Drug exposure during pregnancy in primary care: an algorithm and observational study from SIDIAP database, Catalunya, Spain


ABSTRACT

Objectives To develop an algorithm to identify pregnancy episodes in women at childbearing age using SIDIAP (Information System for the Improvement of Research in Primary Care) data (Catalunya, Spain). To describe drugs dispensed during gestation.

Design Construction of an algorithm to identify all pregnancy episodes occurred from January 2011 to June 2020 in women aged 12–50. The variables used to create the algorithm include first day of last menstrual period, reasons for pregnancy termination and diagnoses registered in the primary healthcare records. Population-based cohort study including the pregnancy episodes identified by the algorithm.

Setting Catalunya, Spain.

Participants All women aged 12–50 with at least one pregnancy episode occurred during January 2011–June 2020.

Interventions No interventions performed.

Primary and secondary outcome measures Identification of pregnancy episodes through an algorithm and description of drug exposure.

Results We identified 327 865 pregnancy episodes in 250 910 people with a mean age of 31.3 years. During the study period, 83.4% of the episodes were exposed to at least one drug. The most frequent groups dispensed were iron preparations (48% of pregnancy episodes), iodine therapy (19.7%), analgesics and antipyretics (28%), vitamins B12 plus folic acid (19.8%), and non-steroidal anti-inflammatory drugs (NSAIDs, 15.1%). The supplements were more frequently dispensed at least twice, and the drugs for acute conditions were mainly dispensed only once during the pregnancy episode.

Conclusions We developed an algorithm to automatically identify the pregnancy periods in SIDIAP. We described prescription drugs used during pregnancy. The most used ones were supplements, analgesics, NSAIDs, and antibiotics. SIDIAP might be an efficient database to study drug safety during pregnancy and the consequences of drug use in the offspring.

Trial registration number EUPAS37675.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ One limitation of the algorithm development is the lack of information on women whose pregnancies are followed up in the hospital or in private settings.

⇒ Other limitation of the algorithm is that it has not been validated. Nevertheless, the number and distribution of pregnancies are in line with the previous algorithms published.

⇒ Despite these limitations, Information System for the Improvement of Research in Primary Care database might be a valid and efficient resource to study drug safety during pregnancy, as it captures clinical information of most of the Catalan population.

INTRODUCTION

Drug use during pregnancy and breast feeding has not been widely analysed due to the exclusion of pregnant and lactating people from the participation in clinical trials for ethical reasons and due to the generalised lack of investigation in women’s health.1–3 However, the use of drugs during gestation and breast feeding may entail risk of different health problems for these women and for the fetus and the newborn.1 4 and also the non-use of certain drugs can lead to worsening of chronic and acute conditions in the mother or to complications for the infant, such as drugs to treat asthma,5 6 autoimmune disorders,7 8 diabetes9 10 or epilepsy.11 Deciding and planning on pharmacotherapy based on the safety profile of a drug during those periods can be challenging for prescribers. For this reason, it is necessary to assess drug safety through postapproval observational studies for drugs which cannot be discontinued during pregnancy, drugs to treat pregnancy-related conditions and drugs with preclinical evidence of risk for the offspring.12 13
Apart from birth cohorts established worldwide to gain knowledge on perinatal health, in the last decades, database studies are offering numerous advantages to study medicines safety in pregnant and lactating women and in their offspring, such as the large number of people included, the availability of mother–child linked data, long follow-up periods, information on consequences derived from drug use or on confounders or the possibility to design algorithms through machine learning methods. Algorithms are often necessary because electronic health records (EHR) lack of a unique register to unambiguously identify the gestation periods. Thus, it is usual to design algorithms to identify pregnancies with the purpose of studying drug use during gestation through database studies.

We aimed to develop an algorithm to identify pregnancy episodes for all women aged from 12 to 50 between January 2011 and June 2020 using data from the EHR of primary healthcare (PHC) in Catalonia, Spain. With the pregnancies identified by the algorithm we aimed to describe the drugs dispensed during gestation.

**METHODS**

**Study design**

First, construction of an algorithm to identify all pregnancy episodes occurred from January 2011 to June 2020 in women at childbearing age (12–50 years) in Catalonia, Spain. And second, population-based cohort study including the pregnancy episodes identified by the algorithm.

**Data source**

The study data source is the Information System for the Improvement of Research in Primary Care (SIDIAP), which captures clinical information of approximately 5.8 million Catalan citizens (around 80% of the Catalan population). This information is pseudonymised and it is originated from different data sources: (1) EHR in PHC of the Catalan Health Institute; (2) specialist referrals, clinical parameters, toxic habits, sickness leave, date of death, laboratory test data and drug prescriptions issued in PHC, registered as Anatomical Therapeutic Chemical (ATC) classification system codes; (3) sexual and reproductive healthcare (ASSIR) records, which is the EHR module used by gynaecologists and midwives to register variables related with the sexual and reproductive health of women and follow-up of pregnancies, as date of last menstrual period (LMP), gestational week, date of delivery or pregnancy termination—based on ultrasound results—or termination outcomes. (3) Pharmacy invoice data corresponding to the PHC drug prescriptions, classified according to the ATC classification.

We used pregnancy-related ATC classification. Variables collected to describe the study population at the pregnancy start date (PSD) were: age, socioeconomic status by MEDEA (Mortalidad en áreas pequeñas Españolas y Desigualdades socioEconómicas y Ambiente) index, body mass index, comorbidities, smoking status and alcohol intake; for this study the latter two correspond to records of less than 12 months before PSD.

We used the pharmacy invoice dispensing data to assess the drug exposure. Pregnancy episodes were classified as exposed to drugs when there was at least one dispensing from 30 days before PSD up to 30 days after delivery date, and per pregnancy trimester as follows: first trimester: from 30 days before PSD to 120 days after PSD; second trimester: from 120 days after PSD to 210 days after PSD and third trimester: from 210 days after PSD to 30 days after delivery date. This definition of the pregnancy trimesters was done for exposure definition purposes in order to consider the particularities of the pharmacy invoice register.

We described the most dispensed pharmacological groups (ATC) in overall pregnancies and by trimester, the frequency of groups with one or more than one dispensing during the pregnancy episode, and the most frequent active principles dispensed more than once.

**Algorithm development**

We created a three-step sequential algorithm to identify the pregnancy episodes occurring during the study period, define their duration and the outcome of the pregnancy.

**Step 1: identification of potential pregnancies**

We carried out a hierarchical and mutually exclusive search in our data source of records linked to gestation. We used the specific record of LMP to identify a potential pregnancy registered in ASSIR. If LMP was not available, the algorithm searched for the following records (figure 1): positive pregnancy test, gestational week, fetal death and PHC diagnoses indicating pregnancy or abortion (ICD-10 codes). See online supplemental table 1 for full list of codes.

**Step 2: length of pregnancy**

We determined the length of pregnancy using the LMP, the pregnancy episode end date (delivery or abortion), the diagnostic codes dates suggestive of pregnancy end and the weeks of gestation registries in ASSIR. If dates were unknown, they were imputed according to the criteria established (figure 1). Pregnancy episodes shorter than 4 or longer than 43 weeks were excluded.

**Step 3: outcome of pregnancy**

Reasons for pregnancy termination were identified by the ASSIR end label and completed by ICD-10 diagnoses related to childbirth or the puerperium. We imputed an outcome for episodes with unknown outcome and known length: abortion for length <24 weeks and live birth if ≥24 weeks (threshold for fetal viability/feasibility).
During the development of the algorithm, pregnancy episodes were scrutinised for potential inconsistencies, such as pregnancy duration discrepancy with pregnancy outcomes, overlapping episodes, those starting or terminating out of the study period after imputation dates and finally a minimum time lapse of 4 weeks was established between the dates of the different episodes.

**Study size**

The study size included all pregnancy episodes identified through the algorithm in women aged 12–50 during the study period.

**Statistical analysis**

After identifying all women with pregnancy episodes defined by the algorithm, we described the study...
population and the drug exposure, using frequencies and percentages for categorical variables and mean and SD for continuous variables.

Missing values on baseline variables are reported in table 1. Missing data for PSD, end date and labels of pregnancy termination outcomes were imputed as explained in the Algorithm development.

All analyses were conducted with R software (V.4.1 or superior).

**Patient and public involvement statement**

The Research Ethics Committee of IDIAPJGol has patient and public representation who participated in reviewing and approving the study protocol.

**RESULTS**

From January 2011 to June 2020, our algorithm identified 327865 pregnancy episodes in 250910 women (1.3 episodes per woman), and a total sum of follow-up of 210,463.1 years (1.6 episodes/year). These women had a mean age of 31.3 years, and 56.8% of the episodes occurred between 30 and 39 years of age. When we analysed the data from the registry of the pregnancy episodes defined by the algorithm, 76.8% of them had complete data on the start and end of the pregnancy and the results of the episode in the EHR. Out of them, 80.6% of pregnancy episodes ended with live births, 19.2% resulted in abortion, 0.15% were stillbirths and 0.08% were ectopic or molar pregnancies (online supplemental figure).

Baseline sociodemographic and clinical characteristics of pregnant women at PSD are included in table 1. Among the most frequent comorbidities at baseline, 24.5% of women had an active diagnosis of anxiety, 14.5% were obese and 9.1% had a respiratory disease. Smoking and alcohol habits had a high number of missing values at PSD.

During the period studied, 274799 (83.4%) of the pregnancy episodes were exposed to at least one drug. The most dispensed pharmacological groups during the overall pregnancies and by trimester are shown in figure 3. The most dispensed substances from these frequent ATC groups are shown in figure 3, being the combination of iodide, vitamin B₁₂, and folic acid (42.5% of episodes), ferrous sulfate (37.6%) and paracetamol (31.1%) the most used drugs. Other frequent drugs dispensed were antibiotics (eg, fosfomycin: 14.8%).

**DISCUSSION**

In the absence of a unique register to identify pregnancies in our setting, we designed the algorithm in the current research context where many algorithms have been published aiming to identify pregnancies and assess drug exposure risks during pregnancy.17 19 20 26 27 Once generated, we identified 327865 pregnancy episodes in a cohort of women at childbearing age attended in PHC from 2011 to 2020, accounting for 1.3 episodes per woman, which matches the observed numbers in Catalonia (1.21) and Spain (1.19).28 29 The type of pregnancy end identified with our algorithm was similar to BIFAP algorithm, which identified 21.5% pregnancy losses, 0.8% ectopic pregnancies and 0.2% stillbirths.17 The first step of our algorithm consisted in searching for the LMP, which is a pregnancy-related specific date which is only registered in the ASSIR records at the pregnancy monitoring initiation, so we considered it as the highest quality registry to identify PSD.

Regarding the clinical characteristics of the population studied, it is remarkable the high number of missing records for tobacco and alcohol habits in pregnant women in our database (91.2% of smoking missing values the prior 12 months vs 22.1% of smoking missing values any time before PSD). It may be recommendable to reinforce the need to record these variables more accurately so health professionals attending pregnant people may initiate smoking cessation interventions when necessary.30

With regard to the most common comorbidities at PSD, it is noteworthy that nearly 25% of pregnant women in our study had a diagnosis of anxiety, which is frequently described in women at childbearing age with other concomitant mental disorders such as depression,31–34 which also showed a significant prevalence in our cohort (5.7%). Respiratory diseases or migraine were frequent in our cohort, these disorders have commonly been reported during pregnancy with the challenges associated to their management.35 36

Dispensation of drugs was frequent in the identified pregnancy episodes. Among the most prescribed, supplements are recommended during pregnancy.30, 37 Drugs used for acute conditions, such as NSAID, analgesics or antibiotics, were also identified, as during pregnancy infectious diseases or pain conditions are also frequently reported. We need to point out that supplements or analgesics counts might be under-reported due to the availability of over the counter (OTC) medicines, not captured in our database. Other OTC drugs that can be underestimated include antacids or laxatives, which are frequently used by gestating women.19 38 39 Among the
### Table 1  Baseline sociodemographic and clinical characteristics of the women with pregnancy episodes included in the study

<table>
<thead>
<tr>
<th>N (%)</th>
<th>Overall pregnancy episodes N=327865</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age in years, mean (SD)</td>
<td>31.3 (5.8)</td>
</tr>
<tr>
<td>12–14</td>
<td>248 (0.1)</td>
</tr>
<tr>
<td>15–24</td>
<td>44,685 (13.6)</td>
</tr>
<tr>
<td>25–29</td>
<td>74,364 (22.7)</td>
</tr>
<tr>
<td>30–34</td>
<td>107,587 (32.8)</td>
</tr>
<tr>
<td>35–39</td>
<td>78,576 (24.0)</td>
</tr>
<tr>
<td>≥40</td>
<td>22,405 (6.8)</td>
</tr>
<tr>
<td>MEDEA index</td>
<td></td>
</tr>
<tr>
<td>Rural</td>
<td>60,692 (18.5)</td>
</tr>
<tr>
<td>Urban quintiles 1–3</td>
<td>121,014 (36.9)</td>
</tr>
<tr>
<td>Urban quintile 4–5</td>
<td>114,513 (34.9)</td>
</tr>
<tr>
<td>Urban unknown</td>
<td>31,482 (9.6)</td>
</tr>
<tr>
<td>Missing values</td>
<td>166 (0.1)</td>
</tr>
<tr>
<td>BMI categorised</td>
<td></td>
</tr>
<tr>
<td>Underweight (&lt;20 kg/m²)</td>
<td>12,915 (3.9)</td>
</tr>
<tr>
<td>Normal (20-25)</td>
<td>50,627 (15.4)</td>
</tr>
<tr>
<td>Overweight (25-29)</td>
<td>41,645 (12.7)</td>
</tr>
<tr>
<td>Obese (≥30)</td>
<td>29,561 (9.0)</td>
</tr>
<tr>
<td>Missing values</td>
<td>193,117 (58.9)</td>
</tr>
<tr>
<td>Smoking habit*</td>
<td></td>
</tr>
<tr>
<td>Smoker</td>
<td>6,998 (2.1)</td>
</tr>
<tr>
<td>Ex-smoker</td>
<td>5,960 (1.8)</td>
</tr>
<tr>
<td>Non-smoker</td>
<td>15,883 (4.8)</td>
</tr>
<tr>
<td>Missing</td>
<td>299,024 (91.2)</td>
</tr>
<tr>
<td>Alcohol intake*</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>378 (0.1)</td>
</tr>
<tr>
<td>Moderate</td>
<td>17,424 (5.3)</td>
</tr>
<tr>
<td>No intake</td>
<td>46,409 (14.1)</td>
</tr>
<tr>
<td>Missing</td>
<td>263,654 (80.4)</td>
</tr>
<tr>
<td>Comorbidities</td>
<td></td>
</tr>
<tr>
<td>Anxiety</td>
<td>80,429 (24.5)</td>
</tr>
<tr>
<td>Cancer</td>
<td>23,266 (0.7)</td>
</tr>
<tr>
<td>Depression and bipolar disorders</td>
<td>18,774 (5.7)</td>
</tr>
<tr>
<td>Diabetes</td>
<td>17,240 (0.5)</td>
</tr>
<tr>
<td>Eating disorders</td>
<td>12,399 (3.8)</td>
</tr>
<tr>
<td>Epilepsy</td>
<td>17,192 (0.5)</td>
</tr>
<tr>
<td>Hypertension</td>
<td>34,024 (1.0)</td>
</tr>
<tr>
<td>Migraine</td>
<td>23,290 (7.1)</td>
</tr>
<tr>
<td>Obesity (ICD-10 and/or BMI≥30)</td>
<td>47,467 (14.5)</td>
</tr>
<tr>
<td>Respiratory diseases</td>
<td>29,760 (9.1)</td>
</tr>
<tr>
<td>Rheumatoid arthritis</td>
<td>760 (0.2)</td>
</tr>
</tbody>
</table>

*Women can be counted more than once, as their characteristics are counted for each pregnancy episode, Smoking and alcohol habits registered <12 months before the pregnancy start date.

BMI, body mass index; ICD-10, International Classification of Diseases 10th version; MEDEA, Mortalidad en áreas pequeñas Españolas y Desigualdades socioEconómicas y Ambientales.
most frequent antibiotics we found fosfomycin and amoxicillin, which might have been prescribed for common infections during pregnancy such as urinary tract infections.30 40

Despite the high prevalence of anxiety disorders in our cohort, anxiolytic or antidepressant drugs were not among the most used pharmacological groups. This may be related with the absence of evidence and safety concerns on the use of psychotropic medications during pregnancy.41–43 Other studies have reported similar levels of utilisation of these drugs.44–46

One limitation of our study concerning the algorithm development is the lack of information on women whose pregnancies are followed up in the hospital or in private settings. The pregnancies referred to hospitals for follow-up include those women with chronic and autoimmune diseases which can increase the risk of complications during the pregnancy or at delivery. Although each ASSIR has different referral protocols, some of the conditions include: history of miscarriages, history of chromosomal anomalies, history of prematurity, morbid obesity, pre-eclampsia, gestational diabetes, severe anaemia, etc.30

Table 2 Exposure to the most frequent pharmacological groups during pregnancy

<table>
<thead>
<tr>
<th>Pharmacological group</th>
<th>N (%) of pregnancy episodes exposed</th>
<th>N (%) episodes with one dispensing*</th>
<th>N (%) episodes with &gt;1 dispensing*</th>
</tr>
</thead>
<tbody>
<tr>
<td>B03A, iron preparations</td>
<td>157 467 (48.0)</td>
<td>59021 (37.5)</td>
<td>98 446 (62.5)</td>
</tr>
<tr>
<td>H03C, iodine therapy</td>
<td>131 763 (40.2)</td>
<td>40 119 (30.4)</td>
<td>91 644 (69.6)</td>
</tr>
<tr>
<td>N02B, other analgesics and antipyretics</td>
<td>91 686 (28.0)</td>
<td>63 563 (69.3)</td>
<td>28 150 (30.7)</td>
</tr>
<tr>
<td>J01C, penicillins</td>
<td>65 050 (19.8)</td>
<td>51 019 (78.4)</td>
<td>14 031 (21.6)</td>
</tr>
<tr>
<td>B03B, vitamin B₁₂ and folic acid</td>
<td>64 534 (19.7)</td>
<td>31 814 (49.3)</td>
<td>32 720 (50.7)</td>
</tr>
<tr>
<td>M01A, NSAID</td>
<td>49 577 (15.1)</td>
<td>41 130 (83.0)</td>
<td>8447 (17.0)</td>
</tr>
</tbody>
</table>

*Percentages calculated over the number of episodes exposed for each group.

NSAID, non-steroidal anti-inflammatory drug.
Other limitation is the lack of a specific validation of the algorithm. Nevertheless, the number and distribution of pregnancies and the drug use during gestation are in line with previous studies, and similar to the Catalan official data.

As pointed out, some of the drug counts in our study might be underestimated, as we are not able to capture OTC drugs, which are frequently used by pregnant women and are often not registered in EHR, being one of the usual limitations of these type of studies. Despite the limitations, SIDIAP database might be a valid and efficient resource to study drug safety during pregnancy, as it captures clinical information of most of the Catalan population, we have developed an algorithm to automatically identify the pregnancy periods, and an algorithm to link mother and child pairs has also been developed and can be used to study pregnancy outcomes. It is an algorithm which establishes mother and child pairs linked through national insurance number and coinsurance status. We have also mapped SIDIAP pregnancy data to the ConcePTION common data model, which has demonstrated its potential to address questions about utilisation, effectiveness and safety of medicines during pregnancy and lactation. All of this might help to fill the gap in the currently available evidence, as safety concerns on drug use during pregnancy affect not only women, but also the fetus and the newborns, and it is also applicable to the lactation period, which we are also planning to assess within our database.

CONCLUSIONS
We have developed an algorithm to automatically identify the pregnancy periods, which will allow us to study not only the drug use in pregnant people, but its consequences in these women’s health and in their offspring’s.

We have described the use of prescription drugs in a large cohort of pregnant women. The most used drugs during pregnancy were recommended supplements and drugs used for acute conditions, such as analgesics, NSAID or antibiotics. SIDIAP database might be an efficient resource to study drug safety during pregnancy.

Author affiliations
1Àrea del Medicament i Servei de Farmàcia, Gerència d’Atenció Primària Barcelona Ciutat, Institut Català de la Salut, Barcelona, Spain
2Universitat Autònoma de Barcelona, Bellaterra (Cerdanyola del Vallés), Spain
3Fundació Institut Universitari per a la Recerca a l’Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), Barcelona, Spain
4College of Pharmacy, Department of Pharmacotherapy, University of Utah, Salt Lake City, Utah, USA
5Marketing farmacèutico & Investigació clínica, Barcelona, Spain
6Former employee at IDIAPJGol, Barcelona Spain
7RTI Health Solutions Barcelona, Barcelona, Spain
8Unitat de farmàcia, Servei d’Atenció Primària Barcelonès Nord i Maresme, Badalona, Spain

Acknowledgements The authors want to thank María Aragón and Clara Rodríguez for the data acquisition process from SIDIAP database.

Contributors MLV: conceptualisation, study design, data curation, writing (original draft), review and editing. AG-L: conceptualisation, study design, data curation, writing, review and editing. OP-V: study design, data curation, formal analysis and review. CVC: data curation, formal analysis, writing (original draft), review and editing. JR: conceptualisation, study design, writing, review and editing. MG-S: conceptualisation, study design, data curation, writing (original draft), review and editing. RM: conceptualisation, data curation, study design, writing, review and editing. MG-S is responsible for the overall content as the guarantor.

Funding This study received funding from the 8th call for SIDIAP grants, 2018–2019, expedient number 4R18/188; and from Health Department of the Generalitat de Catalunya, in the call corresponding to 2021 for the granting of funding of the Strategic Plan for Research and Innovation in Health (PERIS) 2021–2024, modality Research Projects in Primary Care, expedient number SLT/21/000068.
REFERENCES


