (SIDIAP)	Primary care, inpatient care	EHR	5.95M	2006–2024	All objectives
Sweden	Health Impact - Swedish Population Evidence Enabling Data-linkage (HI-SPEED)	Primary care ³ ; Secondary care	Registries	10.56M	2015–2024	All objectives
The United Kingdom	Clinical Practice Research Datalink GOLD (CPRD GOLD)	Primary care	EHR	2.83M	1987–2025	All objectives, except 4d ⁴

EHR=Electronic health records.

¹ Defined as the maximum number of individuals in observation in the last 6 months of data.

² For NLHR and HI-SPEED, different registries that comprise each data source have different start dates. Current data releases include data up to: December 2023 (NLHR); June 2025 (IMASIS); December 2024 (SIDIAP); August 2024 (HI-SPEED); and May 2025 (CPRD GOLD).

³ Primary care diagnoses and procedures were available for two regions (Stockholm region and Västra Götaland region), which consist of approximately 40% of the total Swedish population.

⁴ CPRD GOLD did not have information on mother-child linkage. Consequently, analyses related to mother-child linkage were not possible in this data source.

Data sources selection

These data sources fulfilled the criteria required in terms of data quality, completeness, timeliness, and representativeness while covering different regions of Europe ([Annex I](#)). More details on the selection of data sources for the DARWIN-PeriNet project can be found in the study report of Objective 1 of the project.

In this study, all selected data sources had already implemented the PET, except for the Institut Municipal Assistència Sanitària Information System (IMASIS), who did so before the conclusion of the implementation stage of this study. Additional data sources within the DARWIN EU® network were informed of the opportunity to participate, contingent upon the implementation of the PET prior to the conclusion of the study's implementation phase. All data sources specified in [Table 1](#), except for IMASIS, have previously participated, or are currently participating, in multi-data source studies on pregnancy using the PET.[5-7]

Regarding Norwegian data, it should be acknowledged that Norwegian Linked Health Registry data (NLHR) is onboarded into the DARWIN EU® Network using two separate data sources, which differ based on the data source providing linkage IDs and the time period covered. For this study, we used NLHR@UiO (hereinafter referred to as NLHR), as it has a greater ability to detect early pregnancies (<12

gestational weeks), and its source population includes the entire population of Norway. Additional information regarding the differences between the NLHR data sources can be found in [Annex II](#).

The PET was implemented by each data partner. Importantly, the Health Impact - Swedish Population Evidence Enabling Data-linkage (HI-SPEED) and NLHR populated the PET using data from national birth registries with mandatory reporting of pregnancies that surpass week 22 and 12, respectively. In NLHR, pregnancies ending before week 12 were identified (and included in the PET) based on the pregnancy algorithm developed by the University of Oslo.[11] No linkage to birth registries was in place for the Information System for Research in Primary Care (SIDIAP) data source and CPRD GOLD. SIDIAP includes detailed pregnancy information from reproductive health clinics in primary care. Linkage to hospital discharge data was requested for this study, which enriched information on conditions recorded during hospitalisations. However, the PET was populated based on primary care data and did not include information from hospital settings. Pregnancy episodes in CPRD GOLD were inferred from primary care data using a validated algorithm.[8] IMASIS populated the PET based on data from the hospital birth registry, which only includes data on pregnancies that resulted in birth, with pregnancy start derived from information on pregnancy end date and gestational length. Important considerations regarding this matter and its potential implications in the study are described in [Section 10.2.](#) and [Annex II](#).

All data sources included in this study, except for Clinical Practice Research Datalink GOLD (CPRD GOLD), have information on mother-child linkage (see [Section 10.2.](#)). It should be noted that mother-child linkage can be available through the perinatal expansion infant table,[4] but can also be recorded using other tables in the OMOP CDM. SIDIAP and NLHR have information on mother-child linkage in the “fact_relationship” table of the OMOP CDM, which contains information on the relationship between facts stored as records in any table of the CDM,[12] allowing the linkage between the ID of the pregnant individual with the ID of the infant. HI-SPEED and IMASIS were the only data sources that had populated the infant table at the time of conducting this study, establishing a link between the infant and the pregnancy episode (instead of the pregnant individual, who can have multiple pregnancies).

8.5. Study period

The study period spanned from 01/01/2010 to end of data availability. The enrolment period for pregnancies (and their infants) spanned from 01/01/2010 to one year prior to end of data availability.

It should be noted that in HI-SPEED, data availability begins in 2015, with prescription data only available from 2018 onwards. In this data source, the study period was restricted to 2018 onwards for Objective 4c to ensure complete capture of relevant information for characterisation (see [Section 9](#)). For consistency, the same restriction was applied for HI-SPEED in Objective 4d. In NLHR, data are available starting in 2018, and all objectives included data from that point forward.

8.6. Variables

8.6.1. Covariates

Objective 4a:

Objective 4a focused on required fields of the PET,[4] which included:

- Gestational length (in days)
- Pregnancy outcomes: Livebirth, miscarriage (<20 weeks), stillbirth (≥20 weeks), elective termination of pregnancy, discordant (different outcomes in a multiple gestation), unknown
- Mode of delivery: Vaginal, C-section
- Plurality (singleton versus multiple gestation)
- Parity (nulliparous versus multiparous)

- Gravidity (number of previous pregnancies)

Regarding mother-child linkage, we described:

- Number of pregnancies with mother-child linkage

Objective 4b:

Time-trends of pre-defined covariates included as part of Objective 4a, stratified by calendar year.

- Number of pregnancy episodes
- Pregnancy outcomes
- Mode of delivery
- Number of pregnancies with mother-child linkage
- Calendar year

Objective 4c:

Maternal covariates for this objective are described below. The proposed time windows were selected for their potential to offer a broader view of the recorded health data and population characteristics and to support future feasibility assessments. In most cases, the time windows included: 90 days prior to the pregnancy start date, the period between pregnancy start and end, and 90 days following the end of pregnancy. Trimester-specific windows were only considered for pregnancy complications with the aim to assess the timing of the recording of these events. The 90-day post-pregnancy window was also included for pregnancy complications to capture events that may have occurred prior to the end of pregnancy but were recorded afterward. For medication use, an additional time window representing the first trimester of pregnancy was included.

- Demographics (at index date):
 - Age
 - Age groups: 12–24; 25–34; 35–55 years.
 - Prior and future observation (i.e. data visibility before and after index date)
- Pregnancy outcome and duration:
 - Pregnancy outcomes (as recorded in the PET, same as for Objective 4a)
 - Gestational length (as recorded in the PET, same as for Objective 4a)
- Health-related behaviours (where possible):
 - Smoking, assessed as recorded in the PET (if available) or in the 2 years prior to pregnancy start and up to 90 days after (i.e. Days: [-730,90])
 - Pre-pregnancy body mass index (BMI), as recorded in the PET
 - Number of health visits (any type), assessed in days: [-365, -1], [0, pregnancy end]
 - Vaccination during pregnancy, assessed [0, pregnancy end]:
 - Coronavirus disease 2019 (COVID-19)
 - Influenza
 - Tetanus, Diphtheria, & Pertussis (Tdap)
 - Use of folic acid, assessed [0, pregnancy end], as recorded in the PET

- Baseline comorbidities:
 - Pre-specified conditions included:
 - Alcohol misuse
 - Anxiety
 - Asthma
 - Cancer (any malignancy)
 - Chronic kidney disease
 - Chronic liver disease
 - Diabetes mellitus
 - Epilepsy/seizures
 - Essential hypertension
 - Heart disease (excluding hypertension)
 - Human Immunodeficiency Virus (HIV) infection
 - Inflammatory auto-immune diseases of the intestine (composite of celiac disease and inflammatory bowel disease)
 - Mood disorders (composite of depression and bipolar disorders)
 - Obesity
 - Thyroid disorders
 - Assessment window, in days: [-Inf, -1]
- Pregnancy complications:
 - Pre-specified complications other than miscarriage and stillbirth:
 - Antepartum Haemorrhage
 - Eclampsia
 - Ectopic pregnancy
 - Gestational diabetes
 - Hemolysis with a microangiopathic blood smear, elevated liver enzymes, and low platelets (HELLP) Syndrome
 - Maternal death
 - Pre-eclampsia
 - Preterm labour
 - Postpartum haemorrhage
 - TORCH (toxoplasmosis, others, rubella, cytomegalovirus, herpes simplex) infection. Others included syphilis, chickenpox, and Zika virus.
 - Assessment window (in days): [0, pregnancy end]), [pregnancy end, 90 days after pregnancy end]

- Additional assessment windows (in days): [0, 90], [91,180], [181, pregnancy end]
- Concomitant medications:
 - Medications were assessed using Anatomical Therapeutic Chemical (ATC) Classification System, using 3rd level groups (i.e. pharmacological subgroup)[13]
 - Assessment window, in days: [-90, -1], [0,90], [0, pregnancy end], [pregnancy end, pregnancy end + 90 days]
- Large-scale characterisation:
 - Conditions and medications (assessed at ingredient-level)
 - Assessment window, in days: [-90, -1], [0,90] (for medications only), [0, pregnancy end], [pregnancy end, pregnancy end + 90 days].

A summary of the covariates that were used for characterisation, along with time windows for assessment, can be found in [Table 2](#). Prior and future observation were assessed to provide information on the amount of data visibility before and after index date. This is included in the demographics section as, in some cases, it reflects the time period of enrolment with a given healthcare system. Codes to identify health visits are described in [Annex III](#).

Regarding smoking, it should be noted that this variable has only recently been incorporated into the PET in some data sources, despite not being part of its original design.[4] In these data sources, smoking was directly derived from the PET, where it reflects smoking status as reported at the first pregnancy visit. In other data sources, smoking was captured using records spanning the 2 years prior to pregnancy start up to 90 days after (see [Section 8.10](#)). If individuals had multiple records of smoking status within the assessment window, the record closest to the index date was used. Codes used to assess this variable varied based on how each data source recorded smoking status (see [Annex III](#)). Codes indicating smoking status were categorised into smoker or non-smoker, and therefore, the absence of a record was not interpreted as indicating non-smoking. In SIDIAP, information on former smokers was also available and was described as a separate category.

All baseline comorbidities, pregnancy complications (except pre-term labour and maternal death), and vaccinations studied are not included as pre-defined variables in the PET and were identified using concept sets. Concept sets used for the identification of these covariates are described in [Annex III](#). Some of these definitions are built upon previous work developed for an ongoing pregnancy study on vaccine safety using NLHR, SIDIAP, HI-SPEED, and CPRD GOLD.[6] These codes were refined during the study execution following the DARWIN EU® phenotyping standard processes, which involve the review of code lists by clinical experts, and the review of phenotypes after their execution in the participating data sources.

Pre-term labour and maternal death did not require concept sets and were defined based on the time of recording of events. Pre-term labour was defined as a live birth occurring before 37 completed weeks of gestation. Maternal death was defined as a recorded death occurring during pregnancy or within 90 days following the end of pregnancy.

Table 2. Covariates included in the characterisation of pregnant individuals.

Characteristic	Details	Type of variable	Assessment window ¹	Code type	Diagnosis position
Demographics	Age	Numerical	0	N/A	N/A
	Age groups (12–24; 25–34; 35–44; 45–55 years)	Categorical	0	N/A	N/A
	Prior and future observation (i.e. data visibility)	Numerical	[-Inf,0]; [0, Inf]	N/A	N/A
Health-related behaviours	Smoking ²	Categorical	As recorded in the PET (if available) or [-730, 90]	As recorded in the PET (if available) or SNOMED	N/A
	Pre-pregnancy BMI	Numerical	N/A	As recorded in the PET	N/A
	Health visits	Numerical	[-365, -1], [0, PED]	SNOMED	N/A
	Vaccination (COVID-19, Influenza, Tdap)	Binary	[0, PED]	RxNorm	N/A
	Folic acid ³	Binary	[0, PED]	As recorded in the PET	N/A
Baseline comorbidities	Pre-specified comorbidities ⁴	Binary	[-Inf, -1]	SNOMED	Any
Concomitant medications	Medications before and during pregnancy (assessed using the ATC 3 rd level classification system)	Binary	[-90, -1], [0, 90], [0, PED], [PED + 90]	RxNorm	N/A
Large-scale characterisation	Conditions	Binary	[-90, -1], [0, PED], [PED, PED + 90]	SNOMED	Any
	Medications (assessed at ingredient-level)	Binary	[-90, -1], [0, 90], [0, PED], [PED, PED + 90]	RxNorm	N/A
Pregnancy complications	Pre-specified pregnancy complications ⁵	Binary	[0, PED], [PED, PED + 90]	SNOMED	Any
Pregnancy outcomes	Pre-specified outcomes ⁶	Categorical	N/A	As recorded in the PET.	N/A

ATC=Anatomical Therapeutic Chemical, BMI=Body Mass Index, COVID-19=coronavirus disease 2019; Inf=All available data, N/A=Not applicable, PED=Pregnancy end date, PET=Pregnancy Expansion Table, RxNorm=Medical prescription normalised, SNOMED=Systemised Nomenclature of Medicine; Tdap=Tetanus, Diphtheria, & Pertussis.

¹ Relative to index date, in days.

² NLHR and HI-SPEED have added smoking into the PET. In these data sources, this variable reflects smoking status as reported at the first pregnancy visit. For other data sources, this information was obtained from observation/measurements records, depending on data availability. The specific categories used to define smoking status were determined during the study implementation and included smokers and non-smokers. In SIDIAP, information on former smokers was available and classified as a separate category.

³ If not available in the PET, folic acid was obtained assessing drug records containing folic acid (RxNorm codes).

⁴ Based on concept sets. These included: anxiety, alcohol misuse, asthma, cancer (any malignancy), chronic kidney disease, chronic liver disease, diabetes mellitus, epilepsy/seizures, essential hypertension, heart disease (excluding hypertension), human immunodeficiency virus infection, inflammatory auto-immune diseases of the intestine, mood disorders, obesity, thyroid disorders.

⁵ Based on concept sets. These included: antepartum haemorrhage, eclampsia, ectopic pregnancy, gestational diabetes, HELLP syndrome, maternal death, pre-eclampsia, preterm labour, postpartum haemorrhage, TORCH (toxoplasmosis, others, rubella, cytomegalovirus, herpes simplex) infections. Pre-term labour was defined as a live birth occurring before 37 completed weeks of gestation. Maternal death was defined as a recorded death occurring during pregnancy or within 90 days following the end of pregnancy.

⁶ Livebirth, miscarriage, stillbirth, and elective termination of pregnancy, discordant, unknown.

Objective 4d:

Covariates for infants with mother-child linkage included ([Table 3](#)):

- Sex
- Age at end of follow up
- Future observation (i.e. data visibility before and after index date)
- Gestational age at birth
- Birth weight
- APGAR score (obtained 5 minutes after birth)
- Small for gestational age
- Breastfeeding (where available)
- Congenital malformations
- Mortality

Among the selected variables for characterisation, birth weight, APGAR score, small for gestational age, and congenital malformations are optional fields of the infant table.^[4] Where available, we used information recorded in the infant table. For data sources that did not have this table populated, we derived this information from other tables in the OMOP CDM (e.g. conditions, measurements, observations), where available. Certain variables, not included in infant tables, along with their corresponding assessment windows, were defined during the study implementation based on diagnostic results (see [Section 8.10.](#)).

Values for birth weight and APGAR score were restricted to exclude implausible values, with ranges established as 0–6,000g for birth weight and 0–10 for APGAR score. Small for gestational age was assessed using concepts indicating this trait. In the case of NLHR, this was assessed using birth z-score as a proxy, with values < -1.28 (10th percentile) considered indicative of small for gestational age (see [Section 8.10.](#)).^[14]

Among selected data sources, breastfeeding information was only available in SIDIAP, where this variable is recorded in routine paediatric assessments during the first months of life. In this study, breastfeeding status was assessed using records indicating breastfeeding at follow-up visits at month 1, 6, 12 after birth. Breastfeeding was assessed using two definitions: any breastfeeding (including exclusive and partial breastfeeding) and exclusive only. Mortality was defined as a recorded death occurring 0 to 30 and 0 to 365 days relative to index date.

Codes to define covariates of interest are described in [Annex III](#).

Table 3. Covariates included in the characterisation of infants with mother-child linkage.

Characteristic	Details	Type of variable	Assessment window ¹	Code type	Diagnosis position
Infant characteristics	Sex	Binary	[0,0]	N/A	N/A
	Age	Numerical	Inf (end of follow up or available data)	N/A	N/A
	Future observation	Numerical	[0, Inf]	N/A	N/A
	Birth weight	Numerical	[0,0]	As recorded in the infant table, or as recorded in the data ²	N/A
	APGAR score (5 minutes)	Numerical	[0,0]	As recorded in the infant table, or as recorded in the data ²	N/A
	Small for gestational age	Categorical	[0,0]	As recorded in the infant table, or as recorded in the data ²	N/A
	Breastfeeding	Categorical	Follow-up visits at month 1, 6, 12.	As recorded in the data ³	N/A
	Congenital malformations ⁴	Categorical	[0,0]	As recorded in the infant table, or as recorded in the data ^{2,4}	N/A
	Mortality	Categorical	[0,30], [0,365]	As recorded in the data ⁵	N/A

APGAR=Appearance, Pulse, Grimace, Activity, and Respiration; N/A=Not applicable; Inf=All available data; PED=Pregnancy end date; PET=Pregnancy Expansion Table; RxNorm=Medical prescription normalised; SNOMED=Systemised Nomenclature of Medicine.

¹ Relative to index date, in days if not stated otherwise.

² Obtained from the infant Table. If not available, this information was derived from other records, depending on data availability.

³ Breastfeeding (only available in SIDIAP) was derived from observation records. The specific categories and time windows used to define breastfeeding were determined during the study implementation. Two definitions were used: any breastfeeding (exclusive or partial) and exclusive breastfeeding only.

⁴ For data sources where congenital malformations were not recorded in the infant table, this information was derived from other records using the time window [0,365] for assessment.

⁵ Mortality was defined as a recorded death occurring 0 to 30 and 0 to 365 days relative to index date.

8.7. Study size

No sample size was calculated, as this was a characterisation (descriptive) study which did not test a specific hypothesis.

8.8. Data transformation

Analyses were conducted separately for each data source. Before study initiation, test runs of the analyses were performed on a subset of the data sources and quality control checks were performed. Once all the tests passed (see [Annex IV](#)), the final study codes package was released in the version-controlled Study Repository for execution against all the participating data sources.

The data partners locally executed the analytics against the OMOP CDM in R Studio and reviewed and approved the, by default, aggregated results.

The study results of all data sources were checked, after which they were made available to the team, and the dissemination phase started. All results were locked and timestamped for reproducibility and transparency

8.9. Statistical methods

8.9.1. Main summary measures

Information was summarised using counts and percentages for categorical variables, and medians with interquartile ranges for continuous variables.

8.9.2. Main statistical methods

All analyses were performed using OMOP data. The R packages *CohortCharacteristics* and *PatientProfiles* were used to characterise pregnancy episodes and individuals of interest.[15, 16] The R package *CodelistGenerator* was used to generate ATC-based code lists for medication use, and *IncidencePrevalence* was used to obtain counts of covariates of interest over time.[17, 18] All packages mentioned have been developed by DARWIN EU®.

The list of pre-specified covariates that were described, along with the time windows for assessment, is specified in [Section 8.6.3](#). Characteristics were summarised using counts and percentages for categorical variables, and medians with interquartile ranges for continuous variables.

In Objective 4a, we described covariates of interest for all pregnancy episodes included. Regarding mother-child linkage, we described the number of pregnancy episodes that could be linked to an infant. For data sources where this information was available in the “fact_relationship” table, the link is recorded between the mother and the infant, rather than between the pregnancy episode and the infant. The pregnancy episode for each infant was identified using the date of birth, where possible.

For Objective 4b, the number of included pregnancies, overall and by pregnancy outcomes, mode of delivery were reported by calendar year. The number of pregnancy episodes (or number of infants, if not possible) with mother-child linkage were presented by calendar year. Covariates were presented as counts and proportions (%) relative to the total number of pregnancies per year.

The analysis of Objective 4c and 4d focused on characterising individuals with pregnancy episodes and infants with mother-child linkage, respectively.

Cell counts <5 were suppressed when reporting results to comply with the data source’s privacy protection regulations.

8.9.3. Missing values

Analyses for all objectives were conducted using available data without imputation. As such, no specific methods were applied to address missingness. For covariates based on conditions, medications/vaccines, or healthcare visits, the absence of a recorded entry was interpreted as the individual not having the event of interest (e.g. condition). For other variables where a value was expected, the number of missing observations were reported. These variables included all covariates included in the PET (i.e. covariates of objective 4a and pre-pregnancy BMI), smoking status, and infant characteristics (e.g. birth weight, small for gestational age, APGAR score).

8.10. Deviations from the protocol

Deviation number ¹	Protocol version	Date	Section of study protocol	Deviation	Reason
1	2.0	13/01/2026	8.2. Follow up	We clarified the definition of index date for infants with mother-child linkage (objective 4d). In the approved protocol, this was defined as the “recorded date of birth.” In this report, we have specified that this refers to the date of birth as documented in the maternal records. This resulted in some execution issues in SIDIAP, where we had to add an additional filter to exclude infants whose observation start did not fall within 30 days before or after the birth date recorded in maternal records (see Section 9.1.).	The full date of birth recorded in the infants’ records was not available across data sources due to privacy restrictions. To improve the assessment of infant variables, index date was instead defined based on date of birth recorded in the maternal records.
2	2.0	13/01/2026	8.3. Study population	We applied additional exclusion criteria to define the study population, all of which have been clarified in Section 8.3. For pregnancy episodes, additional filters were applied to exclude pregnancies with conflicting dates (i.e. with overlapping dates) and pregnancies not observed at the start or end of the pregnancy. Pregnancies with an end date earlier than the start date were also added to the exclusion criteria. For infants with mother-child linkage (objective 4d), the study population was restricted to those born from pregnancy episodes identified in objectives 4a, 4b, and 4c, which resulted in live births.	Additional filters were applied to improve the data quality and completeness of the pregnancy episodes studied. For infants, we chose to use the same pregnancies that were previously filtered as part of the other objectives to ensure consistency.
3	2.0	20/02/2026	8.4. Study setting and data sources	IMASIS, SIDIAP, and CPRD GOLD updated data releases after the study protocol was approved. The calendar period covered for those data sources was updated accordingly in Table 1.	New data releases became available during the conduction of the study.
4	–	–	8.6.3. Covariates	Further details are provided to explain how smoking was ascertained (objective 4c). We modified the time window for assessment in data sources where this variable was not available as an additional variable of the PET (i.e. CPRD GOLD, SIDIAP, IMASIS). This time window was defined as [-730,90] days, deviating from the time window suggested in the original protocol ([-730,0] days).	Additional information became available during the study implementation phase. The time window was expanded 90 days after index date to better align with the definition used NLHR and HI-SPEED, where this information is available in the PET and corresponds to the smoking status as reported at the first pregnancy visit.

Deviation number ¹	Protocol version	Date	Section of study protocol	Deviation	Reason
5	–	–	8.6.3. Covariates	Additional information is provided to explain how infant variables were assessed (objective 4d). These included breastfeeding, small for gestational age, and congenital malformations. For the latter, we modified the window for assessment in data sources with mother-child linkage where the infant table was not available (i.e. NLHR, SIDIAP). This was defined as one year after birth (i.e. [0,365] days), deviating from the time window suggested in the original protocol (i.e. [0,0]). Values for birth weight and APGAR score to exclude implausible values, with ranges established as 0–6,000g for birth weight and 0–10 for APGAR score. Small for gestational age was not recorded as a condition/observation in NLHR and assessed using z-scores based on data availability.	Additional information became available during the study implementation phase. Values for birth weight and APGAR score were restricted anticipating the possibility of capturing implausible values.

APGAR=Appearance, Pulse, Grimace, Activity, and Respiration; CPRD GOLD=Clinical Practice Research Datalink GOLD; HI-SPEED=Health Impact - Swedish Population Evidence Enabling Data-linkage; IMASIS=Institut Municipal Assistència Sanitària Information System; NLHR=Norwegian Linked Health Registry data; PET=Perinatal Expansion Table; SIDIAP=The Information System for Research in Primary Care .

¹In accordance with the original protocol (V2.0), it was outlined that certain variables, not included in the PET or infant tables, along with their corresponding assessment windows, would be defined during the study implementation. As such, these do not constitute deviations. However, these are listed here for transparency.

9. RESULTS

The full set of results for this study is available through an interactive web-application ShinyApp [EUPAS1000000812 Objective 4](#).

9.1. Participants

9.1.1. Pregnancy episodes

We identified a total of 2,146,567 pregnancy episodes eligible for inclusion in the study (n= 367,803 in NLHR; n=15,330 in IMASIS; n= 409,207 in SIDIAP; n=887,612 in HI-SPEED; n=466,615 in CPRD GOLD) from 2010 to one year before the end of data availability in each data source. In HI-SPEED and NLHR, the study period started in 2015 and 2018, respectively.

Table 4 includes the study attrition of pregnancy episodes by data source, with a more detailed breakdown available in the Shiny app. Numerous pregnancies were excluded due to lack of observation at pregnancy start or end and due to the study period restriction. In SIDIAP, the leading cause for exclusion was implausible durations (n=137,717). In CPRD GOLD, the only criterion that resulted in exclusion of >5 episodes was the study period restriction.

For Objectives 4c and 4d, the number of pregnancy episodes included in HI-SPEED was further restricted to those commencing in 2018 or later (see [Section 8.5](#).), resulting in a total of 553,300 pregnancies in 445,288 individuals.

Table 4. Study attrition of pregnancy episodes, by data source.

	Records (N)	Subjects (N)	Excluded records (N)	Excluded subjects (N)
NLHR				
Initial qualifying events ¹	1,170,297	594,192	–	–
In observation at pregnancy start and end	1,085,627	557,214	84,670	36,978
Pregnancies with plausible duration ²	1,085,150	557,165	477	39
Pregnancies with conflicting dates	1,071,910	556,462	13,240	703
Age 12 to 55	1,071,833	556,395	77	67
Female	1,071,804	556,372	29	23
Pregnancy start between 01/01/2018 and 1 year before end of data availability ³	367,803	257,402	704,001	298,970
IMASIS				
Initial qualifying events ¹	23,547	18,254	–	–
In observation at pregnancy start and end	18,414	14,410	5,133	3,844
Pregnancies with plausible durations ²	18,396	14,400	18	10
Pregnancies with conflicting dates	18,396	14,400	0	0
Age 12 to 55	18,394	14,398	<5	<5
Female	18,389	14,393	5	5
Pregnancy start between 01/01/2010 and 1 year before end of data availability ³	15,330	12,214	3,059	2,179
SIDIAP				
Initial qualifying events ¹	757,202	490,244	–	–
In observation at pregnancy start and end	673,616	438,991	83,586	51,253
Pregnancies with plausible durations ²	535,899	370,967	137,717	68,024
Pregnancies with conflicting dates	533,272	370,465	2,627	502
Age 12 to 55	533,241	370,436	31	29
Female	533,200	370,405	41	31
Pregnancy start between 01/01/2010 and 1 year before end of data availability ³	409,207	294,381	123,993	76,024
HI-SPEED				
Initial qualifying events ¹	990,989	679,914	–	–
In observation at pregnancy start and end	887,658	626,592	103,331	53,322
Pregnancies with plausible durations ²	887,640	626,586	18	6
Pregnancies with conflicting dates	887,636	626,586	<5	0
Age 12 to 55	887,631	626,582	5	<5
Female	887,612	626,564	19	18
Pregnancy start between 01/01/2015 and 1 year before end of data availability ³	887,612	626,564	0	0
CPRD GOLD				

	Records (N)	Subjects (N)	Excluded records (N)	Excluded subjects (N)
Initial qualifying events ¹	1,181,917	782,073	–	–
In observation at pregnancy start and end	1,181,917	782,073	0	0
Pregnancies with plausible durations ²	1,181,917	782,073	0	0
Pregnancies with conflicting dates	1,181,917	782,073	0	0
Age 12 to 55	1,181,916	782,072	<5	<5
Female	1,181,916	782,072	0	0
Pregnancy start between 01/01/2010 and 1 year before end of data availability ³	466,615	352,508	715,301	429,564

CPRD GOLD=Clinical Practice Research Datalink GOLD; HI-SPEED=Health Impact - Swedish Population Evidence Enabling Data-linkage; IMASIS=Institut Municipal Assistència Sanitària Information System; NLHR=Norwegian Linked Health Registry data; SIDIAP=The Information System for Research in Primary Care .

¹ Pregnancy episodes recorded in the Perinatal Expansion Table (PET).

² Based on pregnancy start and end date recorded in the PET.

³ End of data availability differed across data sources. NLHR: 31/12/2023; IMASIS: 14/06/2025; SIDIAP: 31/12/2024; HI-SPEED: 30/08/2024; CPRD GOLD: 20/05/2025.

9.1.2. Infants with mother-child linkage

We identified a total of 1,482,391 infants eligible for inclusion (n=257,826 in NLHR; n=15,356 in IMASIS; n=312,291 in SIDIAP; n=896,918 in HI-SPEED). Infants were included if they could be linked to a pregnancy episode identified as part of previous objectives ([Table 4](#)), provided that the pregnancy episode resulted in a live birth ([Table 5](#)).

In IMASIS and HI-SPEED, linkage between infants and pregnancy episodes was available directly in the infant table. In NLHR and SIDIAP, mother-child linkage was provided through the fact_relationship table, which records parental relationships (e.g. linking mothers to their children). In these data sources, linkage between the infant and the pregnancy episode was not available and was inferred by restricting the mother's pregnancy episodes to those occurring in the year of the infant's birth. CPRD GOLD did not have mother-child linkage (neither in the fact_relationship or infant table) and did not participate in analyses involving this linkage (see [Section 10.2](#)).

The number of infants was slightly higher than the number of pregnancy episodes included in IMASIS (15,356 infants vs. 15,330 pregnancy episodes) and HI-SPEED (896,918 infants vs. 887,612 pregnancy episodes), and it was lower in NLHR (257,826 infants vs. 367,803 pregnancy episodes) and SIDIAP (312,291 infants vs. 409,207 pregnancy episodes).

In the context of Objective 4d, restricting the study period in HI-SPEED from 2018 onwards resulted in 692,802 infants being included from this data source.

Table 5. Study attrition of infants with mother-child linkage, by data source.

	Records (N)	Subjects (N)	Excluded records (N)	Excluded subjects (N)
NLHR				
Initial qualifying events ¹	905,424	905,424	–	–
Infants whose pregnancy episode was included in the pregnancy episode cohort ¹	260,873	263,190	644,551	642,234
Filter out duplicate infants	258,579	258,579	2,294	2,294
Filter to live births	257,826	257,826	753	753

	Records (N)	Subjects (N)	Excluded records (N)	Excluded subjects (N)
IMASIS				
Initial qualifying events ²	23,173	23,173	–	–
Infants whose pregnancy episode was included in the pregnancy episode cohort	15,367	15,367	7,806	7,806
Filter out duplicate infants	15,367	15,367	0	0
Filter to live births	15,356	15,356	11	11
SIDIAP				
Initial qualifying events ¹	796,260	802,232	–	–
Infants whose pregnancy episode was included in the pregnancy episode cohort ¹	326,585	327,086	469,6675	475,146
Filter to observation start within 30 days before or after cohort start	312,535	312,555	14,050	14,531
Filter out duplicate infants	312,515	312,515	20	40
Filter to live births	312,291	312,291	224	224
HI-SPEED				
Initial qualifying events ²	1,004,656	1,004,656	–	–
Infants whose pregnancy episode was included in the pregnancy episode cohort	899,909	899,909	104,747	104,747
Filter out duplicate infants	899,909	899,909	0	0
Filter to live births	896,918	896,918	2,991	2,991

CPRD GOLD=Clinical Practice Research Datalink GOLD; HI-SPEED=Health Impact - Swedish Population Evidence Enabling Data-linkage; IMASIS=Institut Municipal Assistència Sanitària Information System; NLHR=Norwegian Linked Health Registry data; SIDIAP=The Information System for Research in Primary Care .

¹In NLHR and SIDIAP, infants were identified based on parental relationships recorded in the fact_relationship table. Linkage between the infant and the pregnancy episode was not available and was inferred by selecting infants whose birthing parent was included in the pregnancy episode cohort, with the episode ending in the same calendar year as the infant's birth.

²In HI-SPEED and IMASIS, infants were identified from the infant table from the Perinatal Expansion, which links each infant to a specific pregnancy episode.

³Applied to exclude infants whose observation start date fell outside the 30-day window before or after the birth date recorded in maternal records. This criterion was not applied to the remaining data sources, as no inconsistencies were identified.

9.2. Main results

9.2.1. Characterisation of pregnancy episodes, overall and by year (Objectives 4a & 4b)

Pregnancies identified corresponded to a total of 1,543,069 individuals (n=257,402 in NLHR; n=12,214 in IMASIS; n=294,381 in SIDIAP; n=626,564 in HI-SPEED; n=352,508 in CPRD GOLD), resulting in approximately 1.4 pregnancies per individual. Median gestational length ranged between 273 to 281 days across data sources. The median was estimated directly from the gestational length recorded in the PET ("gestational_length_in_day" field), and from calculations based on pregnancy start and end dates recorded in the PET. Values for both approaches were consistent, except for SIDIAP, where differences were observed in the maximum values (Table 6). This discrepancy is likely attributed to the exclusion criteria of plausible duration (i.e. pregnancies surpassing 44 weeks of gestation, equivalent to ≥308 days, were excluded), which was based on recorded pregnancy start and end date rather than on the actual gestational length recorded in the PET. Minimum values ranged from 7 to 29 in all data sources except for HI-SPEED, where this value was 154 days (Table 6).

Pregnancy outcomes captured differed across data sources, with livebirths and stillbirths identified in all data sources ([Table 6](#)). The proportion of identified pregnancies that resulted in live births was 69.8% NLHR, 99.6% in IMASIS, 90.4% in SIDIAP, 99.7% in HI-SPEED, and 71.3% in CPRD GOLD. The number of pregnancies with an outcome of stillbirths was similar across data sources and ranged from 0.2% in NLHR to 0.6% in SIDIAP. Miscarriages and elective terminations of pregnancy were observed in NLHR, SIDIAP, and CPRD GOLD, with figures for miscarriages estimated at 14.4%, 8.3%, and 27.8%, respectively. Corresponding figures for elective terminations of pregnancy were 14.3%, 0.5%, and 0.7%. Ectopic and molar pregnancies were only identified as outcomes in the PET in NLHR (<1.1%). Outcomes occurring in early stage of gestation were not recorded in the PET in IMASIS (which includes information on pregnancies that resulted in birth) and HI-SPEED (which includes information on pregnancies surpassing week 22).

Mode of delivery, parity, and plurality were not available in the PET in CPRD GOLD. Vaginal delivery ranged from 66.1% in NLHR to 90.2% in HI-SPEED ([Table 6](#)). Mode of delivery was missing in approximately 29.3% of pregnancies identified in NLHR, which aligned with the percentage of pregnancies that did not result in live births or stillbirths in this data source (30%). Considering only livebirths and stillbirths (n=257,592), the proportion of vaginal deliveries in NLHR was 94.3%. C-sections were recorded in 4.6% (NLHR) to 22.6% (IMASIS) of pregnancies. The proportion of multiple pregnancies was approximately 1.0% to 1.5% in NLHR, IMASIS, and HI-SPEED. Multiple pregnancies were not identified in SIDIAP, consistent with the absence of such records in the source data since 2010. The proportion of episodes that corresponded to nulliparous individuals ranged between 42.7% in HI-SPEED to 69.2% in SIDIAP, with median gravidity (available in all data sources except for HI-SPEED) estimated at 0 or 1 across data sources ([Table 6](#)).

Table 6. Descriptive statistics of pregnancy episodes, as recorded in the PET.

	NLHR	IMASIS	SIDIAP	HI-SPEED	CPRD GOLD
Number of pregnancies, N ¹	367,803	15,330	409,207	887,612	466,615
Number of pregnant individuals, N	257,402	12,214	294,381	626,564	352,508
Gestational length recorded (in days), median [min; q25–q75; max] ²	274 [15;80–283;306]	273 [7;266–280;301]	273 [7;266–280;644]	280 [154;273–286;307]	281 [29;97–282;302]
Gestational length computed (in days), median [min; q25–q75; max] ³	274 [15;80 - 283;306]	273 [7;266 - 280;301]	277 [7;268 - 284;307]	280 [154;273 - 286;307]	280 [28;96 - 281;301]
Pregnancy outcome, N (%) ^{4,5}					
- Livebirth	256,773 (69.8%)	15,275 (99.6%)	369,803 (90.4%)	884,807 (99.7%)	332,606 (71.3%)
- Miscarriage	53,141 (14.4%)	–	34,147 (8.3%)	–	129,464 (27.8%)
- Stillbirth	819 (0.2%)	44 (0.3%)	2,242 (0.6%)	2,651 (0.3%)	1,200 (0.3%)
- Elective termination	52,782 (14.3%)	–	2,125 (0.5%)	–	3,345 (0.7%)
- Ectopic	3,910 (1.1%)	–	–	–	–
- Molar	378 (0.1%)	–	–	–	–
- Discordant	–	11 (0.1%)	–	154 (0.02%)	–

	NLHR	IMASIS	SIDIAP	HI-SPEED	CPRD GOLD
- Unknown	–	–	890 (0.2%)	–	–
Mode of delivery, N (%)					
- Vaginal	243,110 (66.1%)	11,873 (77.4%)	329,838 (80.6%)	800,482 (90.2%)	–
- C-section	16,780 (4.6%)	3,457 (22.6%)	79,369 (19.4%)	87,130 (9.8%)	–
- Unknown	107,913 (29.3%)	–	–	–	466,615 (100%)
Plurality, N (%)					
- Singleton	256,122 (69.6%)	15,093 (98.5%)	409,207 (100%)	875,407 (98.6%)	–
- Multiple	3,768 (1.0%)	236 (1.5%)	–	12,205 (1.4%)	–
- Unknown	107,913 (29.3%)	<5	–	–	466,615 (100%)
Parity, N (%)					
- Nulliparous	149,868 (57.7%)	8,686 (56.7%)	281,483 (69.2%)	378,913 (42.7%)	–
- Multiparous	110,022 (42.3%)	6,644 (43.3%)	125,137 (30.8%)	508,699 (57.3%)	–
- Unknown	–	–	–	–	466,615 (100%)
Gravidity, median [min; q25–q75; max]	1 [0; 0–2; 11]	0 [0; 0–1; 19]	1 [0; 1–2; 27]	–	0 [0 ;0–1; 15]

CPRD GOLD=Clinical Practice Research Datalink GOLD; HI-SPEED=Health Impact - Swedish Population Evidence Enabling Data-linkage; IMASIS=Institut Municipal Assistència Sanitària Information System; NLHR=Norwegian Linked Health Registry data; PET=Perinatal Expansion Table; SIDIAP=The Information System for Research in Primary Care .

“–” Indicates that values for this concept or variable were not found in the PET.

¹ The enrolment period for pregnancies spanned from 01/01/2010 to one year prior to end of data availability. Study period started in 2015 in HI-SPEED (Objectives 4a & 4b) and 2018 in NLHR.

²As recorded in the PET (“gestational_length_in_day” field).

³Calculated based on pregnancy start and end date recorded in the PET.

⁴HI-SPEED includes information on pregnancies that surpassed week 22. IMASIS only includes pregnancies that resulted in birth.

⁵Outcomes of the pregnancy based on the PET implementation included: livebirth, miscarriage, stillbirth (≥ 20 weeks), elective termination of pregnancy, discordant (different outcomes in multiple pregnancies), unknown. Ectopic and molar pregnancies were not part of the original design of the PET but were recorded as outcomes in NLHR.

The number of pregnancy episodes (based on pregnancy start) and infants with mother-child linkage (based on infant observation start) included over time are displayed in [Figure 2](#). Trends should be interpreted considering changes in the active population of the data sources, particularly in the case of CPRD GOLD (see [Section 10.3.1.](#)), and the definition of the study population, as pregnancies starting in the last year of available data were not included (see [Section 9.1.](#)). This also affects infants, as only those who could be linked to an included pregnancy episode were considered. Consequently, the number of pregnancies observed in the last year of available data were lower than those observed for other years. This was particularly pronounced for HI-SPEED ([Figure 2](#)), where only pregnancies that surpassed week 22 of gestation are included. A similar pattern was observed for infants in the first year of available data, as infants were assigned to calendar year based on year of observation start. Considering a median gestational length of approximately 280 days, most pregnancies starting in the first year of the study period would not result in births within the same calendar year. Indeed, only pregnancies starting in the first quarter of the year would be expected to result in the birth within the same calendar year. This timing constraint explains the lower number of infants with mother-child linkage observed in the first year of the study period (2010 for IMASIS and SIDIAP, 2018 for NLHR, 2015 for HI-SPEED), which primarily reflects infants conceived in the first trimester of the year ([Figure 2](#)).

Without considering the first and last year of data, the number of pregnancy episodes starting per year was generally stable in NLHR, IMASIS, and HI-SPEED, with counts ranging approximately 71,000–75,500 in NLHR, 940–1,250 in IMASIS, and 100,000–112,000 in HI-SPEED. In SIDIAP and CPRD GOLD, the number of pregnancies decreased substantially over time. The number of pregnancies in 2023 was 26,674 in SIDIAP (vs. 35,393 in 2010) and 11,821 in CPRD GOLD (vs. 61,281). The number of infants with mother-child linkage mirrored trends observed for pregnancies in IMASIS, SIDIAP, and HI-SPEED, with some discrepancies in the first and last year of available data. Without considering those, the number of infants with mother-child linkage per year was approximately 39,000–54,300 in NLHR, 930–1,240 in IMASIS, 19,300–26,350 in SIDIAP, and 102,600–112,660 in HI-SPEED.

Trends by pregnancy outcome were generally stable over time, with minor fluctuations observed around 2020 ([Figure 3](#)). The proportion of pregnancies that resulted in miscarriage increased after 2020 in SIDIAP (7–8% before vs. 10–11% after), while it decreased in CPRD GOLD (27–29% before vs. 22–24% after). Results by mode of delivery also showed stable trends over time ([Figure S2](#)). In IMASIS, some minor fluctuations were observed, with proportion of C-sections decreasing to approximately 18% in 2018 and 2019 (vs. 24–25% in 2016 and 2021).

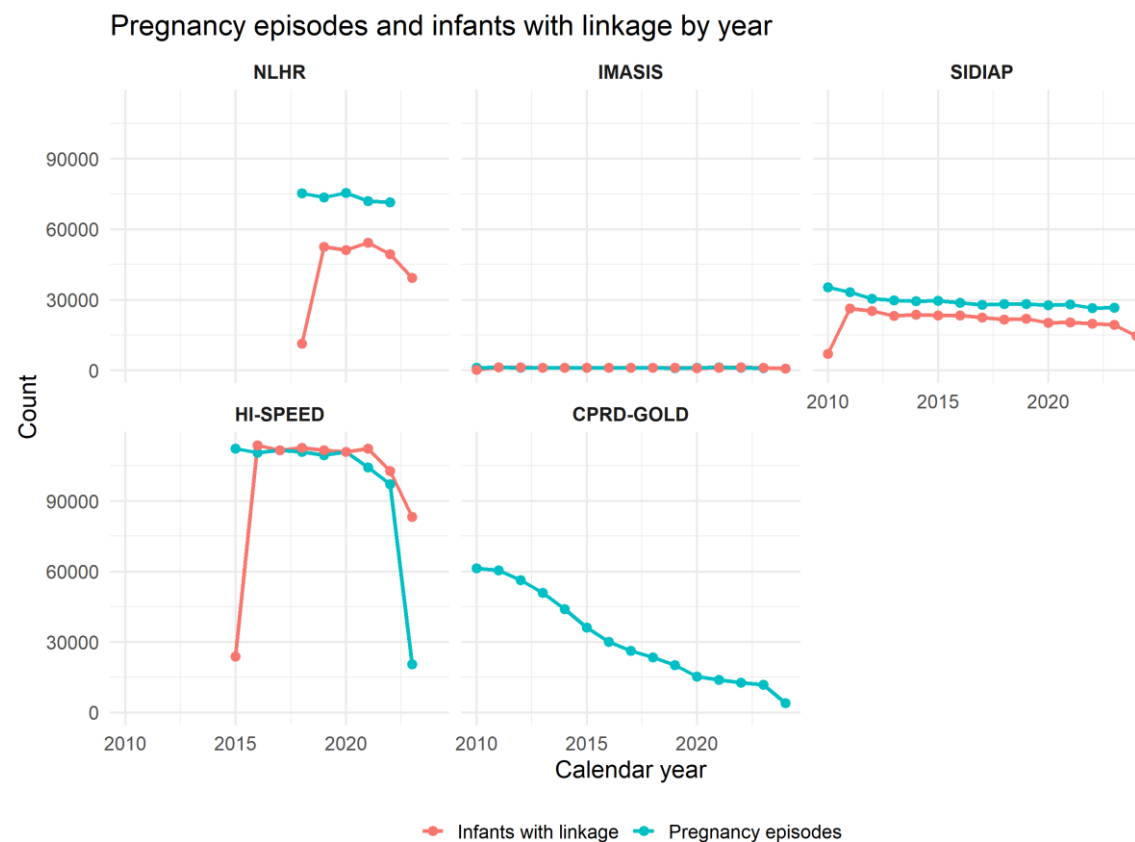


Figure 2. Number of pregnancy episodes by calendar year of pregnancy start (in blue), and number of infants with mother-child linkage by year of infant observation start (in red).

CPRD GOLD=Clinical Practice Research Datalink GOLD; HI-SPEED=Health Impact - Swedish Population Evidence Enabling Data-linkage; IMASIS=Institut Municipal Assistència Sanitària Information System; NLHR=Norwegian Linked Health Registry data; SIDIAP=The Information System for Research in Primary Care .

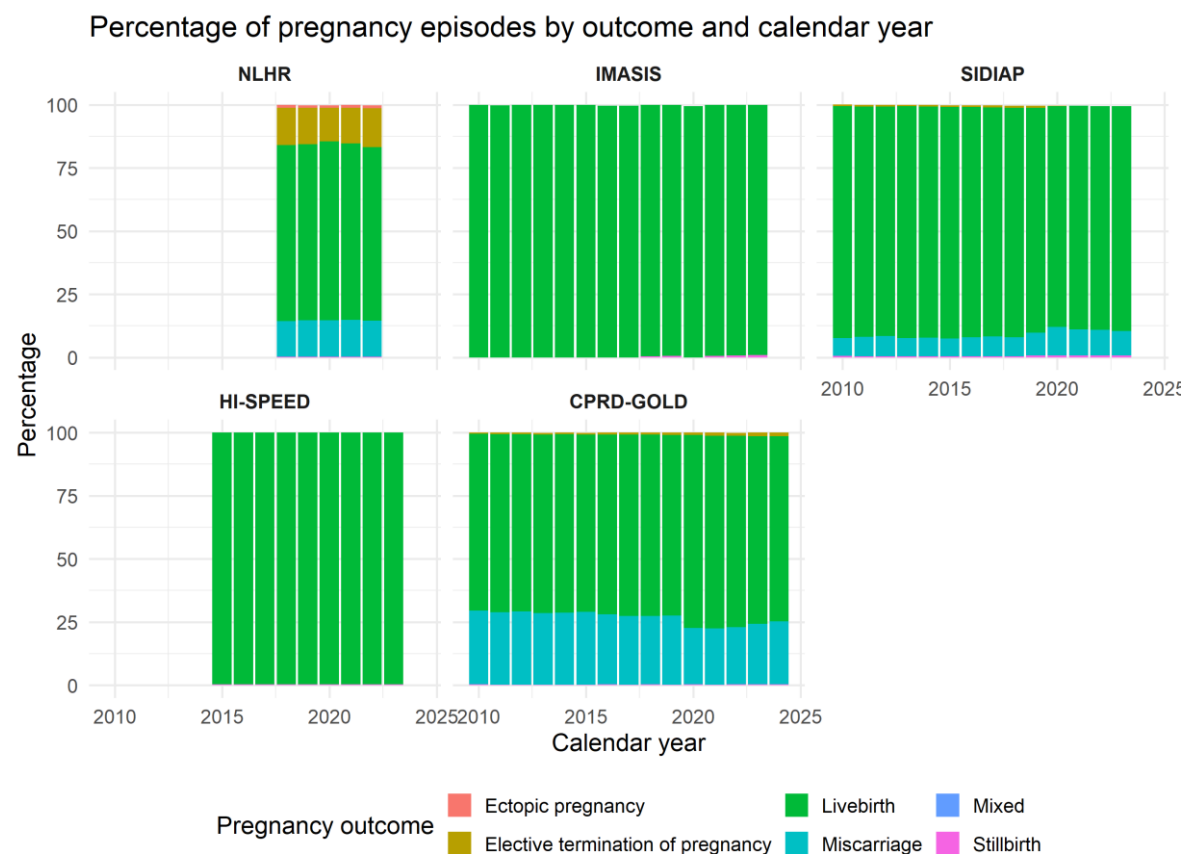


Figure 3. Percentage of pregnancy episodes starting in each calendar year by pregnancy outcome.

CPRD GOLD=Clinical Practice Research Datalink GOLD; HI-SPEED=Health Impact - Swedish Population Evidence Enabling Data-linkage; IMASIS=Institut Municipal Assistència Sanitària Information System; NLHR=Norwegian Linked Health Registry data; SIDIAP=The Information System for Research in Primary Care .

Outcomes assessed as recorded in the PET: livebirth, miscarriage, stillbirth (≥ 20 weeks), elective termination of pregnancy, discordant (i.e. different outcomes in multiple pregnancies; referred as mixed in the plot), unknown (i.e. not shown in the plot; these cases were present in SIDIAP only, and represented 0.2% of pregnancies over the entire study period). Ectopic and molar pregnancies were not part of the original design of the PET but were recorded as outcomes in NLHR. IMASIS only included pregnancies that resulted in birth. HI-SPEED included pregnancies surpassing week 22 of gestation.

9.2.2. Maternal characteristics (Objective 4c)

Demographics, health-related behaviours, and baseline comorbidities

Maternal characteristics by data source are described in [Table 7](#). The number of pregnancy episodes included aligned with those reported in previous sections, except for HI-SPEED, where the study period for Objectives 4c and 4d was restricted to pregnancies starting from 2018 onwards (see [Section 8.5](#)). Median maternal age at pregnancy start was approximately 30 to 31 years, with 53.6% to 69.3% of pregnancies among women aged 25–34 years. The proportion of pregnancy episodes in the 35–55 age group ranged from 19.6% to 23.1% in NLHR, HI-SPEED, and CPRD GOLD, but higher in data sources from Spain (approximately 29%). Median patient-level observation (i.e. data visibility) across sources extended over 5 years prior to the start of pregnancy and more than 3 years afterward.

The recording of smoking status and its assessment differed across data sources (see [Section 8.6.3](#)). Current smoking was observed in 3.8% pregnancy episodes in NLHR, 7.4% in IMASIS, 4.1% in SIDIAP, 9.6% in HI-SPEED, and 25.2% in CPRD GOLD. Records indicating non-smoking ranged from 11.2% in SIDIAP to 85.7% in HI-SPEED. Information on former smokers was only available in SIDIAP and was observed in 4.6% of pregnancies. Pre-pregnancy BMI (as recorded in the PET) was estimated to have a mean value ranging from 24 Kg/m² in HI-SPEED to 25.5 Kg/m² in IMASIS and was not available in CPRD GOLD ([Table 7](#)).

The number of median health visits ranged from 1 to 9 in the year prior to pregnancy start and from 6 to 20 during pregnancy and should be interpreted in light of the health care settings covered by each data source ([Table 1](#)). The lowest number of visits were observed in IMASIS and HI-SPEED, while the highest were observed in NLHR and CPRD GOLD. Vaccination against COVID-19 during pregnancy ranged from 1.7% in IMASIS to 18.3% in HI-SPEED and should be interpreted in light of the study period start (2010), which differed in NLHR and HI-SPEED (2018). For other vaccines (available in all data sources except for HI-SPEED), the proportion of pregnancy episodes exposed ranged from 8.5% in IMASIS to 29.7% in CPRD GOLD against influenza and 0.1% in NLHR to 52.8% in SIDIAP against Tdap. Vaccination records against Tdap in NLHR were consistent with records in the source data (see [Section 10.3.1](#)). Prescription of folic acid during pregnancy ranged from 24.2% to 28.2% in IMASIS, SIDIAP, and CPRD GOLD, and was below 5% in NLHR and HI-SPEED ([Table 7](#)).

Among the included pre-specified baseline comorbidities, anxiety and mood disorders were the most frequently observed, ranging from 2.2% in IMASIS to 33.3% in NLHR and 1.3% in IMASIS to 16.0% in CPRD GOLD, respectively. These were followed by asthma (range: 2.7% to 12.8%), obesity (assessed as a condition; range: 1.9% to 9.5%), thyroid disorders (range: 2.1% to 5.4%), and diabetes (range: 1.6% to 4.4%). Other baseline comorbidities were below 2% across data sources, except for heart disease (excluding hypertension) and essential hypertension in NLHR (5% and 2.2%, respectively) ([Table 8](#)).

Table 7. Demographics and health-related behaviours of pregnant individuals (assessed at pregnancy episode level).

	NLHR	IMASIS	SIDIAP	HI-SPEED	CPRD GOLD
Number of pregnancies, N	367,803	15,330	409,207	553,300	466,615
Number of pregnant individuals, N	257,402	12,214	294,381	445,288	352,508
Cohort start date, range (min, max)	2018-01-01 to 2022-12-30	2010-01-01 to 2023-11-23	2010-01-01 to 2023-12-31	2018-01-01 to 2023-07-24	2010-01-01 to 2024-05-19
Cohort end date, range (min, max)	2018-01-01 to 2022-12-30	2010-01-01 to 2023-11-23	2010-01-01 to 2023-12-31	2018-01-01 to 2023-07-24	2010-01-01 to 2024-05-19
Age (in years), median [min; q25 – q75; max]	30 [12; 27–34; 55]	31 [13;27–35; 51]	31 [12; 27–35; 55]	30 [12; 27–34; 55]	30 [12; 25–34; 55]
Age group, N (%)					
- 12–24 years	48,572 (13.2%)	2,586 (16.9%)	57,173 (14%)	61,383 (11.1%)	103,214 (22.1%)
- 25–34 years	240,946 (65.5%)	8,219 (53.6%)	232,301 (56.8%)	383,581 (69.3%)	255,568 (54.8%)
- 35–55 years	78,285 (21.3%)	4,525 (29.5%)	119,733 (29.3%)	108,336 (19.6%)	107,833 (23.1%)
Prior observation (in days), median [min; q25 – q75; max]	4,305 [0; 3,752–4,881; 5,477]	3,013 [0; 1,124–6,766; 12,323]	2,886 [0; 1,757–4,378; 6,573]	1,890 [0; 1,402– 2,404; 3,126]	2,445 [0; 887–4,574; 12,798]
Future observation or follow-up (in days), median [min; q25 – q75; max]	1,246 [56; 799–1,704; 2,190]	2,380 [17; 1,172.25– 3,702.75; 5,633]	2,854 [49; 1,564– 4,173; 5,478]	1,497 [160; 1,035–1,957; 2,433]	1,567 [33; 796 –2,904; 5,616]
Smoking, N (%) ¹					
- Smoker	13,891 (3.8%)	1,141 (7.4%)	16,644 (4.1%)	53,334 (9.6%)	117,711 (25.2%)
- Non-smoker	229,755 (62.5%)	–	45,637 (11.2%)	474,295 (85.7%)	212,180 (45.5%)
- Former smoker	–	–	18,926 (4.6%)	–	–
- Missing	124,157 (33.8%)	*	328,000 (80.1%)	25,671 (4.6%)	136,726 (29.3%)
Pre-pregnancy BMI (in Kg/m ²) mean (SD) ²	24.9 (5)	25.5 (8.9)	24.2 (6.7)	24 (5.1)	–
Health visits, median [min; q25 – q75; max] ³					
- In the year prior to pregnancy	8 [0; 3 –16;372]	1 [0; 0 –3; 365]	5 [0; 2–10; 267]	2 [0; 0–7; 524]	9 [0; 5 –17; 364]

	NLHR	IMASIS	SIDIAP	HI-SPEED	CPRD GOLD
- During pregnancy	20 [0; 9–29;261]	6 [0; 3–10; 273]	15 [0;9 –22; 244]	7 [0; 3–18; 721]	13 [1; 7–20; 278]
Vaccinations during pregnancy, N (%) ⁴					
- COVID-19	40,958 (11.1%)	267 (1.7%)	15,890 (3.9%)	101,262 (18.3%)	10,792 (2.3%)
- Tdap	207 (0.1%)	216 (1.4%)	215,916 (52.8%)	–	168,047 (36%)
- Influenza	82,468 (22.4%)	1,300 (8.5%)	83,846 (20.5%)	–	138,741 (29.7%)
Prescribed folic acid, N (%)	10,982 (3.0%)	4,327 (28.2%)	101,378 (24.8%)	25,176 (4.6%)	113,089 (24.2%)

BMI=Body Mass Index; COVID-19=Coronavirus disease 2019; CPRD GOLD=Clinical Practice Research Datalink GOLD; HI-SPEED=Health Impact - Swedish Population Evidence Enabling Data-linkage; IMASIS=Institut Municipal Assistència Sanitària Information System; NLHR=Norwegian Linked Health Registry data; SD=standard deviation; SIDIAP=The Information System for Research in Primary Care ; Tdap=Tetanus, Diphtheria, & Pertussis.

“–” Indicates not available.

“*” Indicates not ascertained.

¹ The recording of smoking status and its assessment differed across data sources and was based on available data. In NLHR and HI-SPEED, this was recorded in the PET and reflects smoking status as reported at the first pregnancy visit. For other data sources, this was obtained from other observation/measurements records, and included codes related to current and never smoker. Information on former smokers was only available in SIDIAP. Smoking status was assessed as recorded in the data, no algorithm was used to re-code non-smokers with a historical record of “current smoking” status as former smokers. Missing data were interpreted as the absence of a record indicating smoking status. In IMASIS, missing data were not reported, as no codes for non-smokers were identified. Therefore, we cannot determine whether the absence of a record reflects missing data or non-smoking status.

² As recorded in the PET. Not available in SIDIAP and CPRD GOLD.

³ Health visits captured depend on data type and health care settings covered in each data source.

⁴ In HI-SPEED, vaccination data was limited to vaccination against COVID-19. The proportion of all pregnancy episodes with COVID-19 vaccination was influenced by different start dates of follow-up (2010 in all data sources except for NLHR and HI-SPEED, where follow up started in 2018).

Table 8. Baseline comorbidities of pregnant individuals prior to pregnancy start (assessed at pregnancy episode level).

Pre-specified conditions ¹	NLHR (n=367,803)	IMASIS (n=15,440)	SIDIAP (n=409,207)	HI-SPEED (n=553,300)	CPRD GOLD (n=466,615)
Anxiety	122,523 (33.3%)	339 (2.2%)	83,733 (20.5%)	86,065 (15.6%)	45,448 (9.8%)
Alcohol misuse	3,038 (0.8%)	69 (0.5%)	1,945 (0.5%)	2,114 (0.4%)	3,651 (0.8%)
Asthma	47,141 (12.8%)	409 (2.7%)	17,760 (4.3%)	21,078 (3.8%)	41,551 (8.9%)
Cancer (any malignancy)	4,541 (1.2%)	126 (0.8%)	2,716 (0.7%)	3,338 (0.6%)	1,290 (0.3%)
Chronic kidney disease	1,840 (0.5%)	10 (0.1%)	184 (0.04%)	557 (0.1%)	935 (0.2%)
Chronic liver disease	1,702 (0.5%)	62 (0.4%)	1,690 (0.4%)	2,184 (0.4%)	228 (0.1%)
Diabetes mellitus	16,229 (4.4%)	277 (1.8%)	13,143 (3.2%)	11,720 (2.1%)	7,333 (1.6%)
Epilepsy/seizures	6,507 (1.8%)	112 (0.7%)	2,368 (0.6%)	4,370 (0.8%)	5,317 (1.1%)
Essential hypertension	7,928 (2.2%)	106 (0.7%)	4,180 (1.0%)	3,127 (0.6%)	2,218 (0.5%)
Heart disease (excluding hypertension)	18,498 (5.0%)	192 (1.3%)	5,888 (1.4%)	5,921 (1.1%)	2,946 (0.6%)
HIV infection	323 (0.1%)	38 (0.3%)	256 (0.1%)	307 (0.1%)	127 (0.03%)
Inflammatory auto-immune diseases of the intestine ²	7,192 (2.0%)	52 (0.3%)	2,866 (0.7%)	6,234 (1.1%)	3,290 (0.7%)
Mood disorders ³	40,255 (11.0%)	200 (1.3%)	19,078 (4.7%)	45,253 (8.2%)	74,646 (16.0%)
Obesity	22,482 (6.1%)	328 (2.1%)	39,026 (9.5%)	31,357 (5.7%)	8,716 (1.9%)
Thyroid disorders	19,680 (5.4%)	360 (2.4%)	21,680 (5.3%)	21,339 (3.9%)	9,724 (2.1%)

CPRD GOLD=Clinical Practice Research Datalink GOLD; HI-SPEED=Health Impact - Swedish Population Evidence Enabling Data-linkage; HIV=Human Immunodeficiency Virus; IMASIS=Institut Municipal Assistència Sanitària Information System; NLHR=Norwegian Linked Health Registry data; SIDIAP=The Information System for Research in Primary Care .

¹ Assessed based on diagnostic codes, considering all prior history.

² Composite of celiac disease and inflammatory bowel disease.

³ Mood disorders (composite of depression and bipolar disorders).

Pregnancy complications

Complications assessed during pregnancy are described in [Table 9](#). When interpreting study results, it should be noted that HI-SPEED only included pregnancies that surpassed week 22, and IMASIS only included pregnancies that resulted in birth (see [Section 10.2](#)).

Antepartum haemorrhage, gestational diabetes, and preterm labour were the most common complications observed during pregnancy. Pregnancies affected ranged from 3.8% to 12.0% for antepartum haemorrhage, 2.8% to 8.0% for gestational diabetes, and 3.8% to 7.0% for preterm labour. Post-partum haemorrhage (as recorded during the pregnancy episode) was recorded in 1.6% to 5.2% of pregnancies. The proportion of pregnancies affected by pre-eclampsia ranged from 2.5% to 4.0%, except for CPRD GOLD, where this value was lower (0.3%). HELLP Syndrome and eclampsia during pregnancy occurred in 0.1% cases or less, with fewer than five cases observed in IMASIS. Ectopic pregnancies ranged between 0.1% to 1.4%, with estimates in NLHR showing consistent findings with those reported in the PET (see [Table 5](#)). TORCH infections were observed in 1.1% of cases or less. In CPRD GOLD, no cases of HELLP Syndrome or TORCH infections were observed. Maternal death assessed during the 90 days following birth was rare and affected less than 0.01% pregnancies across data sources.

Table 9. Characteristics of pregnant individuals in terms of pregnancy complications (assessed at pregnancy episode level).

Pre-specified complications ¹	NLHR (n=367,803)	IMASIS ² (n=15,440)	SIDIAP (n=409,207)	HI-SPEED ³ (n=553,300)	CPRD GOLD (n=466,615)
Antepartum Haemorrhage	44,165 (12.0%)	587 (3.8%)	17,416 (4.3%)	45,968 (8.3%)	22,020 (4.7%)
Eclampsia	129 (0.04%)	<5	309 (0.1%)	215 (0.04%)	111 (0.02%)
Ectopic pregnancy	5,184 (1.4%)	33 (0.2%)	1,297 (0.3%)	496 (0.1%)	3,844 (0.8%)
Gestational diabetes	20,476 (5.6%)	588 (3.8%)	32,830 (8.0%)	32,804 (5.9%)	13,116 (2.8%)
HELLP Syndrome	511 (0.1%)	<5	189 (0.05%)	789 (0.1%)	0 (0%)
Maternal death	19 (<0.01%)	<5	12 (<0.01%)	34 (<0.01%)	48 (<0.01%)
Pre-eclampsia	14,516 (3.9%)	613 (4.0%)	10,286 (2.5%)	20,819 (3.8%)	1,529 (0.3%)
Preterm labour	13,843 (3.8%)	1,083 (7.0%)	22,477 (5.5%)	27,535 (5.0%)	21,024 (4.5%)
Postpartum haemorrhage	19,041 (5.2%)	252 (1.6%)	8,688 (2.1%)	26,819 (4.8%)	24,492 (5.2%)
TORCH infection	3,956 (1.1%)	109 (0.7%)	4,477 (1.1%)	2,009 (0.4%)	0 (0%)

CPRD GOLD=Clinical Practice Research Datalink GOLD; HI-SPEED=Health Impact - Swedish Population Evidence Enabling Data-linkage; IMASIS=Institut Municipal Assistència Sanitària Information System; NLHR=Norwegian Linked Health Registry data; SIDIAP=The Information System for Research in Primary Care ; TORCH=toxoplasmosis, others, rubella, cytomegalovirus, herpes simplex.

¹ Assessed based on diagnostic codes, assessed during pregnancy, except for maternal death (i.e. defined as a recorded death occurring during pregnancy or within 90 days following the end of pregnancy) and pre-term labour (i.e. defined as a live birth occurring before 37 completed weeks of gestation)

² IMASIS only included pregnancies that resulted in birth.

³ HI-SPEED included pregnancies surpassing week 22 of gestation.

Results stratified by trimester and in the 90 days following pregnancy end are presented in [Figure 4](#) and can be explored in detail in the Shiny app. Results reflect the presence of a recorded diagnosis during a specific trimester only and should not be interpreted as representing cumulative cases or prevalence across trimesters. For example, a pregnancy with a diagnosis of gestational diabetes recorded in the second trimester may not appear in the third trimester if no additional record of the condition is captured during that period. The proportion of pregnancy episodes with a diagnosis of antepartum haemorrhage was more common in the first trimester compared to the second or third in all data sources, except for SIDIAP, where figures were consistent across trimesters (approximately 2.9%). In other data sources, figures ranged from 2.9% to 8.5% in the first trimester, and from 0.9% to 2.7% in the third. Post-partum haemorrhage was more common in the third trimester and in the 90 days following birth. In these time windows, post-partum haemorrhage ranged from 0.9% to 5.0% and from 1.4% to 6.3%, respectively, across data sources. While less frequently recorded, cases of post-partum haemorrhage were also seen during the first and second trimesters, affecting less than 0.5% of pregnancies, except for CPRD GOLD, where this was observed to affect 3.6% and 0.9% of pregnancies in the first and second trimester, respectively, obtaining the same percentages as those observed for antepartum haemorrhage. Such results are likely attributed to limitations surrounding the phenotypes used to identify haemorrhage complications, with some broad, non-specific codes related to vaginal bleeding overlapping between the two definitions (see [Section 10.2.](#)). Ectopic pregnancies were more common in the first trimester, affecting 0.06% to 1.4% of pregnancies across data sources. In the second and third trimesters, these were detected in equal or less than 0.2% of pregnancies. The proportion of pregnancies affected by gestational diabetes also increased across trimesters of gestation, ranging across data sources from 0.07% to 0.9% in the first trimester, 0.2% to 2.2% in the second, and 2.3% to 7.8% in the third. Cases of pre-eclampsia, HELLP syndrome, and eclampsia were more common in the third trimester. The proportion of patients affected by pre-eclampsia in the third trimester ranged between 2.5% and 3.8% in all data sources, except for CPRD GOLD, where figures were

lower (0.3%). HELLP syndrome and eclampsia were rare, and while they increased in numbers during the third trimester, they represented less than 0.1% across data sources in the last trimester of pregnancy. Although uncommon, cases of TORCH infection also increased during the third trimester, especially in NLHR and HI-SPEED (where estimates doubled compared to the first trimester) and IMASIS (where they increased from 0.1% in the first trimester to 0.5% in the third).

In the 90 days following pregnancy end (inclusive), cases of complications were observed across complications studied, affecting up to 7.4% of pregnancies. In this window, post-partum haemorrhage was the most common complication and was followed by gestational diabetes and pre-eclampsia, with values across data sources ranging from 0.3% to 7.4% and from 0.3% to 3.3%, respectively.

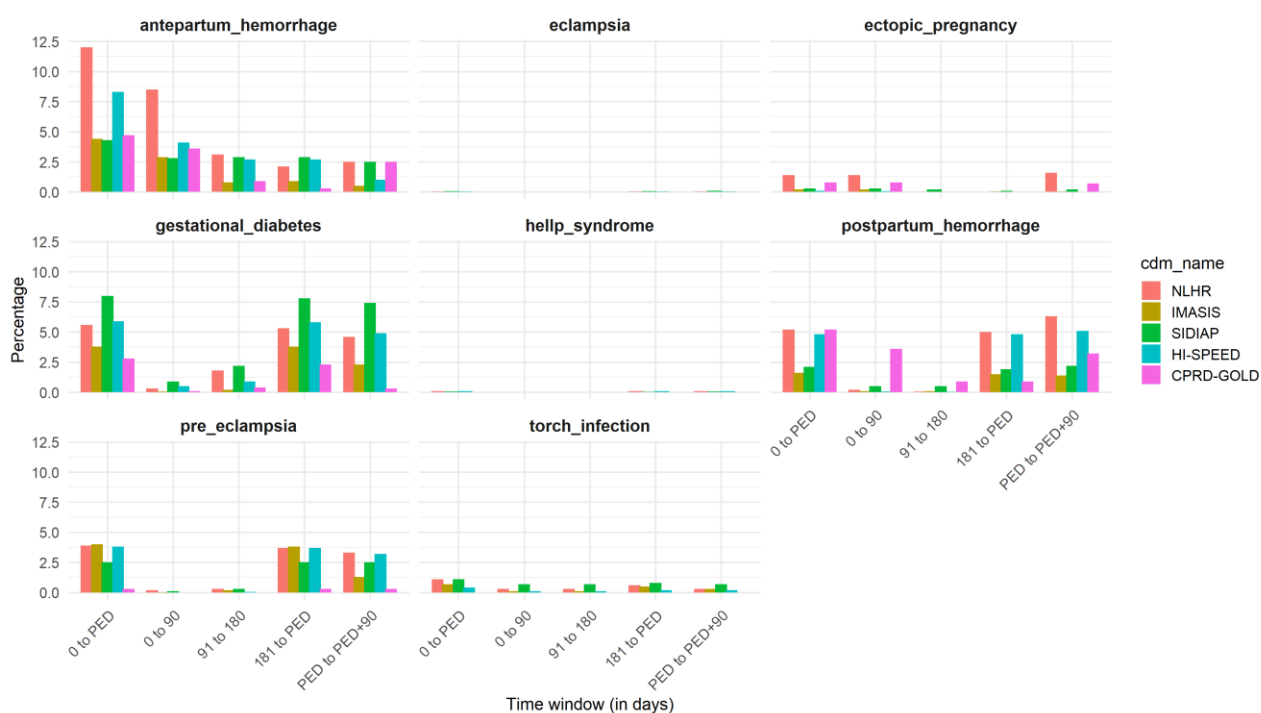


Figure 4. Percentage of pregnancy complications, overall and by trimester.

CPRD GOLD=Clinical Practice Research Datalink GOLD; HI-SPEED=Health Impact - Swedish Population Evidence Enabling Data-linkage; IMASIS=Institut Municipal Assistència Sanitària Information System; NLHR=Norwegian Linked Health Registry data; PED=Pregnancy end date; SIDIAP=The Information System for Research in Primary Care ; TORCH=toxoplasmosis, others, rubella, cytomegalovirus, herpes simplex.

“0” indicates index date (pregnancy start). Numbers correspond to days relative to index date or pregnancy end date (PED). IMASIS only included pregnancies that resulted in birth. HI-SPEED included pregnancies surpassing week 22 of gestation. Occurrences of antepartum or postpartum haemorrhage outside the clinically expected time windows likely reflect limitations of the phenotypes used to identify haemorrhage complications, with some broad, non-specific vaginal bleeding codes overlapping between the two definitions.

Medication use based on ATC 3rd level groups

The 10 most frequent medications based on pharmacological subgroups (i.e. ATC 3rd level) with an ongoing prescription during pregnancy are described in [Table 10](#). Most of these subgroups belonged to the anatomical groups (i.e. ATC 1st level) of “anti-infective for systemic use (J)”, “genitourinary system and sex hormones (G)”, “blood and blood forming organs (B)”, “respiratory system (R)”, and “nervous system (N)”.

Vaccines (i.e. J07) ranked among the top 10 groups in all data sources, except for HI-SPEED, where information on vaccines was limited to those against COVID-19. In other data sources, vaccine codes appeared in 2–3 positions within the top 10, with J07 subgroups (J07A, J07B, J07C) captured in 4.6% to 8.2% pregnancy episodes included ([Table 10](#)).

Subgroups found in three of the five data sources within the top 10 included: “other analgesics and antipyretics” (N02B: 3.3% in NLHR; 4.6% in IMASIS; 6.6% in SIDIAP; 4.6% in IMASIS), “vitamin b12 and folic acid” (B03B: 3.3% in SIDIAP; 4.0% in HI-SPEED; 5.2% in CPRD GOLD), “iron preparations” (B03A: 3.3% in IMASIS; 8.4% in SIDIAP; 4.9% CPRD GOLD), “beta lactam antibacterials penicillin” (J01C: 6.5% in NLHR, 2.8% in SIDIAP, 4.3% in CPRD GOLD), and “other diagnostic agents” (V04C: 2.4% SIDIAP; 2.9% HI-SPEED; 5.1% CPRD GOLD) ([Table 10](#)).

Subgroups identified in at least two data sources in the top 10 were: “anti-inflammatory and antirheumatic products non steroids” (M01A: 4.6% in IMASIS; 3.1% in SIDIAP), “antihistamines for systemic use” (R06A: 7.4% in NLHR; 3.8% in HI-SPEED), “anti-infectives and antiseptics, excl. combinations with corticosteroids” (G01A: 3% in SIDIAP, 3.3% in CPRD GOLD), and “decongestants and other nasal preparations for topical use” (R01A: 3.7% in NLHR; 3% in HI-SPEED). Other subgroups for respiratory system drugs included “throat preparations” (R02A: 5% in IMASIS) and “cough suppressants and expectorants combinations” (R02F: 3.3% in SIDIAP). Antiacids and drugs for peptic ulcer and gastro oesophageal reflux disease also ranked amongst the most common in SIDIAP (A02A: 4%) and CPRD GOLD (A02B: 2.8%), respectively ([Table 10](#)).

In NLHR, subgroups within “Sex hormones and modulators of the genital system (G03)” were common, appearing up to four times in the top 10 subgroups ([Table 10](#)). These included hormonal contraceptives (G03A: 8.2%; G03B: 3.6%), oestrogens (G03C: 4%), and progestogens (G03D: 3.4%). Although other subgroups from the broader “Genitourinary system and sex hormones (G)” category were observed in other data sources, these were not related to sex hormones (e.g. G01, G02, G04).

Pharmacological subgroups were also assessed 90 days prior and after pregnancy start, and 90 days after pregnancy end, with all results available in the Shiny app. Overall, the subgroups identified in the top 10 for these time windows were similar to those observed during pregnancy. Contraceptives (G03A, G03B), oestrogens, and progestogens alone or in combination (G03B, G03C, G03D), and “hormonal and related agents” (L02A) were captured amongst the most frequent subgroups in the 90 days prior to pregnancy start. Specifically, “hormonal contraceptives for systemic use” (G03A) ranked within the top 10 subgroups across all data sources in the 90 days prior to pregnancy start (accounting for 2.5%–13.0% of pregnancies) and in the 90 days following pregnancy end (accounting for 1.7%–8.1% of pregnancies across data sources, except in IMASIS, where it did not appear in the top 10). In NLHR, the G03A subgroup ranked first across all time windows studied, decreasing from 13% in the 90 days prior to pregnancy start to 5.2% in the 90 days following pregnancy start, and then increasing to 8.1% in the 90 days after pregnancy end.

Additional subgroups appearing in the top 10 in the 90 days prior to pregnancy start included “anxiolytics” (N05B: 2.3% in SIDIAP), “antidepressants” (N06A: 3.1% in HI-SPEED and 6.8% in CPRD GOLD), “opioids” (N02A: 3.1%), “gonadotropins and other ovulation stimulants” (G03G: 2.4% in IMASIS), and “androgens and female sex hormones in combination” (G03E: 2.9%), among others. In the 90 days following pregnancy start, additional subgroups (distinct from those identified before or during pregnancy) included “antithrombotic agents” (B01A: 1.5% NLHR; 1.6% HI-SPEED), “topical products for joint and muscular pain” (5.9% IMASIS), and “stomatological preparations” (A01A: 3.8% IN IMASIS).

Table 10. Top 10 pharmacological subgroups (ATC 3rd level groups) used during pregnancy.

Top	NLHR (n=367,803)	IMASIS (n=15,440)	SIDIAP (n=409,207)	HI-SPEED (n=553,300)	CPRD GOLD (n=466,615)
1	g03a hormonal contraceptives for systemic use; 72,860 (8.2%)	j07b viral vaccines; 8,588 (5.5%)	b03a iron preparations; 237,260 (8.4%)	n02b other analgesics and antipyretics; 172,009 (4.6%)	j07a bacterial vaccines; 171,658 (7.9%)
2	r06a antihistamines for systemic use; 66,151 (7.4%)	j07c bacterial and viral vaccines combined; 8,429 (5.4%)	j07b viral vaccines; 232,747 (8.3%)	b03b vitamin b12 and folic acid; 150,926 (4.0%)	j07c bacterial and viral vaccines combined; 171,627 (7.9%)
3	j01c beta lactam antibacterials penicillins; 57,913 (6.5%)	j07a bacterial vaccines; 8,417 (5.4%)	j07c bacterial and viral vaccines combined; 230,224 (8.2%)	r06a antihistamines for systemic use; 143,341 (3.8%)	j07b viral vaccines; 158,887 (7.3%)
4	j07b viral vaccines; 41,412 (4.6%)	r02a throat preparations; 7,773 (5.0%)	j07a bacterial vaccines; 230,153 (8.2%)	c05a agents for treatment of haemorrhoids and anal fissures for topical use; 127,966 (3.4%)	b03b vitamin b12 and folic acid; 111,623 (5.2%)
5	g03c oestrogens; 39,580 (4.4%)	m01a anti-inflammatory and antirheumatic products non steroids; 7,403 (4.8%)	n02b other analgesics and antipyretics; 185,329 (6.6%)	r05f cough suppressants and expectorants combinations; 124,353 (3.3%)	v04c other diagnostic agents; 110,724 (5.1%)
6	l02a hormones and related agents; 35,479 (4.0%)	n02b other analgesics and antipyretics; 7,157 (4.6%)	b03b vitamin b12 and folic acid; 93,285 (3.3%)	p01a agents against amoebiasis and other protozoal diseases; 117,376 (3.1%)	b03a iron preparations; 105,039 (4.9%)
7	r01a decongestants and other nasal preparations for topical use; 32,774 (3.7%)	a02a antacids; 6,257 (4.0%)	m01a anti-inflammatory and antirheumatic products non steroids; 87,480 (3.1%)	n05c hypnotics and sedatives; 116,790 (3.1%)	j01c beta lactam antibacterials penicillins; 91,191 (4.2%)
8	g02b contraceptives for topical use; 31,730 (3.6%)	g04b urologicals; 6,175 (4.0%)	g01a anti-infective and antiseptics excluding combinations with corticosteroids; 84,445 (3.0%)	a03d antispasmodics in combination with analgesics; 114,359 (3.0%)	g01a anti-infective and antiseptics excluding combinations with corticosteroids; 70,779 (3.3%)

Top	NLHR (n=367,803)	IMASIS (n=15,440)	SIDIAP (n=409,207)	HI-SPEED (n=553,300)	CPRD GOLD (n=466,615)
9	g03d progestogens; 30,222 (3.4%)	b03a iron preparations; 5,168 (3.3%)	j01c beta lactam antibacterials penicillins; 79,626 (2.8%)	r01a decongestants and other nasal preparations for topical use; 111,474 (3.0%)	d01a antifungals for topical use; 70,014 (3.2%)
10	n02b other analgesics and antipyretics; 29,708 (3.3%)	g02c other gynaecological; 4,492 (2.9%)	v04c other diagnostic agents; 67,239 (2.4%)	v04c other diagnostic agents; 108,891 (2.9%)	a02b drugs for peptic ulcer and gastro oesophageal reflux disease gord; 60,137 (2.8%)

CPRD GOLD=Clinical Practice Research Datalink GOLD; HI-SPEED=Health Impact - Swedish Population Evidence Enabling Data-linkage; IMASIS=Institut Municipal Assistència Sanitària Information System; NLHR=Norwegian Linked Health Registry data; SIDIAP=The Information System for Research in Primary Care .

Large-scale characterisation

In the large-scale characterisation, all diagnoses and medications recorded within pre-defined time windows relative to pregnancy start. To limit the extensiveness of this report, only the top 10 conditions and medications are described in the report. The results presented in this section include those captured during pregnancy. All results can be explored in the Shiny app, including results beyond the top 10, and results obtained using other time windows for assessment, including 90 days prior to pregnancy start, 90 days after pregnancy start (medications only), and 90 days following pregnancy end.

The top 10 conditions identified during pregnancy mainly included diagnostic codes related to pregnancy (e.g. pregnancy, finding related to pregnancy, vomiting of pregnancy) or labour (e.g. livebirth, delivery normal, complication occurring during labour and delivery) (**Table 11**). First degree perineal laceration appeared in the top 10 conditions in NLHR (22.0%) and SIDIAP (15.5%), and trauma to perineum and/or vulva during delivery in CPRD GOLD (5.9%). Foetal condition affecting obstetrical care of mother appeared in the top 10 conditions in NLHR (13.6%) and HI-SPEED (18.4%), and post-term pregnancy appeared in IMASIS (38.4%) and SIDIAP (30.6%).

In the 90 days prior to pregnancy start, the top 10 conditions were mostly unrelated to pregnancy and included common medical conditions (e.g. acute upper respiratory infection, abdominal pain, urinary tract infection). Codes related to miscarriage were observed in all data sources, ranging from 1.1% in IMASIS to 3.8% in NLHR. Female infertility ranked as the most common code in NLHR (6.2%) and HI-SPEED (3.7%) and also ranked among the top 10 conditions in SIDIAP (1.1%).

Table 11. Top 10 conditions identified through large-scale characterisation during pregnancy.

Top	NLHR (n=367,803)	IMASIS (n=15,440)	SIDIAP (n=409,207)	HI-SPEED (n=553,300)	CPRD GOLD (n=466,615)
1	Livebirth (4092289); 256,377 (69.7%)	Livebirth (4092289); 15,286 (99.7%)	Birth (4216316); 287,556 (70.3%)	Delivery normal (441641); 407,254 (73.6%)	Blood pressure finding (4214962); 254,633 (54.6%)
2	First degree perineal laceration (197343); 80,994 (22.0%)	Single live birth (4014295); 14,874 (97.0%)	Single live birth (4014295); 150,486 (36.8%)	Finding related to pregnancy (444094); 248,688 (45.0%)	Pregnancy (4299535); 232,168 (49.8%)
3	Symptomatic disorders in pregnancy (4034097); 80,369 (21.9%)	Vaginal delivery (4125611); 11,873 (77.5%)	Third trimester pregnancy (4218813); 128,843 (31.5%)	Failed induction of labour (442333); 130,183 (23.5%)	Finding of size of gravid uterus (443289); 120,058 (25.7%)
4	Vomiting of pregnancy (440785); 71,305 (19.4%)	Mother delivered (433260); 72,05 (47.0%)	Post-term pregnancy (432695); 125,112 (30.6%)	Second degree perineal laceration (198492); 103,695 (18.7%)	Miscarriage (4067106); 60,859 (13.0%)
5	Haemorrhage more than 500 ml during delivery [500-1500 mL] (2000000022); 70,030 (19.0%)	Delivery normal (441641); 6,843 (44.6%)	Low back pain (194133); 79,658 (19.5%)	Foetal condition affecting obstetrical care of mother (74104); 102,024 (18.4%)	Parity finding (4041279); 46,860 (10.0%)
6	Complication occurring during labour and delivery (440795); 59,861 (16.3%)	Finding related to pregnancy (444094); 6,644 (43.3%)	First degree perineal laceration (197343); 63,382 (15.5%)	Obesity (433736); 618,66 (11.2%)	Spontaneous vertex delivery (4205240); 32,475 (7.0%)

Top	NLHR (n=367,803)	IMASIS (n=15,440)	SIDIAP (n=409,207)	HI-SPEED (n=553,300)	CPRD GOLD (n=466,615)
7	Illness (4192174); 59,160 (16.1%)	Post-term pregnancy (432695); 5,889 (38.4%)	Finding related to pregnancy (444094); 53,305 (13.0%)	Single live birth (4014295); 57,473 (10.4%)	Finding of pulse rate (4103199); 30,776 (6.6%)
8	Abnormality of forces of labour (437342); 54,672 (14.9%)	Unplanned pregnancy (4307820); 5,540 (36.1%)	Endocrine, nutritional and metabolic disease complicating pregnancy, childbirth and puerperium (4062791); 52,987 (12.9%)	Subluxation of symphysis pubis in pregnancy, childbirth and the puerperium (4063159); 51,780 (9.4%)	Abdominal pain (200219); 29,046 (6.2%)
9	Foetal condition affecting obstetrical care of mother (74104); 49,854 (13.6%)	Postpartum finding (4041280); 5,218 (34.0%)	Gestation period, 40 weeks (444098); 43,539 (10.6%)	Disorder of pregnancy (439658); 49,319 (8.9%)	Trauma to perineum and/or vulva during delivery (432975); 27,685 (5.9%)
10	Pregnancy not yet confirmed (45757490); 45,145 (12.3%)	Third trimester pregnancy (4218813); 3,589 (23.4%)	Advanced maternal age gravida (441092); 42,624 (10.4%)	False labour (4062557); 43,552 (7.9%)	Cough (254761); 22,546 (4.8%)

CPRD GOLD=Clinical Practice Research Datalink GOLD; HI-SPEED=Health Impact - Swedish Population Evidence Enabling Data-linkage; IMASIS=Institut Municipal Assistència Sanitària Information System; NLHR=Norwegian Linked Health Registry data; SIDIAP=The Information System for Research in Primary Care .

Ingredients ranking among the most prescribed during pregnancy are presented in [Table 12](#). Although some variation across data sources was observed, the types of ingredients appearing in the top 10 were broadly consistent. These included vaccines (e.g. against tetanus, pertussis, diphtheria, and COVID-19); antibiotics (e.g. amdinocillin pivoxil, penicillin V, fosfomycin); antihistamines (e.g. desloratadine, promethazine); antiemetics (e.g. metoclopramide); antacids (e.g. magnesium hydroxide, alginic acid); supplements (e.g. ferrous sulfate, cholecalciferol, folic acid, vitamin B12); and antifungals (e.g. clotrimazole), among others. Acetaminophen ranked among the top 10 ingredients prescribed across data sources, ranging from 6.1% in HI-SPEED to 42.7% in IMASIS. In IMASIS, the proportion of pregnancies with a record of ibuprofen was similar to that of acetaminophen (44.6%). Progesterone was observed in 7% of pregnancies identified in NLHR. Ethinyl oestradiol were also observed among the top 10 medications in NLHR (12.9%), where levonorgestrel was observed in similar proportions (12.1%), and CPRD GOLD (7.9%) ([Table 12](#)).

Table 12. Top 10 ingredients identified through large-scale characterisation during pregnancy.

Top	NLHR (n=367,803)	IMASIS (n=15,440)	SIDIAP (n=409,207)	HI-SPEED (n=553,300)	CPRD GOLD (n=466,615)
1	ethinyl oestradiol (1549786); 47,373 (12.9%)	tetanus toxoid vaccine, inactivated (529411); 8,406 (54.8%)	tetanus toxoid vaccine, inactivated (529411); 229,918 (56.2%)	promethazine (1153013); 117,622 (21.3%)	folic acid (19111620); 113,089 (24.2%)
2	levonorgestrel (1589505); 44,665 (12.1%)	diphtheria toxoid vaccine, inactivated (529303); 8,402 (54.8%)	diphtheria toxoid vaccine, inactivated (529303); 229,732 (56.1%)	ephedrine (1143374); 97,863 (17.7%)	ferrous fumarate (1595799); 79,465 (17.0%)

Top	NLHR (n=367,803)	IMASIS (n=15,440)	SIDIAP (n=409,207)	HI-SPEED (n=553,300)	CPRD GOLD (n=466,615)
3	SARS-CoV-2 (COVID-19) vaccine, mRNA spike protein (37003432); 40,623 (11.0%)	ibuprofen (1177480); 6,839 (44.6%)	pertussis vaccine (19033193); 215,916 (52.8%)	caffeine (1134439); 89,852 (16.2%)	amoxicillin (1713332); 72,586 (15.6%)
4	amdinocillin pivoxil (19088223); 36,845 (10.0%)	pertussis vaccine (19033193); 6,700 (43.7%)	ferrous sulfate (1396131); 202,728 (49.5%)	SARS-CoV-2 (COVID-19) vaccine, mRNA spike protein (37003432); 81,631 (14.8%)	clotrimazole (1000632); 68,153 (14.6%)
5	acetaminophen (1125315); 27,493 (7.5%)	acetaminophen (1125315); 6,549 (42.7%)	acetaminophen (1125315); 144,664 (35.4%)	aspirin (1112807); 44,820 (8.1%)	alginate acid (19030059); 54,163 (11.6%)
6	metoclopramide (906780); 26,767 (7.3%)	magnesium hydroxide (992956); 6,146 (40.1%)	folic acid (19111620); 91,252 (22.3%)	levothyroxine (1501700); 41,421 (7.5%)	acetaminophen (1125315); 47,290 (10.1%)
7	progesterone (1552310); 25,677 (7.0%)	folic acid (19111620); 4,327 (28.2%)	amoxicillin (1713332); 62,924 (15.4%)	amdinocillin pivoxil (19088223); 39,895 (7.2%)	ethinyl oestradiol (1549786); 36,747 (7.9%)
8	desloratadine (1103006); 21,624 (5.9%)	vitamin B12 (1308738); 3,989 (26.0%)	clotrimazole (1000632); 51,199 (12.5%)	acetaminophen (1125315); 33,977 (6.1%)	codeine (1201620); 35,869 (7.7%)
9	cholecalciferol (19095164); 21,126 (5.7%)	ferrous glycine sulfate (19000753); 3,446 (22.5%)	fosfomycin (956653); 50,360 (12.3%)	penicillin V (1729720); 30,626 (5.5%)	albuterol (1154343); 34,506 (7.4%)
10	penicillin V (1729720); 18,570 (5.0%)	sodium chloride (967823); 2,912 (19.0%)	vitamin B12 (1308738); 36,117 (8.8%)	vitamin B12 (1308738); 27,930 (5.0%)	hydrocortisone (975125); 33,520 (7.2%)

CPRD GOLD=Clinical Practice Research Datalink GOLD; HI-SPEED=Health Impact - Swedish Population Evidence Enabling Data-linkage; IMASIS=Institut Municipal Assistència Sanitària Information System; NLHR=Norwegian Linked Health Registry data; SIDIAP=The Information System for Research in Primary Care .

Ingredients identified in the 90 days before and after pregnancy start were relatively similar to those observed during pregnancy. In both windows, folic acid was captured in approximately 2–3% pregnancies in NLHR and HI-SPEED, and 19–22% pregnancies in IMASIS, SIDIAP, and CPRD GOLD.

In the 90 days prior to pregnancy start, hormonal contraceptives were observed in the top 10 across all data sources, except for IMASIS, with ranges of 6.5% to 15.9% for ethinyl oestradiol, 4.6% to 13.4% for levonorgestrel, and 3.1% to 6.9% for desogestrel. Progesterone was identified in approximately 3.5% of pregnancies in HI-SPEED and NLHR, and chorionic gonadotropin in HI-SPEED (3.1%) and IMASIS (1.0%). Additional ingredients ranking in the top 10 during the 90 days preceding the index date included analgesics (e.g. acetaminophen, dipyrone), non-steroidal anti-inflammatory drugs (e.g. ibuprofen, naproxen, dexketoprofen), corticoids (e.g. fluticasone, beclomethasone), and antihistaminics (e.g. desloratadine).

In the 90 days following pregnancy start, prescriptions of hormonal contraceptives were also observed, but with lower proportions than in the 90 days prior (e.g. ethinyl oestradiol ranged from 3.3% to 12.2%). Progesterone was captured in NLHR (6.6%), IMASIS (2.2%), and HI-SPEED (4.8%). The use of antiemetics (e.g. metoclopramide) or other drugs with antiemetic effects, including antihistamines (e.g. cyclizine, promethazine, meclizine) or pyridoxine (vitamin B6), ranked among the top 10 ingredients across data

sources. In most cases, these were observed in less than 6% of episodes, except for promethazine in HI-SPEED, which was observed in 16.7% of pregnancies in the first trimester.

In the 90 days preceding and following pregnancy start, vaccines and supplements were also recorded, but at lower proportions than those observed during pregnancy. In IMASIS, vaccines against influenza A or B ranked among the top 10 medications captured in the first trimester (1.4%). Citalopram and sertraline were observed in approximately 3–4% of pregnancies in HI-SPEED and NLHR, respectively, in both windows.

Top 10 ingredients captured in the 90 days following pregnancy end included iron formulations (e.g. ferrous sulfate, ferrous fumarate), hormonal contraceptives (e.g. ethinyl estradiol, levonorgestrel, desogestrel), analgesics and non-steroidal anti-inflammatory drugs (e.g. acetaminophen, ibuprofen, diclofenac, benzylamide), antibiotics (e.g. amoxicillin, floxacillin, dicloxacillin, metronidazole), low-molecular-weight heparins (e.g. enoxaparin, bemiparin), and medications for gastrointestinal symptoms or reflux (e.g. promethazine, ranitidine, magnesium hydroxide), among others. Labetalol and oxytocin also ranked in the top 10 ingredients after 90 days following pregnancy end (inclusive) in HI-SPEED (3.8%) and NLHR (5.5%), respectively. Intravenous fluids appeared in the top 10 in IMASIS (e.g. sodium chloride, glucose).

9.2.3. Infant characteristics (Objective 4d)

We identified a total of 1,482,391 infants who could be linked to a pregnancy episode characterised as part of previous objectives (n=257,826 in NLHR; n=15,356 in IMASIS; n=312,291 in SIDIAP; n=692,802 in HI-SPEED) that resulted in a live birth (see [Section 9.1.](#)). Mother-child linkage was not available in CPRD GOLD (see [Section 10.2.](#)).

Infants were equally distributed by sex ([Table 13](#)). Median age at end of follow up ranged from 2 years in NLHR to 7 years in SIDIAP. The minimum values for age showed a negative value in HI-SPEED. Median data visibility after index date ranged from 964 days in NLHR to 3,529 days in SIDIAP.

Median gestational age at birth ranged from 273 to 280 days, and median birth weight ranged from 3,193 to 3,540 grams ([Table 13](#)). Small for gestational age (available in all data sources except for SIDIAP) was based on different definitions across data sources (see [Section 10.2](#)) and was observed in 10.1% in IMASIS and 2.3% in HI-SPEED. In NLHR, small for gestational age was assessed using the birth z-score as a proxy and was estimated to occur in 8.2% of infants included. Median APGAR score 5 minutes after birth was estimated at 10 across all data sources. Congenital malformations were assessed using a broad and unspecific definition and included both major and minor malformations, including structural and non-structural anomalies (see [Section 10.2.](#)). They were observed to be recorded in 18.3% of linked-infants in NLHR, 9.4% in SIDIAP, and 3.2% in HI-SPEED. In IMASIS, this information was not recorded in the infant table and was not assessed. In the month following birth, mortality was observed in 328 (0.1%) infants in NLHR and 323 (0.05%) in HI-SPEED. In the year after birth, corresponding figures were 450 (0.2%) and 3,447 (0.5%), respectively. In IMASIS and SIDIAP, no deaths were observed in the first month after birth, with <5 deaths observed during the first year of life ([Table 13](#)).

Table 13. Characteristics of infants with mother-child linkage.

	NLHR	IMASIS	SIDIAP	HI-SPEED
Number of infants, N	257,826	15,356	312,291	692,802
Number of pregnant individuals	208,165	12,100	238,913	444,823
Sex female, N (%)	125,668 (48.7%)	7,393 (48.1%)	152,462 (48.8%)	342,885 (49.5%)
Age at end of follow up (in years), median [min; q25 – q75; max]	2 [0; 1–3; 5]	5 [0; 2–8; 14]	7 [0; 3 –10; 14]	3 [-7; 2–5; 9]

	NLHR	IMASIS	SIDIAP	HI-SPEED
Future observation or follow-up (in days), median [min; q25 – q75; max]	964 [0; 523–1,421; 1,989]	2,032 [0; 756 – 3,228; 5,367]	980.5 [0; 529.75–1,414.25; 5,212]	3,529 [0; 3,529–3,529; 3,529]
Birth weight (in grams), median [min; q25 – q75; max]	3,540 [210; 3,200–3,870; 5,920]	3,193 [0; 2,920–3,540; 5,480]	3,280 [90; 2,960–3,580; 5,800]	3,520 [240; 3,180–3,850; 5,995]
APGAR score (5 min), median [min; q25 – q75; max] ¹	10 [0; 9–10; 10]	10 [0; 10–10; 10]	10 [0; 10–10; 10]	10 [0; 10–10; 10]
Small for gestational age, yes (%) ^{1,2}	21,095 (8.2%)	1,549 (10.1%)	–	15,991 (2.3%)
Congenital malformations, yes (%) ^{1,2}	47,231 (18.3%)	–	1,302 (0.4%)	21,928 (3.2%)
Mortality, N (%) ²				
- [0, 30 days]	328 (0.1%)	0 (0%)	0 (0%)	323 (0.05%)
- [0, 365 days]	450 (0.2%)	<5	<5	3447 (0.5%)

NLHR=Norwegian Linked Health Registry data; IMASIS=Institut Municipal Assistència Sanitària Information System; SIDIAP=The Information System for Research in Primary Care ; HI-SPEED=Health Impact - Swedish Population Evidence Enabling Data-linkage. “–” Indicates that it was not assessed in the data source (not available in the data source).

¹ Obtained from the infant table in IMASIS and HI-SPEED. If not available, this information was derived from other records, depending on data availability.

² Assessed using birth z-score as in NLHR. In IMASIS and NLHR, this was defined as birth weight below the 10th percentile for gestational age. In HI-SPEED, it was defined as birth weight more than 2 standard deviations below the mean for gestational age. Not available in SIDIAP.

³ Includes minor and major malformations, structural and non-structural anomalies. For data sources that did not have the infant table (i.e. NLHR, SIDIAP), this was assessed 0 to 365 days after birth. Congenital malformations were not recorded in infant table in IMASIS.

³ Time 0 was defined as the date of birth recorded in the maternal records.

The proportion of infants recorded as breastfeeding (only available in SIDIAP) diminished through the first year of life and was estimated at 88.5%, 59.4%, and 28.9% at month 1, 6, and 12 after birth, respectively. When considering exclusive breastfeeding, corresponding figures were 75.6%, 48.0%, and 25.3%.

Table 14. Breastfeeding, assessed in the SIDIAP data source.

Breastfeeding recorded at follow-up visits ¹			
	Month 1	Month 6	Month 12
Exclusive breastfeeding, N (%)	276,331 (88.5%)	185,650 (59.4%)	90,393 (28.9%)
Partial breastfeeding, N (%)	236,045 (75.6%)	149,888 (48.0%)	79,153 (25.3%)

SIDIAP=The Information System for Research in Primary Care .

¹ As recorded by healthcare professionals during follow-up visits. Months 1, 6, and 12 after birth were chosen as relevant time points. However, this information may be recorded at different routine follow-up visits occurring during the first months of life.

10. DISCUSSION

10.1. Key results

This study provides a comprehensive overview characterisation of pregnancy-related health data in five data sources of the DARWIN EU® Data Network that have implemented the OMOP perinatal expansion table, describing 2,146, 567 pregnancy episodes and 1,482,391 infants linked to their respective pregnancy episode.

Importantly, only pregnancies included in the PET were studied. The completeness and scope of the data used to populate the PET differed across the included data sources and, in some cases, did not encompass the full range of information or healthcare settings captured by the data source (see [Section 10.2](#)). All data sourced had the PET populated, with the infant table only available in IMASIS and HI-SPEED. The PET and infant tables were not available in IMASIS at the start of this study and were implemented as part of Objective 3.

Across included pregnancies, gestational length ranged from 273 to 281 days. Pregnancy outcomes differed between data sources, reflecting variation in the completeness of pregnancy capture. Importantly, IMASIS primarily included live births (99.6%), and HI-SPEED included pregnancies that surpassed week 22. Stillbirths (captured in all data sources) occurred in 0.2%–0.6% of pregnancies. Miscarriages and elective pregnancy terminations were recorded in NLHR (14.4% and 14.3%, respectively), SIDIAP (8.3% and 0.5%), and CPRD GOLD (27.8% and 0.7%). NLHR provided the most granular outcome data in the PET, including data on ectopic (1.1%) and molar (0.1%) pregnancies. Multiple pregnancies were observed in approximately 1.0–1.5% of pregnancies, except for SIDIAP, where none were captured. C-sections were recorded in 4.6% of pregnancies in NLHR, 22.6% in IMASIS, 19.4% in SIDIAP, and 9.8% in HI-SPEED. In CPRD GOLD, mother-child linkage was not available in the data source, and information on mode of delivery, parity, plurality was mapped to the OMOP CDM but not available in the PET (see [Section 10.2](#)).

The number of new pregnancy episodes included per year was generally stable in NLHR, IMASIS, and HI-SPEED, with figures ranging approximately 71,000–75,500, 940–1,250, and 100,000–112,000, respectively. The annual number of pregnancies decreased in SIDIAP (2010 vs. 2023: 35,393 vs. 26,674) and CPRD GOLD (2010 vs. 2023: 61,281 vs. 11,821). The number of infants with mother-child linkage per year was approximately 39,000–54,300 in NLHR, 930–1,240 in IMASIS, 19,300–26,350 in SIDIAP, and 102,600–112,660 in HI-SPEED. Artefactual decreases were observed in the first and last year of data are attributable to the criteria used to define the study population (see [Section 8.3](#)). Trends by pregnancy outcome and mode of delivery were generally stable over time.

Median age at pregnancy start was approximately 30 years. Smoking ranged from 3.8% to 9.6%, except for CPRD GOLD, where it was observed in 25.2%. During pregnancy, the median number of visits ranged from 9 to 29 across data sources, vaccination for COVID-19 varied from 1.7% to 18.3%, 8.5% to 29.7% for influenza, and 0.1% to 52.8% for Tdap. Folic acid use was recorded in 24–28% of pregnancies, except in NLHR and HI-SPEED, where it was below 5%. Of those studied, most frequently recorded conditions were anxiety (range: 2.2% to 33.3%) and mood disorders (range: 1.3% to 16.0%). The top 10 conditions identified during pregnancy through large-scale characterisation included diagnostic codes related to pregnancy or labour. The top 10 conditions in the 90 days prior to pregnancy mostly consisted of common medical conditions unrelated to pregnancy (e.g. acute upper respiratory infection), although codes related to miscarriage were observed in the top 10 across data sources (range: 1.1%–3.8%), and female infertility also ranked among the most common conditions in NLHR (6.2%), HI-SPEED (3.7%), and SIDIAP (1.1%).

The most common complications observed during pregnancy were antepartum haemorrhage (range: 3.8% to 12.0%), gestational diabetes (range: 2.8% to 8.0%), and preterm labour (range: 3.8% to 7.0%). The proportion of pregnancies affected by pre-eclampsia ranged from 2.5% to 4.0%, except in CPRD GOLD, where it was 0.3%. TORCH infections were rare (<1.1%), with no event captured in CPRD GOLD. Severe pregnancy complications were uncommon, with HELLP syndrome and eclampsia affecting <0.1% of cases, and maternal death <0.01%. By trimester, antepartum haemorrhage and ectopic pregnancies were more common in the first trimester, whereas other complications were more frequent in the third trimester. Postpartum haemorrhage occurred most often in the 90 following pregnancy end, with figures ranging from 1.4% to 6.3% across data sources.

Pharmacological subgroups that were most frequently represented in the top 10 exposures during pregnancy included vaccines (J07A, J07B, J07C), “other analgesics and antipyretics” (N02B), “vitamin b12

and folic acid” (B03B), “iron preparations” (B03A), “beta lactam antibacterials penicillin” (J01C). In NLHR, hormonal contraceptives (G03A; G03B), oestrogens (G03C), and progestogens (G03D) also ranked amongst the most frequent codes during pregnancy (range: 3.6% to 8.2%). Most frequent ingredients identified through large-scale characterisation included vaccines, antibiotics, antihistaminics, antiemetics, antacids, and supplements. In general, pharmacological subgroups and ingredients identified in the 90 days before and after pregnancy start were similar to those observed during pregnancy. Overall, pharmacological subgroups and individual ingredients showed lower percentages in the 90 days after the start of pregnancy compared with the 90 days before, except for supplements (e.g. folic acid), which increased. Most frequently observed medications in the 90 days following pregnancy end included analgesics and anti-inflammatory drugs, iron preparations, antibiotics, and low-molecular-weight heparins. Medications for the management of postnatal hypertension (e.g. labetalol) and medications to stimulate uterine contractions (e.g. oxytocin) also ranked amongst the top 10 ingredients 90 days following pregnancy end in HI-SPEED and NLHR, respectively.

Infants with mother-child linkage were equally distributed by sex, with a median age at end of follow up ranging from 2 to 7 years. Birth weight ranged from 3,193 to 3,540g, and small for gestational age was observed in 8.2% in NLHR, 10.1% in IMASIS, and 2.3% in HI-SPEED. Median APGAR score 5 minutes after birth was estimated at 10 across data sources. Congenital malformations (minor and major, and not restricted to structural anomalies) were recorded in 18.3% of infants in NLHR, 9.4% in SIDIAP, and 3.2% in HI-SPEED, and were not recorded in the infant table in IMASIS (see [Section 10.2.](#)). Infant mortality was rare and was observed in >5 cases in NLHR (0.1% within 30 days after birth; 0.2% within one year) and HI-SPEED (0.05% and 0.5%, respectively). Breastfeeding (only available in SIDIAP) was observed in 88.5%, 59.4%, and 28.9%, in month 1, 6, and 12 after birth, respectively.

10.2. Strengths and limitations of the research methods

The completeness and breadth of the data used to populate the PET differed across data sources included. Importantly, NLHR and HI-SPEED populated the PET using data from national birth registries with mandatory reporting of pregnancies starting from gestational week 12 and 22, respectively. In NLHR, pregnancies ending before week 12 were identified based on a pregnancy algorithm developed by the University of Oslo (UiO), which uses data from primary and secondary care data to complement information from the birth registry.[11] The UiO algorithm is applied to the source data prior to mapping to the OMOP CDM. Therefore, pregnancies included in the PET in NLHR comprise those identified from the birth registry and those identified through the UiO algorithm. In HI-SPEED, pregnancies that do not surpass week 22 were not recorded in the birth registry and therefore not captured in the PET. In IMASIS, the PET was derived from the hospital birth registry, which only included pregnancies that resulted in birth, whereas in SIDIAP and CPRD GOLD, the PET was derived from primary care data.

The findings of this study need to be interpreted in light of the data used to populate the PET, as the analysis was restricted to pregnancies included therein. Some events or complications may have been underreported, particularly those occurring early in pregnancy (e.g. miscarriage, ectopic pregnancy). While these events are expected to be captured in data sources with hospital data, they were not well captured among pregnancies included in HI-SPEED and IMASIS, which include hospital data, or in SIDIAP, which incorporates linkage to hospital discharge data. Events corresponding to miscarriages or ectopic pregnancies are likely recorded to some extent and mapped in the included data sources, although the completeness of this data is not known. However, because not all sources of data were used to populate the PET (e.g. HI-SPEED relied solely on the birth registry, SIDIAP only considered primary care data), these pregnancies were not included in the PET and were therefore not analysed in this study. In this context, algorithmic approaches may help to identify and infer pregnancies from the full range of available data, as has been done in NLHR by running the UiO algorithm prior to mapping pregnancy episodes in the PET. Among pregnancies included, events occurring in hospital settings are likely to have been incompletely

captured in CPRD GOLD due to the lack of linked-hospitalisation data, whereas events occurring in primary care are likely to be incompletely captured in IMASIS.

Elective terminations of pregnancy were only seen in 14.3% of pregnancies in NLHR, and <1% in SIDIAP and CPRD GOLD, suggesting incomplete capture of this outcome. Further information on the capture of these outcomes and potential phenotypes to identify them is required to determine whether elective terminations of pregnancy are comprehensively captured and can be reliably distinguished from other early-gestation outcomes such as miscarriage. In CPRD GOLD, the algorithm used to construct pregnancy episodes did not adequately distinguish between induced and spontaneous abortions.[8]

Some variables assessed in this study can be found in multiple tables within the OMOP CDM (see [Section 10.3.2](#)). Where available, we prioritised variables from the perinatal expansion tables. This approach streamlined data analysis and ensured that the records used to populate these variables were directly relevant to the pregnancy episodes under study. This approach applied to many relevant covariates, including pregnancy start and end date, pregnancy outcomes, and some variables used for characterisation objectives (4c and 4d). We assumed the validity of data recorded in the PET and did not assess whether events recorded there corresponded with those in other tables of the data source (e.g. tables that specifically record information on conditions or measurements).

Trimester-specific windows during pregnancy were considered for characterising pregnant individuals in terms of medication use (limited to the first trimester) and pregnancy complications (assessed across all three trimesters). Most complications were distributed as expected, except for antepartum haemorrhage, which was more common in the first trimester (except in SIDIAP, where frequencies were similar across trimesters) and showed frequencies similar to those observed for postpartum haemorrhage in the first trimester in CPRD GOLD (see [Section 10.3.1](#)). These findings are likely attributable to the phenotype used to define haemorrhage complications, which included overlapping codes related to vaginal bleeding (see [Annex III](#)). These codes were included with the intention of capturing cases recorded with broad and non-specific codes (e.g. coded as vaginal bleeding rather than specifically as antepartum or postpartum haemorrhage). This may indicate the need for further refinement of the phenotype definition, particularly in primary care data sources, as well as the inclusion of time-window constraints to more specifically define these events. Lastly, we did not examine pregnancy outcomes by gestational week, which could be further assessed to understand how and when these outcomes are recorded and to identify potential data quality issues that could lead to the misclassification of these events.

Some fields of the PET and infant table (with the latter available in IMASIS and HI-SPEED only) were not consistently implemented across all data sources (see [Section 10.3.2](#)). Mode of delivery, parity, and plurality were not available in the PET in CPRD GOLD, although they were mapped to the OMOP CDM in other tables in the data source. Gravidity was not available in the PET in HI-SPEED, which also lacked data on vaccinations other than those against COVID-19. Regarding infant variables, congenital malformations were not available in the infant table in IMASIS, and small for gestational age was not available for SIDIAP. Mother-child linkage was not available in CPRD GOLD, which was consequently excluded from analyses relying on this linkage. While not available for this study, CPRD has developed derived datasets that can be linked to CPRD data sources. The “CPRD Mother-Baby Link” provides data on likely mother-baby pairs and can be requested on a study-by-study basis.[19] This linkage is based on a probabilistic mother-baby link algorithm and it is based on primary care data.[20]

Smoking status was assessed differently across data sources. For NLHR and HI-SPEED, information was recorded at the first antenatal visit, while for others we used a window from 2 years before to 90 days after the start of pregnancy. This wide window was used to enhance ascertainment of smoking status, given its incomplete and non-systematic recording, but may have resulted in misclassification of individuals who had ceased smoking prior to pregnancy. A more detailed evaluation of smoking exposure was beyond the scope of this analysis and should be addressed in future studies.

Importantly, the assessment of congenital malformations in this study relied on a broad and unspecific definition. The definition used did not differentiate between minor and major malformations and was not restricted to structural anomalies (functional anomalies and genetic changes were also included), which resulted in high estimates in some data sources, limiting the findings obtained. Future research should apply detailed phenotyping tailored to the malformation of interest, accounting for both clinical presentation and age at onset. In studies examining overall congenital malformations rather than specific types, the phenotype should be clearly specified and defined based on the events of interest, and its suitability should be evaluated across participating data sources. In this study, congenital malformations were assessed as recorded in the infant table in HI-SPEED and using a one-year window after birth for data sources without this table populated. The assessment period should be defined on a study-by-study basis according to the events of interest.

Small for gestational age was assessed as recorded in the infant table (in IMASIS and HI-SPEED) or as coded in the data source (using z-scores as a proxy in NLHR). Differences in the proportion of infants born small for gestational age (8.2% in NLHR, 10.1% in IMASIS, and 2.3% in HI-SPEED) may be attributed to differences in the definition used. In IMASIS and NLHR, small for gestational age was defined as a birth weight below the 10th percentile for gestational age, whereas in HI-SPEED it is defined as a birth weight more than 2 standard deviations below the mean for gestational age.

The objective of this work was to provide an overall understanding of the pregnancy-related health data and the characteristics of the pregnant population and infants with mother-child linkage in the selected data sources. Findings obtained in this study support the need for additional conventions on mappings of perinatal data to the OMOP CDM to facilitate future regulatory studies in the perinatal field. More details on the recommendations on future mappings of perinatal data in the context of DARWIN EU® can be found in the report of Objective 3 of the DARWIN-PeriNet study.

10.3. Interpretation

10.3.1. Results in context

The number of live births identified in this study for NLHR and HI-SPEED was consistent with figures reported by national birth statistics, ranging from 52,049 to 56,683 births per year in Norway (2020–2023)[21], and approximately 120,000 per year in Sweden.[22] In contrast, the number of live births captured in SIDIAP and CPRD GOLD was lower than those reported in Catalonia (23,717 vs. 54,217 births in 2023)[23] and the UK (10,649 vs. 694,685 in 2021)[24], taking into account the coverage of these data sources (see [Section 10.4.](#)). However, the decreasing trend in pregnancies observed in SIDIAP is consistent with previous research and national birth statistics.[4, 23] In the case of CPRD GOLD, the observed decline in pregnancies should also be interpreted in the context of dynamic changes in population size and regional coverage in CPRD GOLD, which have substantially fallen over the past ten years.[19]

The proportion of pregnancies ending in stillbirth reported in this study aligns with previous research, indicating a stillbirth rate of 2.9 (2.7–3.0) per 1,000 total births in Western Europe.[25] Regarding mode of delivery, our findings are generally consistent, though slightly lower than those reported in a previous study, which reported C-sections rates of 16% in Norway, 25.7% in Spain, and 17.7% in Sweden, as of 2019.[26] The proportions of ectopic pregnancies and miscarriages observed in NLHR (1.1% and 14.4%, respectively) were broadly consistent with the literature, with ectopic pregnancies representing 1–2% of pregnancies,[27] and approximately 15–25% of clinically recognised pregnancies estimated to end in miscarriage.[28] Percentages of miscarriages and elective terminations varied widely in NLHR, SIDIAP and CPRD GOLD. Miscarriages were observed in a lower proportion of episodes in SIDIAP (8.3%) and higher in CPRD GOLD (27.8%). Elective terminations were observed in fewer than 0.7% of pregnancies in SIDIAP and CPRD GOLD (vs. 14.3% in NLHR) likely reflecting incomplete capture of these outcomes (see [Section 10.2.](#)). In IMASIS and HI-SPEED, outcomes occurring in early pregnancy (e.g. ectopic pregnancies, elective termination, miscarriages) were not observed due to the nature of the pregnancies included, as only those

resulting in birth or surpassing week 22, respectively, were included (see [Section 10.2.](#)). The prevalence of pregnancy complications was consistent with that reported in previous literature for gestational diabetes,[29] preterm birth,[30] and pre-eclampsia[31]. Eclampsia and HELLP syndrome were uncommon in this study, with observed rates lower than global estimates reported in the literature: 0.43% (95% Confidence Interval [CI]: 0.19–0.76%) for eclampsia and 0.39% (95% CI: 0.16–0.72%) for HELLP syndrome.[31] No cases of HELLP Syndrome were captured in CPRD GOLD. A previous study using CPRD GOLD linked to Hospital Episodes Statistics data did capture cases of this condition, although absolute numbers were small.[32] The absence of linked-hospitalisation data may have contributed to the incomplete capture of some pregnancy outcomes and complications in CPRD GOLD. In general, results stratified by trimester were consistent with the expected timing of these complications, with the highest rates of gestational diabetes and hypertensive disorders of pregnancy observed in the third trimester, and a higher proportion of ectopic pregnancies occurring in the first trimester. Although a few events were recorded in time windows where they would not be typically expected, these instances were rare and represented only a small proportion of pregnancies.

Regarding haemorrhage complications, antepartum haemorrhage is expected to complicate 2–5% of pregnancies surpassing beyond 24 weeks.[33] The prevalence of postpartum haemorrhage is estimated at approximately 3.9% after vaginal birth and 8.2% after caesarean birth, increasing to up to 12.6% following vaginal birth when assessed using objective methods of blood loss measurement.[34] Overall, our estimates were broadly consistent with these figures, with NLHR reporting the highest proportion of antepartum haemorrhage during pregnancy (12% in NLHR vs. 4.3–8.3% in other data sources). This may be explained by the observation of pregnancies from earlier stages following conception, resulting in a higher number of recorded cases that include bleeding in early pregnancy. Results for postpartum haemorrhage indicated that approximately 5% of pregnancy episodes were affected, with the lowest estimates observed in IMASIS and SIDIAP (1–2%), which is less than expected. The proportion of pregnancy episodes with a diagnosis of antepartum haemorrhage was more common in the first trimester, which is unexpected, as this condition is defined as bleeding from the vaginal tract after the 24th (or 20th, depending on the definition) week of gestation up to the onset of labour. In SIDIAP, results were similar across trimesters of gestation (approximately 2.9%). The highest proportions of postpartum haemorrhage were observed within 90 days after the end of pregnancy, except for CPRD GOLD, where estimates in the first trimester were similar (3.6% in the 90 days following pregnancy start vs. 3.2% in the 90 days after pregnancy end), and resembled those seen in the first trimester for antepartum haemorrhage (3.6%). This is likely attributed to the phenotypes used to assess haemorrhage complications, with some broad, non-specific codes related to vaginal bleeding overlapping between the two definitions (see [Section 10.2.](#)).

Characteristics of pregnant women showed anxiety and mood disorders to be the most frequently recorded conditions, which is line with previous research conducted in SIDIAP on drug exposures during breastfeeding, where psychiatric diagnoses were found in 24.9% of lactating mothers.[35] Missing data on smoking status (assessed approximately 2 years before up to 90 days after pregnancy start date) was observed for 80.1% of pregnancies in SIDIAP, and was consistent with this previous study, where smoking (assessed 12 months before pregnancy start date) was missing in 91.2% of pregnancies.[35] Regarding large-scale characterisation results, pregnancy-related findings were identified as the most frequently recorded conditions across data sources, providing support for the validity of the pregnancy episodes captured. The most frequently recorded medications during pregnancy were consistent with expectations and included vaccines, antibiotics, antiemetic medications, antacids, and supplements. (e.g. vitamin B12 and iron preparations). Medications in the 90 days following pregnancy end were consistent with those reported in SIDIAP for lactating mothers, with iron and vitamin supplements, penicillin antibiotics, analgesics, and NSAIDs being the most commonly observed.[35] The proportion of pregnancies with prescription of folic acid or receiving vaccination during pregnancy varied across data sources. Some of these differences may be explained by the availability of folic acid without a prescription (e.g. with only high doses requiring prescription) or fragmented care (e.g. IMASIS not capturing vaccinations administered

in primary care). The low proportion of Tdap vaccination during pregnancy observed in NLHR (0.1%) is consistent with the source data, where exposure is also low suggesting this was not related to a mapping issue, and with vaccination recommendations in Norway, which began actively recommending the pertussis vaccine after the data cut for this study (May 2024) and reached 73% coverage within the first year following implementation.[36] The presence of ongoing prescriptions for hormonal contraceptives (e.g. G03A) was observed in 8.2% pregnancies in NLHR and may reflect unplanned pregnancies with dispensing sufficient to carry over into early pregnancy. Other hormones, including those classified under G03C, G03D, G03G, and L02A, may reflect use in assisted reproductive techniques, while progesterone (G03D) could also indicate use among individuals with threatened miscarriage.[37]

Infants with mother-child linkage had similar median birth weights within the expected range of 2.5–4.0 kg. Regarding small for gestational age, estimates for NLHR (8.2%) and IMASIS (10.1%) were consistent with the prevalence reported in a previous study based on 12 European birth cohorts (range: 4.6–14.3%)[38], except for HI-SPEED, where the proportion of infants with this characteristic recorded in the infant table was lower (2.3%). This figure was similar to that reported in a previous study, which estimated the prevalence of small for gestational age at 3.5% for all first singleton births in Sweden (1982–2006),[39] and to that reported by the national statistics.[40] Differences observed across data sources can also reflect differences in criteria for small for gestational age, with HI-SPEED using a more restrictive definition (see [Section 10.2.](#)). Congenital anomalies affect 3–6% of global live births globally, with a recent systematic review and meta-analysis reporting prevalence estimates of 0.86–3.11 cases per 10,000 births for eight major congenital anomalies.[41] In this study, estimates of congenital malformations were higher than those reported in the literature for NLHR (18.3%) and SIDIAP (9.4%), and should be interpreted in light of the broad and unspecific definition used, which did not distinguish between minor and major malformations and was not restricted to structural anomalies (see [Section 10.2.](#)).

10.3.2. Perinatal data capture

The OMOP perinatal expansion,[4] particularly the table containing pregnancy episodes (referred in this report as PET), was central to the development of this study, as it provided a standardised structure to identify pregnancy episodes. We implemented additional processing steps to ensure that pregnancy episodes reflected plausible pregnancies. These steps resulted in the exclusion of a substantial number of pregnancy episode entries in the PET (see [Table 4](#) and the Shiny app for detailed attrition). Whilst some criteria were based on basic constraints (e.g. non-negative pregnancy duration), others were defined within the context of this study objectives (e.g. in observation at pregnancy start and end) and may be refined in future work depending on the specific research. In CPRD GOLD, very few pregnancies were excluded for reasons other than the study period restriction, which might suggest that part of this data quality filtering was already done as part of the algorithm used to infer pregnancy episodes, which was developed by Matcho et al.[8].

Tables included in the perinatal expansion represent a recent development to support perinatal research and have not been widely adopted or systematically assessed for data quality as part of the DARWIN EU® onboarding process, which has been supported by the DARWIN-PeriNet study (Objective 3 and 4). While some data quality efforts were undertaken as part of DARWIN-PeriNet study (Objective 3), the fact that both objectives were conducted in parallel meant that the preparation of the analytical code and analyses took place before these checks were completed.

Therefore, there were variations in how information was recorded across data sources. These differences included, for example, different variable names to represent the same information (e.g. “pregnancy_bmi”, “pre_pregnancy_bmi”), different representation of variables (e.g. previous parity recorded as a numerical or categorical value), or the addition of columns other than those initially listed in the PET (e.g. “pre_pregnancy_smoking”). We also detected different approaches to record multiple pregnancies (e.g. recorded as a single pregnancy or as separate pregnancies with shared or distinct identifiers) and infants

(e.g. inclusion of stillbirths in the infant table). Inconsistencies were also observed for variables that would be expected to contain values in the PET (even if recorded as “Unknown” or “Missing”), such as mode of delivery and the indication of a single pregnancy in CPRD GOLD, which are described as required fields in the publication outlining the implementation of the perinatal expansion.[4] These findings highlight the need for a harmonised convention on the mapping of perinatal data and for clear guidelines on how to populate the OMOP perinatal expansion. To support this, recommendations and mapping conventions for pregnancy data have been developed as part of Objective 3 of this DARWIN-PeriNet study, which has improved data quality of the PET in the three data sources (HI-SPEED, NLHR, and SIDIAP) that have executed the data quality assessment checks and has resulted in the harmonisation of variable names and data recording within the PET.

Certain information included in the PET were also represented in other tables within the OMOP CDM as expected if the PET was populated using an algorithm for example. This includes conditions, medications, or measurements for example maternal BMI or folic acid use, and, for infants, APGAR score, birth weight, and congenital malformations, among others. For maternal variables, inclusion within the PET facilitated analysis by anchoring values to the pregnancy episode under study (e.g. pre-pregnancy BMI). While this simplified the development of the subsequent analytical code, it also reflected a set of assumptions made by data partners at the time of table creation. If data was available both in the PET and in other tables of the CDM, the study team could choose to use these variables as recorded in the PET or to define how they were operationalised within the context of the study, providing flexibility at the cost of increased analytical complexity. As an example, some data sources included smoking within the PET, recording smoking status at the first antenatal care visit as a binary variable (yes/no). In other data sources, however, smoking status was not included into the PET and had to be derived from information recorded using measurement or observation codes, requiring the definition of specific codes to reflect smoking (which varied across data sources) and a time window to ascertain exposure. Future multi-data source studies would benefit from having these variables already recorded in a standardised manner.

The fact that information could have been mapped to multiple tables of the OMOP CDM also applied to pregnancy outcomes. We observed that certain pregnancy outcomes not recorded in the PET for some data sources (e.g. miscarriage) were instead identified through large-scale characterisation using the condition occurrence table. For example, this was the case for IMASIS, where congenital malformations were not directly available in the infant table. However, these could likely be ascertained using information mapped to the condition table, as it was done for data sources without the infant table. Future studies should consider the use of phenotyping algorithms for the definition of certain information rather than relying solely on fields included in the perinatal expansion tables.

The assessment of mother-infant linkage in this study relied on the linkage records available in the included data sources. In NLHR and HI-SPEED, information on this linkage was readily available through the national birth registers, which include unique identifiers for both mother and infant. In SIDIAP, mothers and infants were linked using a deterministic approach based on identifiers such as date of birth and social security affiliation number, allowing linkage for 83.8% of candidate mothers over a 15-year period.[42] Mother-child linkage in IMASIS also followed a deterministic approach.

In the context of this study, mother-child linkage was defined as the connection between the infant and the pregnancy episode, and therefore, additional processing was required. Assumptions were necessary for data sources where linkage was established in the fact_relationship table through parental relationships (e.g. “mother of”, “child of”), which was the case of NLHR and SIDIAP. In these cases, we linked each infant to the pregnancy episode recorded in maternal data that ended in the same year as the infant’s birth. The exact date of birth could not be used for linkage, as it is typically unavailable for data privacy reasons (e.g. some data sources may record date of birth as either the first or last day of the month for all births occurring within a given month). In addition, we observed that, for some data sources, the start of the infant observation period occurred slightly before or after the recorded date of birth in maternal data. This

likely reflects differences in coding practices, including delays in recording the birth or in registering the infant within the healthcare system. This was particularly evident in SIDIAP, where an additional filter was applied to exclude infants whose observation start did not occur within one month of birth. While this restriction was considered reasonable considering that most infants are expected to have contact with paediatric services in primary care within the first days of life, it led to the exclusion of many infants (n=14,531). While this approach was applied to support ascertainment of variables within the first month after birth, it represents an analytical decision that may not be necessary depending on the research question at hand. Future studies aiming to assess outcomes occurring at birth or shortly thereafter should consider how observability is defined in each data source, as well as the potential delay between birth and registration, which may hinder the ascertainment of variables occurring close to the date of birth.

10.3.3. Implications for future research in DARWIN EU®

The findings of this study provide an overview of pregnancy-related health data and the characteristics of the pregnant population across the selected data sources. They also demonstrate the feasibility of linking infants to their corresponding pregnancy episodes in four of the five included data sources, suggesting opportunities for research within the DARWIN EU® Data Network to assess infant outcomes following maternal exposures during pregnancy. Importantly, this study highlights substantial differences in data capture. Future studies should carefully consider the completeness, scope, and data quality limitations of pregnancy episodes captured within each data source (e.g. those exceeding a given gestational age), as these may introduce selection and immortal time bias and affect study validity if not appropriately addressed. It is also important to consider which data are captured, as differences in the types of healthcare settings covered (e.g. primary care vs. hospital) can affect the ability to capture exposures and outcomes of interest. Ideally, data sources encompassing both settings should be favoured, as they are likely to capture the care pathway more comprehensively.

Although not an objective of this study, the characterisation of pregnancies will help inform future feasibility for DARWIN EU® studies in this area, as it provides counts of all conditions and medications recorded in the data sources. Medications in this study were identified using ATC and ingredient-level codes. This should be considered when interpreting these findings for future feasibility assessments, as information on specific drug combinations and doses was not ascertained. In this study, drug exposure counts were generated based on individuals with a prescription within the specified time windows of interest and should therefore not be interpreted as incident use. In addition, the findings highlight the potential for pre-pregnancy prescriptions to carry over into pregnancy, which can pose challenges when ascertaining the exact timing of exposures. This is particularly relevant for studies that rely on exposure windows based on specific gestational weeks and should be carefully designed to avoid selection and timing-related bias in prenatal exposure studies.[43]

Importantly, the counts presented in this report reflect those generated under the study protocol and based on the data releases used and the number of pregnancies is expected to increase as data accrue (see [Table 1](#)). The data available for future feasibility assessments is described as part of Objective 5 of the DARWIN-PeriNet study, which aims to provide data source aggregates and visualisations for pregnancy and mother-child linkage to support feasibility assessments.

Data sources informing Objective 4 were limited to those with the PET implemented. Results from Objective 2 of the DARWIN-PeriNet study, which evaluated the reliability of inferring pregnancy episodes and their outcomes using an OMOP pregnancy algorithms applied to 15 data partners in the DARWIN EU® Data Network, should enable more data sources to undergo the characterisation undertaken in Objective 4 in the future.

The development of this study also underscores the engagement of data partners in improving perinatal data mapping. In the context of this study objective, IMASIS implemented the perinatal expansion to support Objective 4, and SIDIAP mapped additional variables not available in earlier releases, including

birth weight, APGAR score, and breastfeeding. Future work within DARWIN EU® should further address data quality as part of onboarding, along with approaches for phenotype evaluation of events occurring during pregnancy on a study-by-study basis, consistent with the DARWIN EU® phenotyping process [44] and mapping approach selected as part of Objective 3 of the DARWIN-PeriNet study. Continued collaboration with data partners will be essential to interpret source-specific findings, optimise local implementation, and support the expansion of pregnancy research capacity within DARWIN EU®.

10.4. Generalisability

We consider our findings broadly representative of the populations in Norway and Sweden, as the data sources used from these countries provide nationwide coverage and are linked to well-established birth registries, with mandatory reporting of pregnancies exceeding week 12 in Norway and week 22 in Sweden. In both settings, the PET was populated based on data from the birth registries. For Swedish pregnancies, we consider the HI-SPEED results representative of those surpassing week 22. In NLHR, pregnancies ending before week 12 were identified using a pregnancy algorithm developed by the University of Oslo,[11] which was based on primary and secondary care data and improved the capture of non-live outcomes (miscarriages and elective terminations). Pregnancies identified by the UiO algorithm were also included in the PET in NLHR.

The representativeness of the findings for SIDIAP is also likely to be high, given that it comprises regional-level data from primary care settings (linked to hospital discharged data) covering approximately 75% of the population living in Catalonia (Spain) and is broadly generalisable to the Catalan population in terms of age and sex distribution.[45] In Spain, individuals can have private insurance in addition to the public system, which is free of charge. In Catalonia, around 30% of births in 2020 took place in private hospitals.[46] Antenatal care for these pregnancies is likely to have occurred outside the public system and therefore it is likely that not all pregnancies are fully captured in the SIDIAP data source. Nonetheless, some information on pregnancy or mother-child linkage is still generally recorded in the primary care electronic health record, either through routine follow-up in public primary care for the newborn, or in the context of sick leave or maternity leave, which require engagement with the public healthcare system.

In contrast, IMASIS only includes data from a single public hospital in Spain that may be subject to differences in local practice. For this study, IMASIS has incorporated and mapped data from the hospital birth registry, which captures all deliveries occurring at the hospital. The PET was populated exclusively with data from this registry. In Spain, low-risk pregnancies are typically managed by midwives in primary care settings in the public system and, in absence of complications, receive hospital care only at the time of delivery. Importantly, IMASIS only included pregnancies that resulted in birth in the PET and therefore is not representative of all pregnancies. This distinction is important, as antenatal care data might have been available for individuals who received care at the hospital, many of whom may have had higher risk pregnancies, which could further limit the population representativeness of the data source.

Lastly, CPRD GOLD represents approximately 4.3% of the UK population, and while its geographical distribution has substantially changed over the last 10 years, individuals currently registered in CPRD GOLD are representative of the UK population with respect to age and sex.[19]

Variations in underlying population characteristics, healthcare systems, antenatal care practices, and coding standards across European countries may affect cross-country comparability and limit the generalisability of clinical findings to other settings. Findings of this study reflect the pregnancies captured in the included data sources and may not be generalisable to other DARWIN EU® Data Network data sources.

11. CONCLUSION

This study provides a comprehensive characterisation of pregnancy-related health data in five data sources of the DARWIN EU® Data Network that have implemented the OMOP perinatal expansion table, describing 2,146,567 pregnancy episodes and 1,482,391 infants linked to their respective pregnancy episode. Outcomes, conditions, and medications observed during these pregnancy episodes were those generally expected to occur during pregnancy, supporting their use in future pregnancy studies.

The completeness and scope of the pregnancy-related data varied across the included sources and depended upon the setting, data collection, data quality, and how the OMOP PET was populated. This suggests the need for operational rules aimed at ensuring the validity of pregnancy episodes for future studies based upon plausibility, complemented by consistency in OMOP mapping and populating the PET.

12. REFERENCES

- 1.Guideline on good pharmacovigilance practices (GVP) – Module V (Rev 2). EMA/838713/2011 Rev 2. European Medicines Agency 2017 28/03/2017.
- 2.Postapproval Pregnancy Safety Studies Guidance for Industry. U.S. Department of Health and Human Services Food and Drug Administration; 2019 May 2019.
- 3.Data Quality Framework for EU medicines regulation. EMA/326985/2023. European Medicines Agency; 2023 30–10–2023.
- 4.Abellan A, Burn E, Trinh NTH, Burkard T, Callahan A, Fernández-Bertolín S, et al. Expanding the OMOP Common Data Model to Support Perinatal Research in Network Studies. *Pharmacoepidemiology and Drug Safety*. 2025;34(2).
- 5.Mercade-Besora N, Raventós B, Trinh NT, Li H, Hayati S, Giuliodori A, et al. Effectiveness of COVID-19 mRNA primary and booster vaccination against infection and hospitalisation during pregnancy: a target trial emulation and meta-analysis of data from 4 European countries [pre-print]. 2025.
- 6.Prieto-Alhambra D, Mercadé-Besora N. Safety of mRNA COVID-19 vaccines during pregnancy (EUPAS1000000712) HMA-EMA Catalogues of real-world data sources and studies2025 [updated 18/08/2025. Available from: <https://catalogues.ema.europa.eu/node/4615/administrative-details>.
- 7.Politi J, Duarte-Salles T. DARWIN EU® – Utilisation of commonly used benzodiazepines during pregnancy and the incidence of pregnancy losses (EUPAS1000000536): HMA-EMA Catalogues of real-world data sources and studies; 2025 [updated 10–04–2025. Available from: <https://catalogues.ema.europa.eu/node/4437/administrative-details>.
- 8.Matcho A, Ryan P, Fife D, Gifkins D, Knoll C, Friedman A. Inferring pregnancy episodes and outcomes within a network of observational databases. *PLOS ONE*. 2018;13(2):e0192033.
- 9.Raventós B, Fernández-Bertolín S, Aragón M, Voss EA, Blacketer C, Méndez-Boo L, et al. Transforming the Information System for Research in Primary Care (SIDIAP) in Catalonia to the OMOP Common Data Model and Its Use for COVID-19 Research. *Clinical Epidemiology*. 2023;Volume 15:969–86.
10. Trinh NT, Houghtaling J, Bernal FL, Hayati S, Maglanoc LA, Lupattelli A, et al. Harmonizing Norwegian registries onto OMOP common data model: Mapping challenges and opportunities for pregnancy and COVID-19 research. *International Journal of Medical Informatics*. 2024;191:105602.
11. Nordeng H, Lupattelli A, Hilde, Marleen. Detecting and Dating Early Non-live Pregnancy Outcomes: Generation of a Novel Pregnancy Algorithm From Norwegian Linked Health Registries. *Pharmacoepidemiology and Drug Safety*. 2024;33(9).
12. OMOP Common Data Model 2025 [Available from: https://ohdsi.github.io/CommonDataModel/cdm54.html#fact_relationship.
13. ATC/DDD Index 2025: Structure and principles 2022 [updated 10–11–2022; cited 2025 2025]. Available from: https://atcddd.fhi.no/atc/structure_and_principles/.
14. Hutcheon JA, Jacobsen GW, Kramer MS, Martinussen M, Platt RW. Small Size at Birth or Abnormal Intrauterine Growth Trajectory: Which Matters More for Child Growth? *American Journal of Epidemiology*. 2016;183(12):1107–13.
15. Català M, Guo Y, Du M, Lopez-Güell K, Burn E, Mercadé-Besora N, et al. CohortCharacteristics: Summarise and Visualise Characteristics of Patients in the OMOP CDM. R package version 1.0.0 [Available from: <https://darwin-eu.github.io/CohortCharacteristics/>.

16. Català M, Guo Y, Du M, Lopez-Güell K, Burn E, Mercadé-Besora N, et al. PatientProfiles: Identify Characteristics of Patients in the OMOP Common Data Model. R package version 1.3.0 2025 [Available from: <https://darwin-eu.github.io/PatientProfiles/>].
17. Burn E, Català M, Chen X, Mercadé-Besora N, Du M, Newby D, et al. CodelistGenerator: Identify Relevant Clinical Codes and Evaluate Their Use. R package version 3.5.0 2025 [Available from: <https://darwin-eu.github.io/CodelistGenerator/>].
18. Raventós B, Català M, Du M, Guo Y, Black A, Inberg G, et al. IncidencePrevalence: An R package to calculate population-level incidence rates and prevalence using the OMOP common data model. *Pharmacoepidemiology and Drug Safety*. 2024;33(1).
19. Sanchez-Santos MT, Axson EL, Dedman D, Delmestri A. Data Resource Profile Update: CPRD GOLD. *International Journal of Epidemiology*. 2025;54(4).
20. CPRD Mother-Baby Link: Medicines & Healthcare products Regulatory Agency 2025 [updated 25/11/2025. Available from: <https://www.cprd.com/data/algorithm-derived-data/cprd-mother-baby-link>].
21. F0: Fødtte og fødsler per år og måned: Folkehelseinstituttet (FHI); [updated 16/04/2026; cited 2026. Available from: https://statistikk.fhi.no/mfr/92zKyzsWWp9CeDnulgmj_U5ta3iy9IWc9XtxrJGKXJQ?FODSELSTIDSPUNKT_1967=2018,2019,2020,2021,2022,2023&FODSELSMANED=0000&MEASURE_TYPE=BARN_ANTALL].
22. Rydberg H, Månsson H. Statistical register's production and quality National Medical Birth Register. Socialstyrelsen, Statistics DoRa; 2021 14/09/2021.
23. Indicadors de natalitat: Gencat. Institut d'Estadística de Catalunya; 2026 [Available from: <https://www.idescat.cat/indicadors/?id=aec&n=15265>].
24. Vital statistics in the UK: births, deaths and marriages: Office for National Statistics; 2023 [Available from: <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/vitalstatisticspopulationandhealthreferencetables>].
25. Hug L, You D, Blencowe H, Mishra A, Wang Z, Fix MJ, et al. Global, regional, and national estimates and trends in stillbirths from 2000 to 2019: a systematic assessment. *The Lancet*. 2021;398(10302):772–85.
26. Amyx M, Philibert M, Farr A, Donati S, Smáráson AK, Tica V, et al. Trends in caesarean section rates in Europe from 2015 to 2019 using Robson's Ten Group Classification System: A Euro-Peristat study. *BJOG: An International Journal of Obstetrics & Gynaecology*. 2024;131(4):444–54.
27. Qian H, Qian W, Huang H, Wang Y, Jia R. Ectopic pregnancy epidemiology from 1990 to 2021: A global burden of disease (GBD) analysis of 204 countries and territories. *European Journal of Obstetrics & Gynecology and Reproductive Biology*. 2025;314:114688.
28. Genovese HG, McQueen DB. The prevalence of sporadic and recurrent pregnancy loss. *Fertility and Sterility*. 2023;120(5):934–6.
29. Eades CE, Cameron DM, Evans JMM. Prevalence of gestational diabetes mellitus in Europe: A meta-analysis. *Diabetes Research and Clinical Practice*. 2017;129:173–81.
30. Liang X, Lyu Y, Li J, Li Y, Chi C. Global, regional, and national burden of preterm birth, 1990–2021: a systematic analysis from the global burden of disease study 2021. *eClinicalMedicine*. 2024;76:102840.
31. Vera-Ponce VJ, Loayza-Castro JA, Ballena-Caicedo J, Valladolid-Sandoval LAM, Zuzunaga-Montoya FE, Gutierrez De Carrillo CI. Global prevalence of preeclampsia, eclampsia, and HELLP syndrome: a systematic review and meta-analysis. *Frontiers in Reproductive Health*. 2025;7.

32. Ramakrishnan R, Korb D, Li Y, Knight M, Carson C. Maternal postpartum six-week check and short-term health outcomes for women with hypertensive disorders in pregnancy: An observational study using the Clinical Practice Research Datalink (CPRD). *Acta Obstetrica et Gynecologica Scandinavica*. 2025;104(5):937–47.
33. Calleja-Agius J, Custo R, Brincat P. M, Calleja N. Placental abruption and placenta praevia. *European Journal of Obstetrics and Gynecology and Reproductive Biology*. 2006;2:121–7.
34. Yunas I, Islam MA, Sindhu KN, Devall AJ, Podsek M, Alam SS, et al. Causes of and risk factors for postpartum haemorrhage: a systematic review and meta-analysis. *The Lancet*. 2025;405(10488):1468–80.
35. Giner-Soriano M, Ricart GA, Cartanyà-Hueso À, Gomez-Lumbreras A, Vázquez ML, Morros R, et al. Drug exposure during breastfeeding: a descriptive population-based cohort study with SIDIAP database. *International Breastfeeding Journal*. 2025;21(1).
36. Vaskinasjon i svangerskapet 2024-25. Folkehelseinstituttet (FHI); 2025 16/12/2025.
37. Coomarasamy A, Harb HM, Devall AJ, Cheed V, Roberts TE, Goranitis I, et al. Progesterone to prevent miscarriage in women with early pregnancy bleeding: the PRISM RCT. *Health Technology Assessment*. 2020;24(33):1–70.
38. Ruiz M, Goldblatt P, Morrison J, Kukla L, Švancara J, Riitta-Järvelin M, et al. Mother's education and the risk of preterm and small for gestational age birth: a DRIVERS meta-analysis of 12 European cohorts. *Journal of Epidemiology and Community Health*. 2015;69(9):826–33.
39. Li X, Sundquist K, Sundquist J. Risks of small-for-gestational-age births in immigrants: A nationwide epidemiological study in Sweden. *Scandinavian Journal of Public Health*. 2012;40(7):634–40.
40. Statistikdatabas för graviditeter, förlossningar och nyfödda: Socialstyrelsen; 2025 [updated 04/12/2025]. Available from: https://sdb.socialstyrelsen.se/if_mfr_004/val.aspx.
41. Xie X, Pei J, Zhang L, Wu Y. Global birth prevalence of major congenital anomalies: a systematic review and meta-analysis. *BMC Public Health*. 2025;25(1).
42. Segundo E, Far M, Rodríguez-Casado CI, Elorza JM, Carrere-Molina J, Mallol-Parera R, et al. A mother-child data linkage approach using data from the information system for the development of research in primary care (SIDIAP) in Catalonia. *Journal of Biomedical Informatics*. 2024;159:104747.
43. Latour CD, Edwards JK, Funk MJ, Suarez EA, Boggess K, Wood ME. Pregnancy identification method as a source of bias in studies of prenatal exposures using real-world data. *American Journal of Epidemiology*. 2025.
44. Dernie F, Corby G, Robinson A, Bezer J, Mercade-Besora N, Griffier R, et al. Standardised and Reproducible Phenotyping Using Distributed Analytics and Tools in the Data Analysis and Real World Interrogation Network (<scp>DARWIN EU</scp>). *Pharmacoepidemiology and Drug Safety*. 2024;33(11).
45. Recalde M, Rodríguez C, Burn E, Far M, García D, Carrere-Molina J, et al. Data Resource Profile: The Information System for Research in Primary Care (SIDIAP). *International Journal of Epidemiology*. 2022;51(6):e324–e36.
46. Servei de Salut Maternoinfantil. Dades 2020.: Gencat. Agència de Salut Pública de Catalunya (ASPCAT); 2020 [Available from: <https://salutpublica.gencat.cat/ca/ambits/promocio/embaras-part-puerperi/xifres-dades/infografia-dades-2020/>].

13. ANNEXES

ANNEX I. Description of data sources

Norway: Norwegian Linked Health Registry data (NLHR)

#	Section	Description
1	Database Identification and country	NLHR (Norwegian Linked Health Registry data) Norway
2	Data partner information section	University of Oslo Faculty of Mathematics and Natural Science – Department of Pharmacy
3	Coverage and timespan	Data coverage: 2008-2023 (primary care data), 2018-2023 (secondary care data and drug dispensation data) Extent: Nation-wide. 5.5M active population. Norway has a universal public health care system, consisting of primary and specialist health care services covering a population of approximately 5.4 million inhabitants.
4	Healthcare setting / type of data	Primary care –GPs, and primary care specialists (e.g. paediatricians), and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. The following registries are included: the Medical Birth Registry of Norway (MBRN), the Norwegian Prescription Registry (NorPD), the Norwegian Patient Registry (NPR), Norway Control and Payment of Health Reimbursement (KUHR), the Norwegian Surveillance System for Communicable Diseases (MSIS), the Norwegian Immunisation Registry (SYSVAK), the National Death Registry, and the National Registry (NR).
5	Data collection process	Registries. Many population-based health registries were established in the 1960s, with use of unique personal identifiers facilitating linkage between registries. Data in these health registries are used for health analysis, health statistics, improving the quality of healthcare, research, administration, and emergency preparedness.
6	General representativeness	The NLHR data covers the full Norwegian population.
7	Data content /source coding	NPR: ICD-10 for diagnosis, ATC and some special codes for drug use, Norwegian codes for clinical procedures (surgery (NCSP),) and diagnostic imaging, image-guided intervention, and nuclear medicine (NCRP)). KUHR: ICD-10 and ICPC-2 and ICPC-2B for diagnosis/procedure. NorPD: ATC. SYSVAK and MSIS: national classifications. MBRN: custom classifications by questionnaires (incl. check box variables in Maternity health care card)
8	Data Harmonization	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® Data Network. Linkage between the registries was facilitated using project-specific person IDs generated from unique personal identification assigned at birth or immigration for all legal residents in Norway.
9	Quality control (database specific)	In-house data quality checks of rates of common conditions, drug exposures, and outcomes. We compare obtained rates with official national statistics (e.g. birth statistics, yearly rates of drug dispensing, and diagnosis by age and gender). We also review missing data and outliers and inform registry holders of any unusual patterns.
10	Linkage	The NLHR is, by definition, a linkage of datasets. Helsedata.no is one central portal to apply for 11 national health registries, including all the registries that have been mapped to the OMOP CDM.
11	Vital status	The national death registry is linked.
12	Limitations	(i) Diagnostic codes in the secondary care are limited to the comprehensive list requested from the registries. The list of codes will be revised annually with each data delivery. (ii) ICD-10 codes vary in granularity, ranging from 2-character to full-length codes. The granularity of data will be revised annually with each data delivery

#	Section	Description
		(iii) Drug dispensations to outpatients are included; however, over-the-counter (OTC) medications are not captured in the data. Only a few selected expensive drugs given in hospital, are included.
13	Main references	Trinh et al. Harmonizing Norwegian registries onto OMOP common data model: Mapping challenges and opportunities for pregnancy and COVID-19 research. Int J Med Inform 2024: 191: 105602.
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1000000409 Website: https://www.mn.uio.no/farmasi/english/research/groups/pharma-safe/

Spain: Institut Municipal Assistència Sanitària Information System (IMASIS)

#	Section	Description
1	Database Identification and country	IMASIS (Institut Municipal Assistència Sanitària Information System) Catalunya, Spain
2	Data partner information section	Consorti Mar Parc de Salut Barcelona (PSMar) Direcció de Control de Gestió (Management and Control Department)
3	Coverage and timespan	Data collection since: 1990 Extent: Regional. The catchment area is essentially the city area of Barcelona. However, any visit recorded from a patient is included and might include patients residing elsewhere.
4	Healthcare setting / type of data	Secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. Patients are captured when they interact with any service from any hospital or clinic within Parc de Salut Mar Barcelona (PSMar). This can be taken from inpatient EHRs, hospital outpatient specialist care, as well as long term / skilled nursing facilities.
5	Data collection process	Outpatient electronic health records, and Inpatient hospital electronic health records, and Other. Long term/skill nursing facility care.
6	General representativeness	The data source contains patients of no specific socioeconomic status, indicating that the dataset should be representative of the population of Barcelona.
7	Data content /source coding	Health outcomes, diagnoses, and procedures are recorded using ICD-9-CM and ICD-10-CM codes. Laboratory measurements are recorded in the source data using LOINC terminology. Additionally, a limited set of demographics are available.
8	Data Harmonization	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® Data Network. A patient cannot be registered under different ID numbers. They are registered under the same ID, even when changing practice of re-registration or admission.
9	Quality control (database specific)	There are several quality control actions that we have implemented, and the creation and update of IMASIS-2 includes a set of filters that rule out all the data whose format or content are not matching the target variable. Some examples of the actions applied are: • If there are records containing codes with errors that do not belong to the vocabulary used for codifying, the corresponding domain is reviewed or discarded. • If there are records with a date before the birthdate of a patient, this information is also reviewed. • Before the update of IMASIS-2, the number of records in the files coming from the source are counted. Statistics of counts of the new instance of IMASIS-2 are taken and compared with the counts in the source files and with previous IMASIS-2 instances in order to identify issues in the source data. • In the first instances, if the Data Quality Dashboard (DQD) package detected some quality

#	Section	Description
		issues, such as conditions diagnosed in patients with an implausible gender in the source data, this information was fixed.
10	Linkage	Information of diagnosis and treatment received in outpatient settings and mortality registries.
11	Vital status	The age at death for a patient is recorded, but without a specific cause.
12	Limitations	No database-specific limitations documented. General limitations for the data type applicable.
13	Main references	Mayer MA, Furlong LI, Torre P, Planas I, Cots F, Izquierdo E, Portabella J, Rovira J, Gutierrez-Sacristan A, Sanz F "Reuse of EHRs to Support Clinical Research in a Hospital of Reference." Studies in health technology and informatics (2015): 25991136
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/104316 Website: https://www.parcdesalutmar.cat/en/

Spain: The Information System for Research in Primary Care (SIDIAP)

#	Section	Description
1	Database Identification and country	SIDIAP (The Information System for Research in Primary Care) Catalunya, Spain
2	Data partner information section	IDIAPIGol
3	Coverage and timespan	Data collection since: 2006 Extent: Regional. The SIDIAP database contains records of around 6 million people residing in Catalonia, estimated to be representing around 76% of the Catalan population.
4	Healthcare setting / type of data	Primary care – gps, and hospital inpatient care. SIDIAP captured data includes routine visits, demographics, diagnoses, laboratory tests, drugs (prescribed and dispensed), referrals, and lifestyle information.
5	Data collection process	Outpatient electronic health records, and Inpatient hospital electronic health records, and Other. Data is entered by primary care physicians upon healthcare contact, supplemented with hospital discharge records. The Institut Catala de la Salut is the owner of the data and acts as the data controller.
6	General representativeness	It was previously shown that the captured SIDIAP population is highly representative of the entire Catalan region in terms of geographic, age, and sex distributions.
7	Data content /source coding	SIDIAP data covers all services that occur at the Primary Care Centres, as well as support services, such as sexual and reproductive health or home end-of-life care. Drugs are coded in ATC-WHO terminology in the source data. Health outcomes are captured in ICD-10CM codes. The SIDIAP contains all laboratory tests and results performed in primary health centres. Demographics, geographical, as well as socio-economic factors are recorded for each patient.
8	Data Harmonization	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® Data Network. No.
9	Quality control (database specific)	Internal and external validation processes are carried out to determine the data quality of the SIDIAP information at each data update. These include stratifying the data by geographical regions and year in order to identify differences in data collection that need to be harmonized (e.g. recording of specific information under different codes). The measurement units of variables measuring one characteristic are also homogenized (e.g. transformation of the data from every laboratory that measures haemoglobin to grams per decilitre).

#	Section	Description
		Visual inspection of all data included in the database by week is also conducted, allowing one to see temporal patterns in the registry of a certain variable. With this information, the SIDIAP team can issue recommendations to researchers about the most common variable(s) where certain information is recorded (e.g. there are several variables with information concerning the women's menopausal status and with these visual inspection tools the SIDIAP team can inform the researchers about which related variables have the largest number of records and could be more helpful to capture menopause). Data availability (longitudinally and reliability), plausibility (range checks and unusual values), and consistency are inspected through visualisation tools. In addition, before accessing the data for a requested project, research teams have access to a quality-control report. This document contains counts, years, percentiles, maximums and minimums, incidences, and prevalence of the data requested for the project, allowing detection of inconsistencies in the data extraction prior to data delivery. External validation processes of the SIDIAP database mainly include assessing the data recorded in SIDIAP through linkage to external gold standard data sources, by analysing free text, or by sending questionnaires to health professionals.
10	Linkage	SIDIAP is linked to a hospital discharge database, pharmacy dispensation, and primary care laboratories. It can also be linked to other registries in Catalonia on a project by project basis.
11	Vital status	Mortality is fully captured in SIDIAP. The cause of death is not available but can be linked to the Spanish death registry on a project by project basis.
12	Limitations	The SIDIAP data is not representative of individuals not using public primary care, and conditions that are usually followed by specialist care might not be properly captured. In addition, there is limited information on lifestyle variables. Patients are followed until Death or when transferring to another primary health care centre that does not contribute to SIDIAP.
13	Main references	Recalde M, Rodríguez C, Burn E, Far M, García D, Carrere-Molina J, Benítez M, Moleras A, Pistillo A, Bolívar B, Aragón M, Duarte-Salles T "Data Resource Profile: The Information System for Research in Primary Care (SIDIAP)." International journal of epidemiology (2022): 35415748
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/50190 Website: https://www.sidiap.org/index.php/en

Sweden: Health Impact - Swedish Population Evidence Enabling Data-linkage (HI-SPEED)

#	Section	Description
1	Database Identification and country	HI-SPEED (Health Impact - Swedish Population Evidence Enabling Data-linkage) Sweden
2	Data partner information section	SMPA-GU Läkemedelsverket, Pharmacoepidemiology and Analysis Department (FeA) Box 26, 751 03 Uppsala, Sweden together with University of Gothenburg , School of Public Health and Community Medicine, Institute of Medicine Box 469, 405 30 Gothenburg, Sweden
3	Coverage and timespan	Data collection since: 2015 for most data, except prescribed drug register (from 2018), and some COVID-related data (tests, vaccination) from 2020 Extent: Nation-wide (except primary care, see below). The catchment area includes the whole of Sweden, covering the full population of approximately 11.7 million.

#	Section	Description
4	Healthcare setting / type of data	Secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. Primary care (GPs) is available only for the 2 largest regions (~40% of national population) The following data elements are collected: Socio-demographics, dispensed drug prescriptions, cause of death, diagnoses and procedures from secondary (specialist) care and inpatient visits or clinical events, as well as from primary care visits (40%pop only).
5	Data collection process	Registries. The data is acquired from the relevant Swedish national and regional registries, only once all legislative, GDPR and ethical approvals have been granted. Therefore, only relevant data is passed on, which is then entered and processed by the study team. The data are updated several times annually.
6	General representativeness	The coverage includes all inhabitants in Sweden since 2015 of all sociodemographic characteristics. Therefore it should mirror the source population to a very good extent.
7	Data content /source coding	Medicines are coded with ATC and NPLID (National Product ID), ICD10-SE is used for diagnoses and the Swedish procedure coding system (KVA) is used for clinical procedures.
8	Data Harmonization	The data have been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® Data Network.
9	Quality control (database specific)	The source data are obtained from the relevant Swedish National and Regional Registers. The registers perform some regular quality controls on their data. After receiving the data, we perform additional checks and cleaning. We also run regular quality checks on the data we manage.
10	Linkage	Data on specialist care is acquired from the National Patient Register, mortality information is provided by the Cause-Of-Death Registry. Drug data is provided by the Patient Drug Register. And similarly for other registers. Data are linked very accurately using the national personal ID number, and pseudonymised before delivery to HI-SPEED. All data are updated 2-4 times per year.
11	Vital status	Data on death and underlying + contributing causes-of-death are extracted from the Cause-of-Death registry (i.e. based on death certificates).
12	Limitations	Primary care is only available for a subset. General limitations for the data type applicable. This is a research project where all studies require ethics approval.
13	Main references	No main reference provided.
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/node/4463/

United Kingdom: Clinical Practice Research Datalink GOLD (CPRD GOLD)

#	Section	Description
1	Database Identification and country	CPRD GOLD (Clinical Practice Research Datalink GOLD) the United Kingdom
2	Data partner information section	University of Oxford NDORMS
3	Coverage and timespan	Data collection since: 1987 Extent: Nation-wide. CPRD GOLD consists of patients in contributing practices using Vision software. Historically this covered the whole of the UK, but the number of contributing practices in the England is dropping. In January 2025 only 3 practices from England were a part of CPRD GOLD, while historical patient data were from the whole of the UK, and will continue to be so. In the future,

#	Section	Description
		no practices from England will be present, only practices from Scotland, Wales, and Northern Ireland.
4	Healthcare setting / type of data	Primary care – gps, and primary care specialists (e.g. paediatricians), and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. CPRD GOLD data include patient demographics, biological measurements, clinical symptoms and diagnoses, referrals to specialist/hospital and their outcome, laboratory tests/results, and prescribed medications.
5	Data collection process	Outpatient electronic health records. Data is entered by clinicians into the EHR. Data is processed by CPRD and provides data releases for research.
6	General representativeness	CPRD GOLD has been assessed and found to be broadly representative of the UK general population in terms of age, gender, and ethnicity. In CPRD GOLD in January 2025 there were 2,730,707 current acceptable patients (i.e. registered at currently contributing practices that use Vision software, excluding transferred out, deceased patients, and those flagged by CPRD as not acceptable for clinical research for data quality issues). This equals to 4.07%, based on the UK population estimates of 67,026,300 from the Office of National Statistics (mid-2023). Current patients are only from Scotland, Wales, and Northern Ireland. Historically, GOLD does contain data from England as well.
7	Data content /source coding	Gemscript, Read, dm+d
8	Data Harmonization	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® Data Network. In GOLD, a patient can be registered under different ID numbers upon changing practice or re-registration. Researchers are not able to identify these patients, as the data are anonymised. However, GOLD covers less than 5% of the current UK GP practices and it is unlikely that an individual who does change GP practice ends up in another GP practice which uses the Vision software and accepts the CPRD data collection agreement. The very small number of duplicated IDs will have different observation periods and should not have an impact on the data analyses.
9	Quality control (database specific)	CPRD GOLD only includes practices whose data quality is assessed to be up-to-standard (uts). Each practice is associated to an uts date set when the data quality standards become satisfactory, and CPRD recommend using only longitudinal data starting from this uts date. Every time CPRD collect the EHR from a practice, checks are run for the data quality standards and if they are not adequate, the EHR is not accepted. When the data quality becomes acceptable again, CPRD updates the practice uts date. CPRD also check data quality standards at the patient level and associate each patient to a flag, reporting if its data is acceptable for clinical research. Only patients with acceptable data quality are included in the population to be mapped to CDM.
10	Linkage	CPRD GOLD can be linked to several sources, however our Oxford OMOP CDM is only linked to the CPRD GOLD Ethnicity Record and to the CPRD Townsend Deprivation Index at Practice Level
11	Vital status	Vital status is retrieved from the GP records. Population registry (ONS) data can be requested on a study-by-study basis and linked. This data only covers England and is planned to be mapped to OMOP in the future. The cause of death is not captured.
12	Limitations	The main limitation is due to the fact that CPRD GOLD is limited to GP records, and although it contains information on referrals and discharge letters, it may not fully capture specific hospital information. Events from hospital and specialist care are not covered.
13	Main references	Sanchez-Santos MT, Axson EL, Dedman D, Delmestri A "Data Resource Profile Update: CPRD GOLD." International journal of epidemiology (2025): 40499193
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111113 Website: https://cprd.com

ANNEX II. Fitness for use assessment

Data source justification for inclusion and key characteristics

All selected data sources, except for IMASIS, have previously participated, or are currently participating, in multi-data source studies on pregnancy using the PET.

Norway: Norwegian Linked Health Registry data (NLHR)

The inclusion of NLHR in this study was based on its relevance to the perinatal field. NLHR is a registry data source that has been recognised for its value in providing reliable and comprehensive data for pregnancy research. It includes detailed information related to all pregnancies in Norway, including approximately 60,000 deliveries per year from the Medical Birth Registry of Norway (MBRN).[1]

There were study-specific limitations present in NLHR. Notifications of pregnancies and pregnancy outcomes are mandatory in Norway for all pregnancies past gestational week 12, and therefore, those ending before that threshold may have been missed. However, the UiO Pregnancy algorithm improves the detection and dating of early non-live pregnancy outcomes that would have gone unnoticed if relying solely on the medical birth registry information by linking the MBRN records with hospital and primary care records.[2]

This data source has 2 available data sources onboarded to DARWIN EU®, which are: The Medical Birth Registry of Norway (MBRN) (NLHR@ University of Oslo (UiO): PERINATAL CDM) and NLHR (NLHR@UiO CDM dataset). Both data sources use the UiO pregnancy algorithm and cover the entire Norwegian population. However, they have different source populations, based on the data source that provides the linkage IDs (MBRN vs. national registry) and the data period covered. Specifically:

- **NLHR@UiO: PERINATAL CDM**, IDs come from the medical birth registry, covering all individuals with a pregnancy event >12 gestational weeks, and covers the period 2008–2019.
- **NLHR@UiO CDM**, IDs come from the National Registry covering the entire Norwegian population and covers the period 2018–2022.

Overall, both can identify similar pregnancy events (see [Table S1](#)). However, the NLHR@UiO data source has a greater ability to detect early pregnancies (<12 gestational weeks), as its source population includes the entire Norwegian population. In contrast, the source population in the NLHR@UiO: PERINATAL is based on the MBRN, which only includes individuals with a pregnancy episode lasting beyond gestational week 12. Although the UiO pregnancy algorithm improves the identification of pregnancy events <12 gestational weeks, in this data source it can only improve detection for individuals present in the MBRN. Consequently, differences between the two data sources are likely driven by the NLHR@UiO: PERINATAL CDM's lower ability to detect elective terminations, particularly among younger individuals, who are less likely to have had a prior pregnancy extending beyond 12 gestational weeks, and therefore less likely to appear in the MBRN.

Important: Of the two available data sources, we selected the NLHR@UiO data source for this study because it has a greater ability to detect early pregnancies and its more granular information on drug exposures, making it more suitable for use in future studies. For simplicity, NLHR@UiO is referred as NLHR in this report. In this data source, the median length of observation period is 5,840 days (5,230–5,840).

Table S1. Overview of the two Norwegian CDM datasets available.

Data source	Information	Years	Source study population	N pregnancies	Prevalence pregnancy outcomes					
					LB	SAB	T	SB	ECT	MO
NLHR@UiO, P1380	MBRN**, NPR KUHR	2018–2022	All females in Norway*	412,204	64.8	17.4	16.2	0.2	1.2	0.1
NLHR@UiO:PERINATAL, P704	MBRN, NPR KUHR	2008–2018	Pregnancy episodes in MBRN	859,449	74.8	13.1	11.0	0.3	0.7	0.1

*All females (biological sex at birth) 15–55 years with a filled prescription in NorPD 2004–2020 (source population)

**MBRN data in NLHR@UiO (data delivery 2024) is not complete for 2023 (data delivery expected Q4 2025).

LB: Live births, SAB: spontaneous abortions, T: elective terminations, SB: stillbirths. ECT: ectopic pregnancies, MO: molar pregnancies. MBRN: medical birth registry, NPR: Patient registry, KUHR: Primary care registry.

Spain: Institut Municipal Assistència Sanitària Information System (IMASIS)

IMASIS was selected for this study, as it collects detailed information on pregnant women and neonatal population. IMASIS contains EHR data collected from a healthcare service organisation (Parc de Salut Mar Barcelona; PSMar) consisting of different health centres, including two general hospitals. IMASIS includes clinical information from around 1 million patients with at least one diagnosis and who have used the services of this healthcare system, and from different settings such as admissions, outpatients, emergency room, and major ambulatory surgery.

The main study-specific limitation of IMASIS was that it only includes data on pregnancies that resulted in birth, with pregnancy start derived from information on pregnancy end date and gestational length.

Data availability in IMASIS was considered sufficient for this study, as data collection began in 1990, and the most recent data extraction date was June 2025. The median length of the observation period was 66 days (IQR: 0–2,070).

IMASIS completed the implementation of the PET during the implementation phase of this study.

Spain: The Information System for Research in Primary Care (SIDIAP)

SIDIAP was included in this study because it contains detailed pregnancy information from reproductive health clinics in primary care, including information related to the perinatal period for approximately 40 000 deliveries per year.[3] In addition, SIDIAP can be linked to hospital discharge data, which was requested as part of this study.[4]

Data availability in SIDIAP was considered sufficient for this study, as data collection began in 2006, and the most recent data extraction date was December 2024. The median length of the observation period was 5760 days (IQR: 2,170–6,940).

There are study-specific limitations present in SIDIAP. First, SIDIAP does not include data from private healthcare providers.[4] Therefore, underreporting of health events may occur among individuals who choose to receive antenatal care through private healthcare settings. While reproductive health clinics in primary care are covered, individuals considered to have high-risk pregnancies are referred to outpatient hospital care. Recording of diagnoses and prescribed medications in outpatient hospital care may be underreported, as hospital data is only available for inpatient diagnoses. Lastly, voluntary pregnancy losses may be underreported.

Sweden: Health Impact - Swedish Population Evidence Enabling Data-linkage (HI-SPEED)

HI-SPEED was included in this study, as it contains high-quality data on specialist outpatient care and inpatient care. Importantly, it includes data from the National Medical Birth Register, which includes information on both mother and child, with 100% coverage containing about 100,000 births per year.[5]

HI-SPEED contains mother-child linkage information, which is mapped in two tables of the CDM: the “fact_relationship” table and the perinatal expansion infant table. The registries used for linking mothers and infants vary. Linkage in the “fact_relationship” table is derived from the National Patient Registry, while linkage in the infant table is sourced from the birth registry.

Data availability in HI-SPEED starts in 2015 (with drug records starting in 2018), and the date of the most recent data extraction was August 2024. The median length of the observation period was 3,890 days (IQR: 3,330–3,890).

There are study-specific limitations present in HI-SPEED. First, it only captures pregnancies that surpass the threshold of 22 weeks gestation, which led to an underestimation of the total number of pregnancies identified in this data source and can introduce bias in the reported number and proportion of individuals experiencing certain health events or complications. Second, data availability begins in 2015 (with drug records available from 2018), which limited the study period for this data source and reduced the amount of prior data available to assess baseline comorbidities and medication use. Lastly, primary care data is only available for 40% of the included patients.

The United Kingdom: Clinical Practice Research Datalink GOLD (CPRD GOLD)

CPRD GOLD was included in this study because it contains all clinical and referral events in primary care. Pregnancy episodes in CPRD are inferred using a previously validated algorithm [6], which was validated for CPRD data mapped to the OMOP CDM.

Data availability and follow-up in CPRD GOLD were considered sufficient for this study. Data availability began in 1987, and the date of the most recent data extraction was May 2025. The median length of the observation period was 2,150 days (IQR: 727–4,930).

There are study-specific limitations present in CPRD GOLD. First, events occurring in hospital settings are likely underreported, as it primarily covers primary care data. Second, mother-child linkage is not available in the OMOP version of the data, and therefore it was not assessed as part of this study.

Fitness for use assessment references:

1. Medical Birth Registry of Norway - NIPH. <https://www.fhi.no/en/ch/medical-birth-registry-of-norway/> (accessed 16 Oct 2025).
2. Nordeng H, Lupattelli A, Hilde, Marleen. Detecting and dating early non-live pregnancy outcomes: generation of a novel pregnancy algorithm from Norwegian Linked Health Registries. *Pharmacoepidemiol Drug Saf* 2024; 33(9).
3. Abellan A, Burn E, Trinh NTH, et al. Expanding the OMOP Common Data Model to Support Perinatal Research in Network Studies. *Pharmacoepidemiol Drug Saf*. 2025;34(2):e70106. doi:10.1002/pds.70106.
4. Recalde M, Rodríguez C, Burn E, et al. Data resource profile: The Information System for Research in Primary Care (SIDIAP). *Int J Epidemiol* 2022; 51(6): e324–36.
5. National Board of Health and Welfare (Socialstyrelsen). National Medical Birth Register. <https://www.socialstyrelsen.se/en/statistics-and-data/registers/national-medical-birth-register/> (accessed 27 Jun 2024)
6. Matcho A, Ryan P, Fife D, Gifkins D, Knoll C, Friedman A. Inferring pregnancy episodes and outcomes within a network of observational databases. *PLOS ONE* 2018; 13(2): e0192033.

ANNEX III. Concept sets

Table S2. List of codes to identify healthcare visits.

Concept ID	Concept name
9201	Inpatient Visit
9202	Outpatient Visit
9203	Emergency Room Visit
581477	Office Visit
38004453	Family Practice
581476	Home Visit
8883	Ambulatory Surgical Center
32037	Intensive Care
262	Emergency Room and Inpatient Visit

Table S3. List of codes used to assess smoking status.

Data source	Concept ID	Concept name
IMASIS	903654	Tobacco or its derivatives user
	903657	Cigarette smoker
	3003421	Cigarettes smoked total (pack per year) reported ¹
	3004518	Cigarettes smoked current (pack per year) reported ¹
SIDIAP	45884037	Current some day smoker
	45881517	Current every day smoker
	45884038	Heavy tobacco smoker
	45878118	Light tobacco smoker
	903654	Tobacco or its derivatives user
	903657	Cigarette smoker
	45879404	Never smoker
	45883458	Former smoker
CPRD GOLD	903657	Cigarette smoker
	903658	Trivial cigarette smoker (0-1 cigarettes per day)
	903659	Light cigarette smoker (1-9 cigarettes per day)
	903660	Moderate cigarette smoker (10-19 cigarettes per day)
	903661	Heavy cigarette smoker (20-39 cigarettes per day)
	40766579	On the average, when you smoke, or on the last day you smoked every day, about how many cigarettes do, or did, you smoke [PhenX] ¹
	903653	Never used tobacco or its derivatives
	903651	Currently doesn't use tobacco or its derivatives

¹ Observation records with values = 0 were interpreted as "No smoker". Values >1 were interpreted as "Smoker".

Table S4. List of condition definitions.

Alcohol misuse:

Concept ID	Concept name
4338024	Absinthe addiction
433753	Alcohol abuse
435243	Alcohol dependence
45757131	Alcohol dependence in childbirth
45757093	Alcohol dependence in pregnancy
4218106	Alcoholism
44788303	Alcohol misuse enhanced service completed
44788279	Alcohol misuse enhanced services administration
439005	Chronic alcoholism in remission
436953	Continuous chronic alcoholism
378726	Dementia associated with alcoholism
4275257	Detoxication psychiatric therapy for alcoholism
435532	Episodic chronic alcoholism
37017329	Mild alcohol dependence
37018356	Moderate alcohol dependence
4152165	Nondependent alcohol abuse
435534	Nondependent alcohol abuse, continuous
440685	Nondependent alcohol abuse, episodic
441276	Nondependent alcohol abuse in remission
4109691	Persistent alcohol abuse
37017563	Severe alcohol dependence

Anxiety:

Concept ID	Concept name
4102965	Acute fugue state due to acute stress reaction
4099956	Acute panic state due to acute stress reaction
4100536	Acute post-trauma stress state
40486521	Acute posttraumatic stress disorder following military combat
4103567	Acute situational disturbance
440083	Acute stress disorder
4102966	Acute stupor state due to acute stress reaction
4321835	Agoraphobia
439786	Agoraphobia with panic attacks
4189538	Agoraphobia without history of panic disorder
4029005	Agoraphobia without history of panic disorder with limited symptom attacks

Concept ID	Concept name
4219627	Agoraphobia without history of panic disorder without limited symptom attacks
4219044	Amphetamine-induced anxiety disorder
442077	Anxiety disorder
37109206	Anxiety disorder caused by drug
4199892	Anxiety disorder due to a general medical condition
3171318	Anxiety disorder due to brain injury
45757095	Anxiety disorder in mother complicating childbirth
4242095	Anxiety disorder of adolescence
433178	Anxiety disorder of childhood OR adolescence
4288011	Anxiety hyperventilation
4236021	Avoidant disorder of adolescence
4242704	Avoidant disorder of childhood OR adolescence
4033049	Bromisodrophobia
443414	Chronic post-traumatic stress disorder
44783184	Chronic post-traumatic stress disorder following military combat
4138454	Chronic stress disorder
4247433	Combat fatigue
40484109	Complex posttraumatic stress disorder
4103278	Compulsive neurosis
40486886	Delayed posttraumatic stress disorder following military combat
4009184	Dream anxiety disorder
1340251	Exacerbation of anxiety disorder
1340414	Exacerbation of obsessional neurosis
1340434	Exacerbation of phobic disorder
434613	Generalized anxiety disorder
4268775	Generalized social phobia
37109052	Mild major depressive disorder co-occurrent with anxiety single episode
4338031	Mixed anxiety and depressive disorder
37109053	Moderate major depressive disorder co-occurrent with anxiety single episode
4178928	Nightmares associated with chronic post-traumatic stress disorder
4098320	Obsessional neurosis
37110440	Obsessive compulsive disorder caused by cocaine
37110473	Obsessive compulsive disorder caused by psychoactive substance
42538593	Obsessive compulsive disorder caused by stimulant
42538739	Obsessive compulsive disorder caused by synthetic cathinone
440374	Obsessive-compulsive disorder
4024443	Occupation-related stress disorder

Concept ID	Concept name
37109942	Olfactory reference disorder
436074	Panic disorder
4147466	Panic disorder with agoraphobia
4233250	Panic disorder with agoraphobia AND mild panic attacks
4217601	Panic disorder with agoraphobia AND moderate panic attacks
4245974	Panic disorder with agoraphobia AND panic attacks in full remission
4273699	Panic disorder with agoraphobia AND panic attacks in partial remission
4204708	Panic disorder with agoraphobia AND severe panic attacks
4297987	Panic disorder with agoraphobia, agoraphobic avoidance in full remission AND mild panic attacks
4338662	Panic disorder with agoraphobia, agoraphobic avoidance in full remission AND moderate panic attacks
4002897	Panic disorder with agoraphobia, agoraphobic avoidance in full remission AND panic attacks in full remission
4008568	Panic disorder with agoraphobia, agoraphobic avoidance in full remission AND panic attacks in partial remission
4143602	Panic disorder with agoraphobia, agoraphobic avoidance in full remission AND severe panic attacks
4136518	Panic disorder with agoraphobia, agoraphobic avoidance in partial remission AND mild panic attacks
4137120	Panic disorder with agoraphobia, agoraphobic avoidance in partial remission AND moderate panic attacks
4329036	Panic disorder with agoraphobia, agoraphobic avoidance in partial remission AND panic attacks in full remission
4135202	Panic disorder with agoraphobia, agoraphobic avoidance in partial remission AND panic attacks in partial remission
4006459	Panic disorder with agoraphobia, agoraphobic avoidance in partial remission AND severe panic attacks
4270890	Panic disorder with agoraphobia, mild agoraphobic avoidance AND mild panic attacks
4175779	Panic disorder with agoraphobia, mild agoraphobic avoidance AND moderate panic attacks
4085666	Panic disorder with agoraphobia, mild agoraphobic avoidance AND panic attacks in full remission
4177015	Panic disorder with agoraphobia, mild agoraphobic avoidance AND panic attacks in partial remission
4054913	Panic disorder with agoraphobia, mild agoraphobic avoidance AND severe panic attacks
4171336	Panic disorder with agoraphobia, moderate agoraphobic avoidance AND mild panic attacks
4220989	Panic disorder with agoraphobia, moderate agoraphobic avoidance AND moderate panic attacks
4272328	Panic disorder with agoraphobia, moderate agoraphobic avoidance AND panic attacks in full remission
4298550	Panic disorder with agoraphobia, moderate agoraphobic avoidance AND panic attacks in partial remission
4307664	Panic disorder with agoraphobia, moderate agoraphobic avoidance AND severe panic attacks
4109930	Panic disorder with agoraphobia, severe agoraphobic avoidance AND mild panic attacks
4146608	Panic disorder with agoraphobia, severe agoraphobic avoidance AND moderate panic attacks
4243153	Panic disorder with agoraphobia, severe agoraphobic avoidance AND panic attacks in full remission
4252417	Panic disorder with agoraphobia, severe agoraphobic avoidance AND panic attacks in partial remission
4246992	Panic disorder with agoraphobia, severe agoraphobic avoidance AND severe panic attacks
4211231	Panic disorder without agoraphobia
4218624	Panic disorder without agoraphobia with mild panic attacks

Concept ID	Concept name
4277739	Panic disorder without agoraphobia with moderate panic attacks
4221253	Panic disorder without agoraphobia with panic attacks in full remission
4205119	Panic disorder without agoraphobia with panic attacks in partial remission
4182119	Panic disorder without agoraphobia with severe panic attacks
44790897	Paruresis
4304010	Phobic disorder
4238682	Physical AND emotional exhaustion state
436676	Posttraumatic stress disorder
4202389	Posttraumatic stress disorder, delayed onset
436390	Psychogenic rumination
35615154	Recurrent major depressive disorder co-occurrent with anxiety in full remission
35615155	Recurrent major depressive disorder in partial remission co-occurrent with anxiety
35615151	Recurrent mild major depressive disorder co-occurrent with anxiety
35615153	Recurrent moderate major depressive disorder co-occurrent with anxiety
35615152	Recurrent severe major depressive disorder co-occurrent with anxiety
434628	Separation anxiety
37109054	Severe major depressive disorder co-occurrent with anxiety single episode
4088141	Shell shock
4182847	Simple phobia
3189734	Situational anxiety
440690	Social phobia
4103275	Social phobia, fear of eating in public
4103276	Social phobia, fear of public speaking
4103277	Social phobia, fear of public washing
44792465	Stranger anxiety
432600	Stress reaction causing mixed disturbance of emotion and conduct
4033046	Syphilophobia
4033937	Venereophobia

Asthma:

Concept ID	Concept name
257581	Acute asthma
45769441	Acute exacerbation of allergic asthma
46273452	Acute exacerbation of asthma co-occurrent with allergic rhinitis
40481763	Acute exacerbation of chronic asthmatic bronchitis
43530693	Acute exacerbation of chronic obstructive airways disease with asthma
37159957	Acute exacerbation of intermittent allergic asthma

Concept ID	Concept name
45773005	Acute exacerbation of intrinsic asthma
37159958	Acute exacerbation of moderate persistent allergic asthma
36684328	Acute severe exacerbation of allergic asthma
45769438	Acute severe exacerbation of asthma
46269767	Acute severe exacerbation of asthma co-occurrent with allergic rhinitis
45769442	Acute severe exacerbation of immunoglobulin E-mediated allergic asthma
45769443	Acute severe exacerbation of intrinsic asthma
46274124	Acute severe exacerbation of mild persistent allergic asthma
46269777	Acute severe exacerbation of mild persistent allergic asthma co-occurrent with allergic rhinitis
45769352	Acute severe exacerbation of mild persistent asthma
46273462	Acute severe exacerbation of moderate persistent allergic asthma
45769351	Acute severe exacerbation of moderate persistent asthma
46269785	Acute severe exacerbation of moderate persistent asthma co-occurrent with allergic rhinitis
46274059	Acute severe exacerbation of severe persistent allergic asthma
45769350	Acute severe exacerbation of severe persistent asthma
46269771	Acute severe exacerbation of severe persistent asthma co-occurrent with allergic rhinitis
37116845	Acute severe refractory exacerbation of asthma
4191479	Allergic asthma
45766728	Allergic asthma due to Dermatophagoides farinae
45766727	Allergic asthma due to Dermatophagoides pteronyssinus
4271333	Allergic asthma without status asthmaticus
46269801	Aspirin exacerbated respiratory disease
4250128	Aspirin-induced asthma
317009	Asthma
4211530	Asthma caused by wood dust
46274062	Asthma-chronic obstructive pulmonary disease overlap syndrome
45772073	Asthma in mother complicating childbirth
42535716	Asthma in pregnancy
4235703	Asthma management
44783610	Asthma self management
4233784	Asthmatic bronchitis
252942	Asthmatic pulmonary eosinophilia
43530745	Asthma with irreversible airway obstruction
4206340	Asthma without status asthmaticus
4143474	Bakers' asthma
4075237	Brittle asthma
4225553	Cheese-makers' asthma

Concept ID	Concept name
4245676	Chemical-induced asthma
4051466	Childhood asthma
256448	Chronic asthmatic bronchitis
44810117	Chronic asthma with fixed airflow obstruction
46269802	Chronic obstructive asthma co-occurrent with acute exacerbation of asthma
4119300	Colophony asthma
313236	Cough variant asthma
4120262	Cryptogenic pulmonary eosinophilia
4217558	Detergent asthma
4244339	Drug-induced asthma
42538744	Exacerbation of allergic asthma
36684335	Exacerbation of allergic asthma due to infection
1340255	Exacerbation of asthma
4138760	Exacerbation of intermittent asthma
45768911	Exacerbation of mild persistent asthma
45772937	Exacerbation of moderate persistent asthma
45768912	Exacerbation of severe persistent asthma
443801	Exercise-induced asthma
4123253	Hay fever with asthma
312950	IgE-mediated allergic asthma
761844	Inhaled steroid-dependent asthma
37310241	Intermittent allergic asthma
4141978	Intermittent asthma
46270030	Intermittent asthma co-occurrent with allergic rhinitis
46270322	Intermittent asthma uncontrolled
4145497	Intrinsic asthma
252658	Intrinsic asthma without status asthmaticus
4225554	Isocyanate induced asthma
4119298	Late onset asthma
35609846	Life threatening acute exacerbation of allergic asthma
42536207	Life threatening acute exacerbation of asthma
35609847	Life threatening acute exacerbation of intrinsic asthma
4057952	Meat-wrappers' asthma
4155468	Mild asthma
4146581	Mild intermittent asthma
46269776	Mild persistent allergic asthma
4143828	Mild persistent asthma

Concept ID	Concept name
46270029	Mild persistent asthma co-occurrent with allergic rhinitis
4022592	Millers' asthma
4110051	Mixed asthma
42536208	Moderate acute exacerbation of asthma
4155469	Moderate asthma
46269784	Moderate persistent allergic asthma
4142738	Moderate persistent asthma
46273454	Moderate persistent asthma co-occurrent with allergic rhinitis
37206717	Near fatal asthma
4309833	Non-IgE mediated allergic asthma
4155470	Occasional asthma
4212099	Occupational asthma
37109103	Oral steroid-dependent asthma
764677	Persistent asthma
2108536	Persistent asthma, preferred long term control medication or an acceptable alternative treatment, prescribed (NMA-No Measure Associated)
764949	Persistent asthma, well controlled
4232595	Platinum asthma
4080516	Printers' asthma
4279553	Pulmonary eosinophilia
40483397	Seasonal asthma
4152913	Severe asthma
37208352	Severe asthma with fungal sensitisation
37108580	Severe controlled persistent asthma
46269770	Severe persistent allergic asthma
4145356	Severe persistent asthma
46270028	Severe persistent asthma co-occurrent with allergic rhinitis
37108581	Severe uncontrolled persistent asthma
4277596	Simple pulmonary eosinophilia
46273635	Steroid dependent asthma
4120261	Sulfite-induced asthma
4301938	Tea-makers' asthma
3661412	Thunderstorm asthma
4081994	Tropical pulmonary eosinophilia
42539549	Uncomplicated allergic asthma
45768910	Uncomplicated asthma
45768963	Uncomplicated mild persistent asthma

Concept ID	Concept name
45768964	Uncomplicated moderate persistent asthma
42536649	Uncomplicated non-allergic asthma
45768965	Uncomplicated severe persistent asthma
4245292	Weavers' cough
257581	Acute asthma
45769441	Acute exacerbation of allergic asthma
46273452	Acute exacerbation of asthma co-occurrent with allergic rhinitis
40481763	Acute exacerbation of chronic asthmatic bronchitis
43530693	Acute exacerbation of chronic obstructive airways disease with asthma
37159957	Acute exacerbation of intermittent allergic asthma
45773005	Acute exacerbation of intrinsic asthma
37159958	Acute exacerbation of moderate persistent allergic asthma
36684328	Acute severe exacerbation of allergic asthma
45769438	Acute severe exacerbation of asthma
46269767	Acute severe exacerbation of asthma co-occurrent with allergic rhinitis
45769442	Acute severe exacerbation of immunoglobulin E-mediated allergic asthma
45769443	Acute severe exacerbation of intrinsic asthma
46274124	Acute severe exacerbation of mild persistent allergic asthma
46269777	Acute severe exacerbation of mild persistent allergic asthma co-occurrent with allergic rhinitis
45769352	Acute severe exacerbation of mild persistent asthma
46273462	Acute severe exacerbation of moderate persistent allergic asthma
45769351	Acute severe exacerbation of moderate persistent asthma
46269785	Acute severe exacerbation of moderate persistent asthma co-occurrent with allergic rhinitis
46274059	Acute severe exacerbation of severe persistent allergic asthma
45769350	Acute severe exacerbation of severe persistent asthma
46269771	Acute severe exacerbation of severe persistent asthma co-occurrent with allergic rhinitis
37116845	Acute severe refractory exacerbation of asthma
4191479	Allergic asthma
45766728	Allergic asthma due to Dermatophagoides farinae
45766727	Allergic asthma due to Dermatophagoides pteronyssinus
4271333	Allergic asthma without status asthmaticus
46269801	Aspirin exacerbated respiratory disease
4250128	Aspirin-induced asthma
317009	Asthma
4211530	Asthma caused by wood dust
46274062	Asthma-chronic obstructive pulmonary disease overlap syndrome
45772073	Asthma in mother complicating childbirth

Concept ID	Concept name
42535716	Asthma in pregnancy
4235703	Asthma management
44783610	Asthma self management
4233784	Asthmatic bronchitis
252942	Asthmatic pulmonary eosinophilia
43530745	Asthma with irreversible airway obstruction
4206340	Asthma without status asthmaticus
4143474	Bakers' asthma
4075237	Brittle asthma
4225553	Cheese-makers' asthma
4245676	Chemical-induced asthma
4051466	Childhood asthma
256448	Chronic asthmatic bronchitis
44810117	Chronic asthma with fixed airflow obstruction
46269802	Chronic obstructive asthma co-occurrent with acute exacerbation of asthma
4119300	Colophony asthma
313236	Cough variant asthma
4120262	Cryptogenic pulmonary eosinophilia
4217558	Detergent asthma
4244339	Drug-induced asthma
42538744	Exacerbation of allergic asthma
36684335	Exacerbation of allergic asthma due to infection
1340255	Exacerbation of asthma
4138760	Exacerbation of intermittent asthma
45768911	Exacerbation of mild persistent asthma
45772937	Exacerbation of moderate persistent asthma
45768912	Exacerbation of severe persistent asthma
443801	Exercise-induced asthma
4123253	Hay fever with asthma
312950	IgE-mediated allergic asthma
761844	Inhaled steroid-dependent asthma
37310241	Intermittent allergic asthma
4141978	Intermittent asthma
46270030	Intermittent asthma co-occurrent with allergic rhinitis
46270322	Intermittent asthma uncontrolled
4145497	Intrinsic asthma
252658	Intrinsic asthma without status asthmaticus

Concept ID	Concept name
4225554	Isocyanate induced asthma
4119298	Late onset asthma
35609846	Life threatening acute exacerbation of allergic asthma
42536207	Life threatening acute exacerbation of asthma
35609847	Life threatening acute exacerbation of intrinsic asthma
4057952	Meat-wrappers' asthma
4155468	Mild asthma
4146581	Mild intermittent asthma
46269776	Mild persistent allergic asthma
4143828	Mild persistent asthma
46270029	Mild persistent asthma co-occurrent with allergic rhinitis
4022592	Millers' asthma
4110051	Mixed asthma
42536208	Moderate acute exacerbation of asthma
4155469	Moderate asthma
46269784	Moderate persistent allergic asthma
4142738	Moderate persistent asthma
46273454	Moderate persistent asthma co-occurrent with allergic rhinitis
37206717	Near fatal asthma
4309833	Non-IgE mediated allergic asthma
4155470	Occasional asthma
4212099	Occupational asthma
37109103	Oral steroid-dependent asthma
764677	Persistent asthma
2108536	Persistent asthma, preferred long term control medication or an acceptable alternative treatment, prescribed (NMA-No Measure Associated)
764949	Persistent asthma, well controlled
4232595	Platinum asthma
4080516	Printers' asthma
4279553	Pulmonary eosinophilia
40483397	Seasonal asthma
4152913	Severe asthma
37208352	Severe asthma with fungal sensitisation
37108580	Severe controlled persistent asthma
46269770	Severe persistent allergic asthma
4145356	Severe persistent asthma
46270028	Severe persistent asthma co-occurrent with allergic rhinitis

Concept ID	Concept name
37108581	Severe uncontrolled persistent asthma
4277596	Simple pulmonary eosinophilia
46273635	Steroid dependent asthma
4120261	Sulfite-induced asthma
4301938	Tea-makers' asthma
3661412	Thunderstorm asthma
4081994	Tropical pulmonary eosinophilia
42539549	Uncomplicated allergic asthma
45768910	Uncomplicated asthma
4245292	Weavers' cough

Cancer (any malignancy):

Concept ID ¹	Concept name
443392	Malignant neoplastic disease

¹All descendants included.

Chronic kidney disease:

Concept ID	Concept name
193782	End-stage renal disease
197921	Renal osteodystrophy
198185	Chronic renal failure
443597	Chronic kidney disease stage 3
443601	Chronic kidney disease stage 2
443611	Chronic kidney disease stage 5
443612	Chronic kidney disease stage 4
443614	Chronic kidney disease stage 1
443961	Anaemia of chronic renal failure
601161	Chronic kidney disease stage 1 due to drug induced diabetes mellitus
601162	Chronic kidney disease stage 2 due to drug induced diabetes mellitus
601163	Chronic kidney disease stage 3 due to drug induced diabetes mellitus
601164	Chronic kidney disease stage 4 due to drug induced diabetes mellitus
601165	Chronic kidney disease stage 5 due to drug induced diabetes mellitus
601166	End stage renal disease on dialysis due to drug induced diabetes mellitus
619679	Chronic kidney disease with osteoporosis
619681	Renal osteodystrophy with high bone turnover
619682	Renal osteodystrophy with normal bone turnover
619683	Renal osteodystrophy with low bone turnover
760850	Benign hypertensive heart disease and chronic renal disease stage 5

Concept ID	Concept name
762000	Benign hypertensive heart disease and chronic renal disease stage 2
762001	Benign hypertensive heart disease and chronic renal disease stage 3
762033	Malignant hypertensive heart disease and chronic renal disease stage 3
762034	Malignant hypertensive heart disease and chronic renal disease stage 4
762973	Hypertensive end stage renal disease
764011	Benign hypertensive heart disease and chronic renal disease
765535	Benign hypertensive heart disease and chronic renal disease stage 1
765536	Benign hypertensive heart disease and chronic renal disease stage 4
1340282	Exacerbation of chronic renal failure
4030520	End stage renal failure on dialysis
4125970	End stage renal failure with renal transplant
4128067	Acute-on-chronic renal failure
4128200	End stage renal failure untreated by renal replacement therapy
4322556	Chronic progressive renal failure
36716184	Chronic kidney disease following donor nephrectomy
36716455	Chronic kidney disease due to traumatic loss of kidney
36716947	Chronic renal insufficiency
36717349	Chronic kidney disease due to systemic infection
36717534	Chronic kidney disease following excision of renal neoplasm
37017104	Chronic kidney disease mineral and bone disorder
37017813	Chronic kidney disease stage 5 on dialysis
37018761	Chronic kidney disease stage 5 with transplant
37018886	End stage renal disease due to hypertension
37019193	Anaemia co-occurrent and due to chronic kidney disease stage 3
42872405	Anaemia, pre-end stage renal disease on erythropoietin protocol
43020437	Malignant hypertensive chronic kidney disease stage 5
43020455	Malignant hypertensive end stage renal disease
43020456	Malignant hypertensive chronic kidney disease stage 3
43020457	Malignant hypertensive chronic kidney disease stage 4
43021835	Hypertensive heart AND chronic kidney disease stage 3
43021836	Hypertensive heart AND chronic kidney disease stage 2
43021852	Malignant hypertensive chronic kidney disease
43021853	Malignant hypertensive chronic kidney disease stage 1
43021854	Malignant hypertensive chronic kidney disease stage 2
43021864	Malignant hypertensive end stage renal disease on dialysis
43531559	Chronic kidney disease stage 1 due to type 2 diabetes mellitus
43531562	Chronic kidney disease stage 5 due to type 2 diabetes mellitus

Concept ID	Concept name
43531566	Chronic kidney disease stage 2 due to type 2 diabetes mellitus
43531577	Chronic kidney disease stage 4 due to type 2 diabetes mellitus
43531578	Chronic kidney disease due to type 2 diabetes mellitus
43531653	Chronic kidney disease stage 3 due to type 2 diabetes mellitus
44782429	Chronic kidney disease due to hypertension
44782689	Chronic kidney disease stage 4 due to hypertension
44782690	Chronic kidney disease stage 5 due to hypertension
44782691	Chronic kidney disease stage 3 due to hypertension
44782692	Chronic kidney disease stage 2 due to hypertension
44782703	Chronic kidney disease stage 1 due to hypertension
44782717	End stage renal disease on dialysis due to hypertension
44784621	Hypertensive heart and chronic kidney disease
44784637	Hypertensive heart AND chronic kidney disease on dialysis
44784638	Hypertensive heart AND chronic kidney disease stage 5
44784639	Hypertensive heart AND chronic kidney disease stage 4
44784640	Hypertensive heart AND chronic kidney disease stage 1
44792226	Chronic kidney disease stage 1 with proteinuria
44792227	Chronic kidney disease stage 1 without proteinuria
44792228	Chronic kidney disease stage 2 with proteinuria
44792229	Chronic kidney disease stage 2 without proteinuria
44792230	Chronic kidney disease stage 3 with proteinuria
44792231	Chronic kidney disease stage 3 without proteinuria
44792232	Chronic kidney disease stage 3A with proteinuria
44792249	Chronic kidney disease stage 3A without proteinuria
44792250	Chronic kidney disease stage 3B with proteinuria
44792251	Chronic kidney disease stage 3B without proteinuria
44792252	Chronic kidney disease stage 4 with proteinuria
44792253	Chronic kidney disease stage 4 without proteinuria
44792254	Chronic kidney disease stage 5 with proteinuria
44792255	Chronic kidney disease stage 5 without proteinuria
45757137	Pre-existing hypertensive heart and chronic kidney disease in mother complicating childbirth
45757139	Pre-existing hypertensive heart and chronic kidney disease in mother complicating pregnancy
45757356	Pre-existing hypertensive chronic kidney disease in mother complicating pregnancy
45757392	Hypertension concurrent and due to end stage renal disease on dialysis due to type 2 diabetes mellitus
45757393	Hypertension concurrent and due to end stage renal disease on dialysis due to type 1 diabetes mellitus
45757444	Hypertension in chronic kidney disease stage 5 due to type 2 diabetes mellitus
45757445	Hypertension in chronic kidney disease stage 4 due to type 2 diabetes mellitus

Concept ID	Concept name
45757446	Hypertension in chronic kidney disease stage 3 due to type 2 diabetes mellitus
45757447	Hypertension in chronic kidney disease stage 2 due to type 2 diabetes mellitus
45763854	Chronic kidney disease stage 3A
45763855	Chronic kidney disease stage 3B
45768813	Anaemia in end stage renal disease
45769901	Chronic kidney disease stage 2 due to type 1 diabetes mellitus
45769902	Chronic kidney disease stage 4 due to type 1 diabetes mellitus
45769903	Chronic kidney disease stage 5 due to type 1 diabetes mellitus
45769904	End stage renal disease on dialysis due to type 1 diabetes mellitus
45769906	End stage renal disease on dialysis due to type 2 diabetes mellitus
45771064	Hypertension in chronic kidney disease due to type 2 diabetes mellitus
45771067	Hypertension in chronic kidney disease due to type 1 diabetes mellitus
45771075	Chronic kidney disease stage 3 due to type 1 diabetes mellitus
45772751	Hypertension concurrent and due to end stage renal disease on dialysis
45773576	Chronic kidney disease stage 1 due to type 1 diabetes mellitus
45773688	Chronic kidney disease due to type 1 diabetes mellitus
46270353	Chronic kidney disease due to benign hypertension
46270354	Chronic kidney disease stage 1 due to benign hypertension
46270355	Chronic kidney disease stage 2 due to benign hypertension
46270356	Chronic kidney disease stage 5 due to benign hypertension
46270969	Mixed renal osteodystrophy
46271022	Chronic kidney disease
46273164	End stage renal disease due to benign hypertension
46273514	Chronic kidney disease stage 4 due to benign hypertension
46273636	Chronic kidney disease stage 3 due to benign hypertension
46284566	CKD G1A1 - chronic kidney disease with glomerular filtration rate category G1 and albuminuria category A1
46284567	CKD G1A2 - chronic kidney disease with glomerular filtration rate category G1 and albuminuria category A2
46284570	CKD G1A3 - chronic kidney disease with glomerular filtration rate category G1 and albuminuria category A3
46284572	CKD G2A1 - chronic kidney disease with glomerular filtration rate category G2 and albuminuria category A1
46284575	CKD G2A3 - chronic kidney disease with glomerular filtration rate category G2 and albuminuria category A3
46284587	CKD G3aA1 - chronic kidney disease with glomerular filtration rate category G3a and albuminuria category A1
46284588	CKD G3aA3 - chronic kidney disease with glomerular filtration rate category G3a and albuminuria category A3
46284591	CKD G3bA1 - chronic kidney disease with glomerular filtration rate category G3b and albuminuria category A1
46284592	CKD G3bA2 - chronic kidney disease with glomerular filtration rate category G3b and albuminuria category A2
46284593	CKD G3bA3 - chronic kidney disease with glomerular filtration rate category G3b and albuminuria category A3
46284597	CKD G4A1 - chronic kidney disease with glomerular filtration rate category G4 and albuminuria category A1
46284598	CKD G4A2 - chronic kidney disease with glomerular filtration rate category G4 and albuminuria category A2

Concept ID	Concept name
46284599	CKD G4A3 - chronic kidney disease with glomerular filtration rate category G4 and albuminuria category A3
46284600	CKD G5A1 - chronic kidney disease with glomerular filtration rate category G5 and albuminuria category A1
46284602	CKD G5A2 - chronic kidney disease with glomerular filtration rate category G5 and albuminuria category A2
46284603	CKD G5A3 - chronic kidney disease with glomerular filtration rate category G5 and albuminuria category A3
46286992	CKD G3aA2 - chronic kidney disease with glomerular filtration rate category G3a and albuminuria category A2
46287169	CKD G2A2 - chronic kidney disease with glomerular filtration rate category G2 and albuminuria category A2
36674698	Alport syndrome

Chronic liver disease:

Concept ID	Concept name
37017151	Acute on chronic alcoholic liver disease
1450217	Acute on chronic liver failure
4048083	Advanced cirrhosis
196463	Alcoholic cirrhosis
37162862	Autoantibody negative autoimmune hepatitis
4055212	Bacterial portal cirrhosis
192675	Biliary cirrhosis
4059289	Biliary cirrhosis of children
4058681	Capsular portal cirrhosis
4252074	Cardiac cirrhosis
4141628	Cardiac portal cirrhosis
4049282	Cholangiolitic cirrhosis
4026125	Chronic active hepatitis
45769525	Chronic active hepatitis C
4173584	Chronic active type B viral hepatitis
4283078	Chronic active viral hepatitis
4009793	Chronic aggressive type B viral hepatitis
4232466	Chronic aggressive viral hepatitis
4146181	Chronic alcoholic hepatitis
37017009	Chronic alcoholic liver disease
36687200	Chronic autoimmune hepatitis
4340390	Chronic hepatic failure
37162893	Chronic hepatic failure due to portosystemic shunt
200763	Chronic hepatitis
3654685	Chronic hepatitis B co-occurrent with hepatitis C and hepatitis D
37175349	Chronic hepatitis B during pregnancy
198964	Chronic hepatitis C

Concept ID	Concept name
35625141	Chronic hepatitis C caused by Hepatitis C virus genotype 1
35625296	Chronic hepatitis C caused by Hepatitis C virus genotype 1a
35625295	Chronic hepatitis C caused by Hepatitis C virus genotype 1b
35625139	Chronic hepatitis C caused by Hepatitis C virus genotype 2
35625040	Chronic hepatitis C caused by Hepatitis C virus genotype 3
35625140	Chronic hepatitis C caused by Hepatitis C virus genotype 4
35624867	Chronic hepatitis C caused by hepatitis C virus genotype 5
35624866	Chronic hepatitis C caused by hepatitis C virus genotype 6
3654682	Chronic hepatitis C co-occurrent with human immunodeficiency virus infection
45766656	Chronic hepatitis C with stage 2 fibrosis
42872885	Chronic hepatitis E
4212540	Chronic liver disease
4238978	Chronic lobular hepatitis
4322067	Chronic lymphocytic cholangitis-cholangiohepatitis
3655440	Chronic necrosis of liver
201613	Chronic nonalcoholic liver disease
200451	Chronic passive congestion of liver
199867	Chronic persistent hepatitis
4296554	Chronic persistent type B viral hepatitis
4247138	Chronic persistent viral hepatitis
4198610	Chronic rejection of liver transplant
194574	Chronic type B viral hepatitis
4012113	Chronic viral hepatitis
192240	Chronic viral hepatitis B with hepatitis D
439674	Chronic viral hepatitis B without delta-agent
763021	Chronic viral hepatitis C with hepatic coma
42536529	Chronic viral hepatitis D
194692	Cirrhosis - non-alcoholic
44805713	Cirrhosis associated with cystic fibrosis
4064161	Cirrhosis of liver
37111265	Cirrhosis of liver caused by amiodarone
37117933	Cirrhosis of liver caused by methotrexate
37111266	Cirrhosis of liver caused by methyl dopa
3656096	Cirrhosis of liver due to and following cardiac procedure
43531723	Cirrhosis of liver due to chronic hepatitis C
45772057	Cirrhosis of liver due to hepatitis B
42539566	Cirrhosis of liver with primary sclerosing cholangitis

Concept ID	Concept name
4153294	Cirrhosis secondary to cholestasis
37110890	Cirrhotic cardiomyopathy
4292401	Clonorchiasis with biliary cirrhosis
4163687	Cruveilhier-Baumgarten syndrome
4232955	Cryptogenic cirrhosis
37396401	Decompensated cirrhosis of liver
4055210	Diffuse nodular cirrhosis
4342774	Drug-induced chronic hepatitis
4143008	Drug-induced cirrhosis of liver
4159158	Early cirrhosis
1340280	Exacerbation of chronic active hepatitis
4058680	Fatty portal cirrhosis
4294539	Florid cirrhosis
4203601	Glissonian cirrhosis
36676898	Growth retardation, mild developmental delay, chronic hepatitis syndrome
46273476	Hepatic ascites co-occurrent with chronic active hepatitis due to toxic liver disease
197493	Hepatitis D superinfection of hepatitis B carrier
4340946	Hypoxia-associated cirrhosis
37396157	Idiopathic copper associated cirrhosis of liver
37164806	Idiopathic ductopenia
4268006	Indian childhood cirrhosis
4340393	Infectious cirrhosis
4144116	Juvenile portal cirrhosis
4340392	Laennec's cirrhosis, non-alcoholic
4049419	Latent cirrhosis
605193	Liver cirrhosis due to classical cystic fibrosis
3185452	Liver cirrhosis secondary to nonalcoholic steatohepatitis
4184779	Macronodular cirrhosis
4071022	Micronodular cirrhosis
4050640	Mixed micro and macronodular cirrhosis
4148254	Multilobular portal cirrhosis
44783142	North American Indian childhood cirrhosis
4048057	Nutritional cirrhosis
4003673	Obstructive biliary cirrhosis
37017654	Occult chronic type B viral hepatitis
4140536	Parasitic cirrhosis
4300060	Pigment cirrhosis

Concept ID	Concept name
4059285	Pigmentary portal cirrhosis
4304584	Portal cirrhosis
4098583	Posthepatic cirrhosis
4313567	Postnecrotic cirrhosis
3183806	Postviral gastroparesis
4135822	Primary biliary cholangitis
607683	Progressive familial intrahepatic cholestasis type 1
607718	Progressive familial intrahepatic cholestasis type 2
37162165	Progressive familial intrahepatic cholestasis type 3
1245024	Progressive familial intrahepatic cholestasis type 4
4253211	Progressive intrahepatic cholestasis
37109889	Pulmonary fibrosis, hepatic hyperplasia, bone marrow hypoplasia syndrome
37399445	Reynolds syndrome
4046123	Secondary biliary cirrhosis
4053079	Syphilitic cirrhosis
4058682	Syphilitic portal cirrhosis
4046016	Toxic cirrhosis
4059287	Toxic portal cirrhosis
4055209	Unilobular portal cirrhosis

Diabetes mellitus:

Concept ID	Concept name
36715051	Atherosclerosis, deafness, diabetes, epilepsy, nephropathy syndrome
36713275	Atypical diabetes mellitus
36715417	DEND syndrome
4129516	Diabetes-deafness syndrome maternally transmitted
201820	Diabetes mellitus
4143529	Diabetes mellitus associated with cystic fibrosis
4245270	Diabetes mellitus associated with genetic syndrome
4240589	Diabetes mellitus associated with hormonal aetiology
4178452	Diabetes mellitus associated with pancreatic disease
4178790	Diabetes mellitus associated with receptor abnormality
42537681	Diabetes mellitus caused by chemical
765478	Diabetes mellitus caused by drug without complication
3178281	Diabetes mellitus complicating Cystic fibrosis
4144583	Diabetes mellitus due to cystic fibrosis
43531011	Diabetes mellitus due to genetic defect in beta cell function

Concept ID	Concept name
43531642	Diabetes mellitus due to genetic defect in insulin action
4192852	Diabetes mellitus due to insulin receptor antibodies
45757077	Diabetes mellitus due to pancreatic injury
4237068	Diabetes mellitus due to structurally abnormal insulin
4058243	Diabetes mellitus during pregnancy, childbirth and the puerperium
4235410	Diabetes mellitus induced by non-steroid drugs
4136889	Diabetes mellitus induced by non-steroid drugs without complication
45757129	Diabetes mellitus in mother complicating childbirth
194700	Diabetes mellitus in mother complicating pregnancy, childbirth AND/OR puerperium
4079850	Diabetes mellitus in neonate small for gestational age
45766050	Diabetes mellitus in remission
43531645	Diabetes mellitus, transient neonatal 1
43531019	Diabetes mellitus, transient neonatal 2
43531020	Diabetes mellitus, transient neonatal 3
37174264	Diabetes mellitus type 1 stage 1
37174265	Diabetes mellitus type 1 stage 2
37174270	Diabetes mellitus type 1 stage 3
45757674	Diabetes mellitus type 1 without retinopathy
45757474	Diabetes mellitus type 2 without retinopathy
36684827	Diabetes mellitus type 2 with periodontal disease
44793113	Diabetes mellitus with multiple complications
4008576	Diabetes mellitus without complication
42535539	Disorder of nerve co-occurrent and due to type 1 diabetes mellitus
4202383	Drug-induced diabetes mellitus
1340306	Exacerbation of diabetes mellitus
766252	Exacerbation of type 1 diabetes mellitus
766253	Exacerbation of type 2 diabetes mellitus
4084643	Extreme insulin resistance with acanthosis nigricans, hirsutism AND abnormal insulin receptors
4099741	Fibrocalculous pancreatic diabetes
37163452	Fulminant type 1 diabetes mellitus
3183244	Gestational diabetes during pregnancy
4024659	Gestational diabetes mellitus
4326434	Gestational diabetes mellitus class A1
4263902	Gestational diabetes mellitus class A2
37018765	Gestational diabetes mellitus complicating pregnancy
45757124	Gestational diabetes mellitus in childbirth
37396524	Gingival disease co-occurrent with diabetes mellitus

Concept ID	Concept name
46274096	Gingivitis co-occurrent with diabetes mellitus
4048202	Houssay's syndrome
4047906	Insulin dependent diabetes mellitus type 1A
4102018	Insulin dependent diabetes mellitus type 1B
4129524	Insulin resistance - type A
4129525	Insulin resistance - type B
4130162	Insulin treated type 2 diabetes mellitus
36717215	Intellectual disability, craniofacial dysmorphism, hypogonadism, diabetes mellitus syndrome
37162799	Intrauterine growth restriction, short stature, early adult-onset diabetes syndrome
37165632	Juvenile-onset diabetes mellitus, central and peripheral neurodegeneration syndrome
604357	Ketosis-prone diabetes mellitus
37163431	Ketosis-resistant diabetes mellitus
4145827	Latent autoimmune diabetes mellitus in adult
4006979	Leprechaunism syndrome
4131907	Lipoatrophic diabetes
44787902	Lipoatrophic diabetes mellitus without complication
4327944	Malnutrition related diabetes mellitus
4030061	Malnutrition-related diabetes mellitus - fibrocalculous
4096042	Malnutrition-related diabetes mellitus without complications
4096671	Malnutrition-related diabetes mellitus with peripheral circulatory complications
4096670	Malnutrition-related diabetes mellitus with renal complications
4166381	Maternal diabetes mellitus with hypoglycaemia affecting foetus OR newborn
44792134	Maternally inherited diabetes mellitus
43531640	Maturity-onset diabetes of the young
43531006	Maturity onset diabetes of the young, type 1
43531017	Maturity-onset diabetes of the young, type 10
43531018	Maturity-onset diabetes of the young, type 11
4130164	Maturity onset diabetes of the young, type 2
43531012	Maturity-onset diabetes of the young, type 3
43531013	Maturity-onset diabetes of the young, type 4
43531014	Maturity-onset diabetes of the young, type 5
43531643	Maturity-onset diabetes of the young, type 6
43531015	Maturity-onset diabetes of the young, type 7
43531644	Maturity-onset diabetes of the young, type 8
43531016	Maturity-onset diabetes of the young, type 9
4034963	Megaloblastic anaemia, thiamine-responsive, with diabetes mellitus and sensorineural deafness
4034961	Muscular atrophy, ataxia, retinitis pigmentosa, and diabetes mellitus

Concept ID	Concept name
37204818	Myopathy and diabetes mellitus
193323	Neonatal diabetes mellitus
36716258	Pancreatic hypoplasia, diabetes mellitus, congenital heart disease syndrome
43531641	Permanent neonatal diabetes mellitus
37110041	Permanent neonatal diabetes mellitus with cerebellar agenesis syndrome
4030066	Photomyoclonus, diabetes mellitus, deafness, nephropathy and cerebral dysfunction
3198118	Poorly controlled diabetes mellitus
3194119	Poorly controlled type 1 diabetes
3193274	Poorly controlled type 2 diabetes
3192767	Poorly controlled type II diabetes with circulatory disease
3194082	Poorly controlled type II diabetes with complications
3191208	Poorly controlled type II diabetes with neuropathy
3194332	Poorly controlled type II diabetes with renal complications
45757789	Postpartum gestational diabetes mellitus
40482883	Posttransplant diabetes mellitus
44793114	Pre-existing diabetes mellitus
45757079	Pre-existing diabetes mellitus in mother complicating childbirth
43531007	Pre-existing diabetes mellitus in pregnancy
4062687	Pre-existing malnutrition-related diabetes mellitus
606039	Pre-existing malnutrition-related diabetes mellitus in pregnancy
4063042	Pre-existing type 1 diabetes mellitus
43531008	Pre-existing type 1 diabetes mellitus in pregnancy
4063043	Pre-existing type 2 diabetes mellitus
43531010	Pre-existing type 2 diabetes mellitus in pregnancy
43531009	Pregnancy and type 1 diabetes mellitus
4129519	Pregnancy and type 2 diabetes mellitus
37204277	Primary microcephaly, epilepsy, permanent neonatal diabetes syndrome
37204232	Primary microcephaly, mild intellectual disability, young-onset diabetes syndrome
1340515	Progression of diabetes mellitus
4212631	Protein-deficient diabetes mellitus
4140808	Rabson-Mendenhall syndrome
195771	Secondary diabetes mellitus
4034960	Secondary endocrine diabetes mellitus
1340528	Slowly progressive insulin dependent diabetes mellitus type 1A
4099334	Steroid-induced diabetes
4099653	Steroid-induced diabetes mellitus without complication
37116379	Stimmler syndrome

Concept ID	Concept name
4129378	Transient neonatal diabetes mellitus
201254	Type 1 diabetes mellitus
4099215	Type 1 diabetes mellitus maturity onset
443412	Type 1 diabetes mellitus without complication
4099214	Type 1 diabetes mellitus with ulcer
201826	Type 2 diabetes mellitus
4230254	Type 2 diabetes mellitus in nonobese
4304377	Type 2 diabetes mellitus in obese
4193704	Type 2 diabetes mellitus without complication
4099651	Type 2 diabetes mellitus with ulcer
45766051	Type I diabetes mellitus in remission
45766052	Type II diabetes mellitus in remission
37116960	Wolfram-like syndrome
4322638	Wolfram syndrome
37168689	Wolfram syndrome type 1
37168690	Wolfram syndrome type 2
37311329	Woodhouse Sakati syndrome
4034963	Megaloblastic anaemia, thiamine-responsive, with diabetes mellitus and sensorineural deafness
4034961	Muscular atrophy, ataxia, retinitis pigmentosa, and diabetes mellitus
37204818	Myopathy and diabetes mellitus
193323	Neonatal diabetes mellitus
36716258	Pancreatic hypoplasia, diabetes mellitus, congenital heart disease syndrome
43531641	Permanent neonatal diabetes mellitus
37110041	Permanent neonatal diabetes mellitus with cerebellar agenesis syndrome
4030066	Photomyoclonus, diabetes mellitus, deafness, nephropathy and cerebral dysfunction
3198118	Poorly controlled diabetes mellitus
3194119	Poorly controlled type 1 diabetes
3193274	Poorly controlled type 2 diabetes
3192767	Poorly controlled type II diabetes with circulatory disease
3194082	Poorly controlled type II diabetes with complications
3191208	Poorly controlled type II diabetes with neuropathy
3194332	Poorly controlled type II diabetes with renal complications
45757789	Postpartum gestational diabetes mellitus
40482883	Posttransplant diabetes mellitus
44793114	Pre-existing diabetes mellitus
45757079	Pre-existing diabetes mellitus in mother complicating childbirth
43531007	Pre-existing diabetes mellitus in pregnancy

Concept ID	Concept name
4062687	Pre-existing malnutrition-related diabetes mellitus
606039	Pre-existing malnutrition-related diabetes mellitus in pregnancy
4063042	Pre-existing type 1 diabetes mellitus
43531008	Pre-existing type 1 diabetes mellitus in pregnancy
4063043	Pre-existing type 2 diabetes mellitus
43531010	Pre-existing type 2 diabetes mellitus in pregnancy
43531009	Pregnancy and type 1 diabetes mellitus
4129519	Pregnancy and type 2 diabetes mellitus
37204277	Primary microcephaly, epilepsy, permanent neonatal diabetes syndrome
37204232	Primary microcephaly, mild intellectual disability, young-onset diabetes syndrome
1340515	Progression of diabetes mellitus
4212631	Protein-deficient diabetes mellitus
4140808	Rabson-Mendenhall syndrome
195771	Secondary diabetes mellitus
4034960	Secondary endocrine diabetes mellitus
1340528	Slowly progressive insulin dependent diabetes mellitus type 1A
4099334	Steroid-induced diabetes
4099653	Steroid-induced diabetes mellitus without complication
37116379	Stimmler syndrome
4129378	Transient neonatal diabetes mellitus
201254	Type 1 diabetes mellitus
4099215	Type 1 diabetes mellitus maturity onset
443412	Type 1 diabetes mellitus without complication
4099214	Type 1 diabetes mellitus with ulcer
201826	Type 2 diabetes mellitus
4230254	Type 2 diabetes mellitus in nonobese
4304377	Type 2 diabetes mellitus in obese
4193704	Type 2 diabetes mellitus without complication
4099651	Type 2 diabetes mellitus with ulcer
45766051	Type I diabetes mellitus in remission
45766052	Type II diabetes mellitus in remission
37116960	Wolfram-like syndrome
4322638	Wolfram syndrome
37168689	Wolfram syndrome type 1
37168690	Wolfram syndrome type 2
37311329	Woodhouse Sakati syndrome
4099653	Steroid-induced diabetes mellitus without complication

Concept ID	Concept name
37116379	Stimmler syndrome
4129378	Transient neonatal diabetes mellitus
201254	Type 1 diabetes mellitus
4099215	Type 1 diabetes mellitus maturity onset
443412	Type 1 diabetes mellitus without complication
4099214	Type 1 diabetes mellitus with ulcer
201826	Type 2 diabetes mellitus
4230254	Type 2 diabetes mellitus in nonobese
4304377	Type 2 diabetes mellitus in obese
4193704	Type 2 diabetes mellitus without complication
4099651	Type 2 diabetes mellitus with ulcer
45766051	Type I diabetes mellitus in remission
45766052	Type II diabetes mellitus in remission
37116960	Wolfram-like syndrome
4322638	Wolfram syndrome
37168689	Wolfram syndrome type 1
37168690	Wolfram syndrome type 2
37311329	Woodhouse Sakati syndrome

Epilepsy/Seizures:

Concept ID	Concept name
35625535	5q31.3 microdeletion syndrome
4309044	Abdominal seizure
4194232	Absence seizure
4078335	Absence seizure with atonic components
4086180	Absence seizure with automatisms
4078337	Absence seizure with autonomic components
4086179	Absence seizure with impairment of consciousness only
4078334	Absence seizure with mild clonic components
4078336	Absence seizure with tonic components
4046219	Acquired epileptic aphasia
765610	Acquired epileptic aphasia, non-refractory
763039	Acquired epileptic aphasia, refractory
35607997	Action myoclonus renal failure syndrome
35623629	Acute encephalopathy with biphasic seizures and late reduced diffusion
37208113	Acute repetitive seizure
3654658	Acute seizure due to infection of central nervous system

Concept ID	Concept name
3661779	Adverse reaction to benzodiazepine
3661831	Adverse reaction to clorazepate
442022	Adverse reaction to oxazolidine derivative
4316986	Adverse reaction to paramethadione
4317250	Adverse reaction to pentobarbitone
4319413	Adverse reaction to phensuximide
436603	Adverse reaction to succinimide
4315644	Adverse reaction to trimethadione
40485492	Afebrile seizure
4079057	Affective seizure
4165104	Akinetic seizure without atonia
4202330	Alcohol withdrawal-induced convulsion
4176652	Alcohol-induced epilepsy
4176651	Alcohol-related fit
37312387	Alopecia, epilepsy, intellectual disability syndrome Moynahan type
4095669	Alprazolam adverse reaction
4098987	Aminoglutethimide adverse reaction
4044083	Amygdalo-hippocampal epilepsy
4231042	Anoxic epileptic seizure
4216135	Anoxic seizure
4044220	Anterior frontopolar epilepsy
432809	Antiepileptic adverse reaction
4046223	Aquagenic epilepsy
36715051	Atherosclerosis, deafness, diabetes, epilepsy, nephropathy syndrome
43531582	Atonic epilepsy
4173765	Atonic seizure
762830	Atonic seizure, non-refractory
762829	Atonic seizure, refractory
43530704	Atypical absence epilepsy
4050738	Atypical absence seizure
762980	Atypical absence seizure, non-refractory
762979	Atypical absence seizure, refractory
3190189	Atypical seizure
35622969	Audiogenic epilepsy
4267146	Auditory seizure
36675122	Autism epilepsy syndrome due to branched chain ketoacid dehydrogenase kinase deficiency
37116742	Autism spectrum disorder, epilepsy, arthrogryposis syndrome

Concept ID	Concept name
37205088	Autosomal dominant epilepsy with auditory features
44783830	Autosomal dominant nocturnal frontal lobe epilepsy
763527	Autosomal dominant nocturnal frontal lobe epilepsy, non-refractory
763526	Autosomal dominant nocturnal frontal lobe epilepsy, refractory
36676588	Autosomal recessive cerebellar ataxia, epilepsy, intellectual disability syndrome due to TUD deficiency
36674914	Autosomal recessive cerebellar ataxia, epilepsy, intellectual disability syndrome due to WWOX deficiency
4097344	Barbiturate antiepileptic adverse reaction
4097343	Beclamide adverse reaction
37397175	Benign adult familial myoclonic epilepsy
4044079	Benign atypical partial epilepsy in childhood
762981	Benign childhood epilepsy with centrotemporal spikes, non-refractory
765908	Benign childhood epilepsy with centrotemporal spikes, refractory
36674689	Benign familial mesial temporal lobe epilepsy
762706	Benign familial neonatal seizures, non-refractory
762705	Benign familial neonatal seizures, refractory
4048424	Benign focal epilepsy of childhood
37395845	Benign focal seizure of adolescence
4043404	Benign frontal epilepsy of childhood
36674783	Benign infantile focal epilepsy with midline spikes and waves during sleep
35623412	Benign infantile seizure with mild gastroenteritis syndrome
4101742	Benign myoclonic epilepsy in infancy
765786	Benign myoclonic epilepsy in infancy, non-refractory
762728	Benign myoclonic epilepsy in infancy, refractory
4244383	Benign neonatal convulsions
4043411	Benign neonatal familial convulsions
4046209	Benign non-familial neonatal convulsions
4043405	Benign occipital epilepsy of childhood - early onset variant
4044081	Benign occipital epilepsy of childhood - late onset variant
762788	Benign occipital epilepsy of childhood - late onset, non-refractory
762787	Benign occipital epilepsy of childhood - late onset, refractory
36674784	Benign occipital lobe epilepsy
36674785	Benign partial epilepsy of infancy with complex partial seizures
36674988	Benign partial epilepsy with secondarily generalized seizures in infancy
4044078	Benign psychomotor epilepsy of childhood
4185733	Benign Rolandic epilepsy
4245285	Brief atonic seizure
4097497	Bromazepam adverse reaction

Concept ID	Concept name
4099416	Carbamazepine adverse reaction
36716394	Celiac disease with epilepsy and cerebral calcification syndrome
4243340	Central convulsion
4171694	Centrencephalic epilepsy
4179936	Childhood absence epilepsy
762734	Childhood absence epilepsy, non-refractory
762733	Childhood absence epilepsy, refractory
4044080	Childhood epilepsy with occipital paroxysms
40481467	Childhood seizure
4097498	Chlordiazepoxide adverse reaction
4099427	Chlormethiazole adverse reaction
4219724	Chronic progressive epilepsia partialis continua
4047894	Chronic progressive epilepsia partialis continua of childhood
4047889	Cingulate epilepsy
4097351	Clobazam adverse reaction
4099417	Clonazepam adverse reaction
4265182	Clonic seizure
765510	Clonic seizure, non-refractory
765497	Clonic seizure, refractory
3188590	Cognitive dysfunction with epilepsy
4086177	Cognitive seizure
45771300	Complete partial seizure of uncertain origin
374168	Complex febrile seizure
762914	Complex febrile seizure, non-refractory
765872	Complex febrile seizure, refractory
4081162	Complex part seizure with impairment of consciousness only
4236312	Complex partial epileptic seizure
4254546	Complex partial seizure + impairment consciousness at onset
4198103	Complex partial seizure evolving to generalized seizure
46273398	Complex partial seizure of frontal lobe
46273399	Complex partial seizure of occipital lobe
46273397	Complex partial seizure of parietal lobe
37018623	Complex partial seizure of temporal lobe
4165417	Complex partial seizure with impairment of consciousness
4043559	Complex partial status epilepticus
762978	Complex partial status epilepticus, non-refractory
762977	Complex partial status epilepticus, refractory

Concept ID	Concept name
37204238	Congenital muscular dystrophy with intellectual disability and severe epilepsy
380533	Convulsions in the newborn
4229670	Convulsive syncope
4132128	Coordinate convulsion
36674989	Cortical dysplasia with focal epilepsy syndrome
4046211	Cryptogenic generalized epilepsy
36680607	Cryptogenic late-onset epileptic spasms
4044224	Cryptogenic Lennox-Gastaut syndrome
4047900	Cryptogenic myoclonic epilepsy
4043412	Cryptogenic West syndrome
4102347	Cursive (running) epilepsy
36676367	Cyclin-dependent kinase-like 5 deficiency
4046222	Decision-making epilepsy
36715417	DEND syndrome
4097501	Diazepam adverse reaction
37204217	Diffuse cerebral and cerebellar atrophy, intractable seizures, progressive microcephaly syndrome
4098316	Dissociative convulsions
4047890	Dorsolateral epilepsy
4043554	Drug-induced epilepsy
36685474	Drug-induced seizure
4107690	Dysmnestic seizure
4086176	Dysphasic seizure
4043414	Early infantile epileptic encephalopathy with suppression bursts
762727	Early infantile epileptic encephalopathy, non-refractory
762726	Early infantile epileptic encephalopathy, refractory
36674712	Early-onset epileptic encephalopathy and intellectual disability due to GRIN2A mutation
36676621	Early-onset epileptic encephalopathy, cortical blindness, intellectual disability, facial dysmorphism syndrome
37116386	Early-onset Lafora body disease
36675141	Early-onset spastic ataxia, myoclonic epilepsy, neuropathy syndrome
4046221	Eating epilepsy
443700	Eclampsia
133816	Eclampsia - delivered
138811	Eclampsia - delivered with postnatal complication
4289142	Eclampsia added to pre-existing hypertension
4034096	Eclampsia in labour
137613	Eclampsia in pregnancy
4116344	Eclampsia in puerperium

Concept ID	Concept name
4058536	Eclampsia with postnatal complication
45757119	Eclampsia with pre-existing hypertension in childbirth
4145352	Eclamptic seizure
374023	Epilepsia partialis continua
762832	Epilepsia partialis continua, non-refractory
380378	Epilepsy
4015832	Epilepsy associated problems
40483585	Epilepsy characterized by intractable complex partial seizures
37110673	Epilepsy co-occurrent and due to degenerative brain disorder
37110677	Epilepsy co-occurrent and due to dementia
37110676	Epilepsy co-occurrent and due to demyelinating disorder
37110674	Epilepsy co-occurrent and due to mesial temporal sclerosis
3178390	Epilepsy complicating pregnancy
4151917	Epilepsy control good
443379	Epilepsy control poor
4015959	Epilepsy drug side effect
37110522	Epilepsy due to cerebrovascular accident
37110675	Epilepsy due to immune disorder
37110329	Epilepsy due to infectious disease of central nervous system
3655274	Epilepsy due to infectious encephalitis
3655276	Epilepsy due to infectious meningitis
37110524	Epilepsy due to intracranial tumour
3655285	Epilepsy due to neonatal central nervous system infection
37119155	Epilepsy due to perinatal anoxic-ischemic brain injury
37110521	Epilepsy due to perinatal stroke
45757114	Epilepsy in mother complicating childbirth
45757050	Epilepsy in mother complicating pregnancy
37118656	Epilepsy of infancy with migrating focal seizures
37116355	Epilepsy telangiectasia syndrome
4046218	Epilepsy undetermined whether focal or generalized
4043551	Epilepsy with continuous spike wave during slow-wave sleep
4047897	Epilepsy with grand mal seizures on awakening
37116354	Epilepsy, microcephaly, skeletal dysplasia syndrome
4310999	Epilepsy, not refractory
4104857	Epileptic automatism
4251337	Epileptic cry
37018608	Epileptic dementia with behavioural disturbance

Concept ID	Concept name
4096378	Epileptic drop attack
36716897	Epileptic encephalopathy
37111648	Epileptic encephalopathy with global cerebral demyelination
4338020	Epileptic psychosis
4197485	Epileptic seizure
4101741	Epileptic seizures - akinetic
4105360	Epileptic seizures - atonic
4104853	Epileptic seizures - clonic
4102344	Epileptic seizures - tonic
4286178	Epileptic vertigo
4097346	Ethosuximide adverse reaction
1340285	Exacerbation of complex partial seizure with impairment of consciousness
1340323	Exacerbation of epilepsy
1340467	Exacerbation of seizure
4009655	Extratemporal epilepsy
4046220	Eyelid myoclonus with absences
4047902	Familial febrile convulsions
35608023	Familial focal epilepsy with variable foci
37205064	Familial infantile myoclonic epilepsy
37205085	Familial mesial temporal lobe epilepsy with febrile seizures
4089691	Familial neonatal seizures
37204833	Familial temporal lobe epilepsy
444413	Febrile convulsion
37110888	Febrile infection related epilepsy syndrome
37396778	Female restricted epilepsy with intellectual disability syndrome
4171110	Fifth day fits
4102730	Finding of measure of fit (seizure)
4103829	Finding relating to fit
43530759	First generalized onset seizure
4097350	Flunitrazepam adverse reaction
4099421	Flurazepam adverse reaction
35622869	Focal epilepsy, intellectual disability, cerebro-cerebellar malformation syndrome
4027257	Focal motor seizure
40486152	Focal seizure with experiential sensory symptoms
37399364	Folinic acid responsive seizure syndrome
4047888	Frontal lobe epilepsy
4097488	Gabapentin adverse reaction

Concept ID	Concept name
4232618	Gelastic seizure
374924	Generalized convulsive epilepsy
765787	Generalized convulsive epilepsy, non-refractory
4055361	Generalized epilepsy
37395981	Generalized epilepsy and paroxysmal dyskinesia syndrome
44783583	Generalized epilepsy with febrile seizures plus
763656	Generalized epilepsy with febrile seizures plus, non-refractory
763655	Generalized epilepsy with febrile seizures plus, refractory
44784590	Generalized non-convulsive absence epilepsy
376680	Generalized non-convulsive epilepsy
4078333	Generalized seizure
4263736	Generalized-onset seizures
36716652	GM3 synthase deficiency
373748	Grand mal status
762956	Grand mal status epilepticus, non-refractory
762955	Grand mal status epilepticus, refractory
4179207	Gustatory seizure
4044222	Hemiplegia-hemiconvulsion-epilepsy syndrome
35622175	Hot water reflex epilepsy
37206121	Hyperekplexia epilepsy syndrome
4203424	Hypoglycaemia-induced convulsion
35624272	Hypothalamic hamartoma with gelastic seizure
37395921	ICCA syndrome
4249451	Ideational partial seizure
4274575	Idiopathic generalized epilepsy
762976	Idiopathic generalized epilepsy, non-refractory
4047901	Idiopathic myoclonic epilepsy
762088	Idiopathic partial status epilepticus
37110163	Infant epilepsy with migrant focal crisis
36674269	Infantile epileptic dyskinetic encephalopathy
36674719	Infantile spasm and broad thumb syndrome
46270369	Infantile spasms co-occurrent with status epilepticus
762793	Infantile spasms, non-refractory
36676518	Infantile-onset mesial temporal lobe epilepsy with severe cognitive regression
36674191	Intellectual disability, seizures, hypotonia, ophthalmologic, skeletal anomalies syndrome
43531630	Intractable absence seizures
762350	Intractable atypical absence epilepsy

Concept ID	Concept name
46273400	Intractable complex partial parietal lobe epilepsy
42535414	Intractable idiopathic partial epilepsy
761783	Intractable juvenile myoclonic status epilepticus
46273401	Intractable partial frontal lobe epilepsy with impairment of consciousness
46274016	Intractable partial occipital lobe epilepsy with impairment of consciousness
43531718	Intractable partial temporal lobe epilepsy with impairment of consciousness
43530628	Intractable simple partial epilepsy
4046216	Isolated seizures
4145896	Jacksonian, focal or motor epilepsy
37396462	Jeavons syndrome
4046210	Juvenile absence epilepsy
762759	Juvenile absence epilepsy, non-refractory
762758	Juvenile absence epilepsy, refractory
4267274	Juvenile myoclonic epilepsy
763041	Juvenile myoclonic epilepsy, non-refractory
36680579	KCNQ2-related epileptic encephalopathy
4095670	Ketazolam adverse reaction
4033911	Kohlschutter's syndrome
4046214	Lafora disease
4095659	Lamotrigine adverse reaction
4043408	Lateral temporal epilepsy
4046213	Lennox-Gastaut syndrome
762730	Lennox-Gastaut syndrome, non-refractory
762729	Lennox-Gastaut syndrome, refractory
4047895	Localization-related cryptogenic epilepsy
374915	Localization-related epilepsy
4150299	Localization-related idiopathic epilepsy
4044082	Localization-related symptomatic epilepsy
4044221	Localization-related symptomatic epilepsy with specific precipitant
4337361	Long atonic seizure
4095663	Loprazolam adverse reaction
4099424	Lorazepam adverse reaction
4097496	Lormetazepam adverse reaction
4078529	Lowered convulsive threshold
4097182	Magnesium sulfate adverse reaction
37205067	Malignant migrating partial seizures of infancy
4097499	Medazepam adverse reaction

Concept ID	Concept name
4047903	Menstrual epilepsy
36674804	Mesial temporal lobe epilepsy with hippocampal sclerosis
4104856	Mesiobasal limbic epilepsy
4095661	Methylphenobarbitone adverse reaction
37118888	Microcephaly, seizure, intellectual disability, heart disease syndrome
35622355	Micturition induced epilepsy
4097500	Midazolam adverse reaction
762790	Migrating partial seizures in infancy
762792	Migrating partial seizures in infancy, non-refractory
762791	Migrating partial seizures in infancy, refractory
4147503	Motor cortex epilepsy
4139559	Motor epilepsy
4023976	Movement partial seizure
37205773	Multiple congenital anomalies, hypotonia, seizures syndrome
36674383	Multiple congenital anomalies, hypotonia, seizures syndrome type 2
4056835	Musicogenic seizure
4044225	Myoclonic absence epilepsy
762732	Myoclonic absence epilepsy, non-refractory
762731	Myoclonic absence epilepsy, refractory
4043413	Myoclonic astatic epilepsy
763309	Myoclonic astatic epilepsy, non-refractory
763308	Myoclonic astatic epilepsy, refractory
4104714	Myoclonic encephalopathy
36680599	Myoclonic epilepsy in non-progressive encephalopathy
4044223	Myoclonic epilepsy of early childhood
4044227	Myoclonic epilepsy with ragged red fibers
4290772	Myoclonic seizure
762804	Myoclonic seizure, non-refractory
762963	Myoclonic seizure, refractory
4043555	Narcotic withdrawal epilepsy
3188298	New onset seizure
36674194	New-onset refractory status epilepticus
4095667	Nitrazepam adverse reaction
4044229	Nocturnal epilepsy
4043558	Non-convulsive simple partial status epilepticus
4044231	Non-convulsive status epilepticus with three per second spike wave
4043557	Non-convulsive status epilepticus without three per second spike wave

Concept ID	Concept name
44811346	Non-epileptic attack disorder
4047892	Non-progressive Kozhevnikov syndrome
607637	Nonconvulsive seizure
40483620	Nonconvulsive status epilepticus
4196709	Observations of measures of seizures - finding
4046207	Occipital lobe epilepsy
4336820	Olfactory seizure
4043410	Opercular epilepsy
4043409	Orbitofrontal epilepsy
35622368	Orgasm induced epilepsy
36716191	Osteogenesis imperfecta, retinopathy, seizures, intellectual disability syndrome
4099423	Oxazepam adverse reaction
35607971	Pachygyria, intellectual disability, epilepsy syndrome
443342	Paraldehyde adverse reaction
4132780	Paranoid-hallucinatory epileptic psychosis
4046206	Parietal lobe epilepsy
36676587	Partial corpus callosum agenesis, cerebellar vermis hypoplasia with posterior fossa cysts syndrome
4104859	Partial epilepsy with autonomic symptoms
439654	Partial epilepsy with impairment of consciousness
43530676	Partial frontal lobe epilepsy
43530706	Partial occipital lobe epilepsy
43530708	Partial parietal lobe epilepsy
4106574	Partial seizure
4078332	Partial seizure evolving to secondary generalized seizure
4174399	Partial seizure with illusions and hallucinations
4029783	Partial seizure with impaired consciousness
4244738	Partial seizure with multiple symptoms
4245301	Pattern sensitive seizure
37109948	Perceptual disturbances and seizures co-occurrent and due to alcohol withdrawal
37110424	Perceptual disturbances and seizures co-occurrent and due to hypnotic withdrawal
37110471	Perceptual disturbances and seizures co-occurrent and due to psychoactive substance withdrawal
37110423	Perceptual disturbances and seizures co-occurrent and due to sedative withdrawal
35624201	Perioral myoclonia with absences
381842	Petit mal status
765902	Petit mal status, non-refractory
765661	Petit mal status, refractory
44801879	Petit-mal epilepsy

Concept ID	Concept name
4097345	Phenobarbitone adverse reaction
4097491	Phenytoin adverse reaction
4086175	Phonatory seizure
4316060	Photogenic epilepsy
4030066	Photomyoclonus, diabetes mellitus, deafness, nephropathy and cerebral dysfunction
4095660	Piracetam adverse reaction
36676719	Pitt Hopkins-like syndrome
45765412	Pitt-Hopkins syndrome
44784593	Post infectious grand mal epilepsy
44784595	Post-cerebrovascular accident epilepsy
4209118	Post-ictal drowsiness
4135349	Post-ictal state
4326435	Post-traumatic epilepsy
762827	Post-traumatic epilepsy, non-refractory
762826	Post-traumatic epilepsy, refractory
44784592	Postoperative status epilepticus
4195615	Postseizure confusion
4177642	Postseizure delirium
4028205	Postseizure headache
4233724	Posttraumatic seizure
4086174	Postural seizure
4095671	Prazepam adverse reaction
3174006	Pregnancy complicated by epilepsy
43530757	Primary generalized absence epilepsy
37312576	Primary hyperaldosteronism, seizures, neurological abnormalities syndrome
4043406	Primary inherited reading epilepsy
37204277	Primary microcephaly, epilepsy, permanent neonatal diabetes syndrome
4095662	Primidone adverse reaction
40483619	Progressive encephalopathy with edema, hypsarrhythmia and optic atrophy syndrome
36674825	Progressive encephalopathy with edema, hypsarrhythmia, and optic atrophy-like syndrome
4147501	Progressive myoclonic epilepsy
37204409	Progressive myoclonic epilepsy type 3
37204403	Progressive myoclonic epilepsy type 5
37203831	Progressive myoclonic epilepsy type 6
37204476	Progressive myoclonic epilepsy type 8
35622037	Progressive myoclonic epilepsy with dystonia
45773528	Progressive myoclonus epilepsy with ataxia

Concept ID	Concept name
4261957	Psychomotor epilepsy
4104855	Psychosensory epilepsy
4237826	Puerperal convulsion
35625463	PURA syndrome
37110349	Pyridoxal 5-phosphate dependent epilepsy
42535907	Pyridoxine-dependent epilepsy
4168478	Raised convulsive threshold
4041672	Rasmussen syndrome
763040	Rasmussen syndrome, refractory
4297994	Reading seizure
35624480	Recurrent benign focal seizures of childhood
42534952	Recurrent complex partial epilepsy
4143868	Recurrent febrile convulsion
37208117	Recurrent seizure
3171021	Recurrent seizures
4234029	Reflex anoxic seizure
4194366	Reflex epilepsy
43530619	Refractory complex partial seizure with impairment of consciousness
4311977	Refractory epilepsy partialis continua
40483317	Refractory epilepsy
4312246	Refractory frontal lobe epilepsy
43530734	Refractory generalized convulsive epilepsy
40480055	Refractory generalized nonconvulsive epilepsy
46271795	Refractory idiopathic generalized epilepsy
4306496	Refractory infantile spasms
46270370	Refractory infantile spasms co-occurrent with status epilepticus
44784591	Refractory juvenile myoclonic epilepsy
4309257	Refractory localization-related epilepsy
44782522	Refractory myoclonic epilepsy
4314174	Refractory occipital lobe epilepsy
4314726	Refractory parietal lobe epilepsy
4255159	Retropulsion petit mal
4046111	Rhinencephalic epilepsy
35622873	Rolandic epilepsy, speech dyspraxia syndrome
4058001	Salaam spasm
3657974	Scar epilepsy
35622929	SCN8A-related epilepsy with encephalopathy

Concept ID	Concept name
4070660	Secondarily generalized seizures
4043552	Secondary reading epilepsy
37110420	Sedative withdrawal with seizure
377091	Seizure
4079058	Seizure causing illusions
37110422	Seizure co-occurrent and due to anxiolytic withdrawal
42536419	Seizure co-occurrent and due to drug withdrawal
37110421	Seizure co-occurrent and due to hypnotic withdrawal
37110470	Seizure co-occurrent and due to psychoactive substance withdrawal
4029498	Seizure disorder
43530665	Seizure disorder as sequela of stroke
4196708	Seizure related finding
4174747	Seizure undetermined whether focal or generalized
4078330	Seizure with structured hallucinations
36715509	Seizure, sensorineural deafness, ataxia, intellectual disability, electrolyte imbalance syndrome
4159752	Seizures complicating infection
4186827	Seizures complicating infection in the newborn
4159150	Seizures complicating intracranial haemorrhage
4159149	Seizures complicating intracranial haemorrhage in the newborn
4154695	Seizures due to metabolic disorder
762709	Seizures in the newborn, non-refractory
762579	Seizures in the newborn, refractory
4047907	Self-induced non-photosensitive epilepsy
37118973	Severe intellectual disability, epilepsy, anal anomaly, distal phalangeal hypoplasia syndrome
4043550	Severe myoclonic epilepsy in infancy
762805	Severe myoclonic epilepsy in infancy, non-refractory
762806	Severe myoclonic epilepsy in infancy, refractory
36675039	Severe neonatal onset encephalopathy with microcephaly
377101	Simple febrile seizure
762948	Simple febrile seizure, non-refractory
762947	Simple febrile seizure, refractory
44789005	Simple partial epileptic seizure
4078331	Simple partial onset of seizure with automatisms
4171667	Simple partial onset seizure followed by impaired consciousness
43530626	Simple partial seizure
4264143	Simple partial seizure evolving to complex partial seizure evolving to generalized seizure
4245044	Simple partial seizure evolving to secondary generalized seizure

Concept ID	Concept name
4081161	Simple partial seizure followed by impaired consciousness
4251921	Simple partial seizure with autonomic dysfunction
4144967	Simple partial seizure with disturbance of higher cerebral function
4081160	Simple partial seizure with focal motor signs with march
4078328	Simple partial seizure with focal motor signs without march
4219625	Simple partial seizure with motor dysfunction
763151	Simple partial seizure with motor dysfunction, non-refractory
763150	Simple partial seizure with motor dysfunction, refractory
4079056	Simple partial seizure with somatosensory or special sensory dysfunction
4215937	Simple partial seizure with special sensory symptoms
4194056	Simple partial seizure, consciousness not impaired
4313035	Simple partial status epilepticus
763096	Simple partial status epilepticus, non-refractory
42873167	Single epileptic seizure
45766063	Single seizure
36716898	Single seizure due to remote cause
36716899	Single unprovoked seizure
4046215	Situation-related seizures
4097489	Sodium valproate adverse reaction
4104858	Somatosensory epilepsy
4146378	Somatosensory seizure
45766380	Spinal muscular atrophy with progressive myoclonic epilepsy
35622261	Startle epilepsy
4200704	Startle partial seizure
4332304	Status epilepticus
37016270	Status epilepticus due to complex partial epileptic seizure
46270365	Status epilepticus due to generalized idiopathic epilepsy
46270366	Status epilepticus due to intractable idiopathic generalized epilepsy
46270364	Status epilepticus due to refractory complex partial seizures
46270367	Status epilepticus due to refractory epilepsy
46270368	Status epilepticus due to refractory simple partial epilepsy
43531626	Status epilepticus in benign Rolandic epilepsy
4000171	Stress seizure
4133780	Stress-induced epilepsy
35625621	STXBP1 encephalopathy with epilepsy
4044084	Supplementary motor epilepsy
4216711	Symptomatic generalized epilepsy

Concept ID	Concept name
4047898	Symptomatic Lennox-Gastaut syndrome
4043549	Symptomatic myoclonic epilepsy
4046212	Symptomatic West syndrome
36683256	SYNGAP1-related intellectual disability
4095672	Temazepam adverse reaction
4102345	Temporal lobe epilepsy
764684	Temporal lobe epilepsy with mesial temporal sclerosis
762814	Temporal lobe epilepsy, non-refractory
4323821	Tetanic convulsion
35622254	Thinking epilepsy
4282301	Todd's paresis
4198075	Tonic seizure
762813	Tonic seizures, non-refractory
762812	Tonic seizures, refractory
4232071	Tonic-clonic epilepsy
4183856	Tonic-clonic seizure
762957	Tonic-clonic seizure, non-refractory
765514	Tonic-clonic seizure, refractory
4043556	Toothbrushing epilepsy
4193675	Transient epileptic amnesia
4095668	Triazolam adverse reaction
762809	Typical absence seizure
762834	Typical absence seizure, non-refractory
765511	Typical absence seizure, refractory
4044226	Unverricht-Lundborg syndrome
4170562	Uremic convulsion
4086173	Versive seizure
4097490	Vigabatrin adverse reaction
4317514	Visceral epilepsy
4213017	Visual epilepsy
4101744	Visual reflex epilepsy
4032157	Visual seizure
376105	West syndrome
40482060	Witnessed epileptic seizure
4047904	Writing epilepsy
37397174	X-linked epilepsy with learning disability and behaviour disorder syndrome
36714528	X-linked intellectual disability with seizure and psoriasis syndrome

Concept ID	Concept name
37110783	X-linked spasticity, intellectual disability, epilepsy syndrome

Essential hypertension:

Concept ID	Concept name
312648	Benign essential hypertension
4215640	Benign essential hypertension complicating AND/OR reason for care during childbirth
4034031	Benign essential hypertension complicating AND/OR reason for care during pregnancy
4148205	Benign essential hypertension complicating AND/OR reason for care during puerperium
4269358	Benign essential hypertension in obstetric context
320128	Essential hypertension
4083723	Essential hypertension complicating AND/OR reason for care during childbirth
4302591	Essential hypertension complicating AND/OR reason for care during pregnancy
4321603	Essential hypertension complicating AND/OR reason for care during puerperium
4217486	Essential hypertension in obstetric context
4058987	High-renin essential hypertension
4159755	Labile essential hypertension
4263067	Low-renin essential hypertension
317898	Malignant essential hypertension
45757787	Postpartum pre-existing essential hypertension
4180283	Systolic essential hypertension

Heart disease (excluding hypertension):

Concept ID ¹	Concept name
321588	Heart disease

¹All descendants included. Given the large number of descendants, only concepts excluded are listed.

Codes excluded (with descendants):	
Concept ID	Concept name
42599731	Ventricular septal defect with right-to-left shunt
42599730	Ventricular septal defect with left-to-right shunt
42573220	Unclassified cardiomyopathy
36518281	T-cell/histiocyte rich large B-cell lymphoma of heart
36538587	T-cell large granular lymphocytic leukemia of heart
36555797	Systemic EBV positive T-cell lymphoproliferative disease of childhood of heart
44501907	Synovial sarcoma, spindle cell of heart
36523685	Synovial sarcoma, NOS, of heart
44500812	Synovial sarcoma, biphasic of heart
42573212	Symmetrical hypertrophic cardiomyopathy

Codes excluded (with descendants):	
Concept ID	Concept name
42595751	Subvalvular fibrous band stenosis
42595753	Subepicardial hemorrhage
42595752	Subendocardial hemorrhage
42595736	Subendocardial edema
36530304	Subcutaneous panniculitis-like T-cell lymphoma of heart
954104	Steroid-induced cardiomyopathy
42598803	Spontaneous cardiomyopathy
36554367	Splenic marginal zone B-cell lymphoma of heart
36542014	Spindle cell sarcoma of heart
42512386	Spindle cell carcinoma, NOS, of heart
42600137	Small heart
36519529	Sezary syndrome of heart
42599762	Segmental thinning of left ventricular posterior wall
42599763	Segmental thinning of cardiac apex
42599765	Segmental hypokinesis of ventricular septum
44499519	Sarcoma, NOS, of heart
954105	Ruptured left atrium
42597038	Rupture of aortic root
43020905	Right ventricular outflow tract obstruction due to neoplasm
36531218	Rhabdomyosarcoma, NOS, of heart
43021621	Restriction of coronary artery orifice due to neoplasm
954121	Reentrant left atrial tachycardia
4247841	Primary malignant neoplasm of myocardium
316644	Primary malignant neoplasm of heart
4247826	Primary malignant neoplasm of epicardium
4247237	Primary malignant neoplasm of endocardium
36528133	Primary effusion lymphoma of heart
36541002	Primary cutaneous anaplastic large cell lymphoma of heart
316429	Premature beats
36548342	Precursor T-cell lymphoblastic leukemia of heart
36536091	Precursor cell lymphoblastic lymphoma, NOS, of heart
1553525	Post-transplant lymphoproliferative disorder, NOS, of heart
36517876	Pleomorphic liposarcoma of heart
36561286	Plasmablastic lymphoma of heart
42599749	Physiologic mitral insufficiency
42599746	Physiologic aortic insufficiency

Codes excluded (with descendants):	
Concept ID	Concept name
44500668	Osteosarcoma, NOS, of heart
43020616	Obstruction of tricuspid valve due to neoplasm
43020620	Obstruction of pulmonary valve due to neoplasm
43020618	Obstruction of mitral valve due to neoplasm
43020626	Obstruction of aortic valve due to neoplasm
36539052	NK/T-cell lymphoma, nasal and nasal type of heart
4316512	Neoplasm of uncertain behavior of myocardium
4317431	Neoplasm of uncertain behavior of heart
4316380	Neoplasm of uncertain behavior of epicardium
4317415	Neoplasm of uncertain behavior of endocardium
4131458	Neoplasm of myocardium
4308814	Neoplasm of heart
4054529	Neoplasm of epicardium
4128892	Neoplasm of endocardium
42597033	Neoplasm of base of heart
443523	Neonatal tachycardia
37162093	Neonatal encephalomyopathy, cardiomyopathy, respiratory distress syndrome
4173170	Neonatal dysrhythmia
4110550	Neonatal cardiorespiratory arrest
37110330	Neonatal cardiac failure due to pulmonary overperfusion
36716748	Neonatal cardiac failure due to decreased left ventricular output
4172864	Neonatal cardiac failure
4106274	Neonatal cardiac arrest
443522	Neonatal bradycardia
36527721	Myxosarcoma of heart
36549549	Myxoma, NOS, of heart
42599754	Myxoid transformation of pulmonary valve
42599759	Myxoid transformation of mitral valve ACVIM stage D
42599758	Myxoid transformation of mitral valve ACVIM stage C
42599757	Myxoid transformation of mitral valve ACVIM stage B2
42599756	Myxoid transformation of mitral valve ACVIM stage B1
42599755	Myxoid transformation of mitral valve ACVIM stage A
42512590	Myxofibrosarcoma of heart
36562876	Myofibroblastic tumor, NOS, of heart
42512404	Myoepithelial carcinoma of heart
42596125	Myocardial mineralization

Codes excluded (with descendants):	
Concept ID	Concept name
42600077	Myocardial fibrosis
42596123	Myocardial carnitine concentration decreased
36532643	Myeloproliferative neoplasm, unclassifiable of heart
36535843	Myeloid or lymphoid neoplasm with PDGFRA rearrangement of heart
36535075	Myeloid or lymphoid neoplasm with FGFR1 abnormalities of heart
36532342	Mycosis fungoides of heart
42599624	Mulberry heart disease
42593956	Mobitz type Ib incomplete atrioventricular block
42593955	Mobitz type Ia incomplete atrioventricular block
42599795	Mitral valve regurgitation during diastole
42596119	Mitral valve leaflet flailing
44500749	Mesothelioma, malignant of heart
36529461	Mediastinal large B-cell lymphoma of heart
36532191	Mature T-cell lymphoma, NOS, of heart
42599790	Mass of tricuspid valve
36542195	Marginal zone B-cell lymphoma, NOS, of heart
36530187	Mantle cell lymphoma of heart
36525630	Malignant lymphoma, NOS, of heart
36537888	Malignant lymphoma, non-Hodgkin, NOS, of heart
36555771	Malignant lymphoma, mixed small and large cell, diffuse of heart
36563698	Malignant lymphoma, large B-cell, diffuse, immunoblastic, NOS, of heart
3655615	Malignant germ cell neoplasm of heart
44502891	Malignant fibrous histiocytoma of heart
36560539	Lymphoplasmacytic lymphoma of heart
42597337	Lymphoma of heart
36548975	Lipoma, NOS, of heart
36551933	Leiomyosarcoma, NOS, of heart
43020907	Left ventricular outflow tract obstruction due to neoplasm
36526264	Langerhans cell histiocytosis, disseminated of heart
954184	JDCM - Juvenile dilated cardiomyopathy
36531832	Intravascular large B-cell lymphoma of heart
42511975	Intimal sarcoma of heart
36540849	Intestinal T-cell lymphoma of heart
42596126	Infiltrative myocardial disease
37395937	Idiopathic neonatal atrial flutter
42573213	Hypertrophic cardiomyopathy with papillary muscle hypertrophy

Codes excluded (with descendants):	
Concept ID	Concept name
36524786	Hodgkin sarcoma of heart
36536539	Hodgkin lymphoma, NOS, of heart
36563875	Hodgkin lymphoma, nodular sclerosis, NOS, of heart
36532229	Hodgkin lymphoma, nodular sclerosis, grade 2 of heart
36562670	Hodgkin lymphoma, nodular sclerosis, grade 1 of heart
36557496	Hodgkin lymphoma, nodular sclerosis, cellular phase of heart
36517564	Hodgkin lymphoma, nodular lymphocyte predominant of heart
36531663	Hodgkin lymphoma, mixed cellularity, NOS, of heart
36517582	Hodgkin lymphoma, lymphocyte-rich of heart
36535390	Hodgkin lymphoma, lymphocyte depletion, reticular of heart
36551920	Hodgkin lymphoma, lymphocyte depletion, NOS, of heart
36547333	Hodgkin lymphoma, lymphocyte depletion, diffuse fibrosis of heart
36535224	Hodgkin granuloma of heart
44503065	Histiocytic sarcoma of heart
36555464	HHV8 positive diffuse large B-cell lymphoma of heart
36527246	Hepatosplenic T-cell lymphoma of heart
42596239	Hemangiosarcoma of heart
44498979	Hemangiosarcoma of heart
42511756	Hemangioendothelioma, malignant of heart
42599350	Heartworm disease without clinical signs
42599368	Heartworm disease with clinical signs
1074906	Heart failure following administration of antineoplastic agent
42596127	Granulomatous inflammation of myocardium
42512799	Giant cell sarcoma of heart
42599769	German shepherd ventricular arrhythmia
36559929	Follicular lymphoma, NOS, of heart
36531989	Follicular lymphoma, grade 3 of heart
36522198	Follicular lymphoma, grade 2 of heart
36541693	Follicular lymphoma, grade 1 of heart
42595741	Focal myocarditis
954119	Focal left atrial tachycardia
42512705	Fibrous mesothelioma, malignant of heart
42573350	Fibrosing cardiomyopathy
36539627	Fibroma, NOS, of heart
36565617	Fibroblastic reticular cell tumor of heart
44504638	Fetal rhabdomyoma of heart

Codes excluded (with descendants):	
Concept ID	Concept name
4334638	Fetal heart disorder
4172863	Fetal dysrhythmia
42573033	Feline restrictive cardiomyopathy
42573032	Feline intergrade cardiomyopathy
42511863	Epithelioid mesothelioma, malignant of heart
36550734	Epithelioid hemangioendothelioma, NOS, of heart
42574210	Endomyocarditis
42597026	Endocarditis of tricuspid valve
42597025	Endocarditis of pulmonic valve
42597024	Endocarditis of aortic valve
42573927	Endocardiosis of mitral valve
42595737	Endocardiosis
44500088	Embryonal rhabdomyosarcoma, NOS, of heart
42599747	Dynamic mid-left ventricular obstruction
42573200	Dynamic left ventricular outflow tract obstruction
42599777	Drug induced arrhythmia
44501609	Diffuse large B-cell lymphoma, NOS, of heart
42599750	Diastolic tricuspid valve regurgitation
36558174	Cutaneous T-cell lymphoma, NOS, of heart
45766238	Coronary artery occlusion due to neoplastic disease
42599736	Congenital valvular stenosis of pulmonary valve
42599737	Congenital valvular aortic stenosis
42599735	Congenital supra-ventricular stenosis of pulmonary valve
42599734	Congenital subvalvular stenosis of pulmonary valve
42599733	Congenital dynamic stenosis of pulmonary valve
36536953	Composite Hodgkin and non-Hodgkin lymphoma of heart
37399545	Childhood neoplasm of heart
42599761	Cardiomyopathy secondary to doxorubicin administration
42597039	Cardiomyopathy secondary to deficiency of selenium and/or vitamin E
42594384	Cardiomegaly of right ventricle
42594383	Cardiomegaly of left ventricle
42599748	Cardiac hypertrophy secondary to systemic hypertension
42599293	Canine restrictive cardiomyopathy
42599292	Canine intergrade cardiomyopathy
36567183	Burkitt lymphoma, NOS, of heart
4243439	Benign neoplasm of myocardium

Codes excluded (with descendants):	
Concept ID	Concept name
317814	Benign neoplasm of heart
4241031	Benign neoplasm of epicardium
4241646	Benign neoplasm of endocardium
36529212	B-cell chronic lymphocytic leukemia/small lymphocytic lymphoma of heart
36535836	B lymphoblastic leukemia/lymphoma, NOS, of heart
36548356	B lymphoblastic leukemia/lymphoma with t(v;11q23); MLL rearranged of heart
36566674	B lymphoblastic leukemia/lymphoma with t(9;22)(q34;q11.2); BCR-ABL1 of heart
36557442	B lymphoblastic leukemia/lymphoma with t(5;14)(q31;q32); IL3-IGH of heart
36549645	B lymphoblastic leukemia/lymphoma with t(1;19)(q23;p13.3); E2A-PBX1 (TCF3-PBX1) of heart
36566949	B lymphoblastic leukemia/lymphoma with t(12;21)(p13;q22); TEL-AML1 (ETV6-RUNX1) of heart
36555453	B lymphoblastic leukemia/lymphoma with hypodiploidy (Hypodiploid ALL) of heart
36547067	B lymphoblastic leukemia/lymphoma with hyperdiploidy of heart
954120	Atrial reentrant tachycardia - right
42593919	Atrial arrest
42599794	Asymptomatic dilated cardiomyopathy
42573214	Asymmetrical hypertrophic cardiomyopathy with hypertrophy of septum greater than free wall
42606061	Asymmetrical hypertrophic cardiomyopathy with basilar septal hypertrophy
42606060	Asymmetrical HCM, free wall greater than septum
42599797	Aorta overriding ventricular septal defect
42606062	Anomalous moderator band
36531889	Angioimmunoblastic T-cell lymphoma of heart
36529014	Anaplastic large cell lymphoma, T-cell and Null-cell type of heart
42597034	Amyloidosis of heart
36527483	ALK positive large B-cell lymphoma of heart
36530358	Adult cellular rhabdomyoma of heart

Human Immunodeficiency Virus (HIV) infection:

Concept ID	Concept name
432554	Human immunodeficiency virus II infection
439727	Human immunodeficiency virus infection
600510	Human immunodeficiency virus Centers for Disease Control and Prevention stage 1
600511	Human immunodeficiency virus Centers for Disease Control and Prevention stage 2
600512	Human immunodeficiency virus Centers for Disease Control and Prevention stage 3
600513	Human immunodeficiency virus Centers for Disease Control and Prevention stage unknown
604781	Human immunodeficiency virus Centers for Disease Control and Prevention dual infection with HIV-1 and HIV-2
604782	Human immunodeficiency virus Centers for Disease Control and Prevention undifferentiated type

Concept ID	Concept name
604783	Human immunodeficiency virus Centers for Disease Control and Prevention stage 0
604976	World Health Organization human immunodeficiency virus infection clinical stage
606040	Primary human immunodeficiency virus infection
606047	Human immunodeficiency virus modified skin disease
1340516	Progression of Human immunodeficiency virus infection
3654682	Chronic hepatitis C co-occurrent with human immunodeficiency virus infection
3654900	Encephalitis caused by human immunodeficiency virus type 2
3654950	Encephalitis caused by human immunodeficiency virus type 1
4008081	Acute HIV infection
4013106	HIV detected
4045982	Vacuolar myelopathy
4047624	Human immunodeficiency virus leukoencephalopathy
4048033	Neuropathy caused by human immunodeficiency virus
4078242	Human immunodeficiency virus
4087603	Human immunodeficiency virus infection constitutional disease
4087604	Human immunodeficiency virus infection with neurological disease
4092686	Human immunodeficiency virus infection with secondary clinical infectious disease
4128060	Acquired immune deficiency syndrome-related nephropathy
4161950	Human immunodeficiency virus encephalitis
4171124	Congenital acquired immune deficiency syndrome
4171125	Congenital human immunodeficiency virus positive status syndrome
4172009	HIV infection with infectious mononucleosis-like syndrome
4180254	Congenital human immunodeficiency virus infection
4201627	Aseptic meningitis due to human immunodeficiency virus infection
4211956	Neutropenia with AIDS (acquired immunodeficiency syndrome)
4220603	Skin rash with AIDS (acquired immunodeficiency syndrome)
4221489	Acquired immunodeficiency syndrome-associated disorder
4221503	Pneumococcal pneumonia with AIDS (acquired immunodeficiency syndrome)
4221607	Mycobacteriosis with AIDS (acquired immunodeficiency syndrome)
4221911	Fatigue with AIDS (acquired immunodeficiency syndrome)
4222062	Candidiasis of lung with AIDS (acquired immunodeficiency syndrome)
4222224	Hepatomegaly with AIDS (acquired immunodeficiency syndrome)
4222250	Encephalitis with AIDS (acquired immunodeficiency syndrome)
4222383	Intestinal malabsorption with AIDS (acquired immunodeficiency syndrome)
4222402	Hematopoietic system disease with AIDS (acquired immunodeficiency syndrome)
4222405	Acute endocarditis with AIDS (acquired immunodeficiency syndrome)
4222712	Cryptococcosis with AIDS (acquired immunodeficiency syndrome)

Concept ID	Concept name
4222717	Subacute adenoviral encephalitis with AIDS (acquired immunodeficiency syndrome)
4223032	Bacterial pneumonia with AIDS (acquired immunodeficiency syndrome)
4223276	Central nervous system demyelinating disease with AIDS (acquired immunodeficiency syndrome)
4223597	Strongyloidiasis with AIDS (acquired immunodeficiency syndrome)
4223738	Noninfectious gastroenteritis with AIDS (acquired immunodeficiency syndrome)
4223918	Retinal vascular changes with AIDS (acquired immunodeficiency syndrome)
4224078	Myelopathy with AIDS (acquired immunodeficiency syndrome)
4224096	Kaposi's sarcoma with AIDS (acquired immunodeficiency syndrome)
4224240	Presenile dementia with AIDS (acquired immunodeficiency syndrome)
4224425	Subacute myocarditis with AIDS (acquired immunodeficiency syndrome)
4224553	Agranulocytosis with AIDS (acquired immunodeficiency syndrome)
4224566	Skin disorder with AIDS (acquired immunodeficiency syndrome)
4224578	Nocardiosis with AIDS (acquired immunodeficiency syndrome)
4224740	Coccidioidomycosis with AIDS (acquired immunodeficiency syndrome)
4224860	Organic dementia with AIDS (acquired immunodeficiency syndrome)
4225017	Salmonella infection with AIDS (acquired immunodeficiency syndrome)
4225027	Malaise with AIDS (acquired immunodeficiency syndrome)
4225170	Histoplasmosis with AIDS (acquired immunodeficiency syndrome)
4225193	Microsporidiosis associated with acquired immunodeficiency syndrome
4225318	Viral pneumonia with AIDS (acquired immunodeficiency syndrome)
4225643	Respiratory disorder with AIDS (acquired immunodeficiency syndrome)
4225810	Aplastic anemia with AIDS (acquired immunodeficiency syndrome)
4225989	Infectious gastroenteritis with AIDS (acquired immunodeficiency syndrome)
4226110	Central nervous disorder with AIDS (acquired immunodeficiency syndrome)
4226357	Malignant neoplasm with AIDS (acquired immunodeficiency syndrome)
4226388	Reticulosarcoma with AIDS (acquired immunodeficiency syndrome)
4226502	Pneumocystosis with AIDS (acquired immunodeficiency syndrome)
4226629	Salivary gland disease with AIDS (acquired immunodeficiency syndrome)
4226631	Progressive multifocal leukoencephalopathy with AIDS (acquired immunodeficiency syndrome)
4226658	Tuberculosis with AIDS (acquired immunodeficiency syndrome)
4226662	Polyneuropathy with AIDS (acquired immunodeficiency syndrome)
4226790	Blindness with AIDS (acquired immunodeficiency syndrome)
4226792	Candidiasis of mouth with AIDS (acquired immunodeficiency syndrome)
4226905	Thrombocytopenia with AIDS (acquired immunodeficiency syndrome)
4227053	Myocarditis with AIDS (acquired immunodeficiency syndrome)
4227072	Radiculitis with AIDS (acquired immunodeficiency syndrome)
4227077	Nutritional deficiency with AIDS (acquired immunodeficiency syndrome)

Concept ID	Concept name
4227205	Organic brain syndrome with AIDS (acquired immunodeficiency syndrome)
4227229	Dermatomycosis with AIDS (acquired immunodeficiency syndrome)
4227350	Subacute endocarditis with AIDS (acquired immunodeficiency syndrome)
4227380	Disseminated candidiasis with AIDS (acquired immunodeficiency syndrome)
4227519	Neuritis with AIDS (acquired immunodeficiency syndrome)
4227539	Herpes zoster with AIDS (acquired immunodeficiency syndrome)
4227969	Primary lymphoma of brain with AIDS (acquired immunodeficiency syndrome)
4227974	Myelitis with AIDS (acquired immunodeficiency syndrome)
4228115	Retinopathy with AIDS (acquired immunodeficiency syndrome)
4228133	Dementia with AIDS (acquired immunodeficiency syndrome)
4228272	Failure to thrive in infant with AIDS (acquired immunodeficiency syndrome)
4228277	Pneumonia with AIDS (acquired immunodeficiency syndrome)
4228290	Infective arthritis with AIDS (acquired immunodeficiency syndrome)
4228440	Encephalopathy with AIDS (acquired immunodeficiency syndrome)
4228612	Encephalomyelitis with AIDS (acquired immunodeficiency syndrome)
4228745	Neuralgia with AIDS (acquired immunodeficiency syndrome)
4228754	Dyspnea with AIDS (acquired immunodeficiency syndrome)
4228757	Low vision with AIDS (acquired immunodeficiency syndrome)
4228760	Hyperhidrosis with AIDS (acquired immunodeficiency syndrome)
4230974	Human immunodeficiency virus type I
4236860	Pediatric human immunodeficiency virus infection
4239722	Asymptomatic human immunodeficiency virus infection in pregnancy
4241530	Asymptomatic human immunodeficiency virus infection
4253472	Human immunodeficiency virus I infection
4262297	Human immunodeficiency virus encephalopathy
4262565	Human immunodeficiency virus type 2
4267414	AIDS
4276281	Human immunodeficiency virus antibody titer - finding
4276586	Finding of HIV status
4295638	HIV CDC stage finding
4314426	HIV infection with acute lymphadenitis
4320032	Persistent generalized lymphadenopathy
4340791	Human immunodeficiency virus enteropathy
4347288	Human immunodeficiency virus myopathy
36674252	Infection caused by Coccidia co-occurrent with acquired immunodeficiency syndrome
36674253	Infection caused by Toxoplasma gondii co-occurrent with acquired immunodeficiency syndrome
36674254	Infection caused by Isospora co-occurrent with acquired immunodeficiency syndrome

Concept ID	Concept name
36687122	Human immunodeficiency virus infection with cognitive impairment
36714339	Candidiasis of upper respiratory tract co-occurrent with human immunodeficiency virus infection
36714477	Human immunodeficiency virus 1 antibody indeterminate
36714505	Human immunodeficiency virus 2 antibody indeterminate
36714516	Human immunodeficiency virus 2 antibody detected and Human immunodeficiency virus 1 antibody cross-reactivity
36715476	Human immunodeficiency virus complicating pregnancy childbirth and the puerperium
36716524	Parkinsonism due to human immunodeficiency virus infection
37016864	Human immunodeficiency virus p24 antigen detected
37017071	Subacute adenoviral encephalitis co-occurrent with human immunodeficiency virus infection
37017082	Splenomegaly co-occurrent with human immunodeficiency virus infection
37017092	Candidiasis of esophagus co-occurrent with human immunodeficiency virus infection
37017093	Heart disease co-occurrent with human immunodeficiency virus infection
37017094	Disorder of gastrointestinal tract co-occurrent with human immunodeficiency virus infection
37017106	Eruption of skin co-occurrent with human immunodeficiency virus infection
37017108	Myocarditis co-occurrent with human immunodeficiency virus infection
37017112	Primary cerebral lymphoma co-occurrent with human immunodeficiency virus infection
37017124	Infection caused by Strongyloides co-occurrent with human immunodeficiency virus infection
37017125	Disorder of skin co-occurrent with human immunodeficiency virus infection
37017126	Myelitis co-occurrent with human immunodeficiency virus infection
37017132	Anemia co-occurrent with human immunodeficiency virus infection
37017209	Hemophagocytic syndrome with human immunodeficiency virus infection
37017210	Chronic infection caused by herpes simplex virus co-occurrent with human immunodeficiency virus infection
37017243	Reticulosarcoma co-occurrent with human immunodeficiency virus infection
37017244	Disorder of respiratory system co-occurrent with human immunodeficiency virus infection
37017246	Progressive multifocal leukoencephalopathy co-occurrent with human immunodeficiency virus infection
37017247	Presenile dementia co-occurrent with human immunodeficiency virus infection
37017248	Infection caused by Pneumocystis co-occurrent with human immunodeficiency virus infection
37017249	Organic brain syndrome co-occurrent with human immunodeficiency virus infection
37017254	Candidiasis of mouth co-occurrent with human immunodeficiency virus infection
37017259	Disorder of spinal cord co-occurrent with human immunodeficiency virus infection
37017260	Disorder of kidney co-occurrent with human immunodeficiency virus infection
37017261	Gastrointestinal malabsorption syndrome co-occurrent with human immunodeficiency virus infection
37017262	Neuritis co-occurrent with human immunodeficiency virus infection
37017263	Lymphadenopathy co-occurrent with human immunodeficiency virus infection
37017265	Enlargement of liver co-occurrent with human immunodeficiency virus infection
37017266	Acute endocarditis co-occurrent with human immunodeficiency virus infection

Concept ID	Concept name
37017276	Cardiomyopathy co-occurent with human immunodeficiency virus infection
37017278	Recurrent bacterial pneumonia co-occurent with human immunodeficiency virus infection
37017279	Disorder of peripheral nervous system co-occurent with human immunodeficiency virus infection
37017282	Agranulocytosis co-occurent with human immunodeficiency virus infection
37017283	Visual impairment co-occurent with human immunodeficiency virus infection
37017284	Infective arthritis co-occurent with human immunodeficiency virus infection
37017285	Acquired hemolytic anemia co-occurent with human immunodeficiency virus infection
37017294	Demyelinating disease of central nervous system co-occurent with human immunodeficiency virus infection
37017295	Infection caused by Nocardia co-occurent with human immunodeficiency virus infection
37017296	Isosporiasis co-occurent with human immunodeficiency virus infection
37017318	Infectious gastroenteritis co-occurent with human immunodeficiency virus infection
37017319	Disorder of central nervous system co-occurent with human immunodeficiency virus infection
37017320	Malignant neoplastic disease co-occurent with human immunodeficiency virus infection
37017424	Nephrotic syndrome co-occurent with human immunodeficiency virus infection
37017425	Renal failure syndrome co-occurent with human immunodeficiency virus infection
37017442	Diffuse non-Hodgkin immunoblastic lymphoma co-occurent with human immunodeficiency virus infection
37017446	Infection caused by Cytomegalovirus co-occurent with human immunodeficiency virus infection
37017453	Coccidiosis co-occurent with human immunodeficiency virus infection
37017454	Infection caused by herpes simplex virus co-occurent with human immunodeficiency virus infection
37017455	Pyrexia of unknown origin co-occurent with human immunodeficiency virus infection
37017456	Aspergillosis co-occurent with human immunodeficiency virus infection
37017457	Infection caused by herpes zoster virus co-occurent with human immunodeficiency virus infection
37017465	Human immunodeficiency virus antibody detected
37017467	Reactive for both human immunodeficiency virus antibodies and p24 antigen
37017549	Dementia co-occurent with human immunodeficiency virus infection
37017550	Infection caused by Cryptosporidium co-occurent with human immunodeficiency virus infection
37017579	Opportunistic mycosis co-occurent with human immunodeficiency virus infection
37017580	Microsporidiosis co-occurent with human immunodeficiency virus infection
37017586	Focal segmental glomerulosclerosis co-occurent with human immunodeficiency virus infection
37017595	Burkitt lymphoma co-occurent with human immunodeficiency virus infection
37017652	Multidermatomal infection caused by Herpes zoster co-occurent with human immunodeficiency virus infection
37017655	Disseminated atypical infection caused by Mycobacterium co-occurent with human immunodeficiency virus infection
37018063	Immune reconstitution inflammatory syndrome caused by human immunodeficiency virus infection
37018711	Disorder of eye proper co-occurent with human immunodeficiency virus infection
37018714	Radiculitis co-occurent with human immunodeficiency virus infection

Concept ID	Concept name
37018721	Polyneuropathy co-occurent with human immunodeficiency virus infection
37018755	Recurrent salmonella sepsis co-occurent with human immunodeficiency virus infection
37018935	Dermatophytosis co-occurent with human immunodeficiency virus infection
37019034	Neuralgia co-occurent with human immunodeficiency virus infection
37019042	Infection caused by Salmonella co-occurent with human immunodeficiency virus infection
37019052	Disseminated infection caused by Strongyloides co-occurent with human immunodeficiency virus infection
37019055	Aplastic anemia co-occurent with human immunodeficiency virus infection
37019058	Bacterial pneumonia co-occurent with human immunodeficiency virus infection
37116830	Invasive carcinoma of uterine cervix co-occurent with human immunodeficiency virus infection
37116831	Extrapulmonary tuberculosis co-occurent with human immunodeficiency virus infection
37162220	Infectious mononucleosis caused by human immunodeficiency virus
37162221	Infectious mononucleosis caused by human immunodeficiency virus type I
37166798	AIDS (acquired immunodeficiency syndrome) wasting syndrome
40493262	Infectious disease with acquired immune deficiency syndrome
42536590	Human immunodeficiency virus World Health Organization 2007 stage 1 co-occurent with tuberculosis
42536591	Human immunodeficiency virus World Health Organization 2007 stage 1 co-occurent with malaria
42536592	Human immunodeficiency virus World Health Organization 2007 stage 2 co-occurent with tuberculosis
42536593	Human immunodeficiency virus World Health Organization 2007 stage 2 co-occurent with malaria
42536594	Human immunodeficiency virus World Health Organization 2007 stage 3 co-occurent with malaria
42536595	Human immunodeficiency virus World Health Organization 2007 stage 4 co-occurent with tuberculosis
42536596	Human immunodeficiency virus World Health Organization 2007 stage 4 co-occurent with malaria
42537802	World Health Organization 2007 Human immunodeficiency virus infection clinical stage 1
42537803	World Health Organization 2007 Human immunodeficiency virus infection clinical stage 2
42537804	World Health Organization 2007 Human immunodeficiency virus infection clinical stage 3
42537805	World Health Organization 2007 Human immunodeficiency virus infection clinical stage 4
42538959	Asymptomatic human immunodeficiency virus A1 infection
42538960	Asymptomatic human immunodeficiency virus A2 infection
42539031	Human immunodeficiency virus World Health Organization 2007 stage 3 co-occurent with tuberculosis
43531586	Symptomatic human immunodeficiency virus infection
44783780	Cholangitis with AIDS (acquired immunodeficiency syndrome)
44784137	Sepsis with acquired immunodeficiency syndrome
45757102	Acquired immune deficiency syndrome complicating childbirth
45757132	Human immunodeficiency virus in mother complicating childbirth
45769864	Symptomatic human immunodeficiency virus (HIV) I infection
46284256	HIV (human immunodeficiency virus) disease resulting in haematological and immunological abnormalities
44783356	Human immunodeficiency virus carrier

Inflammatory auto-immune diseases of the intestine:

Concept ID	Concept name
46269838	Abscess of intestine co-occurrent and due to chronic ulcerative pancolitis
46269848	Abscess of intestine co-occurrent and due to chronic ulcerative rectosigmoiditis
46269888	Abscess of intestine co-occurrent and due to Crohn's disease
46274073	Abscess of intestine co-occurrent and due to Crohn's disease of large intestine
46269878	Abscess of intestine co-occurrent and due to Crohn's disease of small and large intestine
46269883	Abscess of intestine co-occurrent and due to Crohn's disease of small intestine
46269951	Abscess of intestine co-occurrent and due to ulcerative colitis
602594	Acute left-sided ulcerative colitis
4029372	Acute ulcerative colitis
602593	Acute ulcerative pancolitis
602607	Acute ulcerative rectosigmoiditis
4241413	Adult form of celiac disease
1245349	ALPI-related inflammatory bowel disease
36716986	Arthritis co-occurrent and due to Crohn's disease
4218370	Celiac crisis
194992	Celiac disease
4267871	Celiac disease with diffuse intestinal ulceration
36716394	Celiac disease with epilepsy and cerebral calcification syndrome
40481367	Chronic left-sided ulcerative colitis
40479837	Chronic ulcerative colitis
194077	Chronic ulcerative enterocolitis
78799	Chronic ulcerative ileocolitis
40482241	Chronic ulcerative pancolitis
77317	Chronic ulcerative rectosigmoiditis
37162889	Colorectal Crohn disease
37162756	Combined immunodeficiency, enteropathy spectrum
36715915	Crohn disease of anal canal
37116446	Crohn disease of appendix
37172426	Crohn disease of colon in remission
37170088	Crohn disease of rectum in remission
37172546	Crohn disease of small intestine and large intestine in remission
37170274	Crohn disease of small intestine in remission
42537666	Crohn disease of upper gastrointestinal tract
201606	Crohn's disease
4142544	Crohn's disease in remission
4177488	Crohn's disease of colon

Concept ID	Concept name
4210469	Crohn's disease of duodenum
4340114	Crohn's disease of esophagus
36716695	Crohn's disease of gastrointestinal anastomosis
4122617	Crohn's disease of gingivae
4244235	Crohn's disease of ileum
4246693	Crohn's disease of intestine
4239382	Crohn's disease of jejunum
194684	Crohn's disease of large bowel
4055884	Crohn's disease of oral soft tissues
37162884	Crohn's disease of parastomal skin
4299294	Crohn's disease of penis
4323289	Crohn's disease of pyloric antrum
4266370	Crohn's disease of pylorus
4242392	Crohn's disease of rectum
4292361	Crohn's disease of scrotum
4290835	Crohn's disease of skin
195575	Crohn's disease of small and large intestines
195585	Crohn's disease of small intestine
36686095	Crohn's disease of small intestine with stenosis
4342643	Crohn's disease of stomach
4055020	Crohn's disease of terminal ileum
4297643	Crohn's disease of vulva
37173975	Crohn-like disease of ileal pouch
4086978	Eosinophilic ulcerative colitis
1340297	Exacerbation of Crohn's disease
4212991	Exacerbation of Crohn's disease of large intestine
4212992	Exacerbation of Crohn's disease of small intestine
4187900	Exacerbation of ulcerative colitis
1340490	Exacerbation of ulcerative colitis
1340491	Exacerbation of ulcerative proctocolitis
37111373	Extraintestinal Crohn's
4264850	Gastrointestinal Crohn's disease
37161069	IL21-related infantile inflammatory bowel disease
37162890	Ileojejunal Crohn disease
37162096	Infantile inflammatory bowel disease with neurological involvement
4074815	Inflammatory bowel disease
36715903	Inflammatory polyps co-occurrent and due to inflammatory bowel disease

Concept ID	Concept name
37312266	Iritis co-occurrent with inflammatory bowel disease
4253620	Iritis with Crohn's disease
4259504	Iritis with ulcerative colitis
4115377	Juvenile arthritis in Crohn disease
4114447	Juvenile arthritis in ulcerative colitis
4079736	Juvenile arthritis of inflammatory bowel disease
40482865	Left sided ulcerative colitis
37170100	Left sided ulcerative colitis in remission
4297644	Metastatic Crohn's disease of skin
37162887	Microscopic enteritis
4301738	Mild chronic ulcerative colitis
4302002	Moderate chronic ulcerative colitis
3184226	Pancolonic ulcerative colitis
4340812	Perianal Crohn's disease
36674761	Refractory celiac disease
4031048	Severe chronic ulcerative colitis
4225691	Toxic megacolon due to ulcerative colitis
1245350	TRIM22-related inflammatory bowel disease
81893	Ulcerative colitis
44783784	Ulcerative colitis in remission
4342656	Ulcerative enterocolitis
4025853	Ulcerative ileocolitis
40479839	Ulcerative pancolitis
37172721	Ulcerative pancolitis in remission
4166480	Ulcerative proctocolitis
37172295	Ulcerative rectosigmoiditis in remission

Mood disorders:

Concept ID	Concept name
37016718	Acute depression
4308866	Agitated depression
37312479	Antenatal depression
438727	Atypical depressive disorder
439254	Bipolar affective disorder, current episode depression
439251	Bipolar affective disorder, currently depressed, in full remission
439253	Bipolar affective disorder, currently depressed, mild
437528	Bipolar affective disorder, currently depressed, moderate

Concept ID	Concept name
4107538	Bipolar I disorder, most recent episode depressed with atypical features
4103574	Chronic depression
40481798	Chronic depressive personality disorder
4031328	Chronic major depressive disorder, single episode
4094358	Chronic recurrent major depressive disorder
4131545	Depression requiring intervention
4333687	Depressive conduct disorder
440383	Depressive disorder
37209503	Depressive disorder caused by amphetamine
4103126	Depressive disorder caused by drug
37309680	Depressive disorder caused by methamphetamine
45757213	Depressive disorder in mother complicating childbirth
37018656	Depressive disorder in mother complicating pregnancy
44782943	Depressive disorder in remission
433440	Dysthymia
4096229	Early onset dysthymia
4114950	Endogenous depression
4333679	Endogenous depression first episode
4168858	Endogenous depression - recurrent
1340305	Exacerbation of depressive disorder
1340392	Exacerbation of major depressive disorder
4336980	Generalized neuromuscular exhaustion syndrome
4154805	Involucional depression
4057218	Late onset dysthymia
4269493	Major depression in full remission
4148630	Major depression in partial remission
4176002	Major depression in remission
4154391	Major depression, melancholic type
4282096	Major depression, single episode
4323418	Major depression single episode, in partial remission
37111697	Major depression with psychotic features
4152280	Major depressive disorder
45757195	Major depressive disorder in mother complicating childbirth
45757196	Major depressive disorder in mother complicating pregnancy
4181807	Major depressive disorder, single episode with atypical features
4287238	Major depressive disorder, single episode with catatonic features
4270907	Major depressive disorder, single episode with melancholic features

Concept ID	Concept name
4093584	Major depressive disorder, single episode with postpartum onset
4338029	Masked depression
4133073	Maternity blues
4223090	Menopausal depression
4149320	Mild depression
4336957	Mild major depression
4195572	Mild major depression, single episode
37109052	Mild major depressive disorder co-occurrent with anxiety single episode
4129842	Mild postnatal depression
4228802	Mild recurrent major depression
36713698	Minimal depression
36715000	Minimal major depression
36714999	Minimal major depression single episode
36714997	Minimal recurrent major depression
4174987	Minor depressive disorder
4338031	Mixed anxiety and depressive disorder
4151170	Moderate depression
36717092	Moderately severe depression
36714389	Moderately severe major depression
36717389	Moderately severe major depression single episode
36714998	Moderately severe recurrent major depression
4307111	Moderate major depression
4049623	Moderate major depression, single episode
37109053	Moderate major depressive disorder co-occurrent with anxiety single episode
4077577	Moderate recurrent major depression
3654788	Mood disorder with depressive symptoms caused by amphetamine and/or amphetamine derivative
37110429	Mood disorder with depressive symptoms caused by anxiolytic
37110438	Mood disorder with depressive symptoms caused by cocaine
42538604	Mood disorder with depressive symptoms caused by dissociative drug
37164792	Mood disorder with depressive symptoms caused by ethanol
42538596	Mood disorder with depressive symptoms caused by hallucinogen
37110428	Mood disorder with depressive symptoms caused by hypnotic
42538584	Mood disorder with depressive symptoms caused by opioid
37117211	Mood disorder with depressive symptoms caused by sedative
42538590	Mood disorder with depressive symptoms caused by stimulant
42538736	Mood disorder with depressive symptoms caused by synthetic cathinone
42538599	Mood disorder with depressive symptoms caused by volatile inhalant

Concept ID	Concept name
443864	Multi-infarct dementia with depression
36712668	Perinatal depression
607543	Persistent depressive disorder
37166876	Postmenopausal depression
4305966	Postoperative depression
4239471	Postpartum depression
42534817	Postpartum major depression in remission
4332994	Post-schizophrenic depression
4102973	Postviral depression
4242733	Premenstrual dysphoric disorder
4145216	Premenstrual dysphoric disorder in remission
4307951	Primary dysthymia
4243308	Primary dysthymia early onset
4195680	Primary dysthymia late onset
35622934	Psychosis and severe depression co-occurrent and due to bipolar affective disorder
35609845	Reactive depression, first episode
35609842	Reactive depression, prolonged single episode
35609844	Reactive depression, recurrent
35609843	Reactive depression, single episode
4314692	Reactive depression (situational)
435520	Reactive depressive psychosis
43020483	Reactive depressive psychosis, single episode
4226155	Recurrent brief depressive disorder
4098302	Recurrent depression
35610109	Recurrent depression with current moderate episode
35610097	Recurrent depression with current severe episode and psychotic features
35610108	Recurrent depression with current severe episode without psychotic features
4282316	Recurrent major depression
4263748	Recurrent major depression in full remission
4141454	Recurrent major depression in partial remission
433991	Recurrent major depression in remission
35615154	Recurrent major depressive disorder co-occurrent with anxiety in full remission
35615155	Recurrent major depressive disorder in partial remission co-occurrent with anxiety
4304140	Recurrent major depressive disorder with atypical features
4220023	Recurrent major depressive disorder with catatonic features
4205471	Recurrent major depressive disorder with melancholic features
4324959	Recurrent major depressive disorder with postpartum onset

Concept ID	Concept name
432285	Recurrent major depressive episodes
44805549	Recurrent major depressive episodes, in partial remission
44813499	Recurrent major depressive episodes, in remission
438998	Recurrent major depressive episodes, mild
432883	Recurrent major depressive episodes, moderate
44805542	Recurrent major depressive episodes, severe
434911	Recurrent major depressive episodes, severe, with psychosis
44805669	Recurrent major depressive episodes, severe, with psychosis, psychosis in remission
35615151	Recurrent mild major depressive disorder co-occurrent with anxiety
35615153	Recurrent moderate major depressive disorder co-occurrent with anxiety
35609824	Recurrent reactive depressive episodes, severe, with psychosis
35615152	Recurrent severe major depressive disorder co-occurrent with anxiety
4224940	Schizoaffective disorder, depressive type
4092239	Seasonal affective disorder
4224639	Secondary dysthymia
4150047	Secondary dysthymia early onset
4263770	Secondary dysthymia late onset
4149321	Severe depression
42872722	Severe major depression
42872411	Severe major depression, single episode
441534	Severe major depression, single episode, without psychotic features
438406	Severe major depression, single episode, with psychotic features
4299785	Severe major depression, single episode, with psychotic features, mood-congruent
4067409	Severe major depression, single episode, with psychotic features, mood-incongruent
4327337	Severe major depression without psychotic features
4250023	Severe major depression with psychotic features
4144233	Severe major depression with psychotic features, mood-congruent
4243822	Severe major depression with psychotic features, mood-incongruent
37109054	Severe major depressive disorder co-occurrent with anxiety single episode
4129184	Severe postnatal depression
43531624	Severe recurrent major depression
435220	Severe recurrent major depression without psychotic features
4154309	Severe recurrent major depression with psychotic features
4141292	Severe recurrent major depression with psychotic features, mood-congruent
4034842	Severe recurrent major depression with psychotic features, mood-incongruent
44782720	Severe seasonal affective disorder
4025677	Single episode of major depression in full remission

Concept ID	Concept name
44805550	Single major depressive episode, in remission
439259	Single major depressive episode, severe, with psychosis
44805668	Single major depressive episode, severe, with psychosis, psychosis in remission
4197222	Stuporous depression
607540	Treatment resistant depression
439255	Bipolar affective disorder, currently manic, in full remission
439254	Bipolar affective disorder, current episode depression
439253	Bipolar affective disorder, currently depressed, mild
437528	Bipolar affective disorder, currently depressed, moderate
439251	Bipolar affective disorder, currently depressed, in full remission
439250	Mixed bipolar affective disorder
3172581	Bipolar disease in pregnancy
439249	Mixed bipolar affective disorder, mild
439248	Mixed bipolar affective disorder, moderate
439246	Mixed bipolar affective disorder, severe, with psychosis
439245	Mixed bipolar affective disorder, in full remission
435226	Bipolar affective disorder, current episode mixed
1340265	Exacerbation of bipolar disorder
4000165	Severe bipolar II disorder, most recent episode major depressive with psychotic features, mood-congruent
44788753	Psychosis, schizophrenia and bipolar affective disorder resolved
4051448	Severe bipolar II disorder, most recent episode major depressive with psychotic features, mood-incongruent
4031928	Severe mixed bipolar I disorder with psychotic features, mood-incongruent
4071442	Bipolar I disorder, most recent episode depressed with catatonic features
439001	Severe mixed bipolar I disorder with psychotic features
435225	Depressed bipolar I disorder in full remission
4009648	Mixed bipolar I disorder in full remission
4001733	Chronic bipolar II disorder, most recent episode major depressive
4330846	Bipolar II disorder, most recent episode major depressive with catatonic features
4333670	Organic bipolar disorder
4045263	Severe bipolar II disorder, most recent episode major depressive, in full remission
43020451	Severe manic bipolar I disorder
42872413	Severe depressed bipolar I disorder
4094507	Severe depressed bipolar I disorder with psychotic features, mood-incongruent
43021847	Rapid cycling bipolar I disorder
4131027	Severe bipolar disorder with psychotic features, mood-incongruent
42872412	Severe mixed bipolar I disorder
4028027	Mild bipolar disorder

Concept ID	Concept name
4102603	Severe manic bipolar I disorder with psychotic features
440067	Moderate bipolar I disorder, single manic episode
4107538	Bipolar I disorder, most recent episode depressed with atypical features
4030102	Severe bipolar I disorder, single manic episode with psychotic features, mood-congruent
436665	Bipolar disorder
4147991	Severe bipolar II disorder, most recent episode major depressive with psychotic features
4030856	Severe bipolar I disorder, single manic episode without psychotic features
4148934	Bipolar II disorder, most recent episode major depressive with postpartum onset
4033390	Bipolar I disorder, single manic episode with postpartum onset
443797	Severe manic bipolar I disorder without psychotic features
436086	Manic bipolar I disorder in full remission
4150985	Bipolar I disorder, most recent episode hypomanic
37209504	Bipolar disorder caused by drug
4141603	Severe manic bipolar I disorder with psychotic features, mood-incongruent
4144519	Bipolar II disorder, most recent episode major depressive with melancholic features
4037669	Bipolar II disorder, most recent episode major depressive
4148842	Bipolar I disorder, single manic episode, in full remission
433743	Mixed bipolar I disorder in remission
443906	Mixed bipolar I disorder
4262272	Severe bipolar II disorder, most recent episode major depressive, in remission
4262111	Moderate bipolar II disorder, most recent episode major depressive
437529	Mixed bipolar I disorder in partial remission
4073401	Bipolar I disorder, most recent episode manic with catatonic features
432876	Bipolar I disorder
440078	Bipolar affective disorder, current episode manic
4154283	Severe bipolar I disorder
441834	Bipolar affective disorder, currently manic, mild
4155798	Severe bipolar disorder
433992	Bipolar affective disorder, currently manic, moderate
4161200	Severe bipolar II disorder
439256	Bipolar affective disorder, currently manic, severe, with psychosis
4244078	Schizoaffective disorder, bipolar type
439785	Moderate mixed bipolar I disorder
432290	Mild bipolar I disorder, single manic episode
4195158	Severe bipolar disorder with psychotic features
4220617	Severe bipolar I disorder, single manic episode with psychotic features
4220618	Bipolar disorder in full remission

Concept ID	Concept name
4185096	Bipolar II disorder, most recent episode major depressive with atypical features
4166701	Manic bipolar I disorder in remission
440079	Mild mixed bipolar I disorder
44791461	Psychosis resolved
372599	Severe mixed bipolar I disorder without psychotic features
44791460	Schizophrenia resolved
4172156	Bipolar II disorder, most recent episode hypomanic
441836	Depressed bipolar I disorder
4177651	Depressed bipolar I disorder in partial remission
4197669	Chronic bipolar I disorder, most recent episode depressed
3185163	History of bipolar disorder depressed
4200385	Severe bipolar disorder without psychotic features
4201739	Depressed bipolar I disorder in remission
4182998	Severe depressed bipolar I disorder with psychotic features, mood-congruent
4210024	Bipolar I disorder, most recent episode manic with postpartum onset
4209560	Infection by <i>Bipolaris hawaiiense</i>
4083107	<i>Bipolaris</i>
4128935	Bipolar
436072	Bipolar disorder in partial remission
4190107	Referral for bipolar disorder
436386	Severe depressed bipolar I disorder with psychotic features
37399377	<i>Bipolaris cynodontis</i>
42537545	<i>Bipolaris sorokiniana</i>
442570	Severe depressed bipolar I disorder without psychotic features
442600	Manic bipolar I disorder in partial remission
42537978	<i>Bipolaris cactivora</i>
4276670	Severe mixed bipolar I disorder with psychotic features, mood-congruent
4214890	<i>Bipolaris australiensis</i>
4339383	<i>Bipolaris spicifera</i>
4274957	Bipolar I disorder, most recent episode mixed with postpartum onset
4280361	Moderate depressed bipolar I disorder
4283219	Severe bipolar II disorder, most recent episode major depressive, in partial remission
4287544	Manic bipolar I disorder
4322477	Severe bipolar disorder with psychotic features, mood-congruent
4324945	Mild bipolar II disorder, most recent episode major depressive
4215917	Mild manic bipolar I disorder
37109940	Bipolar type I disorder currently in full remission

Concept ID	Concept name
37117177	Bipolar type II disorder currently in full remission
4251178	Bipolar I disorder, most recent episode mixed with catatonic features
437250	Mild depressed bipolar I disorder
4327669	Bipolar I disorder, single manic episode, in remission
4192865	Bipolar I disorder, most recent episode depressed with melancholic features
44804961	Mixed bipolar affective disorder, in partial remission
44805540	Mixed bipolar affective disorder, severe
35622934	Psychosis and severe depression co-occurrent and due to bipolar affective disorder
35624743	Bipolar disorder, most recent episode depression
35624744	Bipolar disorder, most recent episode manic
35624745	Bipolar affective disorder, most recent episode mixed
35624747	Bipolar I disorder, most recent episode manic
35624748	Bipolar I disorder, most recent episode depression
4301106	Bipolar I disorder, single manic episode, in partial remission
438129	Severe manic bipolar I disorder with psychotic features, mood-congruent
37312578	Rapid cycling bipolar II disorder
4194222	Moderate bipolar disorder
4217940	Severe bipolar II disorder, most recent episode major depressive without psychotic features
4307804	Moderate manic bipolar I disorder
4307956	Bipolar II disorder
4310821	Bipolar disorder in remission
4312736	Severe bipolar I disorder, single manic episode with psychotic features, mood-incongruent
4336405	Bipolar I disorder, most recent episode depressed with postpartum onset
4338812	Bipolar I disorder, single manic episode with catatonic features
432866	Bipolar I disorder, single manic episode

Obesity:

Concept ID	Concept name
604591	6q16 microdeletion syndrome
4271317	Adiposogenital dystrophy
4171972	Adult-onset obesity
4270189	Alstrom syndrome
4079899	Android obesity
4235799	Buffalo obesity
4087487	Central obesity
40481140	Childhood obesity
36713437	Choroideremia with deafness and obesity syndrome

Concept ID	Concept name
36678790	Colobomatous microphthalmia, obesity, hypogenitalism, intellectual disability syndrome
45763687	Congenital leptin deficiency
4097929	Constitutional obesity
4097996	Drug-induced obesity
4182506	Endogenous obesity
4100857	Extreme obesity with alveolar hypoventilation
4160821	Familial obesity
4029276	Generalized obesity
37166819	Genetic non-syndromic obesity
4029900	Gynecoid obesity
36717154	Hydrocephalus with obesity and hypogonadism syndrome
4005991	Hyperinsular obesity
4163032	Hyperplastic-hypertrophic obesity
4185912	Hyperplastic obesity
4171147	Hypertrophic obesity
4177337	Hypogonadal obesity
4220527	Hypothalamic obesity
4203289	Hypothyroid obesity
35622038	Intellectual disability, obesity, brain malformation, facial dysmorphism syndrome
36674490	Intellectual disability, obesity, prognathism, eye and skin anomalies syndrome
36674893	Intellectual disability, seizures, macrocephaly, obesity syndrome
4171317	Lifelong obesity
37208175	Lower body obesity
37164247	MAGEL2-related Prader-Willi-like syndrome
42872398	Maternal obesity complicating pregnancy, childbirth and the puerperium, antepartum
4216214	Mauriac's syndrome
36716144	MEHMO syndrome
37110069	MOMO syndrome
434005	Morbid obesity
37395980	MORM syndrome
433736	Obesity
4212443	Obesity by adipocyte growth pattern
4215969	Obesity by contributing factors
4189665	Obesity by fat distribution pattern
36716555	Obesity caused by energy imbalance
36717199	Obesity, colitis, hypothyroidism, cardiac hypertrophy, developmental delay syndrome
37204685	Obesity due to CEP19 deficiency

Concept ID	Concept name
37206117	Obesity due to leptin receptor gene deficiency
37397209	Obesity due to melanocortin 4 receptor deficiency
37162364	Obesity due to pituitary disease
36716151	Obesity due to prohormone convertase I deficiency
37204815	Obesity due to SIM1 deficiency
37311904	Obesity in adolescence
45757112	Obesity in mother complicating childbirth
4183240	Obesity of endocrine origin
4093860	Peripheral obesity
37163354	PHIP-related behavioural problems, intellectual disability, obesity, dysmorphic features syndrome
36674827	Prader-Willi-like syndrome
3199162	Pregnancy complicated by obesity
45771307	Proopiomelanocortin deficiency syndrome
36676689	Rapid-onset childhood obesity, hypothalamic dysfunction, hypoventilation, autonomic dysregulation syndrome
37204691	Severe early-onset obesity insulin resistance syndrome due to SH2B1 deficiency
37018860	Severe obesity
42539192	Severe obesity complicating pregnancy
37164244	SIM1-related Prader-Willi-like syndrome
4217557	Simple obesity
37166818	Spastic paraplegia, intellectual disability, nystagmus, obesity syndrome
4211019	Steatopygia
36714072	Syndromic X-linked intellectual disability type 7
36714548	Wilson Turner syndrome
37165655	X-linked intellectual disability, short stature, overweight syndrome

Thyroid disorders:

Concept ID	Concept name
36717595	Acquired central hypothyroidism
37161451	Acquired central hypothyroidism due to pituitary disorder
37161450	Acquired central hypothyroidism due to Sheehan syndrome
138384	Acquired hypothyroidism
607903	Adrenocorticotrophic hormone resistance syndrome
4184957	Adult myxedema
45773358	Allan-Herndon-Dudley syndrome
4185526	Athyrotic hypothyroidism sequence
4034815	Autoimmune hypothyroidism
36716385	Bamforth Lazarus syndrome

Concept ID	Concept name
4132435	Central hypothyroidism
36716764	Congenital central hypothyroidism
37110922	Congenital central hypothyroidism due to thyrotropin-releasing hormone receptor deficiency
133728	Congenital hypothyroidism
36713742	Congenital hypothyroidism due to absence of thyroid gland
37155411	Congenital hypothyroidism due to congenital anomaly of thyroid gland
37161472	Congenital hypothyroidism due to dual oxidase maturation factor 2
36716765	Congenital hypothyroidism due to iodine deficiency
37204507	Congenital hypothyroidism due to maternal intake of antithyroid drug
37161473	Congenital hypothyroidism due to peripheral resistance to thyroid hormone
37161468	Congenital hypothyroidism due to symporter mutation
37161471	Congenital hypothyroidism due to thyroglobulin mutation
37161467	Congenital hypothyroidism due to thyroid deiodinase mutation
37161465	Congenital hypothyroidism due to thyroid peroxidase mutation
37164395	Congenital hypothyroidism due to thyroid stimulating hormone receptor mutation
37397258	Congenital hypothyroidism due to transplacental passage of maternal thyroid stimulating hormone binding inhibitory antibody
4081998	Congenital hypothyroidism with diffuse goiter
4130017	Congenital hypothyroidism without goiter
4130026	Congenital iodine deficiency syndrome of mixed type
4246944	Congenital malposition of the thyroid gland
37161466	Consumptive hypothyroidism caused by type 3 iodothyronine deiodinase
37166448	Dementia due to acquired hypothyroidism
37161455	Drug induced acquired central hypothyroidism
4229107	Dyshormonogenetic goiter AND iodide leak
136933	Dyshormonogenic goiter
4034822	Dysprealbuminemic euthyroidal hyperthyroxinemia
4233438	Endemic congenital iodine deficiency syndrome of myxedematous type
4129362	Euthyroid with thyroid antibodies
1340365	Exacerbation of hyperthyroidism
4030051	Familial dysalbuminemic hyperthyroxinemia
4030053	Familial dyshormonogenetic goiter
35608125	Familial glucocorticoid deficiency
36716767	Familial hyperthyroidism
37396803	Familial non-autoimmune autosomal dominant hyperthyroidism
37397527	Familial thyroid dyshormonogenesis
4130024	Generalized thyroid hormone resistance

Concept ID	Concept name
36674793	Genetic transient congenital hypothyroidism
4163854	Glucocorticoid deficiency with achalasia
4130012	Hashitoxicosis - transient
37396504	Hoffman syndrome
4142479	Hyperthyroidism
4337834	Hyperthyroidism due to ectopic thyroid nodule
4227238	Hyperthyroidism due to hydatidiform mole with thyrotoxic crisis
4259138	Hyperthyroidism due to molar thyrotropin
4094160	Hyperthyroidism due to struma ovarii
45757130	Hyperthyroidism in childbirth
4265061	Hyperthyroidism secondary to amiodarone
4241482	Hyperthyroidism secondary to potassium iodide
4194323	Hyperthyroidism secondary to radio contrast dye
4100629	Hyperthyroidism with Hashimoto disease
4291925	Hypothalamic hypothyroidism
4314386	Hypothyroid dwarfism
4034814	Hypothyroid goiter, acquired
140673	Hypothyroidism
37208050	Hypothyroidism caused by amiodarone
37016342	Hypothyroidism caused by drug
4095914	Hypothyroidism caused by resorcinol
4245619	Hypothyroidism due to amyloidosis
4175837	Hypothyroidism due to cystinosis
4210299	Hypothyroidism due to defect in thyroid hormone synthesis
4192649	Hypothyroidism due to fibrous invasive thyroiditis
4172443	Hypothyroidism due to food stuff
4130018	Hypothyroidism due to Hashimoto's thyroiditis
4147364	Hypothyroidism due to infiltrative disease
4289921	Hypothyroidism due to iodide excess
4034824	Hypothyroidism due to iodide organification defect
4129369	Hypothyroidism due to iodide trapping defect
37397535	Hypothyroidism due to mutation in transcription factor of pituitary development
4303055	Hypothyroidism due to sarcoidosis
4268472	Hypothyroidism due to scleroderma
37167445	Hypothyroidism due to systemic sclerosis
45757058	Hypothyroidism due to thyroiditis
4030047	Hypothyroidism due to TSH receptor blocking antibody

Concept ID	Concept name
4231548	Hypothyroidism following radioiodine therapy
45757193	Hypothyroidism in childbirth
4176015	Hypothyroidism in pregnancy
4095912	Hypothyroidism resulting from para-aminosalicylic acid
4095913	Hypothyroidism resulting from phenylbutazone
134312	Iatrogenic hypothyroidism
4309027	Idiopathic atrophic hypothyroidism
37399723	Idiopathic congenital hypothyroidism
4034983	Idiopathic TSH deficiency
4236858	Infant hypothyroidism
44782482	Infant hypothyroidism due to maternal drug
4236859	Infant hypothyroidism to 24 months of age
4202818	Iodide oxidation defect
4331585	Iodide transport defect
140062	Iodine hypothyroidism
4129357	Iodine-induced hyperthyroidism
4076302	Iodotyrosine deiodination defect
4052693	Iodotyrosyl coupling defect
4230503	Isolated thyrotropin deficiency
4028612	Juvenile myxedema
3171361	Lithium induced hypothyroidism
4182596	Myxedema
4263049	Myxedema heart disease
4275410	Myxedematous form of cretinism
36676796	Neonatal iodine exposure
4071079	Neonatal jaundice with congenital hypothyroidism
4210297	Neonatal transient hyperthyroidism
36717199	Obesity, colitis, hypothyroidism, cardiac hypertrophy, developmental delay syndrome
44783239	Ohdo syndrome, Say-Barber-Biesecker-Young-Simpson variant
37397534	Peripheral resistance to thyroid hormone
4232833	Pituitary thyroid hormone resistance
132583	Postablative hypothyroidism
4030049	Post-infectious hypothyroidism
137820	Postoperative hypothyroidism
40609121	Primary hypothyroidism
37161453	Radiation-induced acquired central hypothyroidism
4006466	Refetoff syndrome

Concept ID	Concept name
37166867	Resistance to thyroid hormone due to mutation in thyroid hormone receptor alpha
37166868	Resistance to thyroid hormone due to mutation in thyroid hormone receptor beta
36716766	Secondary hyperthyroidism
4220368	Secondary hypothyroidism
4223217	Severe hypothyroidism
35622406	Short stature with delayed bone age due to thyroid hormone metabolism deficiency
4225129	Sporadic cretinism
4176932	Subclinical hyperthyroidism
4183422	Subclinical hypothyroidism
4130027	Subclinical iodine deficiency hypothyroidism
4269217	Thyroglobulin synthesis defect
4030054	Thyroid hormone resistance syndrome
4177341	Thyroid hormone responsiveness defect
137519	Thyrotoxicosis due to ectopic thyroid nodule with crisis
138718	Thyrotoxicosis due to ectopic thyroid nodule with no crisis
4236235	Thyrotoxicosis due to pituitary thyroid hormone resistance
606080	Transient congenital hypothyroidism due to dual oxidase 2 mutation
4173932	Transient decreased production of T-3
4091169	Transient decreased production of T-4
4104540	Transient hypothyroidism
4079964	Transient hypothyrotropinemia
4173188	Transient neonatal hypothyroidism
37161463	Traumatic acquired central hypothyroidism
36675175	X-linked central congenital hypothyroidism with late-onset testicular enlargement
4236235	Thyrotoxicosis due to pituitary thyroid hormone resistance
606080	Transient congenital hypothyroidism due to dual oxidase 2 mutation
4173932	Transient decreased production of T-3
4091169	Transient decreased production of T-4
4104540	Transient hypothyroidism
4079964	Transient hypothyrotropinemia
4173188	Transient neonatal hypothyroidism
37161463	Traumatic acquired central hypothyroidism
36675175	X-linked central congenital hypothyroidism with late-onset testicular enlargement
4091169	Transient decreased production of T-4
4104540	Transient hypothyroidism
4079964	Transient hypothyrotropinemia
4173188	Transient neonatal hypothyroidism

Concept ID	Concept name
37161463	Traumatic acquired central hypothyroidism
36675175	X-linked central congenital hypothyroidism with late-onset testicular enlargement

Table S5. List of medicines definitions.

Folic acid:

Concept ID ¹	Concept name
19111620	Folic acid

¹ All descendants included.

Coronavirus disease 2019 (COVID-19) vaccination:

Concept ID	Concept name
1525734	0.3 ML SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2 0.05 MG/ML / SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2 OMICRON (BA.4/BA.5) 0.05 MG/ML Injection
36968470	0.5 ML COVID-19 vaccine 0.08 MG/ML Injectable Suspension
36948476	0.5 ML COVID-19 vaccine 0.08 MG/ML Injectable Suspension by Hipra Human Health S.L.U.
36969696	0.5 ML COVID-19 vaccine 1200 UNT/ML Injectable Suspension
36964076	0.5 ML COVID-19 vaccine 1200 UNT/ML Injectable Suspension [CoronaVac]
36962147	0.5 ML COVID-19 vaccine 1200 UNT/ML Injectable Suspension [CoronaVac] by Sinovac Life Sciences
36967747	0.5 ML COVID-19 vaccine 13 UNT/ML Injectable Suspension
36950030	0.5 ML COVID-19 vaccine 13 UNT/ML Injectable Suspension by Beijing Institute of Biological Products
36953390	100 ML COVID-19 vaccine 0.08 MG/ML Injectable Suspension
36923477	100 ML COVID-19 vaccine 0.08 MG/ML Injectable Suspension by Hipra Human Health S.L.U.
36945265	10 ML COVID-19 vaccine 0.08 MG/ML Injectable Suspension
36962290	10 ML COVID-19 vaccine 0.08 MG/ML Injectable Suspension by Hipra Human Health S.L.U.
36962954	1 ML COVID-19 vaccine 1200 UNT/ML Injectable Suspension
36941610	1 ML COVID-19 vaccine 1200 UNT/ML Injectable Suspension [CoronaVac]
36929895	1 ML COVID-19 vaccine 1200 UNT/ML Injectable Suspension [CoronaVac] by Sinovac Life Sciences
36944687	1 ML COVID-19 vaccine 13 UNT/ML Injectable Suspension
36936425	1 ML COVID-19 vaccine 13 UNT/ML Injectable Suspension by Beijing Institute of Biological Products
35894915	COVID-19 vaccine
36935277	COVID-19 vaccine 0.08 MG/ML
36950950	COVID-19 vaccine 0.08 MG/ML Injectable Suspension
36963203	COVID-19 vaccine 1200 UNT/ML
36926164	COVID-19 vaccine 1200 UNT/ML [CoronaVac]
36925409	COVID-19 vaccine 1200 UNT/ML Injectable Suspension
36949344	COVID-19 vaccine 1200 UNT/ML Injectable Suspension [CoronaVac]
36958820	COVID-19 vaccine 13 UNT/ML
36959693	COVID-19 vaccine 13 UNT/ML Injectable Suspension

Concept ID	Concept name
35897994	COVID-19 vaccine Injectable Solution
36939163	COVID-19 vaccine Injectable Suspension
36943268	COVID-19 vaccine Injectable Suspension [CoronaVac]
36118949	COVID-19 vaccine, recombinant, full-length nanoparticle spike (S) protein, adjuvanted with Matrix-M
36929908	COVID-19 vaccine, recombinant, full-length nanoparticle spike (S) protein, adjuvanted with Matrix-M Injectable Solution
36119721	COVID-19 vaccine, recombinant, full-length nanoparticle spike (S) protein, adjuvanted with Matrix-M Injectable Suspension
36126197	COVID-19 vaccine, recombinant, plant-derived Virus-Like Particle (VLP) spike (S) protein, adjuvanted with AS03
36119722	COVID-19 vaccine, recombinant, plant-derived Virus-Like Particle (VLP) spike (S) protein, adjuvanted with AS03 Injectable Suspension
36118948	COVID-19 vaccine, whole virus, inactivated, adjuvanted with Alum and CpG 1018
36119720	COVID-19 vaccine, whole virus, inactivated, adjuvanted with Alum and CpG 1018 Injectable Suspension
716025	SARS-CoV-2 COVID-19 mRNA, bivalent, original/Omicron BA.1, Non-US Vaccine Product (Comirnaty Bivalent), Pfizer-BioNTech
716024	SARS-CoV-2 COVID-19 mRNA, bivalent, original/Omicron BA.1, Non-US Vaccine (Spikevax Bivalent), Moderna
742038	SARS-CoV-2 (COVID-19) vaccine, mRNA-1273 0.025 MG/ML / SARS-CoV-2 (COVID-19) vaccine, mRNA-1273 OMICRON (BA.4/BA.5) 0.025 MG/ML Injectable Suspension
1525543	SARS-CoV-2 (COVID-19) vaccine, mRNA-1273 0.05 MG/ML / SARS-CoV-2 (COVID-19) vaccine, mRNA-1273 OMICRON (BA.4/BA.5) 0.05 MG/ML Injectable Suspension
37003517	SARS-CoV-2 (COVID-19) vaccine, mRNA-1273 0.2 MG/ML
36391504	SARS-CoV-2 (COVID-19) vaccine, mRNA-1273 0.2 MG/ML Injectable Suspension
829421	SARS-CoV-2 (COVID-19) vaccine, mRNA-1273 0.2 MG/ML Injectable Suspension
36398553	SARS-CoV-2 (COVID-19) vaccine, mRNA-1273 0.2 MG/ML Injectable Suspension
37003518	SARS-CoV-2 (COVID-19) vaccine, mRNA-1273 0.2 MG/ML Injectable Suspension
42796198	SARS-CoV-2 (COVID-19) vaccine, mRNA-1273 0.2 MG/ML Injectable Suspension
2998096	SARS-CoV-2 (COVID-19) vaccine, mRNA-1273 0.2 MG/ML Injectable Suspension [Spikevax]
1994903	SARS-CoV-2 (COVID-19) vaccine, mRNA-1273 0.2 MG/ML Injectable Suspension [Spikevax]
2998095	SARS-CoV-2 (COVID-19) vaccine, mRNA-1273 0.2 MG/ML Injectable Suspension [Spikevax]
779679	SARS-CoV-2 (COVID-19) vaccine, mRNA-1273 0.2 MG/ML Injectable Suspension [Spikevax]
1994915	SARS-CoV-2 (COVID-19) vaccine, mRNA-1273 0.2 MG/ML Injectable Suspension [Spikevax]
742037	SARS-CoV-2 (COVID-19) vaccine, mRNA-1273 OMICRON (BA.4/BA.5) 0.025 MG/ML
1525542	SARS-CoV-2 (COVID-19) vaccine, mRNA-1273 OMICRON (BA.4/BA.5) 0.05 MG/ML
742041	SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2 0.0075 MG/ML / SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2 OMICRON (BA.4/BA.5) 0.0075 MG/ML Injectable Suspension
742009	SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2 0.025 MG/ML / SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2 OMICRON (BA.4/BA.5) 0.025 MG/ML Injectable Suspension
1525538	SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2 0.05 MG/ML / SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2 OMICRON (BA.4/BA.5) 0.05 MG/ML Injectable Suspension

Concept ID	Concept name
1525545	SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2 0.05 MG/ML / SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2OMICRON (BA.4/BA.5) 0.05 MG/ML Injection
1219271	SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2 0.1 MG/ML Injectable Suspension
37003436	SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2 0.1 MG/ML Injectable Suspension
1799560	SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2 0.1 MG/ML Injectable Suspension
1799559	SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2 0.1 MG/ML Injectable Suspension
1799558	SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2 0.1 MG/ML Injectable Suspension
1779597	SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2 0.1 MG/ML Injectable Suspension
42797616	SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2 0.1 MG/ML Injectable Suspension
42797615	SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2 0.1 MG/ML Injectable Suspension
36900251	SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2 0.1 MG/ML Injectable Suspension [Comirnaty]
36868891	SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2 0.1 MG/ML Injectable Suspension [Comirnaty]
595835	SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2 0.1 MG/ML Injectable Suspension [Comirnaty]
1759206	SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2 0.1 MG/ML Injectable Suspension [Comirnaty]
1626478	SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2 0.1 MG/ML Injectable Suspension [Comirnaty]
36868892	SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2 0.1 MG/ML Injectable Suspension [Comirnaty]
36900252	SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2 0.1 MG/ML Injectable Suspension [Comirnaty]
742040	SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2OMICRON (BA.4/BA.5) 0.0075 MG/ML
742007	SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2OMICRON (BA.4/BA.5) 0.025 MG/ML
1525544	SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2OMICRON (BA.4/BA.5) 0.05 MG/ML
37003432	SARS-CoV-2 (COVID-19) vaccine, mRNA spike protein
37003435	SARS-CoV-2 (COVID-19) vaccine, mRNA spike protein Injectable Suspension
1759205	SARS-CoV-2 (COVID-19) vaccine, mRNA spike protein Injectable Suspension [Comirnaty]
779678	SARS-CoV-2 (COVID-19) vaccine, mRNA spike protein Injectable Suspension [Spikevax]
739906	SARS-COV-2 (COVID-19) vaccine, vector - Ad26 100000000000 UNT/ML Injectable Suspension
702866	SARS-COV-2 (COVID-19) vaccine, vector non-replicating, recombinant spike protein-Ad26, preservative free, 0.5 mL
724905	SARS-COV-2 (COVID-19) vaccine, vector non-replicating, recombinant spike protein-ChAdOx1, preservative free, 0.5 mL (non-US)
42797615	SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2 0.1 MG/ML Injectable Suspension
36900251	SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2 0.1 MG/ML Injectable Suspension [Comirnaty]
36868891	SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2 0.1 MG/ML Injectable Suspension [Comirnaty]
595835	SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2 0.1 MG/ML Injectable Suspension [Comirnaty]
1759206	SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2 0.1 MG/ML Injectable Suspension [Comirnaty]
1626478	SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2 0.1 MG/ML Injectable Suspension [Comirnaty]
36868892	SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2 0.1 MG/ML Injectable Suspension [Comirnaty]
36900252	SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2 0.1 MG/ML Injectable Suspension [Comirnaty]
742040	SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2OMICRON (BA.4/BA.5) 0.0075 MG/ML

Concept ID	Concept name
742007	SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2 OMICRON (BA.4/BA.5) 0.025 MG/ML
1525544	SARS-CoV-2 (COVID-19) vaccine, mRNA-BNT162b2 OMICRON (BA.4/BA.5) 0.05 MG/ML
37003432	SARS-CoV-2 (COVID-19) vaccine, mRNA spike protein
37003435	SARS-CoV-2 (COVID-19) vaccine, mRNA spike protein Injectable Suspension
1759205	SARS-CoV-2 (COVID-19) vaccine, mRNA spike protein Injectable Suspension [Comirnaty]
779678	SARS-CoV-2 (COVID-19) vaccine, mRNA spike protein Injectable Suspension [Spikevax]
739906	SARS-COV-2 (COVID-19) vaccine, vector - Ad26 100000000000 UNT/ML Injectable Suspension
702866	SARS-COV-2 (COVID-19) vaccine, vector non-replicating, recombinant spike protein-Ad26, preservative free, 0.5 mL
724905	SARS-COV-2 (COVID-19) vaccine, vector non-replicating, recombinant spike protein-ChAdOx1, preservative free, 0.5 mL (non-US)

Influenza vaccination:

Concept ID	Concept name
1361667	0.5 ML influenza A virus A/Brisbane/02/2018 (H1N1) antigen 0.03 MG/ML / influenza A virus A/Kansas/14/2017 (H3N2) antigen 0.03 MG/ML / influenza B virus B/Maryland/15/2016 antigen 0.03 MG/ML Prefilled Syringe
1361668	0.5 ML influenza A virus A/Brisbane/02/2018 (H1N1) antigen 0.03 MG/ML / influenza A virus A/Kansas/14/2017 (H3N2) antigen 0.03 MG/ML / influenza B virus B/Maryland/15/2016 antigen 0.03 MG/ML Prefilled Syringe [Fluad 2019-2020]
1361663	0.5 ML influenza A virus A/Brisbane/02/2018 (H1N1) antigen 0.09 MG/ML / influenza A virus A/Kansas/14/2017 (H3N2) antigen 0.09 MG/ML / influenza B virus B/Maryland/15/2016 antigen 0.09 MG/ML / influenza B virus B/Phuket/3073/2013 antigen 0.09 MG/ML Pre...
1361664	0.5 ML influenza A virus A/Brisbane/02/2018 (H1N1) antigen 0.09 MG/ML / influenza A virus A/Kansas/14/2017 (H3N2) antigen 0.09 MG/ML / influenza B virus B/Maryland/15/2016 antigen 0.09 MG/ML / influenza B virus B/Phuket/3073/2013 antigen 0.09 MG/ML Pre...
35606414	0.5 ML influenza A virus A/California/7/2009 (H1N1) antigen 0.015 MG/ML / influenza A virus A/Switzerland/9715293/2013 (H3N2) antigen 0.015 MG/ML / influenza B virus B/Brisbane/9/2014 antigen 0.015 MG/ML Prefilled Syringe
35606416	0.5 ML influenza A virus A/California/7/2009 (H1N1) antigen 0.015 MG/ML / influenza A virus A/Switzerland/9715293/2013 (H3N2) antigen 0.015 MG/ML / influenza B virus B/Brisbane/9/2014 antigen 0.015 MG/ML Prefilled Syringe [FLUAD 2015-2016]
36249259	0.5 ML influenza A virus A/California/7/2009 (H1N1) antigen 0.03 MG/ML / influenza A virus A/Hong Kong/4801/2014 (H3N2) antigen 0.03 MG/ML / influenza B virus B/Brisbane/60/2008 antigen 0.03 MG/ML Prefilled Syringe
36249411	0.5 ML influenza A virus A/California/7/2009 (H1N1) antigen 0.03 MG/ML / influenza A virus A/Hong Kong/4801/2014 (H3N2) antigen 0.03 MG/ML / influenza B virus B/Brisbane/60/2008 antigen 0.03 MG/ML Prefilled Syringe [Fluad 2016-2017]
40220827	0.5 ML influenza A virus A/California/7/2009 (H1N1) antigen 0.09 MG/ML / influenza A virus A/Hong Kong/4801/2014 (H3N2) antigen 0.09 MG/ML / influenza B virus B/Brisbane/60/2008 antigen 0.09 MG/ML Injection
40220832	0.5 ML influenza A virus A/California/7/2009 (H1N1) antigen 0.09 MG/ML / influenza A virus A/Hong Kong/4801/2014 (H3N2) antigen 0.09 MG/ML / influenza B virus B/Brisbane/60/2008 antigen 0.09 MG/ML Injection [Flublok 2016-2017]

Concept ID	Concept name
46276014	0.5 ML influenza A virus A/California/7/2009 (H1N1) antigen 0.09 MG/ML / influenza A virus A/Switzerland/9715293/2013 (H3N2) antigen 0.09 MG/ML / influenza B virus B/Phuket/3073/2013 antigen 0.09 MG/ML Injection
46276018	0.5 ML influenza A virus A/California/7/2009 (H1N1) antigen 0.09 MG/ML / influenza A virus A/Switzerland/9715293/2013 (H3N2) antigen 0.09 MG/ML / influenza B virus B/Phuket/3073/2013 antigen 0.09 MG/ML Injection [Flublok 2015-2016]
747766	0.5 ML influenza A virus A/Darwin/6/2021 (H3N2) antigen 0.09 MG/ML / influenza A virus A/West Virginia/30/2022 (H1N1) antigen 0.09 MG/ML / influenza B virus B/Austria/1359417/2021 antigen 0.09 MG/ML / influenza B virus B/Phuket/3073/2013 antigen 0.09 M...
747765	0.5 ML influenza A virus A/Darwin/6/2021 (H3N2) antigen 0.09 MG/ML / influenza A virus A/West Virginia/30/2022 (H1N1) antigen 0.09 MG/ML / influenza B virus B/Austria/1359417/2021 antigen 0.09 MG/ML / influenza B virus B/Phuket/3073/2013 antigen 0.09 M...
780304	0.5 ML influenza A virus A/Darwin/6/2021 (H3N2) antigen 0.09 MG/ML / influenza A virus A/Wisconsin/588/2019 (H1N1) antigen 0.09 MG/ML / influenza B virus B/Austria/1359417/2021 antigen 0.09 MG/ML / influenza B virus B/Phuket/3073/2013 antigen 0.09 MG/M...
780305	0.5 ML influenza A virus A/Darwin/6/2021 (H3N2) antigen 0.09 MG/ML / influenza A virus A/Wisconsin/588/2019 (H1N1) antigen 0.09 MG/ML / influenza B virus B/Austria/1359417/2021 antigen 0.09 MG/ML / influenza B virus B/Phuket/3073/2013 antigen 0.09 MG/M...
1146798	0.5 ML influenza A virus A/Hawaii/70/2019 (H1N1) antigen 0.09 MG/ML / influenza A virus A/Minnesota/41/2019 (H3N2) antigen 0.09 MG/ML / influenza B virus B/Phuket/3073/2013 antigen 0.09 MG/ML / influenza B virus B/Washington/02/2019 antigen 0.09 MG/ML...
1146799	0.5 ML influenza A virus A/Hawaii/70/2019 (H1N1) antigen 0.09 MG/ML / influenza A virus A/Minnesota/41/2019 (H3N2) antigen 0.09 MG/ML / influenza B virus B/Phuket/3073/2013 antigen 0.09 MG/ML / influenza B virus B/Washington/02/2019 antigen 0.09 MG/ML...
1146800	0.5 ML influenza A virus A/Hong Kong/2671/2019 (H3N2) antigen 0.03 MG/ML / influenza A virus A/Victoria/2454/2019 (H1N1) antigen 0.03 MG/ML / influenza B virus B/Victoria/705/2018 antigen 0.03 MG/ML Prefilled Syringe
1146801	0.5 ML influenza A virus A/Hong Kong/2671/2019 (H3N2) antigen 0.03 MG/ML / influenza A virus A/Victoria/2454/2019 (H1N1) antigen 0.03 MG/ML / influenza B virus B/Victoria/705/2018 antigen 0.03 MG/ML Prefilled Syringe [Fluad 2020-2021]
1593896	0.5 ML influenza A virus A/Hong Kong/4801/2014 (H3N2) antigen 0.03 MG/ML / influenza A virus A/Singapore/GP1908/2015 (H1N1) antigen 0.03 MG/ML / influenza B virus B/Brisbane/60/2008 antigen 0.03 MG/ML Prefilled Syringe
1593901	0.5 ML influenza A virus A/Hong Kong/4801/2014 (H3N2) antigen 0.03 MG/ML / influenza A virus A/Singapore/GP1908/2015 (H1N1) antigen 0.03 MG/ML / influenza B virus B/Brisbane/60/2008 antigen 0.03 MG/ML Prefilled Syringe [Fluad 2017-2018]
1593520	0.5 ML influenza A virus A/Hong Kong/4801/2014 (H3N2) antigen 0.09 MG/ML / influenza A virus A/Michigan/45/2015 (H1N1) antigen 0.09 MG/ML / influenza B virus B/Brisbane/60/2008 antigen 0.09 MG/ML / influenza B virus B/Phuket/3073/2013 antigen 0.09 MG/M...
1593522	0.5 ML influenza A virus A/Hong Kong/4801/2014 (H3N2) antigen 0.09 MG/ML / influenza A virus A/Michigan/45/2015 (H1N1) antigen 0.09 MG/ML / influenza B virus B/Brisbane/60/2008 antigen 0.09 MG/ML / influenza B virus B/Phuket/3073/2013 antigen 0.09 MG/M...
1593537	0.5 ML influenza A virus A/Hong Kong/4801/2014 (H3N2) antigen 0.09 MG/ML / influenza A virus A/Michigan/45/2015 (H1N1) antigen 0.09 MG/ML / influenza B virus B/Brisbane/60/2008 antigen 0.09 MG/ML Injection
1593876	0.5 ML influenza A virus A/Hong Kong/4801/2014 (H3N2) antigen 0.09 MG/ML / influenza A virus A/Michigan/45/2015 (H1N1) antigen 0.09 MG/ML / influenza B virus B/Brisbane/60/2008 antigen 0.09 MG/ML Injection [Flublok 2017-2018]

Concept ID	Concept name
35201910	0.5 ML influenza A virus A/Michigan/45/2015 (H1N1) antigen 0.09 MG/ML / influenza A virus A/Singapore/INFIMH-16-0019/2016 (H3N2) antigen 0.09 MG/ML / influenza B virus B/Maryland/15/2016 antigen 0.09 MG/ML / influenza B virus B/Phuket/3073/2013 antigen...
35201911	0.5 ML influenza A virus A/Michigan/45/2015 (H1N1) antigen 0.09 MG/ML / influenza A virus A/Singapore/INFIMH-16-0019/2016 (H3N2) antigen 0.09 MG/ML / influenza B virus B/Maryland/15/2016 antigen 0.09 MG/ML / influenza B virus B/Phuket/3073/2013 antigen...
35201914	0.5 ML influenza A virus A/Singapore/GP1908/2015 (H1N1) antigen 0.03 MG/ML / influenza A virus A/Singapore/INFIMH-16-0019/2016 (H3N2) antigen 0.03 MG/ML / influenza B virus B/Maryland/15/2016 antigen 0.03 MG/ML Prefilled Syringe
35201915	0.5 ML influenza A virus A/Singapore/GP1908/2015 (H1N1) antigen 0.03 MG/ML / influenza A virus A/Singapore/INFIMH-16-0019/2016 (H3N2) antigen 0.03 MG/ML / influenza B virus B/Maryland/15/2016 antigen 0.03 MG/ML Prefilled Syringe [Fluad 2018-2019]
1537647	0.5 ML influenza A virus A/Tasmania/503/2020 (H3N2) antigen 0.09 MG/ML / influenza A virus A/Wisconsin/588/2019 (H1N1) antigen 0.09 MG/ML / influenza B virus B/Phuket/3073/2013 antigen 0.09 MG/ML / influenza B virus B/Washington/02/2019 antigen 0.09 MG...
1537643	0.5 ML influenza A virus A/Tasmania/503/2020 (H3N2) antigen 0.09 MG/ML / influenza A virus A/Wisconsin/588/2019 (H1N1) antigen 0.09 MG/ML / influenza B virus B/Phuket/3073/2013 antigen 0.09 MG/ML / influenza B virus B/Washington/02/2019 antigen 0.09 MG...
43810065	0.5 ML influenza A virus (H1N1) antigen 0.03 MG/ML / Influenza A virus vaccine, A-California-7-2009 (H1N1)-like virus 0.03 MG/ML / influenza B virus antigen 0.03 MG/ML Injectable Suspension [Afluria]
43828149	0.5 ML influenza A virus (H1N1) antigen 0.03 MG/ML / Influenza A virus vaccine, A-California-7-2009 (H1N1)-like virus 0.03 MG/ML / influenza B virus antigen 0.03 MG/ML Injectable Suspension [Afluria] by bioCSL
43594165	0.5 ML influenza A virus (H1N1) antigen 0.03 MG/ML Injectable Suspension [Afluria]
43720036	0.5 ML influenza A virus (H1N1) antigen 0.03 MG/ML Injectable Suspension [Afluria] Box of 10
43864462	0.5 ML influenza A virus (H1N1) antigen 0.03 MG/ML Injectable Suspension [Afluria] Box of 10 by bioCSL
43612139	0.5 ML influenza A virus (H1N1) antigen 0.03 MG/ML Injectable Suspension [Afluria] by bioCSL
43612132	0.5 ML Influenza A virus vaccine, A-California-7-2009 (H1N1)-like virus 0.03 MG/ML Injectable Suspension [Afluria]
43702292	0.5 ML Influenza A virus vaccine, A-California-7-2009 (H1N1)-like virus 0.03 MG/ML Injectable Suspension [Afluria] Box of 10
43702293	0.5 ML Influenza A virus vaccine, A-California-7-2009 (H1N1)-like virus 0.03 MG/ML Injectable Suspension [Afluria] Box of 10 by bioCSL
43791911	0.5 ML Influenza A virus vaccine, A-California-7-2009 (H1N1)-like virus 0.03 MG/ML Injectable Suspension [Afluria] by bioCSL
43612144	0.5 ML influenza B virus antigen 0.03 MG/ML Injectable Suspension [Afluria]
43702304	0.5 ML influenza B virus antigen 0.03 MG/ML Injectable Suspension [Afluria] Box of 10
43594169	0.5 ML influenza B virus antigen 0.03 MG/ML Injectable Suspension [Afluria] Box of 10 by bioCSL
43738160	0.5 ML influenza B virus antigen 0.03 MG/ML Injectable Suspension [Afluria] by bioCSL
43259431	0.5 ML Influenza B virus vaccine, B-Brisbane-60-2008-like virus 0.015 MG/ML / Influenza Virus Vaccine, Inactivated A-California-07-2009 X-181 (H1N1) strain 0.015 MG/ML / ... Injection [Fluvirin]
43270331	0.5 ML Influenza B virus vaccine, B-Brisbane-60-2008-like virus 0.015 MG/ML / Influenza Virus Vaccine, Inactivated A-California-07-2009 X-181 (H1N1) strain 0.015 MG/ML / ... Injection [Fluvirin] Box of 1
43297451	0.5 ML Influenza B virus vaccine, B-Brisbane-60-2008-like virus 0.015 MG/ML / Influenza Virus Vaccine, Inactivated A-California-07-2009 X-181 (H1N1) strain 0.015 MG/ML / ... Injection [Fluvirin] Box of 10

Concept ID	Concept name
40225027	influenza A-California-7-2009-(H1N1)v-like virus vaccine 0.03 MG/ML / influenza B virus vaccine B/Brisbane/60/2008 antigen 0.03 MG/ML / influenza virus vaccine, inactivated A-Victoria-210-2009 X-187 (H3N2) (A-Perth-16-2009) strain 0.03 MG/ML Injectable...
40240930	influenza A-California-7-2009-(H1N1)v-like virus vaccine 0.03 MG/ML / influenza B virus vaccine B/Brisbane/60/2008 antigen 0.03 MG/ML / influenza virus vaccine, inactivated A-Victoria-210-2009 X-187 (H3N2) (A-Perth-16-2009) strain 0.03 MG/ML Injectable...
40240929	influenza A-California-7-2009-(H1N1)v-like virus vaccine 0.03 MG/ML / influenza B virus vaccine B/Brisbane/60/2008 antigen 0.03 MG/ML / influenza virus vaccine, inactivated A-Victoria-210-2009 X-187 (H3N2) (A-Perth-16-2009) strain 0.03 MG/ML Injectable...
40240134	influenza A-California-7-2009-(H1N1)v-like virus vaccine 0.03 MG/ML / influenza B virus vaccine B/Brisbane/60/2008 antigen 0.03 MG/ML / influenza virus vaccine, inactivated A-Victoria-210-2009 X-187 (H3N2) (A-Perth-16-2009) strain 0.03 MG/ML Injectable...
40225026	influenza A-California-7-2009-(H1N1)v-like virus vaccine 0.03 MG/ML / influenza B virus vaccine B/Brisbane/60/2008 antigen 0.03 MG/ML / influenza virus vaccine, inactivated A-Victoria-210-2009 X-187 (H3N2) (A-Perth-16-2009) strain 0.03 MG/ML Injectable...
40225007	influenza A-California-7-2009-(H1N1)v-like virus vaccine 0.03 MG/ML / influenza B virus vaccine B/Brisbane/60/2008 antigen 0.03 MG/ML / influenza virus vaccine, inactivated A-Victoria-210-2009 X-187 (H3N2) (A-Perth-16-2009) strain 0.03 MG/ML Injectable...
40240925	influenza A-California-7-2009-(H1N1)v-like virus vaccine / influenza B virus vaccine B/Brisbane/60/2008 antigen / influenza virus vaccine, inactivated A-Victoria-210-2009 X-187 (H3N2) (A-Perth-16-2009) strain Injectable Suspension
40225015	influenza A-California-7-2009-(H1N1)v-like virus vaccine / influenza B virus vaccine B/Brisbane/60/2008 antigen / influenza virus vaccine, inactivated A-Victoria-210-2009 X-187 (H3N2) (A-Perth-16-2009) strain Injectable Suspension [Afluria 2010-2011 Fo...
40225014	influenza A-California-7-2009-(H1N1)v-like virus vaccine / influenza B virus vaccine B/Brisbane/60/2008 antigen / influenza virus vaccine, inactivated A-Victoria-210-2009 X-187 (H3N2) (A-Perth-16-2009) strain Injectable Suspension [Fluvirin 2010-2011 F...
40240926	influenza A-California-7-2009-(H1N1)v-like virus vaccine / influenza B virus vaccine B/Brisbane/60/2008 antigen / influenza virus vaccine, inactivated A-Victoria-210-2009 X-187 (H3N2) (A-Perth-16-2009) strain Injectable Suspension [Fluvirin 2011-2012 F...
40224999	influenza A-California-7-2009-(H1N1)v-like virus vaccine / influenza B virus vaccine B/Brisbane/60/2008 antigen / influenza virus vaccine, inactivated A-Victoria-210-2009 X-187 (H3N2) (A-Perth-16-2009) strain Injectable Suspension [Fluzone 2010-2011 Fo...
40240131	influenza A-California-7-2009-(H1N1)v-like virus vaccine / influenza B virus vaccine B/Brisbane/60/2008 antigen / influenza virus vaccine, inactivated A-Victoria-210-2009 X-187 (H3N2) (A-Perth-16-2009) strain Injectable Suspension [Fluzone 2011-2012 Fo...
1361279	influenza A virus A/Brisbane/02/2018 (H1N1) antigen 0.03 MG/ML / influenza A virus A/Kansas/14/2017 (H3N2) antigen 0.03 MG/ML / influenza B virus B/Maryland/15/2016 antigen 0.03 MG/ML Prefilled Syringe
1361280	influenza A virus A/Brisbane/02/2018 (H1N1) antigen 0.03 MG/ML / influenza A virus A/Kansas/14/2017 (H3N2) antigen 0.03 MG/ML / influenza B virus B/Maryland/15/2016 antigen 0.03 MG/ML Prefilled Syringe [Fluad 2019-2020]
1361154	influenza A virus A/Brisbane/02/2018 (H1N1) antigen 0.09 MG/ML / influenza A virus A/Kansas/14/2017 (H3N2) antigen 0.09 MG/ML / influenza B virus B/Maryland/15/2016 antigen 0.09 MG/ML / influenza B virus B/Phuket/3073/2013 antigen 0.09 MG/ML Prefilled...
1361155	influenza A virus A/Brisbane/02/2018 (H1N1) antigen 0.09 MG/ML / influenza A virus A/Kansas/14/2017 (H3N2) antigen 0.09 MG/ML / influenza B virus B/Maryland/15/2016 antigen 0.09 MG/ML / influenza B virus B/Phuket/3073/2013 antigen 0.09 MG/ML Prefilled...

Concept ID	Concept name
35606417	influenza A virus A/California/7/2009 (H1N1) antigen 0.015 MG/ML / influenza A virus A/Switzerland/9715293/2013 (H3N2) antigen 0.015 MG/ML / influenza B virus B/Brisbane/9/2014 antigen 0.015 MG/ML Prefilled Syringe
35606418	influenza A virus A/California/7/2009 (H1N1) antigen 0.015 MG/ML / influenza A virus A/Switzerland/9715293/2013 (H3N2) antigen 0.015 MG/ML / influenza B virus B/Brisbane/9/2014 antigen 0.015 MG/ML Prefilled Syringe [FLUAD 2015-2016]
36249270	influenza A virus A/California/7/2009 (H1N1) antigen 0.03 MG/ML / influenza A virus A/Hong Kong/4801/2014 (H3N2) antigen 0.03 MG/ML / influenza B virus B/Brisbane/60/2008 antigen 0.03 MG/ML Injectable Suspension
36249272	influenza A virus A/California/7/2009 (H1N1) antigen 0.03 MG/ML / influenza A virus A/Hong Kong/4801/2014 (H3N2) antigen 0.03 MG/ML / influenza B virus B/Brisbane/60/2008 antigen 0.03 MG/ML Injectable Suspension [Afluria 2016-2017]
36249265	influenza A virus A/California/7/2009 (H1N1) antigen 0.03 MG/ML / influenza A virus A/Hong Kong/4801/2014 (H3N2) antigen 0.03 MG/ML / influenza B virus B/Brisbane/60/2008 antigen 0.03 MG/ML Prefilled Syringe
36249412	influenza A virus A/California/7/2009 (H1N1) antigen 0.03 MG/ML / influenza A virus A/Hong Kong/4801/2014 (H3N2) antigen 0.03 MG/ML / influenza B virus B/Brisbane/60/2008 antigen 0.03 MG/ML Prefilled Syringe [Fluad 2016-2017]
46276313	influenza A virus A/California/7/2009 (H1N1) antigen 0.03 MG/ML / influenza A virus A/South Australia/55/2014 (H3N2) antigen 0.03 MG/ML / influenza B virus B/Phuket/3073/2013 antigen 0.03 MG/ML Injectable Suspension
46276315	influenza A virus A/California/7/2009 (H1N1) antigen 0.03 MG/ML / influenza A virus A/South Australia/55/2014 (H3N2) antigen 0.03 MG/ML / influenza B virus B/Phuket/3073/2013 antigen 0.03 MG/ML Injectable Suspension [Afluria 2015-2016]
46275381	influenza A virus A/California/7/2009 (H1N1) antigen 0.03 MG/ML / influenza A virus A/Switzerland/9715293/2013 (H3N2) antigen 0.03 MG/ML / influenza B virus B/Phuket/3073/2013 antigen 0.03 MG/ML Injectable Suspension
46275385	influenza A virus A/California/7/2009 (H1N1) antigen 0.03 MG/ML / influenza A virus A/Switzerland/9715293/2013 (H3N2) antigen 0.03 MG/ML / influenza B virus B/Phuket/3073/2013 antigen 0.03 MG/ML Injectable Suspension [Fluzone 2015-2016]
40220833	influenza A virus A/California/7/2009 (H1N1) antigen 0.09 MG/ML / influenza A virus A/Hong Kong/4801/2014 (H3N2) antigen 0.09 MG/ML / influenza B virus B/Brisbane/60/2008 antigen 0.09 MG/ML Injection
40220834	influenza A virus A/California/7/2009 (H1N1) antigen 0.09 MG/ML / influenza A virus A/Hong Kong/4801/2014 (H3N2) antigen 0.09 MG/ML / influenza B virus B/Brisbane/60/2008 antigen 0.09 MG/ML Injection [Flublok 2016-2017]
46276019	influenza A virus A/California/7/2009 (H1N1) antigen 0.09 MG/ML / influenza A virus A/Switzerland/9715293/2013 (H3N2) antigen 0.09 MG/ML / influenza B virus B/Phuket/3073/2013 antigen 0.09 MG/ML Injection
46276020	influenza A virus A/California/7/2009 (H1N1) antigen 0.09 MG/ML / influenza A virus A/Switzerland/9715293/2013 (H3N2) antigen 0.09 MG/ML / influenza B virus B/Phuket/3073/2013 antigen 0.09 MG/ML Injection [Flublok 2015-2016]
36249426	influenza A virus A/Christchurch/16/2010 (H1N1) antigen 0.03 MG/ML / influenza A virus A/Hong Kong/4801/2014 (H3N2) antigen 0.03 MG/ML / influenza B virus B/Brisbane/60/2008 antigen 0.03 MG/ML Injectable Suspension
36249431	influenza A virus A/Christchurch/16/2010 (H1N1) antigen 0.03 MG/ML / influenza A virus A/Hong Kong/4801/2014 (H3N2) antigen 0.03 MG/ML / influenza B virus B/Brisbane/60/2008 antigen 0.03 MG/ML Injectable Suspension [Fluvirin 2016-2017]

Concept ID	Concept name
46276178	influenza A virus A/Christchurch/16/2010 (H1N1) antigen 0.03 MG/ML / influenza A virus A/Switzerland/9715293/2013 (H3N2) antigen 0.03 MG/ML / influenza B virus B/Phuket/3073/2013 antigen 0.03 MG/ML Injectable Suspension
46276182	influenza A virus A/Christchurch/16/2010 (H1N1) antigen 0.03 MG/ML / influenza A virus A/Switzerland/9715293/2013 (H3N2) antigen 0.03 MG/ML / influenza B virus B/Phuket/3073/2013 antigen 0.03 MG/ML Injectable Suspension [Fluvirin 2015-2016]
746137	influenza A virus A/Darwin/6/2021 (H3N2) antigen 0.09 MG/ML / influenza A virus A/West Virginia/30/2022 (H1N1) antigen 0.09 MG/ML / influenza B virus B/Austria/1359417/2021 antigen 0.09 MG/ML / influenza B virus B/Phuket/3073/2013 antigen 0.09 MG/ML Pr...
746132	influenza A virus A/Darwin/6/2021 (H3N2) antigen 0.09 MG/ML / influenza A virus A/West Virginia/30/2022 (H1N1) antigen 0.09 MG/ML / influenza B virus B/Austria/1359417/2021 antigen 0.09 MG/ML / influenza B virus B/Phuket/3073/2013 antigen 0.09 MG/ML Pr...
780110	influenza A virus A/Darwin/6/2021 (H3N2) antigen 0.09 MG/ML / influenza A virus A/Wisconsin/588/2019 (H1N1) antigen 0.09 MG/ML / influenza B virus B/Austria/1359417/2021 antigen 0.09 MG/ML / influenza B virus B/Phuket/3073/2013 antigen 0.09 MG/ML Prefi...
780105	influenza A virus A/Darwin/6/2021 (H3N2) antigen 0.09 MG/ML / influenza A virus A/Wisconsin/588/2019 (H1N1) antigen 0.09 MG/ML / influenza B virus B/Austria/1359417/2021 antigen 0.09 MG/ML / influenza B virus B/Phuket/3073/2013 antigen 0.09 MG/ML Prefi...
1146508	influenza A virus A/Hawaii/70/2019 (H1N1) antigen 0.09 MG/ML / influenza A virus A/Minnesota/41/2019 (H3N2) antigen 0.09 MG/ML / influenza B virus B/Phuket/3073/2013 antigen 0.09 MG/ML / influenza B virus B/Washington/02/2019 antigen 0.09 MG/ML Prefill...
1146507	influenza A virus A/Hawaii/70/2019 (H1N1) antigen 0.09 MG/ML / influenza A virus A/Minnesota/41/2019 (H3N2) antigen 0.09 MG/ML / influenza B virus B/Phuket/3073/2013 antigen 0.09 MG/ML / influenza B virus B/Washington/02/2019 antigen 0.09 MG/ML Prefill...
1146534	influenza A virus A/Hong Kong/2671/2019 (H3N2) antigen 0.03 MG/ML / influenza A virus A/Victoria/2454/2019 (H1N1) antigen 0.03 MG/ML / influenza B virus B/Victoria/705/2018 antigen 0.03 MG/ML Prefilled Syringe
1146535	influenza A virus A/Hong Kong/2671/2019 (H3N2) antigen 0.03 MG/ML / influenza A virus A/Victoria/2454/2019 (H1N1) antigen 0.03 MG/ML / influenza B virus B/Victoria/705/2018 antigen 0.03 MG/ML Prefilled Syringe [Fluad 2020-2021]
793959	influenza A virus A/Hong Kong/4801/2014 (H3N2) antigen 0.03 MG/ML / influenza A virus A/Singapore/GP1908/2015 (H1N1) antigen 0.03 MG/ML / influenza B virus B/Brisbane/46/2015 antigen 0.03 MG/ML Injectable Suspension
793961	influenza A virus A/Hong Kong/4801/2014 (H3N2) antigen 0.03 MG/ML / influenza A virus A/Singapore/GP1908/2015 (H1N1) antigen 0.03 MG/ML / influenza B virus B/Brisbane/46/2015 antigen 0.03 MG/ML Injectable Suspension [Afluria 2017-2018]
1593904	influenza A virus A/Hong Kong/4801/2014 (H3N2) antigen 0.03 MG/ML / influenza A virus A/Singapore/GP1908/2015 (H1N1) antigen 0.03 MG/ML / influenza B virus B/Brisbane/60/2008 antigen 0.03 MG/ML Injectable Suspension
1593909	influenza A virus A/Hong Kong/4801/2014 (H3N2) antigen 0.03 MG/ML / influenza A virus A/Singapore/GP1908/2015 (H1N1) antigen 0.03 MG/ML / influenza B virus B/Brisbane/60/2008 antigen 0.03 MG/ML Injectable Suspension [Fluvirin 2017-2018]
1593902	influenza A virus A/Hong Kong/4801/2014 (H3N2) antigen 0.03 MG/ML / influenza A virus A/Singapore/GP1908/2015 (H1N1) antigen 0.03 MG/ML / influenza B virus B/Brisbane/60/2008 antigen 0.03 MG/ML Prefilled Syringe
1593903	influenza A virus A/Hong Kong/4801/2014 (H3N2) antigen 0.03 MG/ML / influenza A virus A/Singapore/GP1908/2015 (H1N1) antigen 0.03 MG/ML / influenza B virus B/Brisbane/60/2008 antigen 0.03 MG/ML Prefilled Syringe [Fluad 2017-2018]

Concept ID	Concept name
1593524	influenza A virus A/Hong Kong/4801/2014 (H3N2) antigen 0.09 MG/ML / influenza A virus A/Michigan/45/2015 (H1N1) antigen 0.09 MG/ML / influenza B virus B/Brisbane/60/2008 antigen 0.09 MG/ML / influenza B virus B/Phuket/3073/2013 antigen 0.09 MG/ML Prefi...
1593523	influenza A virus A/Hong Kong/4801/2014 (H3N2) antigen 0.09 MG/ML / influenza A virus A/Michigan/45/2015 (H1N1) antigen 0.09 MG/ML / influenza B virus B/Brisbane/60/2008 antigen 0.09 MG/ML / influenza B virus B/Phuket/3073/2013 antigen 0.09 MG/ML Prefi...
1593538	influenza A virus A/Hong Kong/4801/2014 (H3N2) antigen 0.09 MG/ML / influenza A virus A/Michigan/45/2015 (H1N1) antigen 0.09 MG/ML / influenza B virus B/Brisbane/60/2008 antigen 0.09 MG/ML Injection
1593877	influenza A virus A/Hong Kong/4801/2014 (H3N2) antigen 0.09 MG/ML / influenza A virus A/Michigan/45/2015 (H1N1) antigen 0.09 MG/ML / influenza B virus B/Brisbane/60/2008 antigen 0.09 MG/ML Injection [Flublok 2017-2018]
35200027	influenza A virus A/Michigan/45/2015 (H1N1) antigen 0.09 MG/ML / influenza A virus A/Singapore/INFIMH-16-0019/2016 (H3N2) antigen 0.09 MG/ML / influenza B virus B/Maryland/15/2016 antigen 0.09 MG/ML / influenza B virus B/Phuket/3073/2013 antigen 0.09 M...
35200028	influenza A virus A/Michigan/45/2015 (H1N1) antigen 0.09 MG/ML / influenza A virus A/Singapore/INFIMH-16-0019/2016 (H3N2) antigen 0.09 MG/ML / influenza B virus B/Maryland/15/2016 antigen 0.09 MG/ML / influenza B virus B/Phuket/3073/2013 antigen 0.09 M...
35200049	influenza A virus A/Singapore/GP1908/2015 (H1N1) antigen 0.03 MG/ML / influenza A virus A/Singapore/INFIMH-16-0019/2016 (H3N2) antigen 0.03 MG/ML / influenza B virus B/Maryland/15/2016 antigen 0.03 MG/ML Injectable Suspension
35200054	influenza A virus A/Singapore/GP1908/2015 (H1N1) antigen 0.03 MG/ML / influenza A virus A/Singapore/INFIMH-16-0019/2016 (H3N2) antigen 0.03 MG/ML / influenza B virus B/Maryland/15/2016 antigen 0.03 MG/ML Injectable Suspension [Afluria 2018-2019]
35200047	influenza A virus A/Singapore/GP1908/2015 (H1N1) antigen 0.03 MG/ML / influenza A virus A/Singapore/INFIMH-16-0019/2016 (H3N2) antigen 0.03 MG/ML / influenza B virus B/Maryland/15/2016 antigen 0.03 MG/ML Prefilled Syringe
35200048	influenza A virus A/Singapore/GP1908/2015 (H1N1) antigen 0.03 MG/ML / influenza A virus A/Singapore/INFIMH-16-0019/2016 (H3N2) antigen 0.03 MG/ML / influenza B virus B/Maryland/15/2016 antigen 0.03 MG/ML Prefilled Syringe [Fluad 2018-2019]
1537482	influenza A virus A/Tasmania/503/2020 (H3N2) antigen 0.09 MG/ML / influenza A virus A/Wisconsin/588/2019 (H1N1) antigen 0.09 MG/ML / influenza B virus B/Phuket/3073/2013 antigen 0.09 MG/ML / influenza B virus B/Washington/02/2019 antigen 0.09 MG/ML Pre...
1537508	influenza A virus A/Tasmania/503/2020 (H3N2) antigen 0.09 MG/ML / influenza A virus A/Wisconsin/588/2019 (H1N1) antigen 0.09 MG/ML / influenza B virus B/Phuket/3073/2013 antigen 0.09 MG/ML / influenza B virus B/Washington/02/2019 antigen 0.09 MG/ML Pre...
43625193	influenza A virus (H1N1) antigen 0.03 MG/ML / Influenza A virus vaccine, A-California-7-2009 (H1N1)-like virus 0.03 MG/ML / influenza B virus antigen 0.03 MG/ML Injectable Suspension [Afluria]
43643464	influenza A virus (H1N1) antigen 0.03 MG/ML Injectable Suspension [Afluria]
43607294	influenza A virus (H1N1) antigen 0.03 MG/ML Injectable Suspension [Afluria] Box of 10
46276177	influenza A virus (H1N1) antigen / influenza A virus (H3N2) antigen / influenza B virus antigen Injectable Suspension
46276314	influenza A virus (H1N1) antigen / influenza A virus (H3N2) antigen / influenza B virus antigen Injectable Suspension [Afluria 2015-2016]
36249271	influenza A virus (H1N1) antigen / influenza A virus (H3N2) antigen / influenza B virus antigen Injectable Suspension [Afluria 2016-2017]
793960	influenza A virus (H1N1) antigen / influenza A virus (H3N2) antigen / influenza B virus antigen Injectable Suspension [Afluria 2017-2018]

Concept ID	Concept name
35200052	influenza A virus (H1N1) antigen / influenza A virus (H3N2) antigen / influenza B virus antigen Injectable Suspension [Afluria 2018-2019]
46276181	influenza A virus (H1N1) antigen / influenza A virus (H3N2) antigen / influenza B virus antigen Injectable Suspension [Fluvirin 2015-2016]
36249429	influenza A virus (H1N1) antigen / influenza A virus (H3N2) antigen / influenza B virus antigen Injectable Suspension [Fluvirin 2016-2017]
1593907	influenza A virus (H1N1) antigen / influenza A virus (H3N2) antigen / influenza B virus antigen Injectable Suspension [Fluvirin 2017-2018]
46275384	influenza A virus (H1N1) antigen / influenza A virus (H3N2) antigen / influenza B virus antigen Injectable Suspension [Fluzone 2015-2016]
46275653	influenza A virus (H1N1) antigen / influenza A virus (H3N2) antigen / influenza B virus antigen Injection
46276017	influenza A virus (H1N1) antigen / influenza A virus (H3N2) antigen / influenza B virus antigen Injection [Flublok 2015-2016]
40220830	influenza A virus (H1N1) antigen / influenza A virus (H3N2) antigen / influenza B virus antigen Injection [Flublok 2016-2017]
1593874	influenza A virus (H1N1) antigen / influenza A virus (H3N2) antigen / influenza B virus antigen Injection [Flublok 2017-2018]
46276001	influenza A virus (H1N1) antigen / influenza A virus (H3N2) antigen / influenza B virus antigen Prefilled Syringe
35606412	influenza A virus (H1N1) antigen / influenza A virus (H3N2) antigen / influenza B virus antigen Prefilled Syringe [FLUAD 2015-2016]
36249409	influenza A virus (H1N1) antigen / influenza A virus (H3N2) antigen / influenza B virus antigen Prefilled Syringe [Fluad 2016-2017]
1593899	influenza A virus (H1N1) antigen / influenza A virus (H3N2) antigen / influenza B virus antigen Prefilled Syringe [Fluad 2017-2018]
35200045	influenza A virus (H1N1) antigen / influenza A virus (H3N2) antigen / influenza B virus antigen Prefilled Syringe [Fluad 2018-2019]
1361188	influenza A virus (H1N1) antigen / influenza A virus (H3N2) antigen / influenza B virus antigen Prefilled Syringe [Fluad 2019-2020]
1146532	influenza A virus (H1N1) antigen / influenza A virus (H3N2) antigen / influenza B virus antigen Prefilled Syringe [Fluad 2020-2021]
1593361	influenza A virus (H1N1) antigen / influenza A virus (H3N2) antigen / influenza B virus antigen Prefilled Syringe [Flublok Quadrivalent 2017-2018]
35200025	influenza A virus (H1N1) antigen / influenza A virus (H3N2) antigen / influenza B virus antigen Prefilled Syringe [Flublok Quadrivalent 2018-2019]
1361148	influenza A virus (H1N1) antigen / influenza A virus (H3N2) antigen / influenza B virus antigen Prefilled Syringe [Flublok Quadrivalent 2019-2020]
1146505	influenza A virus (H1N1) antigen / influenza A virus (H3N2) antigen / influenza B virus antigen Prefilled Syringe [Flublok Quadrivalent 2020-2021]
1758617	influenza A virus (H1N1) antigen / influenza A virus (H3N2) antigen / influenza B virus antigen Prefilled Syringe [Flublok Quadrivalent 2021-2022]
780109	influenza A virus (H1N1) antigen / influenza A virus (H3N2) antigen / influenza B virus antigen Prefilled Syringe [Flublok Quadrivalent 2022-2023]
746136	influenza A virus (H1N1) antigen / influenza A virus (H3N2) antigen / influenza B virus antigen Prefilled Syringe [Flublok Quadrivalent 2023-2024]

Concept ID	Concept name
43841366	Influenza A virus vaccine, A-California-7-2009 (H1N1)-like virus 0.03 MG/ML Injectable Suspension [Afluria]
43841367	Influenza A virus vaccine, A-California-7-2009 (H1N1)-like virus 0.03 MG/ML Injectable Suspension [Afluria] Box of 10
45774411	influenza A virus vaccine, A-Texas-50-2012 X-223A (H3N2), inactivated 0.03 MG/ML / influenza A-California-7-2009-(H1N1)v-like virus vaccine 0.03 MG/ML / influenza B virus vaccine, B-Massachusetts-2-2012-like virus 0.03 MG/ML Injectable Suspension
45774413	influenza A virus vaccine, A-Texas-50-2012 X-223A (H3N2), inactivated 0.03 MG/ML / influenza A-California-7-2009-(H1N1)v-like virus vaccine 0.03 MG/ML / influenza B virus vaccine, B-Massachusetts-2-2012-like virus 0.03 MG/ML Injectable Suspension [Flul...
43532864	influenza A virus vaccine, A-Victoria-361-2011 (H3N2)-like virus 0.03 MG/ML / influenza A-California-7-2009-(H1N1)v-like virus vaccine 0.03 MG/ML / influenza B virus vaccine, B-Massachusetts-2-2012-like virus 0.03 MG/ML Injectable Suspension
44818413	influenza A virus vaccine, A-Victoria-361-2011 (H3N2)-like virus 0.03 MG/ML / influenza A-California-7-2009-(H1N1)v-like virus vaccine 0.03 MG/ML / influenza B virus vaccine, B-Massachusetts-2-2012-like virus 0.03 MG/ML Injectable Suspension [Afluria 2...
43532964	influenza A virus vaccine, A-Victoria-361-2011 (H3N2)-like virus 0.03 MG/ML / influenza A-California-7-2009-(H1N1)v-like virus vaccine 0.03 MG/ML / influenza B virus vaccine, B-Massachusetts-2-2012-like virus 0.03 MG/ML Injectable Suspension [Afluria 2...
43532865	influenza A virus vaccine, A-Victoria-361-2011 (H3N2)-like virus 0.03 MG/ML / influenza A-California-7-2009-(H1N1)v-like virus vaccine 0.03 MG/ML / influenza B virus vaccine, B-Massachusetts-2-2012-like virus 0.03 MG/ML Injectable Suspension [Flulaval...
44818419	influenza A virus vaccine, A-Victoria-361-2011 (H3N2)-like virus 0.03 MG/ML / influenza A-California-7-2009-(H1N1)v-like virus vaccine 0.03 MG/ML / influenza B virus vaccine, B-Massachusetts-2-2012-like virus 0.03 MG/ML Injectable Suspension [Fluvirin...
43560216	influenza A virus vaccine, A-Victoria-361-2011 (H3N2)-like virus 0.03 MG/ML / influenza A-California-7-2009-(H1N1)v-like virus vaccine 0.03 MG/ML / influenza B virus vaccine, B-Massachusetts-2-2012-like virus 0.03 MG/ML Injectable Suspension [Fluvirin...
44818422	influenza A virus vaccine, A-Victoria-361-2011 (H3N2)-like virus 0.03 MG/ML / influenza A-California-7-2009-(H1N1)v-like virus vaccine 0.03 MG/ML / influenza B virus vaccine, B-Massachusetts-2-2012-like virus 0.03 MG/ML Injectable Suspension [Fluzone 2...
43532874	influenza A virus vaccine, A-Victoria-361-2011 (H3N2)-like virus 0.03 MG/ML / influenza A-California-7-2009-(H1N1)v-like virus vaccine 0.03 MG/ML / influenza B virus vaccine, B-Massachusetts-2-2012-like virus 0.03 MG/ML Injectable Suspension [Fluzone 2...
42873937	influenza A virus vaccine, A-Victoria-361-2011 (H3N2)-like virus 0.03 MG/ML / influenza A-California-7-2009-(H1N1)v-like virus vaccine 0.03 MG/ML / influenza B virus vaccine, B-Wisconsin-1-2010-like virus 0.03 MG/ML Injectable Suspension
42873939	influenza A virus vaccine, A-Victoria-361-2011 (H3N2)-like virus 0.03 MG/ML / influenza A-California-7-2009-(H1N1)v-like virus vaccine 0.03 MG/ML / influenza B virus vaccine, B-Wisconsin-1-2010-like virus 0.03 MG/ML Injectable Suspension [Flulaval 2012...
42873940	influenza A virus vaccine, A-Victoria-361-2011 (H3N2)-like virus 0.03 MG/ML / influenza A-California-7-2009-(H1N1)v-like virus vaccine 0.03 MG/ML / influenza B virus vaccine, B-Wisconsin-1-2010-like virus 0.03 MG/ML Injectable Suspension [Fluvirin 2012...
42873941	influenza A virus vaccine, A-Victoria-361-2011 (H3N2)-like virus 0.03 MG/ML / influenza A-California-7-2009-(H1N1)v-like virus vaccine 0.03 MG/ML / influenza B virus vaccine, B-Wisconsin-1-2010-like virus 0.03 MG/ML Injectable Suspension [Fluzone 2012-...
44507887	influenza A virus vaccine, A-Victoria-361-2011 (H3N2)-like virus 0.09 MG/ML / influenza A-California-7-2009-(H1N1)v-like virus vaccine 0.09 MG/ML / influenza B virus vaccine, B-Massachusetts-2-2012-like virus 0.09 MG/ML Injectable Solution

Concept ID	Concept name
44818416	influenza A virus vaccine, A-Victoria-361-2011 (H3N2)-like virus 0.09 MG/ML / influenza A-California-7-2009-(H1N1)v-like virus vaccine 0.09 MG/ML / influenza B virus vaccine, B-Massachusetts-2-2012-like virus 0.09 MG/ML Injectable Solution [Flublok 201...
44507888	influenza A virus vaccine, A-Victoria-361-2011 (H3N2)-like virus 0.09 MG/ML / influenza A-California-7-2009-(H1N1)v-like virus vaccine 0.09 MG/ML / influenza B virus vaccine, B-Massachusetts-2-2012-like virus 0.09 MG/ML Injectable Solution [Flublok 201...
43012767	influenza A virus vaccine, A-Victoria-361-2011 (H3N2)-like virus 0.09 MG/ML / influenza A-California-7-2009-(H1N1)v-like virus vaccine 0.09 MG/ML / influenza B virus vaccine, B-Wisconsin-1-2010-like virus 0.09 MG/ML Injectable Solution
43012768	influenza A virus vaccine, A-Victoria-361-2011 (H3N2)-like virus 0.09 MG/ML / influenza A-California-7-2009-(H1N1)v-like virus vaccine 0.09 MG/ML / influenza B virus vaccine, B-Wisconsin-1-2010-like virus 0.09 MG/ML Injectable Solution [Flublok 2012-20...
44507842	influenza A virus vaccine, A-Victoria-361-2011 (H3N2)-like virus / influenza A-California-7-2009-(H1N1)v-like virus vaccine / influenza B virus vaccine, B-Massachusetts-2-2012-like virus Injectable Solution
44507843	influenza A virus vaccine, A-Victoria-361-2011 (H3N2)-like virus / influenza A-California-7-2009-(H1N1)v-like virus vaccine / influenza B virus vaccine, B-Massachusetts-2-2012-like virus Injectable Solution [Flublok 2013-2014 Formula]
44818106	influenza A virus vaccine, A-Victoria-361-2011 (H3N2)-like virus / influenza A-California-7-2009-(H1N1)v-like virus vaccine / influenza B virus vaccine, B-Massachusetts-2-2012-like virus Injectable Solution [Flublok 2014-2015 Formula]
43532648	influenza A virus vaccine, A-Victoria-361-2011 (H3N2)-like virus / influenza A-California-7-2009-(H1N1)v-like virus vaccine / influenza B virus vaccine, B-Massachusetts-2-2012-like virus Injectable Suspension
43532775	influenza A virus vaccine, A-Victoria-361-2011 (H3N2)-like virus / influenza A-California-7-2009-(H1N1)v-like virus vaccine / influenza B virus vaccine, B-Massachusetts-2-2012-like virus Injectable Suspension [Afluria 2013-2014 Formula]
44818097	influenza A virus vaccine, A-Victoria-361-2011 (H3N2)-like virus / influenza A-California-7-2009-(H1N1)v-like virus vaccine / influenza B virus vaccine, B-Massachusetts-2-2012-like virus Injectable Suspension [Afluria 2014-2015 Formula]
43532651	influenza A virus vaccine, A-Victoria-361-2011 (H3N2)-like virus / influenza A-California-7-2009-(H1N1)v-like virus vaccine / influenza B virus vaccine, B-Massachusetts-2-2012-like virus Injectable Suspension [Flulaval 2013-2014 Formula]
44818100	influenza A virus vaccine, A-Victoria-361-2011 (H3N2)-like virus / influenza A-California-7-2009-(H1N1)v-like virus vaccine / influenza B virus vaccine, B-Massachusetts-2-2012-like virus Injectable Suspension [Flulaval 2014-2015 Formula]
43560213	influenza A virus vaccine, A-Victoria-361-2011 (H3N2)-like virus / influenza A-California-7-2009-(H1N1)v-like virus vaccine / influenza B virus vaccine, B-Massachusetts-2-2012-like virus Injectable Suspension [Fluvirin 2013-2014 Formula]
44818112	influenza A virus vaccine, A-Victoria-361-2011 (H3N2)-like virus / influenza A-California-7-2009-(H1N1)v-like virus vaccine / influenza B virus vaccine, B-Massachusetts-2-2012-like virus Injectable Suspension [Fluvirin 2014-2015 Formula]
43532662	influenza A virus vaccine, A-Victoria-361-2011 (H3N2)-like virus / influenza A-California-7-2009-(H1N1)v-like virus vaccine / influenza B virus vaccine, B-Massachusetts-2-2012-like virus Injectable Suspension [Fluzone 2013-2014 Formula]
44818116	influenza A virus vaccine, A-Victoria-361-2011 (H3N2)-like virus / influenza A-California-7-2009-(H1N1)v-like virus vaccine / influenza B virus vaccine, B-Massachusetts-2-2012-like virus Injectable Suspension [Fluzone 2014-2015 Formula]
43012763	influenza A virus vaccine, A-Victoria-361-2011 (H3N2)-like virus / influenza A-California-7-2009-(H1N1)v-like virus vaccine / influenza B virus vaccine, B-Wisconsin-1-2010-like virus Injectable Solution

Concept ID	Concept name
43012764	influenza A virus vaccine, A-Victoria-361-2011 (H3N2)-like virus / influenza A-California-7-2009-(H1N1)v-like virus vaccine / influenza B virus vaccine, B-Wisconsin-1-2010-like virus Injectable Solution [Flublok 2012-2013 Formula]
42873919	influenza A virus vaccine, A-Victoria-361-2011 (H3N2)-like virus / influenza A-California-7-2009-(H1N1)v-like virus vaccine / influenza B virus vaccine, B-Wisconsin-1-2010-like virus Injectable Suspension
42873921	influenza A virus vaccine, A-Victoria-361-2011 (H3N2)-like virus / influenza A-California-7-2009-(H1N1)v-like virus vaccine / influenza B virus vaccine, B-Wisconsin-1-2010-like virus Injectable Suspension [Flulaval 2012-2013 Formula]
42873922	influenza A virus vaccine, A-Victoria-361-2011 (H3N2)-like virus / influenza A-California-7-2009-(H1N1)v-like virus vaccine / influenza B virus vaccine, B-Wisconsin-1-2010-like virus Injectable Suspension [Fluvirin 2012-2013 Formula]
42873923	influenza A virus vaccine, A-Victoria-361-2011 (H3N2)-like virus / influenza A-California-7-2009-(H1N1)v-like virus vaccine / influenza B virus vaccine, B-Wisconsin-1-2010-like virus Injectable Suspension [Fluzone 2012-2013 Formula]
43751233	influenza B virus antigen 0.03 MG/ML Injectable Suspension [Afluria]
43859536	influenza B virus antigen 0.03 MG/ML Injectable Suspension [Afluria] Box of 10
43264065	Influenza B virus vaccine, B-Brisbane-60-2008-like virus 0.015 MG/ML / Influenza Virus Vaccine, Inactivated A-California-07-2009 X-181 (H1N1) strain 0.015 MG/ML / ... Injection [Fluvirin]
43296629	Influenza B virus vaccine, B-Brisbane-60-2008-like virus 0.015 MG/ML / Influenza Virus Vaccine, Inactivated A-California-07-2009 X-181 (H1N1) strain 0.015 MG/ML / ... Injection [Fluvirin] Box of 1
43275078	Influenza B virus vaccine, B-Brisbane-60-2008-like virus 0.015 MG/ML / Influenza Virus Vaccine, Inactivated A-California-07-2009 X-181 (H1N1) strain 0.015 MG/ML / ... Injection [Fluvirin] Box of 10
40213153	Influenza, seasonal, injectable
40213156	influenza virus vaccine, split virus (incl. purified surface antigen)-retired CODE
40213158	influenza virus vaccine, unspecified formulation
40213159	influenza virus vaccine, whole virus
40213152	Seasonal, quadrivalent, recombinant, injectable influenza vaccine, preservative free
40213157	Seasonal trivalent influenza vaccine, adjuvanted, preservative free
40213151	Seasonal, trivalent, recombinant, injectable influenza vaccine, preservative free

Tetanus, Diphtheria & Pertussis (Tdap) vaccination:

Concept ID	Concept name
42800267	0.5 ML Bordetella pertussis filamentous hemagglutinin vaccine, inactivated 0.016 MG/ML / Bordetella pertussis pertactin vaccine, inactivated 0.005 MG/ML / Bordetella pertussis toxoid vaccine, inactivated 0.016 MG/ML / diphtheria toxoid vaccine, inactiv...
42800266	0.5 ML Bordetella pertussis filamentous hemagglutinin vaccine, inactivated 0.016 MG/ML / Bordetella pertussis pertactin vaccine, inactivated 0.005 MG/ML / Bordetella pertussis toxoid vaccine, inactivated 0.016 MG/ML / diphtheria toxoid vaccine, inactiv...
42800473	0.5 ML Bordetella pertussis filamentous hemagglutinin vaccine, inactivated 0.016 MG/ML / Bordetella pertussis pertactin vaccine, inactivated 0.005 UNT/ML / Bordetella pertussis toxoid vaccine, inactivated 0.016 MG/ML / diphtheria toxoid vaccine, inacti...
42800474	0.5 ML Bordetella pertussis filamentous hemagglutinin vaccine, inactivated 0.016 MG/ML / Bordetella pertussis pertactin vaccine, inactivated 0.005 UNT/ML / Bordetella pertussis toxoid vaccine, inactivated 0.016 MG/ML / diphtheria toxoid vaccine, inacti...

Concept ID	Concept name
42800265	0.5 ML Bordetella pertussis filamentous hemagglutinin vaccine, inactivated 0.01 MG/ML / Bordetella pertussis fimbriae 2/3 vaccine, inactivated 0.01 MG/ML / Bordetella pertussis pertactin vaccine, inactivated 0.006 MG/ML / Bordetella pertussis toxoid va...
42800464	0.5 ML Bordetella pertussis filamentous hemagglutinin vaccine, inactivated 0.01 MG/ML / Bordetella pertussis fimbriae 2/3 vaccine, inactivated 0.01 MG/ML / Bordetella pertussis pertactin vaccine, inactivated 0.006 MG/ML / Bordetella pertussis toxoid va...
44791056	Booster diphtheria, tetanus, acellular pertussis and inactivated polio vaccination
44792127	Booster diphtheria, tetanus, acellular pertussis, haemophilus influenzae type b, inactivated polio vaccination
724894	Diphtheria, Tetanus, Poliomyelitis adsorbed (non-US)
40213281	diphtheria, tetanus toxoids and acellular pertussis vaccine
40213282	diphtheria, tetanus toxoids and acellular pertussis vaccine, 5 pertussis antigens
40213289	Diphtheria, tetanus toxoids and acellular pertussis vaccine, and poliovirus vaccine, inactivated
40213283	diphtheria, tetanus toxoids and acellular pertussis vaccine, unspecified formulation
40213291	diphtheria, tetanus toxoids and pertussis vaccine
19033193	pertussis vaccine
4219827	Second low dose diphtheria, tetanus and inactivated polio vaccination
40213230	tetanus toxoid, reduced diphtheria toxoid, and acellular pertussis vaccine, adsorbed
4167723	Third low dose diphtheria, tetanus and inactivated polio vaccination

Table S6. List of pregnancy complications (not included in the PET).

Antepartum Haemorrhage:

Concept ID	Concept name
4228479	Accidental antepartum haemorrhage
434089	Antepartum hemorrhage
438489	Antepartum hemorrhage associated with coagulation defect
45757208	Antepartum hemorrhage due to disseminated intravascular coagulation
44784542	Antepartum hemorrhage due to placenta previa type I
44784543	Antepartum hemorrhage due to placenta previa type II
44784544	Antepartum hemorrhage due to placenta previa type III
44784545	Antepartum hemorrhage due to placenta previa type IV
4129032	Antepartum hemorrhage with afibrinogenemia
443133	Antepartum hemorrhage with coagulation defect - delivered
438203	Antepartum hemorrhage with coagulation defect - not delivered
4129031	Antepartum hemorrhage with hyperfibrinolysis
4034091	Antepartum hemorrhage with hypofibrinogenemia
4143074	Antepartum hemorrhage, abruptio placentae and placenta previa
3404588	Concealed accidental hemorrhage
193535	Fetal-maternal hemorrhage - delivered

Concept ID	Concept name
196757	Fetal-maternal hemorrhage with antenatal problem
435323	Hemorrhage in early pregnancy, delivered
4025198	Hemorrhagic complication of pregnancy
4129827	Indeterminate antepartum hemorrhage
193264	Placenta previa with hemorrhage - delivered
200149	Placenta previa with hemorrhage - not delivered
4327745	Second trimester bleeding
4142581	Third trimester bleeding
37109810	Vaginal bleeding complicating early pregnancy
196751	Placenta previa with hemorrhage
201627	Abnormal vaginal bleeding
4083209	Vaginal vault bleeding
4095942	Vaginal introitus bleeding
4124469	Profuse vaginal bleeding
4126713	Moderate vaginal bleeding
4126908	Scanty vaginal bleeding
4129155	Vaginal bleeding
4199815	Heavy episode of vaginal bleeding
4219594	Non-menstrual vaginal bleeding
4127410	Vaginal vault hematoma
199068	Vaginal hematoma
4126712	Fresh bleeding from vagina
4126906	Passing clots vaginally
4126907	Passing altered blood from vagina
4129156	Passing small clots vaginally
4129157	Passing large clots vaginally
44806343	Minimal vaginal blood loss
4195179	Abnormal uterine bleeding unrelated to menstrual cycle
4169954	Genitourinary tract hemorrhage
4062417	Antepartum hemorrhage with uterine leiomyoma

Eclampsia:

Concept ID	Concept name
4116344	Eclampsia in puerperium
443700	Eclampsia
133816	Eclampsia - delivered
4062550	Pre-eclampsia or eclampsia with pre-existing hypertension with postnatal complication

Concept ID	Concept name
134414	Pre-eclampsia or eclampsia with pre-existing hypertension - delivered with postnatal complication
42537548	Nephropathy following eclampsia
4058536	Eclampsia with postnatal complication
4289142	Eclampsia added to pre-existing hypertension
138811	Eclampsia - delivered with postnatal complication
137613	Eclampsia in pregnancy
4034096	Eclampsia in labour
45757119	Eclampsia with pre-existing hypertension in childbirth
135601	Pre-eclampsia or eclampsia with pre-existing hypertension - delivered
136743	Pre-eclampsia or eclampsia with pre-existing hypertension - not delivered
4034096	Eclampsia in labour

Ectopic pregnancy:

Concept ID	Concept name
37203955	Intramural ectopic pregnancy of myometrium
37209653	Ectopic pregnancy of left ovary
37209654	Ectopic pregnancy of right ovary
42536560	Pelvic infection due to and following ectopic pregnancy
42536561	Genital tract infection due to and following ectopic pregnancy
42537762	Delayed haemorrhage due to and following ectopic pregnancy
42537763	Excessive haemorrhage due to and following ectopic pregnancy
42537767	Damage to pelvic organs and tissues following ectopic pregnancy
42537770	Disorder of vein following ectopic pregnancy
42539701	Embolism due to and following ectopic pregnancy
43530979	Miscarriage of tubal ectopic pregnancy
45757120	Renal failure after ectopic pregnancy
46269806	Salpingitis due to ectopic pregnancy
46269807	Sepsis due to ectopic pregnancy
46269808	Shock due to ectopic pregnancy
46269809	Urinary tract infection due to ectopic pregnancy
195014	Renal failure following molar AND/OR ectopic pregnancy
196177	Genital tract AND/OR pelvic infection following molar AND/OR ectopic pregnancy
196475	Damage to pelvic organs AND/OR tissues following molar AND/OR ectopic pregnancy
432371	Metabolic disorder following molar AND/OR ectopic pregnancy
437611	Ectopic pregnancy
440464	Complication following molar AND/OR ectopic pregnancy
440777	Embolism following molar AND/OR ectopic pregnancy

Concept ID	Concept name
442271	Shock following molar AND/OR ectopic pregnancy
4023382	Postoperative shock following molar AND/OR ectopic pregnancy
4063939	Intravascular haemolysis following molar AND/OR ectopic pregnancy
4066014	Endometritis following molar AND/OR ectopic pregnancy
4069106	Ruptured ectopic pregnancy
4080587	Septic shock following molar AND/OR ectopic pregnancy
4093262	Septic embolism following molar AND/OR ectopic pregnancy
4095115	Renal tubular necrosis following molar AND/OR ectopic pregnancy
4100193	Damage to bladder following molar AND/OR ectopic pregnancy
4103056	Blood clot embolism following molar AND/OR ectopic pregnancy
4107071	Oliguria following molar AND/OR ectopic pregnancy
4118058	Foetal or neonatal effect of ectopic pregnancy
4118263	Foetal or neonatal effect of intraperitoneal ectopic pregnancy
4118419	Foetal or neonatal effect of abdominal ectopic pregnancy
4118420	Foetal or neonatal effect of tubal ectopic pregnancy
4141453	Air embolism following molar AND/OR ectopic pregnancy
4147060	Cerebral anoxia following molar AND/OR ectopic pregnancy
4151214	Uraemia following molar AND/OR ectopic pregnancy
4152444	Antenatal ultrasound confirms ectopic pregnancy
4172629	Amniotic fluid embolism following molar AND/OR ectopic pregnancy
4173946	Damage to cervix following molar AND/OR ectopic pregnancy
4174543	Damage to periurethral tissue following molar AND/OR ectopic pregnancy
4185718	Damage to bowel following molar AND/OR ectopic pregnancy
4187090	Sepsis following molar AND/OR ectopic pregnancy
4194053	Damage to uterus following molar AND/OR ectopic pregnancy
4215648	Acute renal failure following molar AND/OR ectopic pregnancy
4216382	Damage to vagina following molar AND/OR ectopic pregnancy
4232953	Damage to broad ligament following molar AND/OR ectopic pregnancy
4239009	Salpingo-oophoritis following molar AND/OR ectopic pregnancy
4242401	Urinary tract infection following molar AND/OR ectopic pregnancy
4265400	Soap embolism following molar AND/OR ectopic pregnancy
4266839	Defibrination syndrome following molar AND/OR ectopic pregnancy
4276942	Electrolyte imbalance following molar AND/OR ectopic pregnancy
4280067	Pulmonary embolism following molar AND/OR ectopic pregnancy
4296038	Tetanus complicating ectopic AND/OR molar pregnancy
4312095	Afibrinogenemia following molar AND/OR ectopic pregnancy
4314877	Pelvic peritonitis following molar AND/OR ectopic pregnancy

Concept ID	Concept name
4315388	Fat embolism following molar AND/OR ectopic pregnancy
4316051	Salpingitis following molar AND/OR ectopic pregnancy
4336392	Parametritis following molar AND/OR ectopic pregnancy
4135209	Combined pregnancy
193525	Abdominal pregnancy
199076	Tubal pregnancy
200153	Ovarian pregnancy
4066151	Intraperitoneal pregnancy
4112952	Membranous pregnancy
4113218	Mesenteric pregnancy
4113680	Ruptured tubal pregnancy
4113682	Angular pregnancy
4113846	Delivery of viable foetus in abdominal pregnancy
4129027	Viable foetus in abdominal pregnancy
4129712	Unruptured tubal pregnancy
4149029	Intraligamentous pregnancy
4174577	Secondary abdominal pregnancy
4194693	Cervical pregnancy
4227267	Cornual pregnancy
4243803	Mesometric pregnancy
37209655	Pregnancy of left fallopian tube
37209656	Ruptured tubal pregnancy of right fallopian tube
37209657	Ruptured tubal pregnancy of left fallopian tube
37209658	Pregnancy of right fallopian tube
40483581	Combined tubal and intrauterine pregnancy
44783183	Combined intrauterine and ovarian pregnancy
3043153	[#] Ectopic pregnancy
37167330	Metabolic disorder following ectopic pregnancy
37167172	Shock following ectopic pregnancy
4215648	Acute renal failure following molar AND/OR ectopic pregnancy
4216382	Damage to vagina following molar AND/OR ectopic pregnancy
4232953	Damage to broad ligament following molar AND/OR ectopic pregnancy
4239009	Salpingo-oophoritis following molar AND/OR ectopic pregnancy
4242401	Urinary tract infection following molar AND/OR ectopic pregnancy
4265400	Soap embolism following molar AND/OR ectopic pregnancy
4266839	Defibrination syndrome following molar AND/OR ectopic pregnancy
4276942	Electrolyte imbalance following molar AND/OR ectopic pregnancy

Concept ID	Concept name
4280067	Pulmonary embolism following molar AND/OR ectopic pregnancy
4296038	Tetanus complicating ectopic AND/OR molar pregnancy
4312095	Afibrinogenemia following molar AND/OR ectopic pregnancy
4314877	Pelvic peritonitis following molar AND/OR ectopic pregnancy
4315388	Fat embolism following molar AND/OR ectopic pregnancy
4316051	Salpingitis following molar AND/OR ectopic pregnancy
4336392	Parametritis following molar AND/OR ectopic pregnancy
4135209	Combined pregnancy
193525	Abdominal pregnancy
199076	Tubal pregnancy
200153	Ovarian pregnancy
4066151	Intraperitoneal pregnancy
4112952	Membranous pregnancy
4113218	Mesenteric pregnancy
4113680	Ruptured tubal pregnancy
4113682	Angular pregnancy
4113846	Delivery of viable foetus in abdominal pregnancy
4129027	Viable fetes in abdominal pregnancy
4129712	Unruptured tubal pregnancy
4149029	Intraligamentous pregnancy
4174577	Secondary abdominal pregnancy
4194693	Cervical pregnancy
4227267	Cornual pregnancy
4243803	Mesometric pregnancy
37209655	Pregnancy of left fallopian tube
37209656	Ruptured tubal pregnancy of right fallopian tube
37209657	Ruptured tubal pregnancy of left fallopian tube
37209658	Pregnancy of right fallopian tube
40483581	Combined tubal and intrauterine pregnancy
44783183	Combined intrauterine and ovarian pregnancy
3043153	[#] Ectopic pregnancy
37167330	Metabolic disorder following ectopic pregnancy
37167172	Shock following ectopic pregnancy

Gestational diabetes:

Concept ID	Concept name
37018765	Gestational diabetes mellitus complicating pregnancy
43020791	Gestational diabetes mellitus uncontrolled
45757124	Gestational diabetes mellitus in childbirth
45757789	Postpartum gestational diabetes mellitus
4024659	Gestational diabetes mellitus
4263902	Gestational diabetes mellitus, class A>2<
4326434	Gestational diabetes mellitus, class A>1<
4058243	Diabetes mellitus during pregnancy, childbirth and the puerperium

HELLP Syndrome:

Concept ID	Concept name
4316372	HELLP syndrome
37169699	Postpartum hemolysis-elevated liver enzymes-low platelet count syndrome

Pre-eclampsia:

Concept ID	Concept name
4057976	Severe pre-eclampsia with postnatal complication
438490	Severe pre-eclampsia - delivered
132685	Severe pre-eclampsia - not delivered
439393	Pre-eclampsia
4283352	Pre-eclampsia added to pre-existing hypertension
42537547	Nephropathy following pre-eclampsia
35622939	Postpartum pre-eclampsia
604804	Fetal disorder due to maternal pre-eclampsia
1340440	Exacerbation of pre-eclampsia
314090	Mild pre-eclampsia
433536	Severe pre-eclampsia
439077	Severe pre-eclampsia - delivered with postnatal complication
4034094	Moderate proteinuric hypertension of pregnancy
4057976	Severe pre-eclampsia with postnatal complication
438490	Severe pre-eclampsia - delivered
141084	Pre-eclampsia or eclampsia with pre-existing hypertension

Postpartum haemorrhage:

Concept ID	Concept name
37394499	Minor postpartum haemorrhage
37394500	Major postpartum haemorrhage

Concept ID	Concept name
37108740	Postpartum hemorrhage due to total placenta previa
37110289	Postpartum hemorrhage co-occurrent and due to uterine rupture following obstructed labor
194109	Delayed AND/OR secondary postpartum hemorrhage
196488	Atonic postpartum hemorrhage
443929	Postpartum hemorrhage
4006327	Postpartum afibrinogenemia with hemorrhage
4065618	Secondary postpartum hemorrhage - delivered with postnatal problem
4169890	Postpartum coagulation defect with hemorrhage
4266066	Postpartum fibrinolysis with hemorrhage
4194998	Retained products of conception, following delivery with hemorrhage
46274097	Compression suturing of uterus for atonic postpartum hemorrhage
4127264	Surgical control of postpartum hemorrhage
4169003	Pack to control postpartum hemorrhage
606891	Postpartum hemorrhage with retained placenta
4298174	Bleeding reduction: postpartum uterus
201627	Abnormal vaginal bleeding
4083209	Vaginal vault bleeding
4095942	Vaginal introitus bleeding
4124469	Profuse vaginal bleeding
4126713	Moderate vaginal bleeding
4126908	Scanty vaginal bleeding
4129155	Vaginal bleeding
4199815	Heavy episode of vaginal bleeding
4219594	Non-menstrual vaginal bleeding
4127410	Vaginal vault hematoma
199068	Vaginal hematoma
4126712	Fresh bleeding from vagina
4126906	Passing clots vaginally
4126907	Passing altered blood from vagina
4129156	Passing small clots vaginally
4129157	Passing large clots vaginally
44806343	Minimal vaginal blood loss
4195179	Abnormal uterine bleeding unrelated to menstrual cycle
4169954	Genitourinary tract haemorrhage

TORCH (toxoplasmosis, others, rubella, cytomegalovirus, herpes simplex) infection:

Concept ID	Concept name
440032	Cytomegalovirus infection
444429	Herpes simplex
4273323	Rubella
436033	Syphilis
440653	Toxoplasmosis
434272	Varicella
37164853	Lymphadenitis caused by Toxoplasma gondii
4093496	Toxoplasmosis of multiple sites
4093494	Toxoplasma myocarditis
42536597	Disorder of eye caused by Toxoplasma gondii
4089507	Toxoplasma pneumonia
3655831	Myelitis caused by Toxoplasma gondii
4251491	Congenital toxoplasmosis
438975	Multisystemic disseminated toxoplasmosis
4080617	Cutaneous toxoplasmosis
36674253	Infection caused by Toxoplasma gondii co-occurrent with acquired immunodeficiency syndrome
4058590	Acute lymphadenopathic toxoplasmosis
45757273	Toxoplasma tubulointerstitial nephropathy
37163864	Hepatitis caused by Toxoplasma gondii
4103105	Toxoplasma encephalitis
4177963	Syphilis of tendon
435739	Late syphilis
4051481	Nasal syphilis
4291175	Syphilis of synovium
619069	Iris papulosa caused by Treponema pallidum
619070	Iris roseola caused by Treponema pallidum
45757152	Syphilis in mother complicating childbirth
4059744	Syphilitic prostatitis
434279	Congenital syphilis
4160083	Syphilitic skin disorder
74862	Syphilis of bone
380030	Syphilitic acoustic neuritis
605342	Secondary syphilis of eye
40485064	Latent syphilis
4132494	Ocular syphilis
4122760	Oral syphilis

Concept ID	Concept name
4089985	Early symptomatic syphilis
4260509	Syphilitic punched out ulcer
4001478	Syphilitic bursitis
76032	Syphilis of muscle
444294	Syphilis of lung
375738	Neurosyphilis
4108798	Acute syphilitic pericarditis
4083187	Quaternary syphilis
4029910	Syphilis of breast
4080592	Acquired syphilis
442345	Syphilis of kidney
3655840	Perinatal mucocutaneous infection caused by Human herpesvirus 3
4048187	Acute herpes simplex pharyngitis
36715817	Sexually transmissible infection caused by Herpes simplex virus
36715818	Pharyngotonsillitis caused by Human herpes simplex virus
4345462	Chronic herpes simplex
4234499	Herpes simplex type 1 infection
37160805	Herpes simplex gastrointestinal tract infection
45773026	Nongenital herpes simplex
37017454	Infection caused by herpes simplex virus co-occurrent with human immunodeficiency virus infection
4122209	Oral herpes simplex infection
4291634	Recurrent herpes simplex
36716867	Infection of vagina caused by Human herpes simplex virus
4251448	Herpes simplex type 2 infection
37164620	Herpes simplex blepharitis
4087996	Kaposi's varicelliform eruption caused by herpes simplex virus
373404	Ophthalmic herpes simplex
4239065	Congenital herpes simplex
42536745	Perinatal infection caused by Human herpes simplex virus
380324	Herpes simplex otitis externa
4165265	Primary herpes simplex
4051338	Herpes simplex pneumonia
4029319	Disseminated herpes simplex
1340357	Exacerbation of herpes simplex
4093145	Anogenital herpesviral infection
36715970	Infection of skin and/or mucous membrane caused by Herpes simplex virus
36715971	Infection of central nervous system caused by Herpes simplex virus

Concept ID	Concept name
37311936	Infection of urinary tract caused by Herpes simplex
45757252	Hepatitis caused by human herpes simplex virus
74564	Rubella arthritis
433869	Congenital rubella syndrome
444251	Rubella in mother complicating pregnancy, childbirth AND/OR puerperium
4335453	Endocochlear rubella
4008080	Rubella infection of central nervous system
45770900	Rubella pneumonia
4345464	Hemorrhagic rubella
4128374	Cytomegalovirus-induced glomerulonephritis
4171253	Fetal cytomegalovirus syndrome
4179507	Cytomegaloviral gastritis
4237775	Cytomegalovirus infection of the central nervous system
36717503	Infection of intestine caused by Human cytomegalovirus 5
4177199	Disseminated cytomegalovirus infection
37017446	Infection caused by Cytomegalovirus co-occurent with human immunodeficiency virus infection
37160677	Cytomegaloviral dacryoadenitis
37160674	Cytomegaloviral polyneuropathy
252949	Cytomegaloviral pneumonia
761747	Large liver due to infectious mononucleosis caused by Cytomegalovirus
45757222	Cytomegalovirus viremia
4339484	Endocochlear cytomegalovirus infection
4340962	Cytomegaloviral pancreatitis
764846	Gastric ulcer caused by cytomegalovirus
4330151	Cytomegaloviral retinitis
42598956	Inclusion body rhinitis of swine
4298473	Cytomegalovirus infection of skin
436532	Congenital cytomegalovirus infection
4092685	Cytomegalovirus hepatitis
4036612	Cytomegaloviral mononucleosis
37018701	Esophagitis caused by Cytomegalovirus
4144320	Cytomegaloviral enteritis
74855	Genital herpes simplex
37169603	Zika virus infection during pregnancy
4214984	Zika virus disease
443291	Maternal syphilis during pregnancy, childbirth and the puerperium

Table S7. Infant covariates.

Concept ID ¹	Concept name
4264825	Birth weight
3004221	5 minute Apgar Score
4079975	Congenital malformations
37166037	Maternal exclusive breastfeeding
37166036	Partial maternal breastfeeding

¹ Codes were used in data sources where the infant table was not available (i.e. SIDIAP, NLHR). Small for gestational age was not available in SIDIAP and was analysed using z-score at birth as a proxy in NLHR (recorded with the custom Concept ID: 2000000048). Breastfeeding was available in SIDIAP only. All descendants of 4079975 were considered to assess congenital malformations.

ANNEX IV. Operational and reporting considerations

DATA MANAGEMENT

Data management

All data sources have previously mapped their data to the OMOP common data model. This enables the use of standardised analytics and using DARWIN EU® tools across the network, since the structure of the data and the terminology system is harmonised. The OMOP CDM was developed and maintained by the Observational Health Data Sciences and Informatics (OHDSI) initiative and is described in detail on the wiki page of the CDM: <https://ohdsi.github.io/CommonDataModel> and in The Book of OHDSI: <http://book.ohdsi.org>.

The analytic code for this study was written in R and used standardised analytics wherever possible. Each data partner executed the study code against their data source containing patient-level data and then returned the results (csv files), which only contained aggregated data. The results from each of the contributing data sites were then combined in tables and figures for the study report.

Data storage and protection

For this study, participants from various EU member states processed personal data from individuals that is collected in national/regional electronic health record data sources. Due to the sensitive nature of this personal medical data, it is important to be fully aware of ethical and regulatory aspects and to strive to take all reasonable measures to ensure compliance with ethical and regulatory issues on privacy.

All data sources used in this study are already used for pharmaco-epidemiological research and have a well-developed mechanism to ensure that European and local regulations dealing with ethical use of the data and adequate privacy control are adhered to. In agreement with these regulations, rather than combining person level data and performing only a central analysis, local analyses was run, which generate non-identifiable aggregate summary results.

The output files are stored in the DARWIN EU® Remote Research Environment (RRE). These output files do not contain any data that allow identification of subjects included in the study. The RRE implements further security measures to ensure a high level of stored data protection to comply with the local implementation of the General Data Protection Regulation (GDPR) (EU) 679/2016 in the various member states.

QUALITY CONTROL

Data source quality control

When defining conditions of interest, a systematic search of possible codes for inclusion was identified using the *CodelistGenerator* R package (<https://github.com/darwin-eu/CodelistGenerator>). This package allows the user to define a search strategy and will use this to query the vocabulary tables of the OMOP common data model so as to find potentially relevant codes. The *PhenotypeR* R package (<https://github.com/OHDSI/PhenotypeR>) was run to assess the use of different codes across the data sources contributing to the study and identify any codes potentially omitted in error. All medications were assessed at the ingredient level unless stated otherwise. In certain cases, medications were evaluated based on ATC level 3 groupings. Concept sets corresponding to these ATC groupings were generated using the *CodelistGenerator* R package. Given the early nature of the process of standardisation to the pregnancy extension tables, additional bespoke diagnostics were created for the study. These are contained in the diagnostics R package (<https://github.com/darwin-eu-studies/P4-C5-004-Diagnostics>).

The study code was based on DARWIN EU® R packages *CohortCharacteristics* and *PatientProfiles*, which were used to characterise pregnancy episodes and individuals of interest. Counts over time were obtained using the *IncidencePrevalence* R package, also developed by DARWIN EU®. These packages included

numerous automated unit tests to ensure the validity of the codes, alongside software peer review and user testing. The R package will be made publicly available via GitHub.

All analyses were performed using OMOP data.

PLANS FOR DISSEMINATING AND COMMUNICATING STUDY RESULTS

A PDF report including an executive summary, and the specified tables and/or figures is submitted to EMA by the DARWIN EU® CC upon completion of the study.

An interactive dashboard incorporating all the results (tables and figures) is provided alongside the PDF report. The full set of underlying aggregated data used in the dashboard will also be made available, if requested.

ANNEX V. Supplementary tables and figures

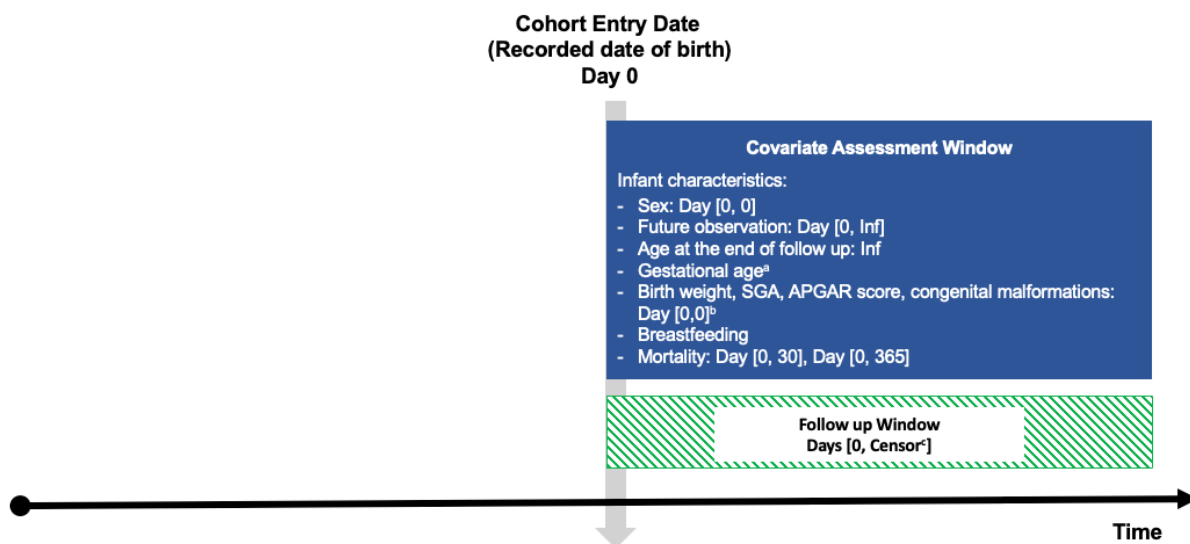


Figure S1. Graphical depiction of the study design for Objective 4d.

SGA=Small for Gestational Age

- a. As recorded in the Perinatal Expansion Table.
- b. As recorded in the infant table (where available). For data sources where congenital malformations were not recorded in the Infant table, this information was derived from other records using the time window [0,365] for assessment.
- c. Defined as loss to follow-up, death, or end of observation period (the latest available data), whichever occurred first.

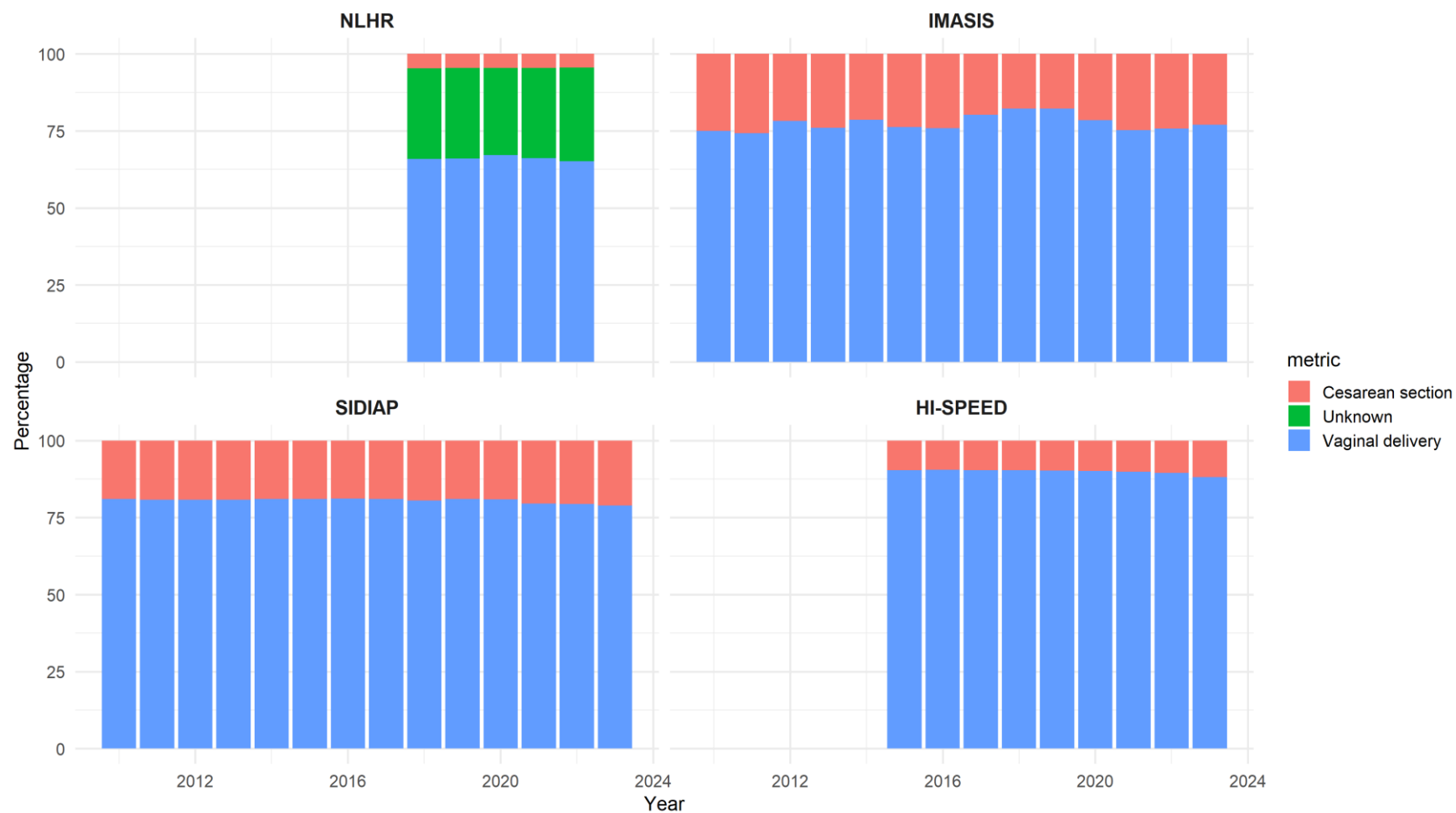


Figure S2. Percentage of pregnancy episodes starting in each calendar year by mode of delivery.

HI-SPEED=Health Impact - Swedish Population Evidence Enabling Data-linkage; IMASIS=Institut Municipal Assistència Sanitària Information System; NLHR=Norwegian Linked Health Registry data; SIDIAP=The Information System for Research in Primary Care .

ANNEX VI. Glossary

Additional definitions are available in the EMA Glossary of terms <https://www.ema.europa.eu/en/about-us/glossaries>.

Aggregated Data

Data collected and combined from multiple sources to generate summary information, typically anonymised.

Benefit-Risk Assessment

Evaluation of the positive therapeutic effects of a medicine compared to its risks (e.g. side effects).

Common Data Model (CDM)

A standardised data structure that enables data from multiple sources to be harmonised, making analysis consistent and reproducible. DARWIN EU® utilises the OMOP CDM maintained by the OHDSI community.

Complex Studies (C3)

Studies requiring the development or customisation of specific study designs, protocols, and Statistical Analysis Plans (SAPs), with extensive collection or extraction of data. Examples include etiological studies measuring the strength and determinants of an association between an exposure and the occurrence of a health outcome in a defined population considering sources of bias, potential confounding factors, and effect modifiers.

Coordination Centre (CC)

The central hub responsible for managing and overseeing the activities within DARWIN EU®. It is based at Erasmus University Medical Center in Rotterdam, the Netherlands.

Data Access

The process of obtaining permission to use specific datasets for regulatory or scientific studies.

Data Quality Framework

A set of standards and procedures to ensure accuracy, completeness, timeliness, and consistency of data used in DARWIN EU®.

Data Source

A database or repository of structured health-related data, such as electronic health records (EHRs), insurance claims, or registries.

DARWIN EU®

The European Medicines Agency's (EMA) federated network of real-world data sources designed to generate evidence to support regulatory decision-making.

EMA (European Medicines Agency)

The regulatory body responsible for the evaluation and supervision of medicinal products in the EU, overseeing DARWIN EU®.

Evidence Generation

The process of analysing real-world data to produce scientific information that can inform healthcare or regulatory decisions.

Federated Network

A data infrastructure where data remain at their original location but can be analysed in a harmonised way across multiple partners using a common model and tools.

GDPR (General Data Protection Regulation)

The EU regulation governing the protection of personal data and privacy, crucial to how DARWIN EU® handles health data.

Health Technology Assessment (HTA)

A systematic evaluation of properties and impacts of health technology, often using DARWIN EU® data to support assessments.

Metadata

Descriptive information about a data source (e.g. its content, quality, and structure), essential for identifying relevant databases in DARWIN EU® studies.

Off-the-Shelf Studies (OTS)

Studies for which a standard protocol per study/analysis type and standardised analytics may be developed and applied or adapted, typically relating to a descriptive research question. This includes studies on disease epidemiology, for example, the estimation of the prevalence or incidence of health outcomes in defined time periods and population groups, or drug utilisation studies at the population or patient level.

OHDSI (Observational Health Data Sciences and Informatics)

An open-science collaborative community that develops tools and standards (including the OMOP CDM) to enable large-scale analytics of observational health data. OHDSI provides the technical and scientific foundation for DARWIN EU®'s analytical ecosystem.

Patient-Level Data

Data related to individuals, de-identified, used for longitudinal or detailed analyses.

OMOP (Observational Medical Outcomes Partnership)

A common data model (CDM) that standardises the structure and content of observational healthcare data, enabling systematic analysis across disparate datasets. DARWIN EU® uses the OMOP CDM to ensure interoperability and consistency in real-world evidence generation.

Real-World Data (RWD)

Data relating to individual health status or healthcare delivery that is collected from routine clinical practice rather than from randomised controlled trials.

Real-World Evidence (RWE)

Clinical evidence derived from the analysis of RWD, used to inform decisions by regulators, payers, or clinicians.

Regulatory Decision-Making

The process by which authorities like EMA assess data to authorise, monitor, or modify the use of medicines in the EU.

Routine Repeated Studies (RR)

Studies that are either Off-the-Shelf or Complex studies repeated on a regular basis, following the same protocol and study code, but with updated data and/or different data partners.

Study Protocol

A detailed plan describing how a specific real-world study will be conducted, including objectives, design, data sources, and analyses.

Very Complex Studies (C4)

Studies which cannot rely only on electronic health care databases, or which would require complex methodological work, for example, due to the occurrence of events that cannot be defined by existing diagnosis codes, including events that do not yet have a diagnosis code, where it may be necessary to combine a diagnosis code with other data such as results of laboratory investigations. These studies might require the collection of data prospectively, or the inclusion of new (not previously onboarded) data sources.



Study Report

P4-C4-001

DARWIN EU® - PeriNet

OBJECTIVE 5

Generation of aggregated counts related to pregnancy and mother-child linkage to support study feasibility assessment

26/05/2026

Version 2.0

Authors: Maxim Moinat, Berta Raventós, Talita Duarte-Salles

Public

Study title	DARWIN EU® - PeriNet Objective 5: Generation of aggregated counts related to pregnancy and mother-child linkage to support study feasibility assessment
Study report version	V2.0
Date	26/05/2026
EUPAS number	EUPAS1000000812
Active substance	NA
Medicinal product	NA
Objectives	To generate a data source with aggregate counts of real-world data elements related to pregnancy and mother-child linkage to support feasibility assessment of future non-interventional studies on pregnancy, including maternal and offspring perinatal health. The specific objectives were: 5a. To define the data elements needed to inform feasibility assessment of studies on pregnancy, including maternal and offspring perinatal health, or requiring linked mother-child data in the DARWIN EU®. 5b. To incorporate the code for the extraction of aggregated counts defined in objective 5a into the DARWIN EU® onboarding packages. 5c. To extract and compile aggregated counts of the data elements defined as needed for feasibility assessment in objective 5a from each participating data source, overall and stratified by age of the mother, sex of the child, and calendar year. 5d. To incorporate the aggregated counts extracted in objective 5c into the DARWIN EU® Portal.
Countries of study	NA
Authors	Maxim Moinat m.moinat@darwin-eu.org Berta Raventós b.raventos@darwin-eu.org Talita Duarte-Salles t.duarte@darwin-eu.org

LIST OF ABBREVIATIONS

Acronyms/term	Description
DP	Data Partner
CC	Coordination Centre
CDM	Common Data Model
CSV	Comma-separated Values
DARWIN EU®	Data Analysis and Real-World Interrogation Network
EHR	Electronic Health Records
EMA	European Medicines Agency
OHDSI	Observational Health Data Sciences and Informatics
OMOP	Observational Medical Outcomes Partnership
RC	Record Count
PC	Person Count
Dist	Value Distribution
IRB	Institutional Review Board
PET	Perinatal Expansion Tables
PIs	Principal Investigators
IMASIS	Institut Municipal Assistència Sanitària Information System

1. TITLE

DARWIN EU® - Objective 5: Generation of aggregated counts related to pregnancy and mother-child linkage to support study feasibility assessment.

2. DESCRIPTION OF THE STUDY TEAM

Study team role	Names	Organisation
Principal Investigator	Maxim Moinat	Erasmus MC
Co-Principal Investigator	Talita Duarte-Salles	Erasmus MC
Data Scientists	Adam Black	Erasmus MC
Epidemiologist	Berta Raventós Katia Verhamme	Erasmus MC
Study Manager	Natasha Yefimenko	Erasmus MC

3. ABSTRACT

Title

DARWIN EU® - Objective 5: Generation of aggregated counts related to pregnancy and mother-child linkage to support study feasibility assessment.

Rationale and background

The DARWIN EU® PeriNet study aims to serve as a foundational step in establishing the DARWIN EU® Data Network's capacity to support high-quality, real-world evidence generation for pregnancy, including maternal and offspring perinatal health, for regulatory, clinical, and epidemiological studies.

For assessing the feasibility of studies, the DARWIN EU® Coordination Centre (CC) collects aggregated counts of variables during onboarding and at each data refresh for all data partners. While this is very valuable to judge feasibility for most DARWIN EU® studies, these aggregated counts are generated from the entire data source and do not provide insights into specific population subgroups, such as the pregnant population. To assess the feasibility of future pregnancy studies, we propose creating new aggregates and visualisations for pregnancy-related data and mother-child linked data in the DARWIN EU® Data Network.

Objectives

To create a data source with aggregated counts of real-world data elements related to pregnancy and mother-child linkage to support feasibility assessment of future non-interventional studies on pregnancy, including maternal and offspring perinatal health.

The specific objectives were:

- 5a. To define the data elements needed to inform feasibility assessment of studies, including maternal and offspring perinatal health, or requiring linked mother-child data in the DARWIN EU®.
- 5b. To incorporate the code for the extraction of aggregated counts defined in objective 5a into the DARWIN EU® onboarding packages.
- 5c. To extract and compile aggregated counts of the data elements defined as needed for feasibility assessment in objective 5a from each participating data source, overall and stratified by age of the mother, sex of the child, and calendar year.
- 5d. To incorporate the aggregated counts extracted in objective 5c into the DARWIN EU® Portal.

Methods

We created a comprehensive list of data elements necessary for assessing the feasibility of future studies on perinatal health. This list was created by several interactions between the European Medicines Agency (EMA) and experts. Using this list, the appropriate software was developed to calculate the aggregate counts from data sources. This software was incorporated into the DARWIN EU® onboarding packages, such that all subsequent Observational Medical Outcomes Partnership Common Data Model (OMOP CDM) releases of Data Partners (DPs) will execute this and share the results with the Coordination Centre (CC). A new section of the DARWIN EU® Portal dashboard was created that makes the aggregates usable for future feasibility assessments.

Data sources

The 12 data sources that participated in the DARWIN EU® PeriNet objective 2 study were included in Objective 5c. Results for other data sources will be obtained once they have a new data release for which quality control results are submitted to the CC.

Results

We developed a new dashboard in the DARWIN EU® Portal that will support the feasibility assessment of future perinatal studies. This dashboard contains a table for each of the data elements identified for use in feasibility assessments and is accessible to the DARWIN EU® team.

Discussion

The work done shows a comprehensive approach to support feasibility assessment with the systematic extraction of aggregated data across multiple sources. By leveraging the DARWIN EU® Portal and onboarding packages, the methodology ensures consistency and reproducibility, allowing for robust stratification by key variables such as maternal age, child sex, and pregnancy outcome. The inclusion of a dashboard accessible to the DARWIN EU® team further enhances transparency and supports informed decision-making for future perinatal studies.

4. AMENDMENTS AND UPDATES

None.

5. MILESTONES

Study deliverable	Timelines (planned)	Timelines (actual)
Final Study Protocol	March 2026	March 2026
Software developed to calculate aggregates	March 2026	April 2026
Software developed to display aggregates	March 2026	April 2026
Draft Study Report	March 2026	May 2026
Final Study Report	April 2026	May 2026

6. RATIONALE AND BACKGROUND

The “DARWIN EU® - Enabling pregnancy and mother-child research in the DARWIN EU® Data Network – DARWIN EU®- PeriNet” study aims to serve as a foundational step in establishing the DARWIN EU® Data Network’s capacity to support high-quality, real-world evidence generation for pregnancy, including maternal and offspring perinatal health, for regulatory, clinical, and epidemiological studies.

Feasibility assessments are a critical component of study development within the DARWIN EU® Data Network and are performed in response to requests from the European Medicines Agency (EMA). To support feasibility assessments, the DARWIN EU® Coordination Centre (CC) collects aggregate counts of variables recorded in the data sources within the network. This process is conducted each time a new data partner (DP) joins the network or a new data release from an existing data partner is being onboarded. These aggregated counts are obtained by DPs through the execution of the DARWIN EU® *DashboardExport* package.[1] Results are shared upon each new data release with DARWIN EU® Network Operations for validation. Subsequently, the aggregate counts are made available to the entire CC through the DARWIN EU® Portal, an interactive webtool that hosts all metadata of DPs. The Portal also displays information on data source characteristics and data collection processes reported by DPs. While this information is very valuable to inform feasibility for most DARWIN EU® studies, these aggregated counts are generated from the entire data source and do not provide insights into specific population subgroups, such as the pregnant population.

To further support the feasibility assessment of future perinatal research studies within the DARWIN EU® Network, this study aimed to create new data source aggregates and visualisations for data elements related to pregnancy and mother-child linkages accessible to the CC through the DARWIN EU® Portal. This work corresponds to the fifth objective of the DARWIN EU®-PeriNet study.

7. RESEARCH QUESTION AND OBJECTIVES

Objectives

To create a data source with aggregated counts related to pregnancy and mother-child linkage to support study feasibility assessment.

The specific objectives were:

- 5a. To define the data elements needed to inform feasibility assessment of studies on pregnancy, including maternal and offspring perinatal health, or requiring linked mother-child data in the DARWIN EU®.

5b. To incorporate the code for the extraction of aggregated counts defined in objective 5a into the DARWIN EU® onboarding packages.

5c. To extract and compile aggregated counts of the data elements defined as needed for feasibility assessment in objective 5a from each participating data source, overall and stratified by age of the mother, sex of the child, and calendar year.

5d. To incorporate the aggregated counts extracted in objective 5c into the DARWIN EU® Portal.

8. RESEARCH METHODS

8.1. Definition of variables (objective 5a)

The selection of variables was developed through a collaborative process involving both EMA and DARWIN EU® experts. During the drafting of the protocol these experts provided their feedback, and the variables were refined over several iterations to ensure relevance and suitability for assessing the feasibility of future perinatal study requests.

During this process three types of variables were distinguished:

- Type I: Variables that can be routinely calculated upon new data source release, as part of the existing common data model (CDM) onboarding and data quality packages.
- Type II: Variables that to be obtained need institutional review board (IRB) approval.
- Type III: Descriptive variables that can be part of the Portal description (which includes metadata collected upon onboarding from DPs, such as type of data, healthcare settings covered, and geographic origin): Pregnancy episode derivation method (from data source or derived through algorithms) and type of parent-child linkage (deterministic or probabilistic).

The type of aggregate reported for each variable included:

- Record count (RC): Counts representing the total number of occurrences of a specific variable.
- Person count (PC): Counts representing the number of unique individuals that have a record of a specific variable.
- Value distribution (Dist): Set of numerical statistics: mean, standard deviation, minimum, maximum, median, 1st, and 3rd quartile.

8.2. Development of software to extract aggregated counts (objective 5b)

The variable aggregates described in objective 5a were calculated by extending the *DashboardExport* package (<https://github.com/darwin-eu/dashboardexport>) that is already run at every new data source onboarded in the DARWIN EU® Data Network or new data source releases of DPs in the network. This package was used to retrieve (mostly) univariate aggregated counts to inform feasibility assessments, with stratifications where specified in the selection of variables. The addition of these variables to the *DashboardExport* package required development of new queries.

The resulting data format was designed to be the same as the current format used in the existing DARWIN EU® Portal; a comma-separated values (CSV) file with a column for identifying the analysis, strata, and the count or distribution value. In line with the other aggregates calculated by the *DashboardExport* package, counts were rounded up to the nearest 100 and censored if count ≤ 5.

The variable aggregates that needed IRB approval were calculated in objectives 2 and 4 of the DARWIN EU®-PeriNet.

8.3. Extraction of aggregated counts (objective 5c)

The software developed as part of objective 5b was incorporated into the onboarding process. As a result, DPs were required to recalculate the aggregates with each new data release. This requirement formed part of the validation process for approving the data release for use in DARWIN EU® studies, ensuring that DARWIN EU® researchers consistently had access to the information on the most recent data releases for feasibility assessments.

Upon receiving results from DPs, the Network Operations team ensured that the results were complete by performing a data quality control check, as described in the PeriNet Objective 3 report and the general data quality process performed upon onboarding of DARWIN EU® Data Partners.

Additionally, results from objectives 2 and 4 were available from 12 and 5 data sources, respectively.

8.4. Creation of new feasibility dashboard (objective 5d)

The approach taken by the Coordination Centre (CC) for feasibility assessments involves leveraging the DARWIN EU® Portal as the central platform for reviewing aggregated data across the network. Initially, the Portal was used to report and explore core feasibility variables, but with the expansion to perinatal research, it became necessary to include additional, specialised perinatal variables within this framework.

To display the aggregates produced by the DPs with the *DashboardExport* package, we added new database views in the backend of the Portal. These views are automatically applied once new results are uploaded to the Portal. On top of these views, tables were created for each variable and displayed in the interactive dashboard of the Portal.

9. RESULTS

9.1. Definition of variables (objective 5a)

Table 1 and **Table 2** contain a list of aggregated type I and type II variables, respectively, that are expected to be required for future feasibility assessment requests on perinatal research.

Table 1 contains univariate aggregates in the pregnant population. Counts are reported overall and stratified as given in Table 1.

Table 1. Overview of type I variables needed to inform future feasibility assessments of perinatal studies included in the onboarding packages.

Name	Stratification	Aggregate Type
Pregnancies	Singleton, Multiple birth	RC, PC
Age at start of pregnancy	10-year bins	RC, Dist
Pregnancies by calendar year	Yearly bins	RC
Length of pregnancies (weeks)		Dist
Pregnancy outcomes	Livebirth, stillbirth, miscarriage, elective termination, unknown	RC, PC
Pregnancy mode of delivery	Vaginal, Caesarean section	RC, PC
Number of pregnancies by person		Dist
Mother-child linkages		RC
Father-child linkages		RC
Completeness of linkage (number of pregnancies with linked children)		RC

Name	Stratification	Aggregate Type
Infant follow-up time (observation period length among linked children)		Dist
Pregnancy episode derivation method	Algorithm, PET, or source	-
Type of parent-child linkage	Probabilistic, Deterministic	-

RC=Record Count, PC=Person Count, Dist=Value Distribution, PET=Perinatal Expansion Tables.

Type I variables include those that can be routinely calculated upon new data source release, as part of the existing CDM onboarding and data quality packages.

Table 2 contains type II variables that are multi-variate, i.e. combine two or more characteristics of the data, and will require subsequent IRB approval to be provided. The aggregate counts for these variables are available in the results explorer of the PeriNet Objective 4 “Characterisation of pregnancies and mother-child linkage in data sources with pregnancy-related data”. This includes clinical events, including drug exposures, 90 days before, during, and 90 days after pregnancy start. See the objective 4 report for more details.

Table 2. Overview of variables for feasibility of pregnancy and perinatal studies to be aligned with DPs.

Name	Stratification	Aggregate Type
Clinical events among pregnant women	Condition Drug Procedure Measurement	RC, PC
Clinical events among linked children	Condition Drug Procedure Measurement	RC, PC
Drug exposure during pregnancy	Drug Exposure	RC, PC

RC=Record Count, PC=Person Count

For Type III, the descriptive metadata, we have collected the following information for the DPs participating in objective 4:

- Pregnancy episode derivation method (source, derived with algorithm, other)
- Available parent-child linkage (mother-child, father-child, other familial relationships)
- Type of parent-child linkage (deterministic, probabilistic)

9.2. Development of software to extract aggregated counts (objective 5b)

Fourteen new queries (**Table 3**) were added to the *DashboardExport* package v1.5.[1] These analyses cover all database elements listed in **Table 1** and are retrieved from the PET pregnancy and infant tables [2] with the exception of the number of parent-child relationships.

The number of parent-child linkages are retrieved from the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM) Fact Relationship table by stratifying by the Relationship Concept. This concept can refer to ‘Mother’, ‘Father’, or ‘Child’ but also hold other familial relationships.

Table 3. List of analyses implemented in the *DashboardExport* package. The ‘stratum’ column contains the OMOP CDM field names that the analysis stratifies by.

Description	Type	Stratum
Number of fact relationship records by relationship concept		relationship_concept_id (e.g. mother-child)
Number of persons by pregnancy outcome	PC	pregnancy_outcome
Number of records by pregnancy outcome	RC	pregnancy_outcome
Distribution of age	dist	
Distribution of length of pregnancies in weeks	dist	
Number of records by start month	RC	pregnancy_start_date
Number of records by age decile	RC	
Number of persons by pregnancy mode delivery	PC	pregnancy_mode_delivery
Number of records by pregnancy mode delivery	RC	pregnancy_mode_delivery
Number of persons by pregnancy single	PC	pregnancy_single
Number of records by pregnancy single	RC	pregnancy_single
Number of pregnancies with linked children	RC	
Distribution of length of observation period among children linked to pregnancy	dist	
Number of pregnancies per person	dist	

RC=Record Count, PC=Person Count, Dist=Value Distribution.

9.3. Execution of aggregate calculation software (objective 5c)

At the time of writing this report, one data source has successfully executed the updated *DashboardExport* package for their new OMOP data release. The results were shared with the CC and made available in the Portal dashboard.

The *DashboardExport* package is part of the DARWIN EU® onboarding packages, and DPs are expected to execute it upon each new data release and share the results with the CC. DARWIN EU® DPs create a new data release every six months, with the exception of six registries that create a new data release every twelve months. This means that gradually the information will become available for all DPs, also for DPs not participating in PeriNet.

In order to already populate dashboard with results that can be used for feasibility, we processed the objective 2 results of the thirteen participating DPs that were able to provide complete results for that objective. Of these, twelve were uploaded into the Dashboard, and results for IMASIS were calculated with the *DashboardExport* package. The results from both sources include a record count and person count for pregnancies stratified by outcome, by mode of delivery and by age decile. The value distribution of age at first pregnancy and duration of pregnancy was also included. All results were uploaded to the DARWIN EU® Portal.

9.4. Creation of dashboard with aggregates for feasibility (objective 5d)

A new dashboard (**Figure 1**) with charts per data element (**Figures 2, 3, and 4**) has been developed in the DARWIN EU® Portal. These charts are automatically updated once new metadata is uploaded to the Portal, which is done by Network Operations at each new data release. The method of pregnancy episode

derivation and parent-child linkage are added to the questionnaire on the Portal and will be manually updated if needed (**Figure 5**).

All results shown here are from objective 2 results as described in **Section 9.3**. The results from IMASIS were extracted directly with the updated *DashboardExport* package, as they already created a new data release after the update of the package.

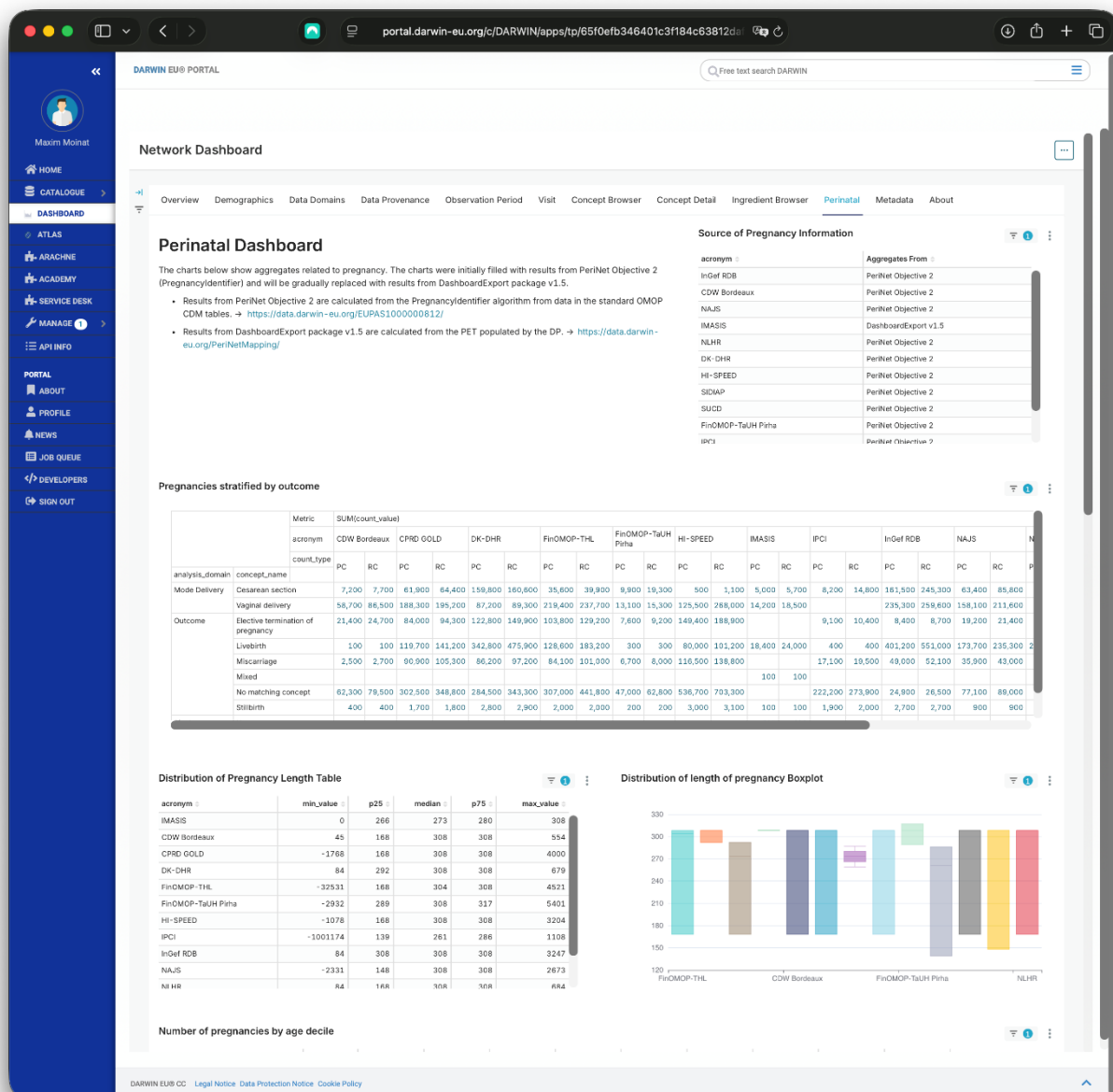


Figure 1. Screenshot of the Perinatal section of the Network Dashboard in the DARWIN EU® Portal.

			Metric	Count			
			acronym	NAJS	NLHR	SIDIAP	SUCD
			count_type				
concept_id	concept_name	analysis_domain		RC	RC	RC	RC
4248584	Mother	Fact Relationship		1.13M			
4277283	Natural mother	Fact Relationship			906k		98.6k
4283070	Father	Fact Relationship		1.06M			
4285883	Child	Fact Relationship		2.19M	888k		
4321888	Natural father	Fact Relationship			888k		
4326600	Natural child	Fact Relationship			906k		98.6k
40478917	Father of subject	Fact Relationship				1.7k	
40478925	Mother of subject	Fact Relationship				801k	
40485452	Child of subject	Fact Relationship				802k	

Figure 2. Screenshot DARWIN EU® Portal showing available familial relationships per data partner. Note that only results from data partners in PeriNet are shared here.

Network Dashboard



		Overview Demographics Data Domains Data Provenance Observation Period Visit Concept Browser Concept Detail Ingredient Browser Perinatal ...															
		Pregnancies stratified by outcome															
		Metric	SUM(count_value)														
		acronym	CDW Bordeaux		CPRD GOLD		DK-DHR		FinOMOP-THL		FinOMOP-TaUH Pirha		HI-SPEED		IMASIS		IPCI
		count_type	PC	RC	PC	RC	PC	RC	PC	RC	PC	RC	PC	RC	PC	RC	PC
Mode Delivery	Cesarean section		7,200	7,700	61,900	64,400	159,800	160,600	35,600	39,900	9,900	19,300	500	1,100	5,000	5,700	8,200
	Vaginal delivery		58,700	86,500	188,300	195,200	87,200	89,300	219,400	237,700	13,100	15,300	125,500	268,000	14,200	18,500	
Outcome	Elective termination of pregnancy		21,400	24,700	84,000	94,300	122,800	149,900	103,800	129,200	7,600	9,200	149,400	188,900			9,100
	Livebirth		100	100	119,700	141,200	342,800	475,900	128,600	183,200	300	300	80,000	101,200	18,400	24,000	400
	Miscarriage		2,500	2,700	90,900	105,300	86,200	97,200	84,100	101,000	6,700	8,000	116,500	138,800			17,100
	Mixed														100	100	
	No matching concept		62,300	79,500	302,500	348,800	284,500	343,300	307,000	441,800	47,000	62,800	536,700	703,300			222,200
	Stillbirth		400	400	1,700	1,800	2,800	2,900	2,000	2,000	200	200	3,000	3,100	100	100	1,900

Figure 3. Screenshot DARWIN EU® Portal showing pregnancy person and record counts stratified by mode of delivery, and by outcome

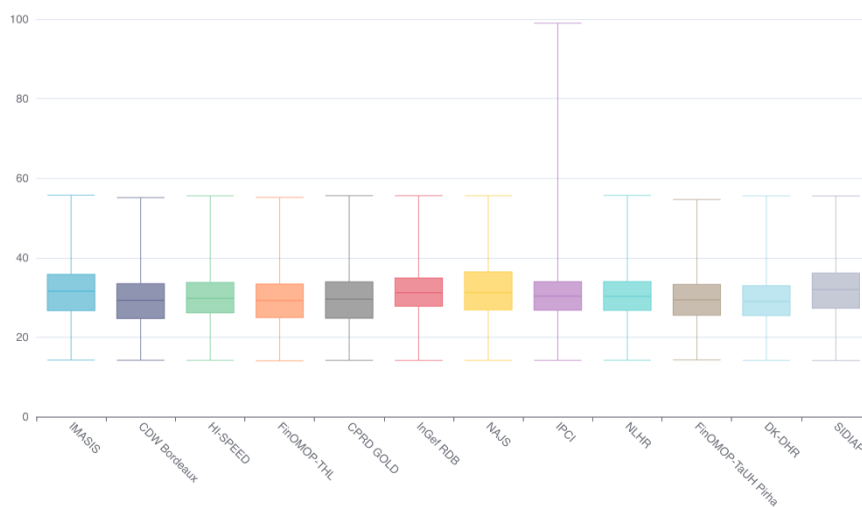


Figure 4. Screenshot of the DARWIN EU® Portal showing a boxplot of the age at start of pregnancy.

3.05. Pregnancy

3.05.01. Pregnancy data available in the source data

- ☒ Yes
☐ No

3.05.02. Has the pregnancy data been mapped to OMOP?

- ☒ Yes
☐ No

3.05.03. Pregnancy episode derivation method

- ☒ From source
☐ Derived with algorithm
☐ Other

3.05.04. Link to PeriNet results if available

<https://data.darwin-eu.org/EUPAS1000000812>

3.05.05. Notes on the availability of pregnancy data

3.06. Parent-child Linkage

3.06.01. Is parent-child linkage available in the source data?

- ☒ Yes
☐ No

3.06.02. Has parent-child linkage been mapped to the OMOP CDM?

- ☒ Yes
☐ No

3.06.03. Types of parent-child linkage

- ☒ Mother-child
☒ Father-child
☐ Other familial relationships

3.06.04. Method of parent-child linkage

- ☐ Deterministic
☒ Probabilistic
☐ Other method

3.06.05. Notes on the availability of parent-child linkage

More information on mother-child linkage can be found here: <https://www.sciencedirect.com/science/article/pii/S1532046424001655>

Figure 5. Screenshot DARWIN EU® Portal showing the questionnaire section for a data partner on Pregnancy availability and Parent-child linkage.

10. DISCUSSION

10.1. Key results

The key result of this work is the operationalisation of a standard set of pregnancy- and mother-child linkage-related feasibility indicators in a format that is directly usable by DARWIN EU® research team. First, we implemented a dedicated perinatal feasibility dashboard in the DARWIN EU® Portal that shows the agreed-upon data elements in tables, providing a consistent overview across data partners. Second, we populated the dashboard with the currently available PeriNet Objective 2 outputs, making results for twelve data partners available at the time of writing and demonstrating how the dashboard can be used immediately for feasibility requests. The current dashboard coverage focuses on routinely extractable (predominantly univariate) aggregates that do not require separate IRB approval, including record and person counts for pregnancies (overall and stratified by key characteristics such as outcome, mode of delivery, and maternal age) and distributions for age at start of pregnancy and pregnancy duration. To ensure sustainability beyond this initial population step, we incorporated the corresponding analyses into the *DashboardExport* onboarding package. The package contains metadata about the data source release that the aggregates were calculated for, which is also shown in the Portal and updated upon each new upload. This integration means that, when data partners create a new data release and execute the updated package as part of the standard onboarding and validation workflow, the same set of feasibility aggregates are regenerated to be shared with the DARWIN EU® CC enabling regular updates without initiating a new study. Over time, this mechanism is expected to broaden coverage to data partners that were not part of PeriNet Objective 2, thereby progressively improving the completeness and timeliness of feasibility information available in the Portal. Finally, feasibility indicators that require IRB approval (for example, counts of specific clinical events or exposures during defined windows around pregnancy) are handled through study-specific analyses and dissemination via Shiny applications. In this way, the Portal dashboard provides an accessible baseline for rapid feasibility screening, while more detailed, study-specific characterisation remains available through the corresponding study outputs for participating data partners.

10.2. Strengths and limitations of the research methods

The calculated aggregates should be considered as approximate numbers to inform the feasibility assessment, as the information provided depends on the data quality and choices made to define the pregnancy episodes, which can vary per DP. Therefore, before adding the aggregates to the dashboard, they should be quality checked. This is part of the existing quality control process and executed by Network Operations upon each new data release received from a DARWIN EU® DP.

The specified *DashboardExport* analyses derive variables from the PET,[2] a set of extension tables to hold the most important variables for perinatal health. This currently serves as the short-term solution for capturing pregnancy data for DARWIN EU® studies as outlined in Objective 3. At the time of writing, the PET has been implemented by five DARWIN EU® DPs. There are restrictions of using the PET for storing perinatal health information (see PeriNet Objective 3), and therefore the long-term solution may be to map this data to the standard OMOP CDM tables, including the new Episode table (PeriNet Objective 3). The calculation of aggregated counts based on the standard OMOP CDM tables, including the Episode table, has been included in earlier versions of *DashboardExport* and will be further extended where needed.

PeriNet Objective 2 results from twelve DPs were used to populate the Portal dashboard to reduce waiting time for the execution of the *DashboardExport* package upon a new data release. It should be noted that Objective 2 utilised the Pregnancy Algorithm (v3) to derive pregnancy episodes and outcomes instead of the PET. Therefore, these findings may deviate from those produced by DPs populating the PET and do not encompass all selected variables. For instance, stratification by single pregnancy and distribution of observation time for linked children are not included.

Variables that are multi-variate, e.g. outcomes after pregnancy, will require subsequent IRB approval. The aggregate counts must be sourced from relevant studies. Accordingly, updates to these variables will only be made following the initiation and completion of new studies and will exclusively contain data from participating DPs. The strength is that all study results are disseminated with a Shiny app that allows DARWIN EU® study team to interactively explore the results. This has been created and made available for PeriNet Objective 4, and provides an in-depth view of the available data and characteristics of the data source. The limitation is that the results are only available for the participating DPs, and updating after a new data release requires a rerun of the study. In some cases, this might require new IRB approval.

11. CONCLUSION

This work delivers the key components needed to strengthen feasibility assessment for future perinatal studies in DARWIN EU®. We defined a set of pregnancy- and mother-child linkage-related data elements and implemented the corresponding aggregate calculations by extending the *DashboardExport* onboarding package, ensuring that updated counts can be generated routinely with new data releases. To make these outputs usable, we integrated the resulting aggregates into a dedicated dashboard in the DARWIN EU® Portal, providing a single point of access for the DARWIN EU® CC team. As additional data partners execute the updated package and submit results with subsequent data releases, the dashboard will progressively expand and remain up to date, enabling quicker and more consistent feasibility assessments across the network.

12. REFERENCES

1. DashboardExport. DARWIN EU® Github. Updated May 4, 2026. Accessed May 4, 2026. <https://github.com/darwin-eu/dashboardexport>
2. Abellan A, Burn E, Trinh NT, Burkard T, Callahan A, Fernández-Bertolín S, Hurley E, Rodriguez C, Segundo E, Morales DR, ME Nordeng H. Expanding the OMOP common data model to support perinatal research in network studies. *Pharmacoepidemiology and Drug Safety*. 2025 Feb;34(2):e70106.

13. ANNEXES

ANNEX I. Glossary

Additional definitions are available in the EMA Glossary of terms <https://www.ema.europa.eu/en/about-us/glossaries>.

Aggregated Data

Data collected and combined from multiple sources to generate summary information, typically anonymised.

Benefit-Risk Assessment

Evaluation of the positive therapeutic effects of a medicine compared to its risks (e.g. side effects).

Common Data Model (CDM)

A standardised data structure that enables data from multiple sources to be harmonised, making analysis consistent and reproducible. DARWIN EU® utilises the OMOP CDM maintained by the OHDSI community.

Complex Studies (C3)

Studies requiring the development or customisation of specific study designs, protocols, and Statistical Analysis Plans (SAPs), with extensive collection or extraction of data. Examples include etiological studies measuring the strength and determinants of an association between an exposure and the occurrence of a health outcome in a defined population considering sources of bias, potential confounding factors, and effect modifiers.

Coordination Centre (CC)

The central hub responsible for managing and overseeing the activities within DARWIN EU®. It is based at Erasmus University Medical Center in Rotterdam, the Netherlands.

Data Access

The process of obtaining permission to use specific datasets for regulatory or scientific studies.

Data Quality Framework

A set of standards and procedures to ensure accuracy, completeness, timeliness, and consistency of data used in DARWIN EU®.

Data Source

A database or repository of structured health-related data, such as electronic health records (EHRs), insurance claims, or registries.

DARWIN EU®

The European Medicines Agency's (EMA) federated network of real-world data sources designed to generate evidence to support regulatory decision-making.

EMA (European Medicines Agency)

The regulatory body responsible for the evaluation and supervision of medicinal products in the EU, overseeing DARWIN EU®.

Evidence Generation

The process of analysing real-world data to produce scientific information that can inform healthcare or regulatory decisions.

Federated Network

A data infrastructure where data remain at their original location but can be analysed in a harmonised way across multiple partners using a common model and tools.

GDPR (General Data Protection Regulation)

The EU regulation governing the protection of personal data and privacy, crucial to how DARWIN EU® handles health data.

Health Technology Assessment (HTA)

A systematic evaluation of properties and impacts of health technology, often using DARWIN EU® data to support assessments.

Metadata

Descriptive information about a data source (e.g. its content, quality, and structure), essential for identifying relevant databases in DARWIN EU® studies.

Off-the-Shelf Studies (OTS)

Studies for which a standard protocol per study/analysis type and standardised analytics may be developed and applied or adapted, typically relating to a descriptive research question. This includes studies on disease epidemiology, for example, the estimation of the prevalence or incidence of health outcomes in defined time periods and population groups, or drug utilisation studies at the population or patient level.

OHDSI (Observational Health Data Sciences and Informatics)

An open-science collaborative community that develops tools and standards (including the OMOP CDM) to enable large-scale analytics of observational health data. OHDSI provides the technical and scientific foundation for DARWIN EU®'s analytical ecosystem.

Patient-Level Data

Data related to individuals, de-identified, used for longitudinal or detailed analyses.

OMOP (Observational Medical Outcomes Partnership)

A common data model (CDM) that standardises the structure and content of observational healthcare data, enabling systematic analysis across disparate datasets. DARWIN EU® uses the OMOP CDM to ensure interoperability and consistency in real-world evidence generation.

Real-World Data (RWD)

Data relating to individual health status or healthcare delivery that is collected from routine clinical practice rather than from randomised controlled trials.

Real-World Evidence (RWE)

Clinical evidence derived from the analysis of RWD, used to inform decisions by regulators, payers, or clinicians.

Regulatory Decision-Making

The process by which authorities like EMA assess data to authorise, monitor, or modify the use of medicines in the EU.

Routine Repeated Studies (RR)

Studies that are either Off-the-Shelf or Complex studies repeated on a regular basis, following the same protocol and study code, but with updated data and/or different data partners.

Study Protocol

A detailed plan describing how a specific real-world study will be conducted, including objectives, design, data sources, and analyses.

Very Complex Studies (C4)

Studies which cannot rely only on electronic health care databases, or which would require complex methodological work, for example, due to the occurrence of events that cannot be defined by existing diagnosis codes, including events that do not yet have a diagnosis code, where it may be necessary to combine a diagnosis code with other data such as results of laboratory investigations. These studies might require the collection of data prospectively, or the inclusion of new (not previously onboarded) data sources.