Endometriosis Phenotype Characterization Across Observational Health Databases Protocol

Authors:
Mollie McKillop, MA, MPhil, MPH, Columbia University
Sharon Gorman, MS, Columbia University
Anastasios Siapos, IQVIA
Chris Knoll, Jansen
Patrick Ryan, PhD, Jansen
Noémie Elhadad, PhD, Columbia University

Date: 30 Nov 2018

Acknowledgement: The analysis is based in part on work from the Observational Health Sciences and Informatics collaborative. OHDSI (http://ohdsi.org) is a multi-stakeholder, interdisciplinary collaborative to create open-source solutions that bring out the value of observational health data through large-scale analytics.

Contents

Background

Endometriosis is a chronic disease in reproductive age-women in which endometrial-like cells grow outside of the uterus. It is characterized through surgical findings, with menstrual pain and infertility described as typical symptoms[1]. The disease is highly prevalent (estimated to affect 1 in 10 women) but also highly enigmatic. There is no known biomarker, the etiology of the disease remains unknown, and there is wide variation in responses to treatments amongst patients[2]. As a result there is a long lag to diagnosis, and on average it takes 10 years to be diagnosed[3]. There exists evidence that the disease is a systemic condition with symptoms beyond dysmenorrhea and infertility, and the natural history of the disease is not well understood[4]. Because of misdiagnosis and under-diagnosis of the disease, there is a need to identify disease patterns prior to diagnosis. As a first step toward better understanding the disease we would like to better characterize it before disease diagnosis using a well-validated disease phenotype. This phenotype was developed at our institution using over 1400 manual chart review annotations of endometriosis patients and achieved high precision and recall. Observational data in the form of electronic health records and insurance claims data have the potential to fill this knowledge gap through retrospective analysis of patients characteristics prior and at the time of disease diagnosis. Specifically, we wish to explore the prevalence of signs and symptoms, treatments, and healthcare utilization patterns in those who will get the disease and for those in a comparison cohort.

Objective

The goal of this study is to understand the prevalence of signs and symptoms, treatments, and healthcare utilization patterns among endometriosis patients before diagnosis and relate these patterns to a comparison cohort.

Data sources

The analyses will be performed across a network of observational healthcare databases. All databases have been transformed into OMOP Common Data Model, version 5. The complete specification for OMOP Common Data Model, version 5 is available at: https://github.com/OHDSI/CommonDataModel. The following databases will be included in this analysis:

- Columbia University Irving Medical Center Data Warehouse (Columbia EHR)
- IBM MarketScan® Commercial Database (MCD)
- IBM MarketScan® Multi-State Medicaid Database (MMMD)
- Optum® Clinformatics® Extended DataMart (Optum)
- IQVIA US Ambulatory EMR
- IQVIA Hospital US Charge Master
- IQVIA Open Claims Non-adjudicated Claims

Columbia University Irving Medical Center Data Warehouse (Columbia EHR)

The Columbia EHR contains electronic health record data extracted from a clinical transaction-based data repository serving Columbia University Irving Medical Center and New York-Presbyterian Hospital. This data source includes clinical observational health data for over 4 million patients including, demographics, inpatient and outpatient encounters, condition occurrences, prescribed medications, procedures, laboratory tests, and clinical measurements.

IBM MarketScan® Commercial Database (MCD)

The MCD database contains data from individuals enrolled in United States employer-sponsored insurance health plans. The database captures administrative claims for inpatient, outpatient visit and pharmacy and laboratory data for a privately insured population.

IBM MarketScan® Multi-State Medicaid Database (MMMD)

The MMD database contains administrative claims data for Medicaid enrollees from multiple states. The data includes administrative claims for inpatient and outpatient, as well as outpatient pharmacy dispensing.

Optum® Clinformatics® Extended DataMart (Optum)

The Optum database contains administrative claims data for members with private health insurance, who are fully insured in commercial plans or in administrative services only, Legacy Medicare Choice prior to January 2006, and Medicare Advantage Prescription Drug coverage starting January 2006. It includes inpatient and outpatient medical services, prescriptions as dispensed, as well as results for outpatient laboratory tests processed by large national lab vendors who participate in data exchange with Optum.

IQVIA US Ambulatory EMR

The IQVIA US Ambulatory EMR contains outpatient data for the general population. It sources data from 2006 until May 2018. Anonymized patient records are collected from patient management software used by general practitioners documenting patients' clinical records.

IQVIA Hospital US Charge Master

The IQVIA Hospital US Charge Master contains data for inpatient and outpatient hospital encounters in the general population, including emergency room visits. It sources data from 2001 until January 2018. Anonymized patient level data are sourced from hospital charge detail masters and collected from resource management software within short-term, acute-care and non-federal hospitals.

IQVIA Open Claims Non-adjudicated Claims

The IQVIA Open Claims Non-adjudicated Claims contains data for institutional and outpatient populations that are insured. It sources data from 1997 until July 2018. It contains pre-adjudicated claims at the anonymized patient-level collected from office-based physicians and specialists via office management software and clearinghouse switch sources for the purpose of reimbursement.

PharMetrics Plus Adjudicated Claims

PharMetrics Plus Adjudicated Claims contains data for commercially-insured populations (medical, pharmacy and enrollment). It sources data from 2007 until March 2018. It is a patient-centric, closed claims database of fully adjudicated pharmacy, hospital and medical claims at the anonymized patient-level sourced from commercial payers.

Data collection

Data collection will be performed by manually running several SQL queries and exporting results to CSV files. This code will be distributed to the data partners, executed locally against data in OMOP CDM format, and results will be returned to the central coordinating site (Columbia University Department of Biomedical Informatics).

Population

Included in the analysis are those who have an initial event of either: 1) an endometriosis-related laparoscopy procedure AND an endometriosis diagnosis 30 days before or after the procedure, or 2) an endometriosis-prevalent procedure AND an endometriosis diagnosis 30 days before or after the procedure. Inclusion criteria consist of having a second endometriosis diagnosis after this initial event as well as being female and between 15 and 49 at the time of the initial event. This will be compared to a cohort of women who are reproductive age (15-49) who do not have endometriosis.

Methods

1. Set your database, server, port, user & password, this example assumes you have set your environment variables to the requried values.

   connectionDetails <- DatabaseConnector::createConnectionDetails(dbms = Sys.getenv("dbms"), server = Sys.getenv("server"), port = as.numeric(Sys.getenv("port")), user = Sys.getEnv("username"), password = Sys.getEnv("password")

2. Run the package to create the cohorts

   library(createCohorts)

3. Run the package to generate the prevalence counts and output to study files.

   library(main)

4. Please email results to Mollie McKillop at mm4234@cumc.columbia.edu

References

1. Vercellini P, Fedele L, Aimi G, Pietropaolo G, Consonni D, Crosignani PG. Association between endometriosis stage, lesion type, patient characteristics and severity of pelvic pain symptoms: a multivariate analysis of over 1000 patients. Human Reproduction. 2006 Aug 26;22(1):266-71.

2. Acién P, Velasco I. Endometriosis: a disease that remains enigmatic. ISRN obstetrics and gynecology. 2013 Jul 17;2013.

3. Hadfield R, Mardon H, Barlow D, Kennedy S. Delay in the diagnosis of endometriosis: a survey of women from the USA and the UK. Human Reproduction. 1996 Apr 1;11(4):878-80.

4. McKillop M, Mamykina L, Elhadad N. Designing in the Dark: Eliciting Self-tracking Dimensions for Understanding Enigmatic Disease. InProceedings of the 2018 CHI Conference on Human Factors in Computing Systems 2018 Apr 21 (p. 565). ACM.