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

Mollie McKillop[1], Sharon Lipsky Gorman[1],
Shadi Safar Goli[1], Chris D'ambrosia[1], Christoper Knoll[2],
Patrick Ryan[2], Noémie Elhadad[1]

[1]Columbia University; [2]Janssen Pharmaceuticals

OHDSI

Columbia University
Irving Medical Center

Looking for OHDSI collaborators

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

Mollie McKillop[1], Sharon Lipsky Gorman[1],
Shadi Safar Goli[1], Chris D'ambrosia[1], Christoper Knoll[2],
Patrick Ryan[2], Noémie Elhadad[1]

[1]Columbia University; [2]Janssen Pharmaceuticals

OHDSI

Columbia University
Irving Medical Center

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

Cohort	Initial Event	Inclusion Rules
A	**endo-related procedures** AND **endo diagnosis** OR **endo-prevalent procedures** AND **endo diagnosis**	Females ages 15-49
B	**endo-related procedures** AND **endo diagnosis** OR **endo-prevalent procedures** AND **endo diagnosis**	Females ages 15-49 AND 2 **endo diagnosis** after index date
C	**endo-related procedures** AND **endo diagnosis** AND **endo-related imaging procedures** AND **endo diagnosis** before index date	Females ages 15-49
D	**endo-related procedures** AND **endo diagnosis** AND **endo-related imaging procedures** AND **endo diagnosis** before index date	Females ages 15-49 AND 2 **endo diagnosis** after index date

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

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

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

Cohort definitions

Best performing cohort we used to characterize

- 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

Methods

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

2. Characterize cohorts pre-diagnosis across databases

Cohort characterization

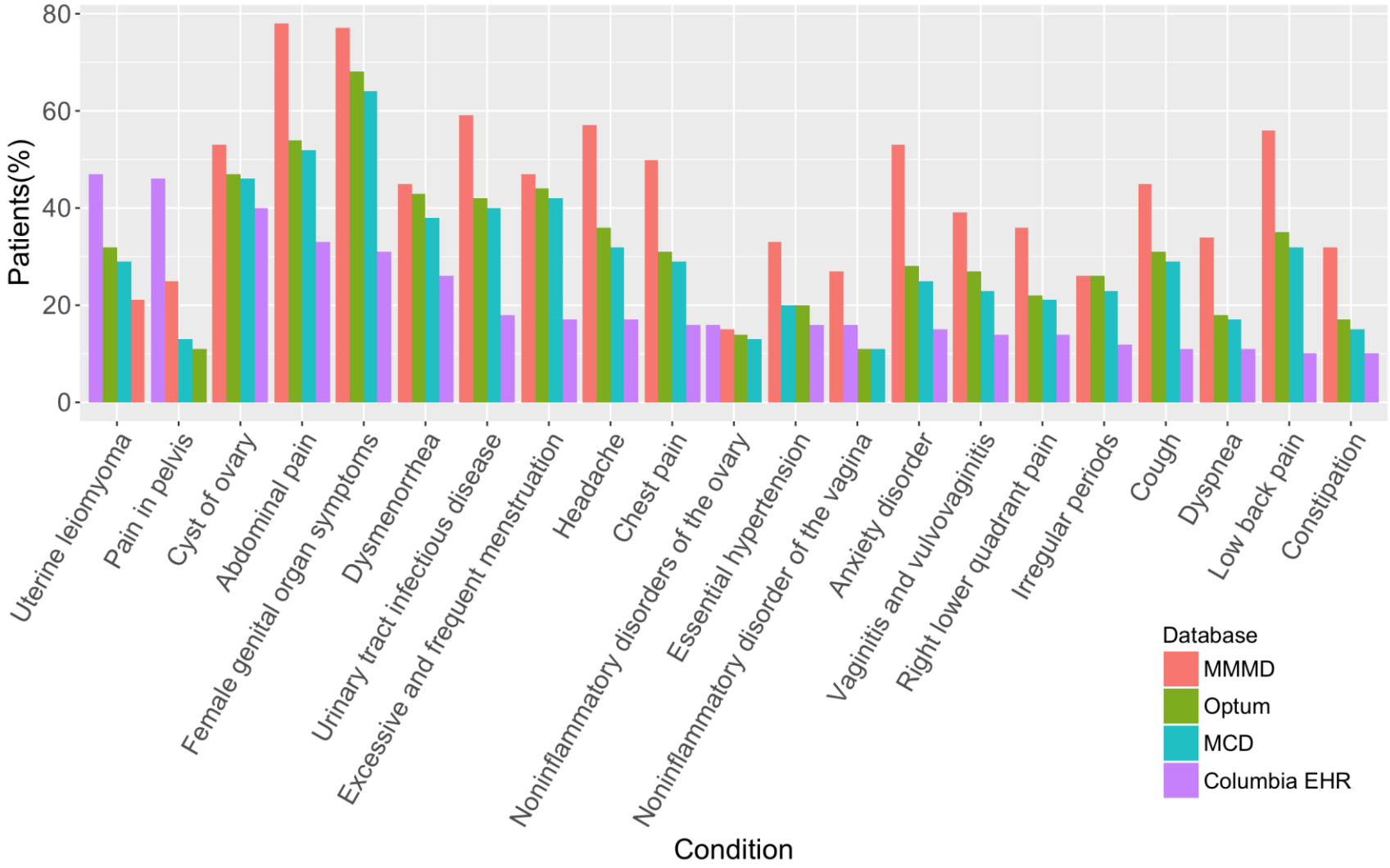
- Select best-performing cohort, Atlas characterization
- Analysis carried out on four databases
  - Columbia EHR (**Columbia EHR**)
  - Optum® Clinformatics® Extended DataMart (**Optum**)
  - IBM MarketScan® Commercial Database (**MCD**)
  - IBM MarketScan® Multi-State Medicaid Database (**MMMD**)
- Report on prevalent (>10% of patients) conditions and drugs
- For comparison, prevalence in general cohort of women of reproductive age

Cohort characterization

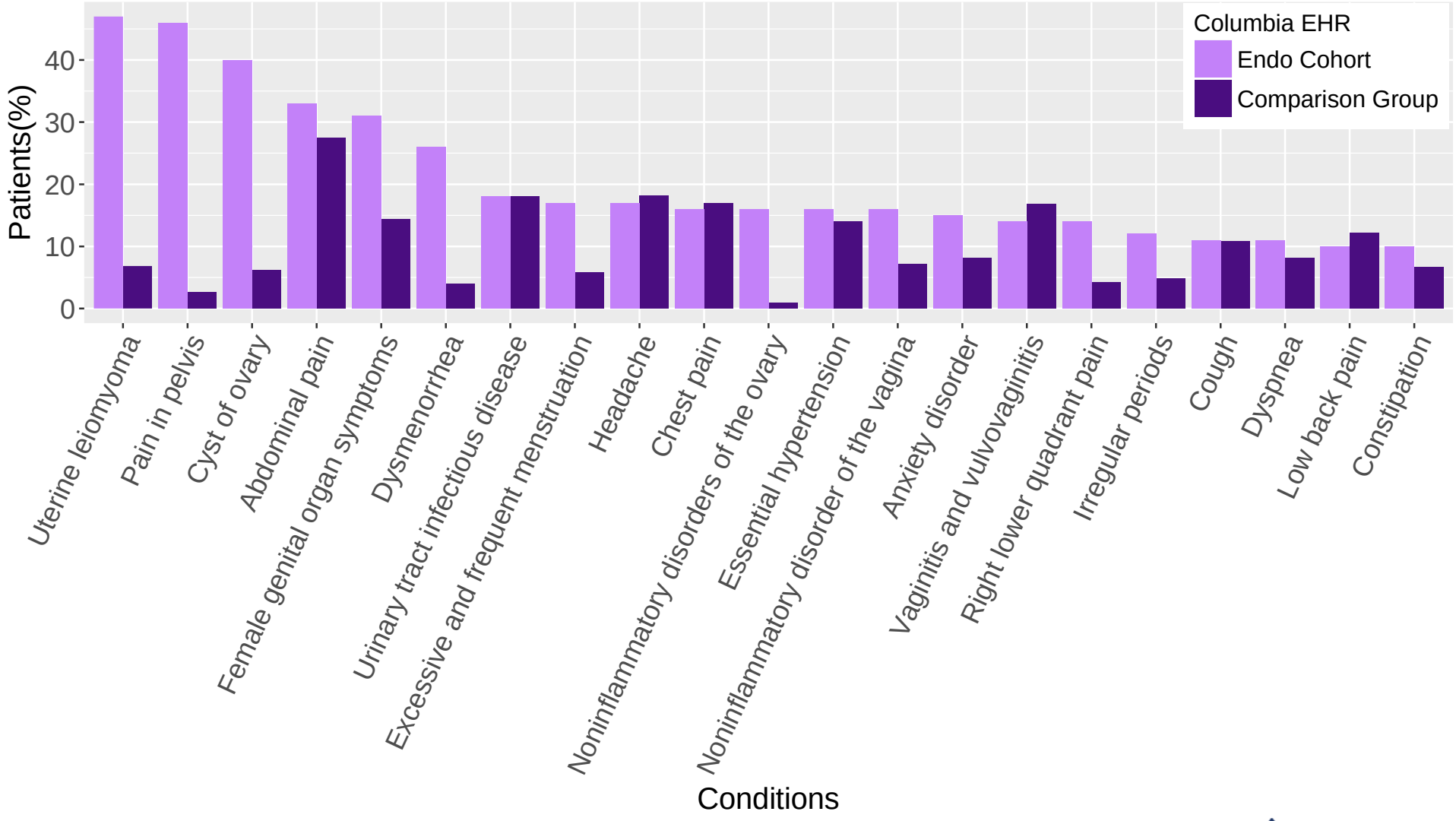
- Overall >480,000 patients in endometriosis cohort

Database	Total records in cohort
Columbia EHR	3,328
Optum	24,725
MMMD	54,609
MCD	398,015

Conditions


Patients(%)
80
60
40
20
0
Uterine leiomyoma
Pain in pelvis
Cyst of ovary
Abdominal pain
Female genital organ symptoms
Dysmenorrhea
Urinary tract infectious disease
Excessive and frequent menstruation
Headache
Chest pain
Noninflammatory disorders of the ovary
Essential hypertension
Noninflammatory disorder of the vagina
Anxiety disorder
Vaginitis and vulvovaginitis
Right lower quadrant pain
Irregular periods
Cough
Dyspnea
Low back pain
Constipation
Condition
Database
MMMD
Optum
MCD
Columbia EHR

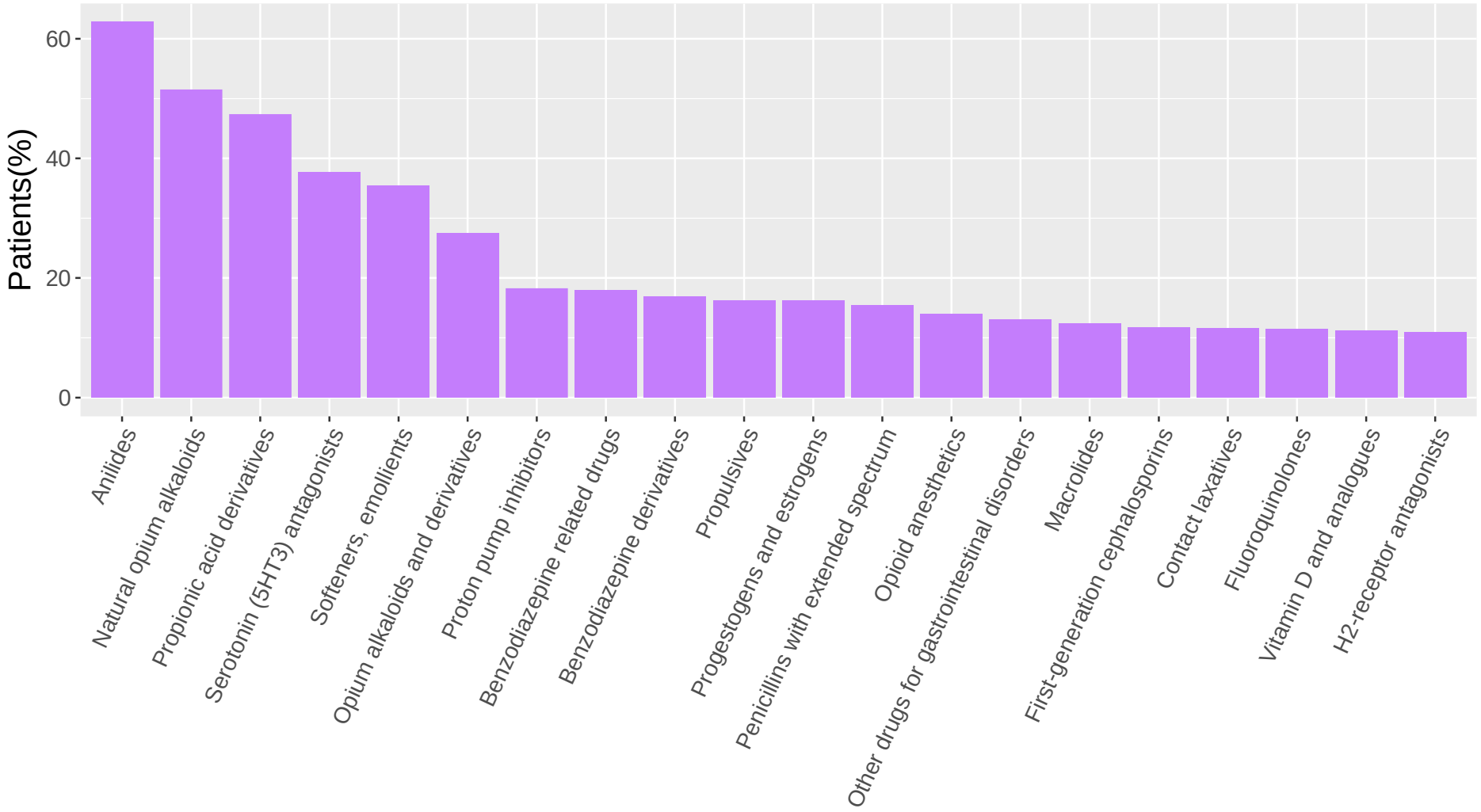
Comparison (not control) Group Conditions



OHDSI

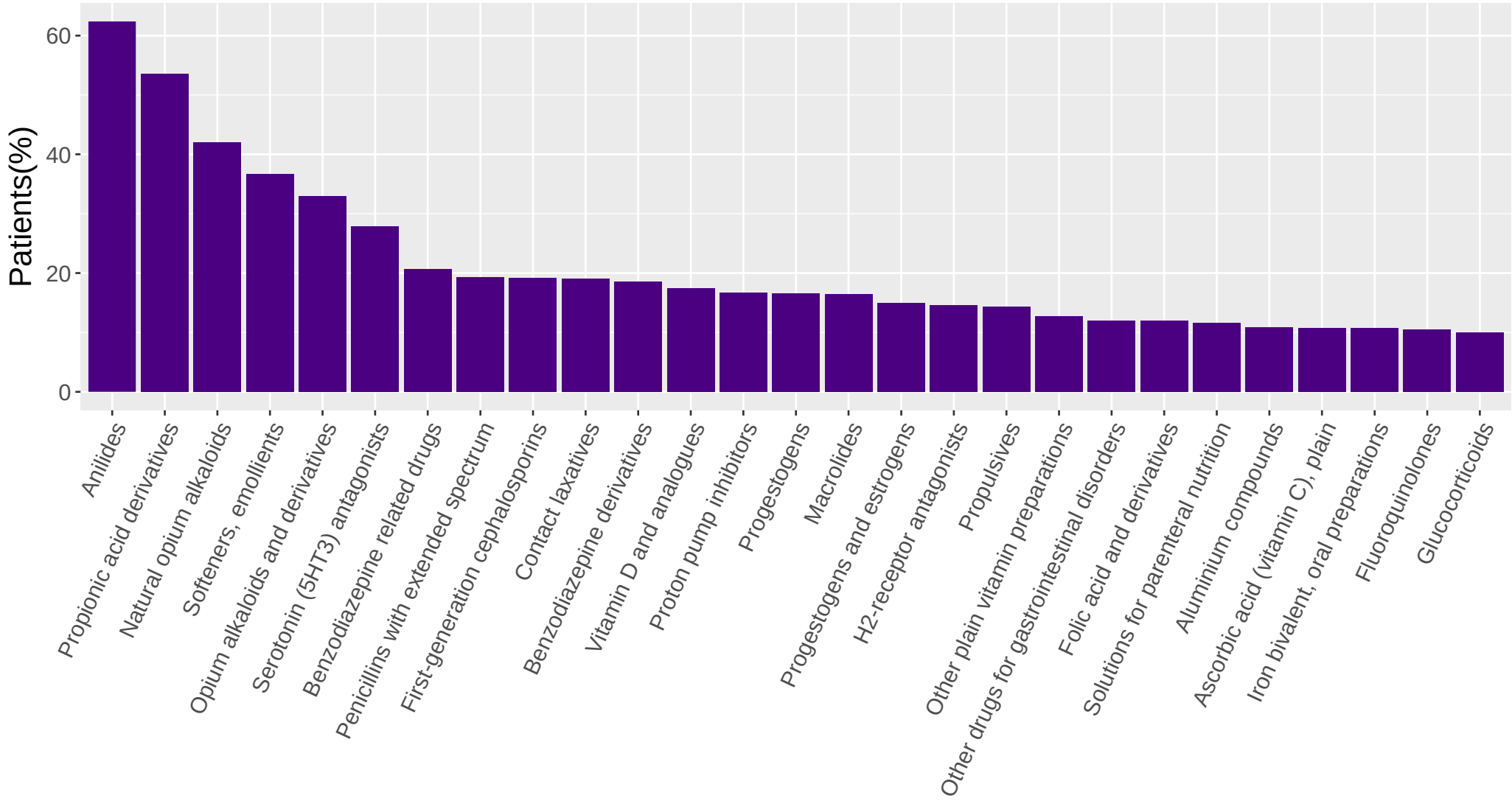
Columbia University Irving Medical Center

Drugs



OHDSI
COLUMBIA UNIVERSITY IRVING MEDICAL CENTER

Comparison Group Drugs



ATC Class 5

OHDSI

COLUMBIA UNIVERSITY
IRVING MEDICAL CENTER

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!

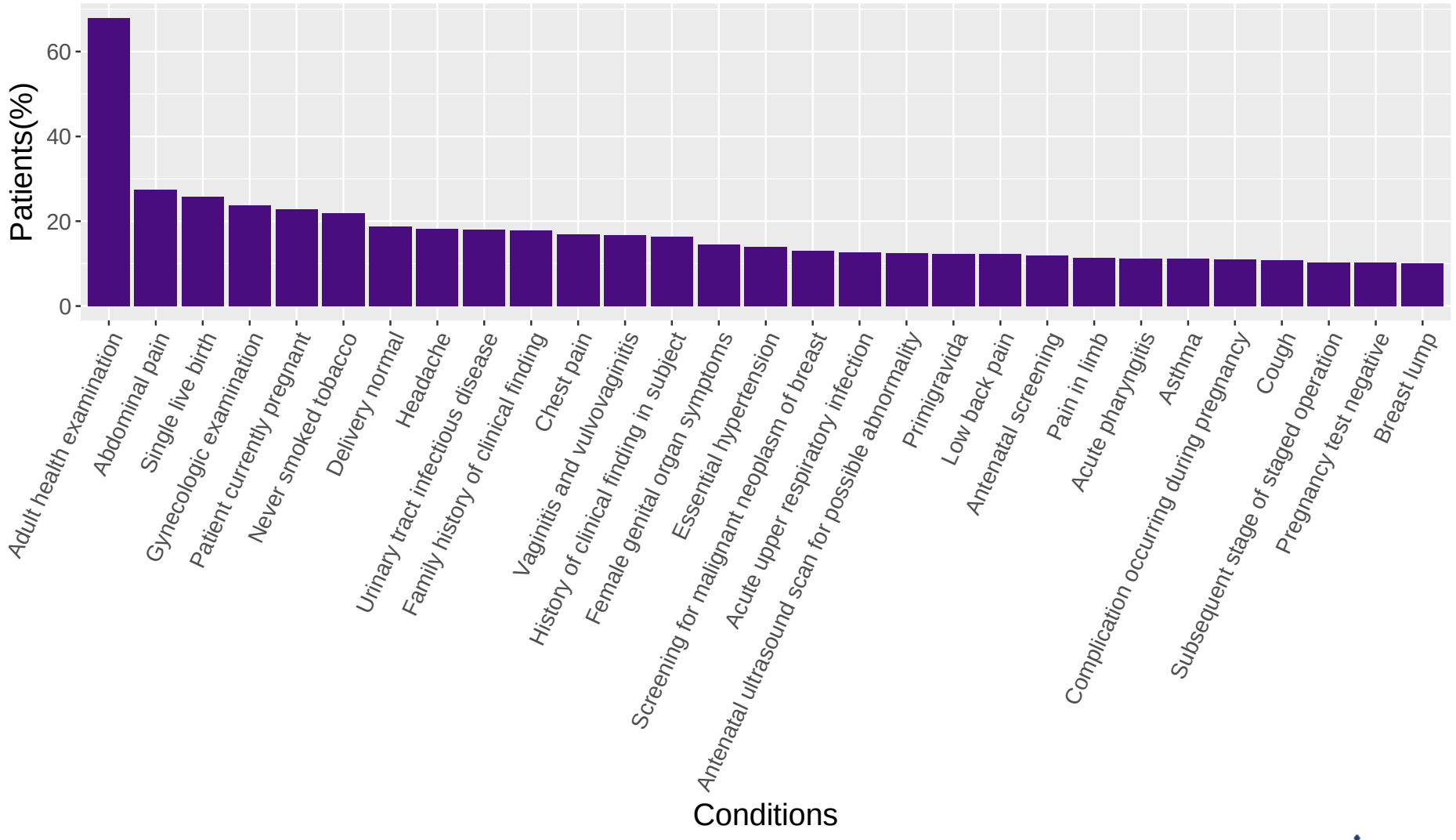
Next steps

- 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 1year+ later?
  - Using PLP modules
  - Get in touch if you'd like to participate

Thanks, questions?

- Columbia OHDSI bootcamp participants

Comparison (not control) Group Conditions



OHDSI
Columbia University Irving Medical Center