Endometriosis Phenotype Development, Validation, and Characterization from Observational Health Databases

Mollie McKillop¹, Sharon Lipsky Gorman¹, Shadi Safar Goli¹, Chris D’ambrosia¹, Christopher Knoll², Patrick Ryan², Noémie Elhadad¹

¹Columbia University; ²Janssen Pharmaceuticals
Endometriosis Phenotype Development, Validation, and Characterization from Observational Health Databases

Mollie McKillop¹, Sharon Lipsky Gorman¹, Shadi Safar Goli¹, Chris D’ambrosia¹, Christoper Knoll², Patrick Ryan², Noémie Elhadad¹

¹Columbia University; ²Janssen Pharmaceuticals
**Endometriosis**

- Chronic disease in reproductive-age women
  - Endometrial cells grow outside uterus
  - Characterized through surgical findings
  - Menstrual pain and infertility as the most common characterization

- Prevalent
  - Estimated to ~10% of women in reproductive age

- Highly enigmatic
  - No known biomarker, etiology, or treatment response
  - Long lag to diagnosis (~10 years)
Research gaps

• Incomplete characterization
  • Systemic condition with symptoms beyond dysmenorrhea and infertility
  • From onset of symptoms to after diagnosis
  • Because misdiagnoses/under-diagnosis, need to identify the patterns of endo patients prior to diagnosis towards earlier/better detection

• Lack of validated phenotype
  • Epidemiological studies rely on single high-level ICD code
  • Types of endometriosis have been proposed but focus on surgical findings and do not correlate with patient experience of disease
Research questions

1. What is an accurate phenotype for endometriosis to identify cohorts from EHR and claims databases?

2. What are the patterns of patients' experiences before diagnosis?
   - Signs/symptoms, treatments, healthcare utilization patterns, etc.

   • Additional desiderata:
     - Phenotype valid across claims and EHR databases to identify a wide range of patients
     - Open access to other OHDSI members
Methods

1. Define and validate an endometriosis phenotype for EHRs / claims databases

2. Characterize cohorts pre-diagnosis across databases
Relevant concepts

- **Endo diagnosis**: endometriosis diagnosis codes (including adenomyosis)
- **Endo-related procedures**: guideline-based procedures for endometriosis diagnosis and treatment (e.g., pelvic laparoscopy)
- **Endo-prevalent procedures**: procedures present >50% of patients with ≥1 endo diagnosis
- **Endo-related imaging procedures**: guidelines-based imaging procedures (e.g. pelvic MRI)
Cohort definition experiments

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Initial Event</th>
<th>Inclusion Rules</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>endo-related procedures AND endo diagnosis OR endo-prevalent procedures AND endo diagnosis</td>
<td>Females ages 15-49</td>
</tr>
<tr>
<td>B</td>
<td>endo-related procedures AND endo diagnosis OR endo-prevalent procedures AND endo diagnosis</td>
<td>Females ages 15-49 AND 2 endo diagnosis after index date</td>
</tr>
<tr>
<td>C</td>
<td>endo-related procedures AND endo diagnosis AND endo-related imaging procedures AND endo diagnosis before index date</td>
<td>Females ages 15-49</td>
</tr>
<tr>
<td>D</td>
<td>endo-related procedures AND endo diagnosis AND endo-related imaging procedures AND endo diagnosis before index date</td>
<td>Females ages 15-49 AND 2 endo diagnosis after index date</td>
</tr>
</tbody>
</table>
Cohort definition validation

• Manual chart review from clinical experts on EHR database (Columbia)
  • Index data between 1/1/2016 to 6/1/2018
  • Reviewed records of patients across all cohort definitions
  • Confirmed endometriosis diagnosis through histological analysis post-laparoscopy

• Gold-standard annotations of 1,406 patient records
  • Two annotators
  • Kappa on 38 records: .89*
    • Determined N=38 provides expected confidence limits between .6 and 1

* p-value <.05
## Cohort definition validation

- **Cohort B** had highest precision/recall
  - Negative predictive value 0.84

<table>
<thead>
<tr>
<th>Cohort name on OHDSI.org</th>
<th>Precision</th>
<th>Recall</th>
<th>Records in cohort at Columbia EHR 2016-2018*</th>
<th>Total records in cohort at Columbia EHR 1999-2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cohort D</td>
<td>0.84</td>
<td>0.26</td>
<td>162</td>
<td>1,248</td>
</tr>
<tr>
<td>Cohort C</td>
<td>0.78</td>
<td>0.28</td>
<td>189</td>
<td>1,950</td>
</tr>
<tr>
<td><strong>Cohort B</strong></td>
<td><strong>0.85</strong></td>
<td><strong>0.70</strong></td>
<td><strong>430</strong></td>
<td><strong>3,328</strong></td>
</tr>
<tr>
<td>Cohort A</td>
<td>0.37</td>
<td>1**</td>
<td>1,406</td>
<td>5,666</td>
</tr>
</tbody>
</table>

* each of these cohorts was reviewed by clinical experts.
** all other cohorts are a subset of this cohort so recall=1.
Cohort definitions

• Available at:
  - http://www.ohdsi.org/web/atlas/#/cohortdefinition/1769393
  - http://www.ohdsi.org/web/atlas/#/cohortdefinition/1769395
  - http://www.ohdsi.org/web/atlas/#/cohortdefinition/1769396
  - http://www.ohdsi.org/web/atlas/#/cohortdefinition/1769397

Best performing cohort we used to characterize
Methods

1. Define and validate an endometriosis phenotype for EHRs / claims databases

2. Characterize cohorts pre-diagnosis across databases
Cohort characterization

• Select best-performing cohort, Atlas characterization

• Analysis carried out on four databases
  • Columbia EHR (Columbia EHR)
  • Optum® Clinformatics® Extended DataMart (Optum)
  • IBM MarketScan® Commercial Database (MCD)
  • IBM MarketScan® Multi-State Medicaid Database (MMMD)

• Report on prevalent (>10% of patients) conditions and drugs

• For comparison, prevalence in general cohort of women of reproductive age
Cohort characterization

• Overall >480,000 patients in endometriosis cohort

<table>
<thead>
<tr>
<th>Database</th>
<th>Total records in cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Columbia EHR</td>
<td>3,328</td>
</tr>
<tr>
<td>Optum</td>
<td>24,725</td>
</tr>
<tr>
<td>MMMD</td>
<td>54,609</td>
</tr>
<tr>
<td>MCD</td>
<td>398,015</td>
</tr>
</tbody>
</table>
Conditions

Database
- MMMD
- Optum
- MCD
- Columbia EHR

OHDSI Call – Oct 30, 2018
Comparison (not control) Group Conditions

OHDSI Call – Oct 30, 2018
Drugs

- Anilides
- Natural opium alkaloids
- Propanoic acid derivatives
- Serotonin (5HT3) antagonists
- Opium alkaloids and derivatives
- Proton pump inhibitors
- Benzodiazepine related drugs
- Prostaglandins and estrogens
- Pencillin with extended spectrum
- Other drugs for gastrointestinal disorders
- Macrolides
- Contact laxatives
- Fluoroquinolones
- Vitamin D and analogues
- H2-receptor antagonists

ATC Class 5

OHDSI Call – Oct 30, 2018
Discussion

• First study to develop well-validated endometriosis phenotype for cohort selection

• Findings across databases consistent with ongoing endometriosis research
  • Primary symptoms present prior to diagnosis related to pelvic pain and heavy pain medication

• Results congruent with new knowledge about endometriosis
  • Beyond dysmenorrhea and pelvic pain, systemic impact of disease (i.e. anxiety, constipation)

• OHDSI essential for characterization
  • Wouldn’t have started to get full picture w/o collaborators!
Next steps

• Characterize in larger, more diverse cohort
  • Get in touch if you’d like to participate
  • All definitions available on ohdsi.org and further queries available on demand

• Use phenotype definition for patient-level prediction
  • In women in reproductive age presenting in ED with abdominal pain, who is likely to be diagnosed with endometriosis 1 year+ later?
  • Using PLP modules
  • Get in touch if you’d like to participate
Thanks, questions?

- Columbia OHDSI bootcamp participants
Comparison (not control) Group Conditions