Thérapie 2014 Janvier-Février; 69 (1): 83–89
DOI: 10.2515/therapie/2014007

The Clinical Practice Research Datalink for Drug Safety in Pregnancy Research: an Overview

Rachel Charlton, Julia Snowball, Cormac Sammon and Corinne de Vries

Department of Pharmacy and Pharmacology, University of Bath, Bath, United Kingdom

Text received October 8th, 2013; accepted November 18th, 2013

Keywords:
electronic health records;
pregnancy;
pregnancy outcome;
congenital abnormalities;
teratogens

Abstract – Medicine use during pregnancy is common; however the safety of medicine use during pregnancy is largely unknown when a medicine comes to market. Electronic healthcare databases, including the Clinical Practice Research Datalink (CPRD), are increasingly being used for post-marketing surveillance in this field. The CPRD contains anonymised, longitudinal medical records routinely collected in primary care. Using CPRD data it is possible to identify medical records indicative of pregnancy, including pregnancy losses. Data on prescriptions issued can be used to determine maternal exposure and for about 80% of pregnancies it is possible to link the mother's medical record to the medical record of the child. Data in the medical records of the mother and child can then be used to identify adverse pregnancy outcomes, including congenital malformations. This paper describes some of the complexities involved in using CPRD data for pregnancy related research and discusses some of its strengths and limitations.

Mots clés :
dossiers de santé électroniques ;
grossesse ;
issues de grossesse ;
anormalités congénitales ;
teratogènes

Résumé – À propos de l'utilisation du *Clinical Practice Reseach Datalink* pour étudier la securité des médicaments pendant la grossesse. L'utilisation de médicaments pendant la grossesse est une pratique courante, néanmoins la sécurité de d'utilisation des médicaments chez la femme enceinte est largement inconnue lorsque un médicament est mis sur le marché. Les bases de données électroniques de santé, y compris la base de recherche clinique *Clinical Practice Research Datalink* (CPRD), sont de plus en plus utilisées pour la surveillance post-commercialisation dans ce domaine. Le CPRD contient des données médicales longitudinales anonymisées, recueillies lors des soins primaires. L'utilisation des données du CPRD, rend possible l'identification des dossiers médicaux indiquant une grossesse, y compris les interruptions de grossesse. Les données sur les ordonnances délivrées peuvent être utilisées pour déterminer l'exposition de la mère et pour environ 80 % des grossesses, il est possible de relier le dossier médical de la mère au dossier médical de l'enfant. Les données contenues dans les dossiers médicaux de la mère et de l'enfant peuvent être utilisées pour identifier les issues défavorables de la grossesse, y compris les malformations congénitales. Cet article décrit les difficultés liées à l'utilisation des données du CPRD pour la recherche liée à la grossesse et traite de certaines de ses forces et limites.

Abbreviations: see end of article.

1. Background

Medication use is common in women of child bearing age. Studies have shown that between 27% and 99% of women take some form of medicine during pregnancy,[1-3] with estimates varying depending on the country and calendar time of study, the types of products included and the specific pregnancy time period of interest. Pregnant women are often excluded from clinical trials for ethical reasons and this means there is little information available on a medicine's safety during pregnancy when a new medicine enters the market. As a result, post-marketing surveillance is required to monitor and evaluate the safety of these products when used by pregnant women.

Following the thalidomide disaster in the 1960s,[4] a number of post-marketing surveillance systems have been created to monitor the safety of medicine use during pregnancy. In recent years, the increase in routinely collected healthcare data has led to an increase in the use and number of electronic healthcare databases available for medication in pregnancy research. This paper aims to describe one electronic healthcare database, the Clinical Practice Research

Datalink (CPRD) and provide an overview of the complexities, as well as the strengths and limitations, of this data source as a tool for studying the safety of medicine use during pregnancy. This paper will also provide a summary of research in this field that has been carried out using the CPRD to date.

2. The Clinical Practice Research Datalink

The Clinical Practice Research Datalink, formerly the General Practice Research Database (GPRD), contains anonymised, longitudinal patient medical and prescribing records collected within United Kingdom general practice.[5] Data is entered into the database by general practice staff as part of routine patient management and includes information relating to pregnancy, symptoms and diagnoses, immunisations, consultations, tests, issued prescriptions and some information on lifestyle factors.[6] In the UK, the GP acts as the gatekeeper to services within the National health service and therefore some information relating to hospital or specialist referrals and admissions as well as outpatient and emergency visits may also be recorded. At present the CPRD captures ~8.5% of the UK population with ~5.4 million patients actively contributing data at any one time, in addition to over 7 million patients for whom historic data are available. The population captured by the CPRD is broadly representative of the UK population in terms of age and sex, although children aged 0-4 years and young adults (particularly males aged 17-30) are slightly under-represented.[7]

The recording of data from each GP practice is subject to quality control checks and each practice is assigned an "up-to-standard" (UTS) date, which is the date the database provider considered the practice to have started contributing data that is of a standard suitable for the purposes of research. The CPRD group has obtained ethical approval from a Multi-Centre Research Ethics Committee (MREC) for all purely observational research using CPRD data. For each study, additional approval is required from the Independent Scientific Advisory Committee (ISAC).[8] Although the CPRD enables links to other data sources (such as hospital episode statistics and cancer registries) this paper focuses primarily on data that is captured and recorded within UK general practice.

3. Identification of pregnancies

Data is largely entered into the CPRD in the form of medical Read codes and there are over 4 000 different Read codes that a GP can enter into a woman's medical record relating to pregnancy. These codes relate to pregnancy tests, the date of the first day of the last menstrual period (LMP), the estimated date of delivery, delivery bookings, antenatal care and pregnancy outcomes, as well as neonatal and postnatal care. Pregnancy care in the UK is led primarily by midwives and obstetricians although occasionally the GP will be the primary point of contact throughout pregnancy. Most details are recorded in a paper file the woman takes with her wherever she goes in the healthcare system during her pregnancy. Therefore, the level of detail recorded in the CPRD is largely dependent on how much of the paper records is transferred into the general practice computer system; this may be more in cases where the GP is the primary point of contact or at practices where there is a more established routine of recording all details on the computer. As a result the level of detail recorded varies between patients.

When carrying out research using CPRD data, most investigators identify the majority of pregnancies based on a Read code for a pregnancy outcome.[9,10] An advantage of the CPRD, compared to some other electronic healthcare databases, is that it captures all types of pregnancy outcome including live deliveries, stillbirths, induced terminations and spontaneous pregnancy losses, although very early spontaneous losses that occur before the pregnancy is clinically recognised are not captured. For some pregnancies identified in the CPRD, in addition to a medical code providing information on the date and type of pregnancy outcome, there is a medical code stating the date of conception or the date of the last menstrual period. For many pregnancies, however, accurate information on the beginning of pregnancy is not available and unlike databases such as "*Évaluation chez la Femme Enceinte des MÉdicaments et de leurs RISques*" (evaluation about drugs and their risks on pregnant women, EFEMERIS)[11] there is often little information on the gestational age at delivery from which this could be inferred. For these pregnancies it is therefore necessary to create an algorithm to try to estimate the date each pregnancy started by evaluating all the Read code entries within each woman's medical record and making assumptions based on the data available. For example, if when a woman becomes pregnant she visits her GP and he enters the expected date of delivery of the pregnancy it is possible to work back from this date to determine an estimate of the date the pregnancy started. Alternatively, if a woman had a record for an antenatal scan at 12 weeks' gestation an assumption can be made that the pregnancy started roughly 12 weeks before that date.[9]

Although the creation of an algorithm can be beneficial in determining the start and end dates of a pregnancy, sometimes there is conflicting or insufficient information available in a woman's medical record and assumptions need to be made regarding the duration of the pregnancy. For a delivery, these assumptions commonly take a defaulted pregnancy duration of between 270 and 280 days whilst a shorter duration of approximately 70 days has been taken for pregnancy losses.[9,12-15] Within the CPRD there are some deliveries where there is an indication that the pregnancy was premature (for example the Read code states "baby premature") but there is no further information relating to the gestational age; here an alternative, shorter defaulted duration of 35 or 36 weeks may be assigned.[9] Assigning defaulted durations is one way to ensure a reasonable estimate of a pregnancy start date is established for all pregnancies, however for those that end in a premature or post-mature delivery, where there is no coded evidence to indicate the delivery did not occur at term, assigning a defaulted duration will result in an over or under-estimate of the duration of pregnancy.[16]