O5 September 2022

Data analysis plan

Drug utilisation study of macrolide-containing medicinal products during pregnancy

Table of Contents

1. Rationale and background

Recently there have been a number of publications investigating risks associated with macrolide use during pregnancy.

To support a thorough review of this issue within regulatory evaluation procedures, a drug utilisation analysis of macrolide use in pregnancy (stratified by specific agent) was considered relevant to inform regulatory evaluation and decision-making.

While identifying clinical codes in electronic health records that indicate events related to pregnancy is simple, reliably detecting the start and end of pregnancy, to be able to assess exposure by trimester, is significantly more complicated.

Recently there has been a few papers attempting to validate phenotyping algorithms to detect gestational age, including by Moll et al, Drug Saf. 2021[(1)], Taylor et al, PDS. 2022[(2)] and Matcho et al, PLoS One. 2018[(3)]. These tend to be specific to one or a small number of databases and data models and thus a direct application of these in new datasets is unfeasible.

In this study, a short validation of a phenotyping algorithm for gestational age will be performed across three EU databases.

2. Research question and objectives

This study will aim to:

1) Develop and validate a phenotyping algorithm that identifies gestational age.
2) Describe the use of macrolides and amoxicillin during pregnancy, specifically:
    a) Prescriptions by year, age, gravidity, and trimester of pregnancy, stratified by substance.
    b) Number of prescriptions and number of distinct substances prescribed by pregnancy.
    c) Indications by substance.
3) Characterise the drug quantity, namely number of units prescribed in tablets or capsules, of Erythromycin by trimester of pregnancy.

3. Research methods

3.1. Study design

This will be a longitudinal observational cohort study.

3.2. Setting and study population

The study period will be from start of data collection for each of the three databases to January 2022. See details of each database in Annex I. The population will include all pregnant women identified according to the phenotyping algorithm as defined in Annex II.

3.3. Variables

3.3.1. Vocabularies

The medical vocabulary used will be OMOP Concept ids. These concept ids use SNOMED CT as standard vocabulary and are mapped to several other vocabularies, including Read, ICD-9 and ICD-10 as well as to the national editions of SNOMED CT.

The medicinal product dictionary used will be RxNorm and extensions.

3.3.2. Pregnancy

A phenotyping algorithm to identify pregnancy start and end dates will be developed in OMOP concept id and assessed using three databases in the OMOP common data model, IQVIA™ Disease Analyzer Germany, IQVIA™ Disease Analyzer France and IQVIA™ Medical Research Data EMIS (IMRD UK). The details of the phenotyping algorithm are in Annex II. Only the databases that allow for consistent and reliable identification of pregnancy trimester will be used as per the algorithm.

OMOP Concept ids for pregnancy and related events will be extracted from the paper by Matcho et al, PLoS One. 2018.

The threshold to separate any two pregnancies will be defined as any two pregnancy related OMOP concept ids separated by more than 365 days or any pregnancy related code 90 days after a OMOP concept id related to abortion or miscarriage.

The main analysis will be conducted using patients that had temporal consistency for codes identifying start of pregnancy and end of pregnancy. A sensitivity analysis will be conducted with patients that had delivery codes in OMOP CDM.

3.3.3. Exposure

3.3.3.1. Antibiotics

OMOP concept codes will be identified for the following macrolides: Azithromycin, Clarithromycin, Erythromycin, Midecamycin, Oleandomycin, Roxithromycin, Spiramycin, Telithromycin. For comparison, the OMOP concept codes for Amoxicillin will also be used.

3.3.3.2. Indications

All observations within +/- seven days of the prescription of macrolide or amoxicillin-containing products will be identified and screened for children concepts of the SNOMED CT Infectious disorders concept, in OMOP concept ids.

3.4. Data sources

The study will be conducted using three European databases: a French database (IQVIA™ Disease Analyzer France December 2021 version) a German database (IQVIA™ Disease Analyzer Germany December 2021 version) and a United Kingdom database (IQVIA™ Medical Research Database January 2022 version) (see Annex I for more details).

3.5. Statistical analysis

3.5.1. Main statistical methods

We will describe the use of macrolides and amoxicillin in pregnant women who will be identified using the phenotyping algorithm as defined in Annex II.

Descriptive statistics stratified by substance will be also presented for prescriptions of macrolides and amoxicillin by year, age group, gravidity (i.e., pregnancy number), trimester and indications. In addition, the number of prescriptions per pregnancy will be determined and stratified by substance.

Trends of use of each substance over time, relative to number of pregnancies will also be plotted.

Drug quantity is reported as pack size (standard pack sizes) in the databases only for each solid oral formulation of erythromycin. Therefore, drug quantity will be described by trimester of pregnancy only for erythromycin.

The main analysis will be conducted including patients that had temporal consistency for codes identifying both start of pregnancy and end of pregnancy.

3.5.2. Sensitivity analysis

A sensitivity analysis will be carried out with patients that had OMOP concept ids for full-term delivery (See Annex II).

Analyses will be conducted by EMA researchers using IHD, SAS, and R statistical software.

3.6. Quality control

The study will be conducted according to the ENCePP code of conduct (European Medicines Agency 2018). Standard operating procedures or internal process guidance will be adhered to for the conduct of the study. These procedures include rules for secure and confidential data storage, quality-control procedures for all aspects of the study from protocol development to the reporting of the results.

All documents undergo at least one round a review by an experienced reviewer, while the results from the statistical analysis are either reviewed or checked via double coding.

The quality control of the data is the responsibility of the data holder.

3.7. Limitations of the research methods

Limitations of this study will include:

Misclassification of gestational age. To identify the gestational age of each patient, a phenotyping algorithm had to be defined and tested. The main phenotyping algorithm required the presence of

codes indicating start of pregnancy and end of pregnancy. Considering that the presence of both sets of codes depleted the sample of pregnant women significantly, it may be that these patients, that have a well described pregnancy are not representative of the overall population of pregnant women as it suggests they had more healthcare interactions than several other pregnant women. In addition, a sensitivity algorithm, the presence of delivery codes to determine day 280 of pregnancy, may not accurately reflect the duration of pregnancy. To mitigate the risk of misclassifying gestational age, all codes for elective delivery were removed.

Missing data on miscarriages. Only successful deliveries are included in the study because it is difficult to establish the gestational age of women with miscarriage. However, the effects of exposure in this sub-group of women are extremely relevant for drug safety studies.

Misclassification of indication. Another limitation is that the indication was inferred, this is because it can happen that the date at which a prescription is made is not the same as the date at which the corresponding diagnosis is made. A seven-day interval was defined whereby any infection indication within +/- 7 days of prescription was considered the indication. The 7-day threshold was defined empirically. Several intervals, from 1 to 60 days were tested, suggesting a linear relationship between the number of days used in the interval and the number of extra indications found.

References

1. Moll K, Wong HL, Fingar K, 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):1151-1164. doi:10.1007/s40264-021-01113-8
2. Taylor LG, Bird ST, Stojanovic D, et al. Utility of fertility procedures and prenatal tests to estimate gestational age for live-births and stillbirths in electronic health plan databases. Pharmacoepidemiol Drug Saf. 2022;31(5):534-545. doi:10.1002/pds.5414
3. 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. Published 2018 Feb 1. doi:10.1371/journal.pone.0192033

Annexes

Annex I – Databases

IQVIA Disease Analyser Germany

IQVIA Disease Analyser (IDA) Germany collects computerised information from specialised and general primary care practices throughout Germany since 1992. Around 3% of general practitioners (GP) practices are included, which covers all patients consulting a practice. Data from IDA Germany have been shown to be reasonably representative of German healthcare statistics for demographics and certain diseases and is considered one of the largest national medical databases worldwide. IDA Germany includes more than 2,500 practices and 3,100 physicians (13 speciality groups) representing over 15,000,000 patients. This database used to be named IMS Germany and some use of this terminology may persist.

The March 2022 OMOP version of this database was used.

IQVIA Disease Analyzer France

IDA France collects anonymised patient medical records since 1997 through a representative panel of GPs. The physician sample represents approximately 2% of physicians and is weighted by age and gender of the physician, doctor region and the SNIR of the physician (National Official Indicator of the GP volume of activity in terms of visits and consultations). Some 99% of the French population is insured, but there are differences regarding level of coverage. IDA France includes around 1,000 GPs and represents more than 4,000,000 of patients and considered representative for the French population. This database used to be named IMS France and some use of this terminology may persist.

The March 2022 OMOP version of this database was used.

IQVIA Medical Research Data EMIS UK

IQVIA Medical Research Data (IMRD) EMIS UK is a primary care database from the UK. GPs play a gatekeeper role in the healthcare system in the UK, as they are responsible for delivering primary health care and specialist referrals. Over 98% of the UK-resident population is registered with a GP, so that GP patient records are broadly representative of the UK population in general. Patients are affiliated to a practice, which centralizes the medical information from GPs, specialist referrals, hospitalizations, and tests.

The May 2022 OMOP version of this database was used.

Annex II – Phenotyping algorithm to identify gestational age

Methodology

Patients with any OMOP concept code from the Matcho *et al.*, PLoS One, 2018[3] puplication will be identified and extracted from IDA France, IDA Germany and IMRD UK (available online and upon request).

Gravidity thresholds will be defined as any two-pregnancy related OMOP concept ids separated by more than 365 days or any pregnancy related code 90 days after a OMOP concept id related to abortion or miscarriage.

Within each pregnancy, the first codes recorded will be profiled and codes likely identifying pregnancy will be extracted.

The duration between codes indicating start of pregnancy and full-term delivery will be determined and profiled. In general, there is a lag time between conception and first medical appointment for pregnancy of about six to seven weeks [3], therefore, the calculated duration will be adjusted to include those days.

Codes indicating gestational age, postnatal care and miscarriage will be also profiled.

Pregnancies that match 280 +/- 7 days between codes indicating start of pregnancy and codes indicating full-term pregnancy, and pregnancies that match their reported gestational age +/- 7 days (e.g., if code was pregnancy at 38 weeks, a valid duration would be 266 +/- 7 days) will be included in the main cohort.

Where codes indicating delivery or end of pregnancy are reasonably consistent , even without a codes indicating start of pregnancy, pregnant women identified by these codes will be included in a sensitivity cohort.

Annex 2 – Codelists

Codelists for pregnancy-related codes are available on the Matcho *et al.*, PLoS One publication[3], as supplementary material , which are enclosed below as embedded files.

2018_Matcho_Inferring pregnancy episc

2018_Matcho_S1Table.xlsx

Codelist for infections were identified in the SNOMED CT (SNOMED CT Infectious disorders). These concepts will be matched to OMOP concept ids.

Codelist for the substances are available upon request given the large number of products included.