ORIGINAL ARTICLE

Characteristics and Outcomes of Over a Million Patients with Inflammatory Bowel Disease in Seven Countries: Multinational Cohort Study and Open Data Resource

Chen Yanover[1,2] · Ramit Magen-Rimon[1,3] · Erica A. Voss[1,4] · Joel Swerdel[1,4] · Anna Sheahan[1,4] · Nathan Hall[1,4] · Jimyung Park[1,5,6] · Rae Woong Park[1,6] · Kwang Jae Lee[1,7] · Sung Jae Shin[1,7] · Seung In Seo[1,8] · Kyung-Joo Lee[1,9] · Thomas Falconer[1,5] · Leonard Haas[1,10] · Paul Nagy[1,10] · Mary Grace Bowring[1,10] · Michael Cook[1,10] · Steven Miller[1,10] · Tal El-Hay[1,2] · Maytal Bivas-Benita[1,2] · Pinchas Akiva[1,2] · Yehuda Chowers[1,11] · Roni Weisshof[1,11]

Abstract

Background and Aims Observational healthcare data are an important tool for delineating patients' inflammatory bowel disease (IBD) journey in real-world settings. However, studies that characterize IBD cohorts typically rely on a single resource, apply diverse eligibility criteria, and extract variable sets of attributes, making comparison between cohorts challenging. We aim to longitudinally describe and compare IBD patient cohorts across multiple geographic regions, employing unified data and analysis framework.

Methods We conducted a descriptive cohort study, using routinely collected healthcare data, from a federated network of data partners in sixteen databases from seven countries (USA, UK, France, Germany, Japan, Korea, and Australia); and computed the prevalence of thousands of attributes, across multiple baseline and follow-up time windows, for full disease cohorts and various strata.

Results Characterizing the disease trajectory of 462,502 Crohn's disease (CD) and 589,118 ulcerative colitis (UC) subjects, we observed a decline over time in the average age at CD diagnosis in Europe and North America but less pronounced shifts in Japan and Korea; an uptick in the proportion of patients with anxiety diagnosis prior to CD diagnosis in European and US datasets; and stable rates of segmental colonic and small bowel resections within one and three years following UC and CD diagnosis, respectively, in most US databases.

Conclusions The study provides a comprehensive characterization of IBD patient cohorts from various countries including insights into disease trends, demographics, and pre-diagnosis symptoms. All characteristics and outcomes are publicly available, providing an unprecedented, comprehensive open resource for clinicians and researchers.

Keywords Crohn's disease · Ulcerative colitis · Routinely collected health data · Cohort study

✉ Chen Yanover
yanover@ohdsi.org

1. OHDSI Collaborators, Observational Health Data Sciences and Informatics (OHDSI), New York, NY, USA

2. KI Research Institute, 11 haZait St., Kfar Malal, Israel

3. Faculty of Medicine, Ruth Children's Hospital of Haifa, Rambam Medical Center, Pediatric Gastroenterology & Nutrition Institute, Technion, Haifa, Israel

4. Janssen Research & Development LLC, Raritan, NJ, USA

5. Department of Biomedical Informatics, Columbia University Irving Medical Center, New York, NY, USA

6. Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, Korea

7. Department of Gastroenterology, Ajou University School of Medicine, Suwon, Korea

8. Department of Internal Medicine, Kangdong Sacred Heart Hospital, Hallym University College of Medicine, Seoul, Korea

9. Kangdong Sacred Heart Hospital, Seoul, Korea

10. Division of Pediatric Gastroenterology, Hepatology and Nutrition, Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, MD, USA

11. Department of Gastroenterology, Rambam Health Care Campus and Bruce Rappaport School of Medicine, Technion Israel Institute of Technology, Haifa, Israel

Introduction

Crohn's disease (CD) and ulcerative colitis (UC) are chronic inflammatory bowel diseases (IBD) with consistently increasing incidence rates in both newly industrialized and developed countries [1–3]. IBD has a significant impact on the patients' and their families' quality of life, imposing a substantial healthcare financial burden due to chronic medication use, hospitalizations, and surgical procedures [2]. Population-based studies are an important tool for exploring the epidemiology and identifying risk factors as well as complications associated with IBD. Such studies analyze large datasets to determine the prevalence and incidence of IBD and identify trends and patterns, elucidate the natural history and long-term outcomes of the disease, and identify environmental and genetic determinants associated with IBD. Typically, however, these efforts focus on data sources from a limited geographical area, implement diverse algorithms to identify patient cohorts, and extract different sets of attributes [4, 5]. Consequently, integrating the reported findings may be challenging.

The Observational Health Data Sciences and Informatics (OHDSI) community (http://www.ohdsi.org/) is a global, collaborative network of clinicians, researchers, and data scientists whose mission is to improve the use of observational health data for research and healthcare decision-making [6]. The OHDSI community develops and maintains a common data model, standardized vocabulary and terminology systems, and software tools for data analysis and visualization. Harmonizing an increasingly large number of data sources to employ standardized data structure, content, and semantics, allows systematic application of research methods across different healthcare settings and geographies, permitting production of comparable and reproducible results in large-scale, multi-national studies [7, 8]. A unique advantage of such efforts is the ability to identify region-specific disease characteristics on the one hand, and unifying disease motifs on the other.

We describe an open OHDSI network study that characterizes IBD patient cohorts to better understand their natural history, risk factors, symptoms, associated comorbidities (including malignancies and extra-intestinal manifestations), treatment pathways, and outcomes. We compute the prevalence and aggregate counts of thousands of attributes, in multiple baseline and follow-up time windows, for IBD, CD, UC, and unclassified IBD (IBDU) patient cohorts as well as various strata (e.g., by age group, gender, race). Characterization results for approximately 1.2 million IBD patients in sixteen databases from seven countries—USA, UK, France, Germany, Japan, Korea, and Australia—are freely available in an interactive website (and in downloadable formats) and can serve as a resource

for further exploration and analysis. Below we discuss representative analyses to confirm the validity of the presented resource and demonstrate its utility.

Methods

Study Design and Setting

We characterized the disease trajectory of patients with IBD in a descriptive cohort study, using routinely collected healthcare data, from a federated network of data partners in the USA, Europe (France, Germany, and the UK), Asia (South Korea and Japan), and Australia. Standardizing data sources to the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) [6] allowed all contributing partners to locally run the same open-source analysis package and, upon completion, share only aggregated statistics.

The study protocol and analysis package are available through GitHub (https://github.com/ohdsi-studies/IbdCharacterization). All the data partners obtained Institutional Review Board (IRB) or equivalent governance approval and consented to sharing the results on the OHDSI Analysis Viewer (https://data.ohdsi.org/).

Data Sources

We report IBD characterization statistics from primary care, outpatient, and inpatient healthcare data in sixteen databases (DBs):

- Nine DBs from the USA:

  o Insurance claims data from Merative™ MarketScan® Commercial Database (CCAE), Merative™ MarketScan® Multi-State Medicaid Database (MDCD), Merative™ MarketScan® Medicare Database (MDCR), Optum's de-identified Clinformatics® Extended Data Mart (Optum-CDM), and IQVIA™ Adjudicated Health Plan Claims (PharMetrics+)
  o Electronic health records (EHRs) from Optum® de-identified Electronic Health Record (Optum-EHR), Columbia University Irving Medical Center (CUIMC), and Johns Hopkins Medicine (JHM)
  o Both claims and EHR data from IQVIA Adjudicated Health Plan Claims Data (AMB-EMR)

- Three DBs from Europe:

  o EHRs from IQVIA™ Disease Analyzer – France, IQVIA™ Disease Analyzer – Germany, and

IQVIA™ Medical Research Data – UK (IMRD-UK; THIN, A Cegedim Database)

- Three DBs from Asia:

  o Insurance claims from Japan in JMDC
  o South Korean EHRs from Ajou University School of Medicine (AUSOM) and Kangdong Sacred Heart Hospital (KDH)

- One Australian DB:

  o IQVIA™ Australian Longitudinal Patient Data (Australia)

For more details and database characteristics, see Supplementary Methods 1.

Study Participants and Follow-Up

IBD cohorts included individuals with at least two diagnoses of IBD or with an IBD diagnosis and a prescription for an IBD medication (conceptually following Friedman et al. [4]; for a list of medications, see the project GitHub repository: https://github.com/ohdsi-studies/IbdCharacterization). CD and UC cohorts further required at least one diagnosis of the corresponding disease and none of the other. IBD unclassified (IBDU) cohorts include patients with IBD (as defined above), who had diagnoses of both CD and UC, or neither. In this report we focus on incident cohorts, which also required that individuals have a minimum observation of 365 days prior to the first IBD-related diagnosis or prescription (index date).

For CD and UC, we case-reviewed, in the IMRD-UK database, random samples of one hundred adults and fifty children and observed positive predictive values ≥ 85%. In addition, we validated the robustness of the unified cohort definitions by comparing the aggregate patient data from the JHU OMOP CDM to Epic's SlicerDicer EHR reporting tool. Overall, cohort identification was similar between OMOP and SlicerDicer [9]. Local validation of data extraction through SlicerDicer queries can allow for rapid validation of queries for big data research. Case reviews of patient cohorts in the remaining DBs were infeasible due to the nature of the study; indeed, one of the goals of this work is to confirm the validity of the collected data.

Each target cohort was analyzed in full and sub-grouped based on follow-up time (1, 3, 5, and 10 years), sex, specific age groups (very early onset: 2–5 years; early onset: 6–9 years; pediatrics: 10–17 years; adults: 18–64 years; elderly: 65 years and older), race (black or African American, white, Asian), index date year (in 5-year windows), body mass index (BMI) categories (Underweight: BMI less than 18.5 kg/m^2; Normal weight: BMI between 18.5 and 25 kg/m^2; Overweight: BMI between 25 and 30 kg/m^2; Obesity: BMI greater than 30 kg/m^2), and pregnancy; for efficiency, cohorts and strata with less than 140 individuals were not characterized. The definition of all concept-sets and cohorts are available through the project GitHub repository, following the links listed in Supplementary Methods 2.

Patient Characteristics, Outcomes, and Stratifications

In addition to OHDSI predefined features (demographics: age group in 5-year bins, sex; SNOMED-CT hierarchical condition groups; and prescribed drug ingredient groups), we defined a large set of IBD-specific attributes. These include risk factors (e.g., family history of IBD, smoking), related comorbidities (e.g., autoimmune diseases, malignancies), symptoms (e.g., diarrhea, abdominal pain, rectal bleeding), abnormal lab test results (e.g., anemia; high C-Reactive Protein, CRP; high calprotectin), complications (e.g., hospitalization; extra-intestinal manifestations; death), medications (e.g., immunomodulators; steroids; biologics; antibiotics; non-steroidal anti-inflammatory drugs, NSAIDs), procedures (e.g., partial or total colectomy, colostomy, drainage of abscess), and vaccines (e.g., pneumococcal; influenza; Human papillomavirus, HPV; Measles, Mumps, and Rubella, MMR). All features are binary, indicating whether specific attributes have been recorded for each given patient; hence, by definition, the extracted characteristics include no missing values. For a full list of attributes and links to their definitions, see Supplementary Methods 2.

We extracted baseline characteristics during subjects' entire history and the year and month before the index date. Outcomes and treatments were identified during the 1 month, 1, 3, 5, 10 years, and all-time following each subject's index date.

Data Analysis

The IBD characterization analysis package was run on each data source independently, sharing only aggregated information, i.e., proportion of subjects having each specific attribute out of the corresponding cohort (or stratum); attributes with less than five individuals have not been shared, to minimize the risk of re-identification. To provide context, we also report the basic characteristics of the population covered by each data source. And, as these populations may differ considerably, we primarily focus on temporal trends, which investigate changes within a given data source.

To compute average age, we assumed that the age of individuals in each 5-year bin is uniformly distributed, therefore weighing each bin's center by its prevalence. We used simple linear regression to estimate the temporal trend in attribute

values, e.g., average age at diagnosis or percent females, across index date year strata, and reported R^2 as a measure of the goodness of fit. We used paired t-test to assess the statistical significance of attribute (e.g., symptom prevalence) differences between two strata (e.g., males versus females) across multiple DBs.

Results

Collectively, we characterized the disease trajectory of over one million patients with IBD, including 462,502 CD and 589,118 UC individuals. Tables 1, 2 and Supplementary Tables 2–4 present information on the size, age and sex distributions, and observation time of individuals in the CD, UC, IBD and IBDU cohorts, respectively, in each database (and, as context, compare to each database characteristic in Supplementary Table 1); extended tables are available online (https://data.ohdsi.org/IbdTable1/). Figure 1 shows the percentage of individuals with first IBD-related event (diagnosis or prescription) within each index-year stratum, indicating that most patients, in nearly all DBs (with the exception of MDCR and KDH), had their recorded onset event no earlier than 2010. Of note, temporal and age coverage differ considerably between DBs (see Supplementary Tables 1).

Demographics of IBD Cohorts

The percentage of female patients with incident CD and UC in Asia (Japan's JMDC, South Korea's AUSOM and KDH: 27.4%–38.6% and 35.6%–39.8%, respectively; Table 1 and 2) is considerably lower than in the European and USA cohorts (53.8%–65.2% and 49.1%–65.9%; and see Supplementary Tables 5 and 6 for female percentage in index-age strata); as well as in Asian sub-populations at western countries (CD: 53.2% in AMB-EMR and 57.5% in Optum-EHR; UC: 47.3% in IMRD-UK, 52.8% in AMB-EMR, and 54.1% Optum-EHR). CD female proportion decreased over time in Europe, USA, and Australia (e.g., CCAE: 59.1% and 51.5% in 2000 and 2020, respectively) but increased in Japan (JMDC: 21.4% in 2010 to 30.5% in 2020; Supplementary Table 7); UC female to male ratio decreased in most US databases (noticeably, MDCD and JHM) but remained stable in Germany, UK, and Japan (Supplementary Table 8).

The most common cohort index event, in both CD and UC, is the corresponding primary diagnosis, accounting for 25–80% and 28–84.6% of cases, respectively (Supplementary File 2); followed by more specific diagnoses, e.g., CD of small intestine or CD of large bowel (e.g., up to 19.5% and 16%, respectively, in AUSOM) and chronic ulcerative pancolitis (19.7% in JHM) or ulcerative proctocolitis (18.1% in IMRD-UK). Specific formulations—in particular

of mesalamine, prednisone, and prednisolone—have been prescribed, in some databases, to at most 25% of individuals on the index date.

Overall, there has been a noticeable decrease in the average age at which CD is diagnosed across various populations and regions, as indicated in Supplementary Table 9. This trend is particularly prominent in Europe and North America (e.g., average age-of-diagnosis decreased from 47.4 years in the 2005–2010 period to 44.3 years after 2020 in the Optum-EHR database). In Japan, age at CD diagnosis has been stable (average age of 33.7 years in the 2010–2015 period and 33.9 years after 2020). A similar, though weaker, decrease in diagnosis age has been observed in the UC cohorts (Supplementary Table 10); e.g., average age of diagnosis decreased from 52.5 years in the 2005–2010 period to 49.5 years after 2020 in the Optum-EHR database and 52.2 to 45.1 years in 2005–2010 and after 2020, respectively, in the MDCD database.

Symptoms Prior to IBD Diagnosis

Next, we investigated the symptoms recorded for patients in the year leading up to their IBD diagnoses. In US insurance claims databases, as well as in IMRD-UK, about one-third of patients with CD had a diagnostic code of abdominal pain during that timeframe (e.g., 38.1% in the Optum-CDM dataset in 2020), while much lower rates have been recorded in all other databases (e.g., 15.5% in GERMANY; Supplementary Table 11). Overall, abdominal pain prevalence increased over time, though in some cases only modestly. Similar trends have been observed in UC, with the exception of IMRD-UK abdominal pain rates being much lower in patients with UC than CD (15.1% versus 28.8% in 2015–2020; Supplementary Table 12).

For rectal bleeding among patients with UC, considerable differences were noted between individual datasets (Supplementary Table 13), such as 31.8% and 34.7% in IMRD-UK versus 11.7% to 11.1% in the MDCR, during 2000–2005 and 2015–2020, respectively. While rates increased in most databases over the years, some showed an opposite trend (e.g., Optum-CDM decreasing from 20.7% in 2000–2005 to 14.8% in 2015–2020).

Psychological symptoms prior to IBD diagnosis were also examined. Across most European and US datasets, there was an increase in the proportion of patients receiving an anxiety diagnosis before their CD diagnosis (Supplementary Table 14). In the IMRD dataset, the prevalence rose from 13.8% in 2000 to 27.3% during 2015–2020, and in the Optum-CDM dataset it increased from 21.3% to 29.3% during the same periods. Conversely, the increase in depression diagnosis rates was less pronounced (23.9% to 35.4% and 26.5% to 29.2%, respectively, for the same populations;

Table 1 Baseline characteristics of Crohn's disease incident cohorts

	Database	N	History (years)[*]	Follow-up (years)[*]	Female	Age at first IBD indication[¶,†]				
						2–6years	6–10years	10–18years	18–65years	≥65years
USA	CCAE	84,959	2.27 [1.48, 3.84]	2.99 [1.3, 5.95]	55.6%	0.5%	1.4%	9.0%	88.6%	0.5%
	MDCD	16,538	2.17 [1.44, 3.69]	3.44 [1.54, 5.88]	65.2%	1.4%	3.1%	13.8%	69.9%	11.5%
	MDCR	12,809	2.41 [1.55, 4.04]	3.96 [1.82, 6.84]	58.5%	–	–	–	2.7%	97.3%
	Optum-CDM	61,376	2.28 [1.48, 3.83]	3 [1.26, 5.99]	56.1%	0.3%	0.9%	5.5%	66.0%	27.2%
	PharMetrics+	93,217	2.42 [1.5, 3.46]	2.55 [1.12, 4.65]	53.8%	0.3%	1.0%	7.2%	82.6%	8.9%
	Optum-EHR	118,610	4.33 [2.45, 6.72]	4.7 [2.18, 7.15]	59.5%	0.2%	0.7%	5.6%	75.5%	18.0%
	CUIMC	4438	8.15 [3.81, 14.29]	5.37 [2.01, 9.37]	56.6%	0.9%	1.6%	10.2%	63.7%	23.6%
	JHM	1128	2.27 [1.53, 3.44]	2.24 [0.95, 3.63]	62.5%	0.8%	1.8%	10.1%	63.3%	23.9%
	AMB-EMR	51,328	2.94 [1.79, 4.66]	4.41 [2.32, 7.02]	61.6%	0.2%	0.5%	4.2%	71.4%	23.8%
Europe	FRANCE	320	1.71 [1.3, 2.42]	2.11 [1.22, 3.04]	62.5%	–	–	5.6%	86.3%	8.1%
	GERMANY	8321	3.53 [2.07, 5.53]	3.6 [1.58, 5.81]	59.3%	0.4%	0.6%	5.3%	75.8%	17.8%
	IMRD_UK	6936	6.09 [3.18, 10.02]	5.07 [2.42, 8.84]	54.4%	0.2%	1.3%	10.9%	73.7%	13.9%
Asia	JMDC	1837	3.27 [1.96, 5.39]	2.39 [1.03, 4.39]	27.4%	0.6%	0.9%	13.7%	82.9%	1.9%
	AUSOM	431	8.21 [3.48, 12.57]	4.57 [1.45, 9.68]	35.7%	<1.2%	2.3%	17.6%	72.6%	6.7%
	KDH	44	6.56 [3.68, 14.84]	4.79 [0.25, 9.26]	38.6%	–	–	<11.4%	88.6%	<11.4%
	AUSTRALIA	210	1.81 [1.35, 2.63]	1.64 [0.55, 2.71]	16.2%[‡]	–	–	4.8%	81.0%	14.3%

[*]History and follow-up median; interquartile range is shown in brackets

[¶]Age at first IBD-related diagnosis or prescription

[†]Age groups correspond to very early onset (2–6 years), early onset (6–10 years), pediatrics (10–18 years), adults (18–65 years) and elderly (≥65 years)

[‡]In the Australian database, 59% of subjects have no designated sex

13.3% of CD subjects identified as males

Table 2 Baseline characteristics of ulcerative colitis incident cohorts

	Database	N	History (years)	Follow-up (years)	Female	Age at first IBD indication				
						2–6y	6–10y	10–18y	18–65y	≥65y
USA	CCAE	109,500	2.28 [1.48, 3.86]	3.08 [1.34, 6.12]	55.2%	0.3%	0.5%	3.8%	94.8%	0.6%
	MDCD	16,431	2.32 [1.5, 3.94]	3.34 [1.43, 5.87]	65.9%	0.7%	1.6%	6.8%	69.0%	21.8%
	MDCR	26,242	2.52 [1.6, 4.18]	3.94 [1.73, 6.89]	58.7%	–	–	–	1.9%	98.1%
	Optum-CDM	93,009	2.35 [1.51, 3.98]	3.08 [1.28, 6.01]	56.9%	0.2%	0.3%	2.0%	59.3%	38.2%
	PharMetrics+	123,149	2.44 [1.52, 3.61]	2.62 [1.16, 4.65]	53.7%	0.2%	0.4%	3.0%	82.9%	13.5%
	Optum-EHR	119,188	4.35 [2.49, 6.76]	4.76 [2.25, 7.22]	58.2%	0.1%	0.3%	2.4%	71.9%	25.3%
	CUIMC	3922	7.67 [3.54, 14.27]	4.97 [1.92, 8.91]	59.0%	0.2%	0.7%	3.3%	65.4%	30.3%
	JHM	892	2.3 [1.59, 3.46]	2 [0.76, 3.56]	57.5%	1.1%	0.8%	6.7%	60.4%	30.8%
	AMB-EMR	58,404	2.96 [1.81, 4.66]	4.4 [2.34, 7.03]	58.6%	0.3%	0.4%	2.2%	67.3%	29.8%
Europe	FRANCE	299	1.72 [1.31, 2.4]	2.03 [0.87, 2.95]	51.5%	<1.7%	–	2.7%	79.3%	17.7%
	GERMANY	10,668	3.55 [2.04, 5.59]	3.51 [1.4, 5.79]	53.2%	0.3%	0.5%	2.8%	68.1%	28.1%
	IMRD_UK	13,924	5.68 [2.96, 9.49]	5.55 [2.63, 9.46]	49.1%	0.2%	0.5%	3.7%	72.8%	22.9%
Asia	JMDC	12,527	3.28 [1.99, 5.29]	2.38 [1.04, 4.25]	35.6%	0.2%	0.2%	4.3%	92.8%	2.4%
	AUSOM	660	6.45 [3, 11.64]	5.14 [1.79, 10.45]	36.4%	–	<0.8%	5.5%	85.8%	8.5%
	KDH	93	7.28 [2.96, 11.26]	4.45 [1.84, 11.04]	39.8%	<5.4%	–	<5.4%	79.6%	16.1%
	AUSTRALIA	210	1.85 [1.34, 2.75]	1.49 [0.64, 2.65]	15.7%[‡]	–	<2.4%	<2.4%	76.7%	22.4%

See Table 1 for more information

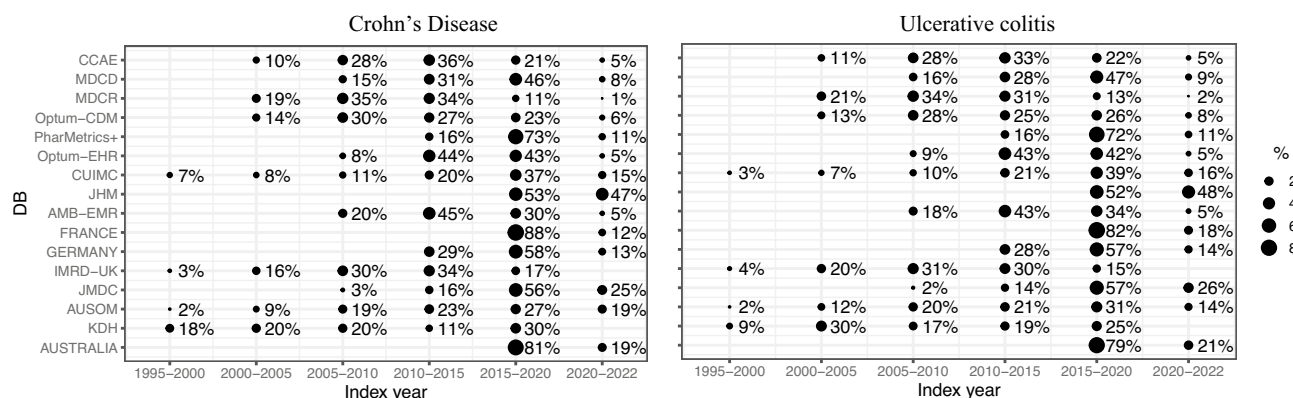[‡]18.1% of the Australian cohort's subjects identified as males

Fig. 1 Percentage of patients with incident IBD, per database (y-axis), within each index-year stratum (x-axis); left: Crohn's disease, right: ulcerative colitis

Supplementary Table 15) and several datasets exhibited a decrease in depression rates over time.

Anxiety diagnosis rates also increased among patients with UC since the early 2000s across various regions and populations (Supplementary Table 16), with some databases nearly doubling in prevalence. Despite substantial differences among datasets, this upward trend remained consistent: 5.1% and 21.6% in 2000–2005 to 12.3% and 33.6% after 2020, in the CUIMC and Optum-CDM datasets, respectively. Depression rates, though less prominent, also showed an increase among these populations over the past two decades (6.9% and 27.2% to 14.9% and 34.1% from 2000 to 2020 in CUIMC and Optum-CDM, respectively; Supplementary Table 17).

Sex differences in diagnosis rates of both physical and psychological symptoms were evident among patients with IBD across the datasets. Specifically, for CD (Supplementary Table 18), anxiety was diagnosed in 8.7% to 44.8% of females, whereas in males, it ranged from 6.9% to 32.8%, and the proportion of females with anxiety was consistently higher across all datasets (Paired t-test P-value < 0.001). Similarly, for depression, the range was 10.1% to 46.4% in females and 6.9% to 32.9% in males, again showing a higher proportion among females in all datasets (Paired t-test P-value < 0.001). Recorded complaints of abdominal pain ranged from 7.1% to 51.2% among females and 7.2% to 44.4% among males, with most datasets indicating higher proportions among females (Paired t-test P-value < 0.001). Diagnosis of diarrhea occurred at similar proportions between sexes, ranging from 0.7% to 29.9% in females and 0.5% to 25.9% in males, with no significant differences noted in the individual datasets (Paired t-test P-value = 0.1).

Similar patterns were identified in the UC population (Supplementary Table 19), wherein anxiety and depression diagnosis rates ranged from 10.7% to 48.8% and 8.4% to 50.7% among females, and 7.5% to 39.8% and 7.9% to 41.2%

among males, respectively. In all datasets, a higher proportion of anxiety and depression diagnoses were observed in females (for both diagnoses, paired t-test P-value < 0.001). Abdominal pain occurrence varied from 6.5% to 48.7% in females and 4.8% to 42.6% in males across the datasets, consistently demonstrating a higher proportion among females (Paired t-test P-value < 0.001). Conversely, diarrhea diagnosis rates were similar between sexes across all datasets, ranging from 0.8% to 31.6% in females and 0.9% to 32.8% in males (Paired t-test P-value = 0.3).

Surgical Procedures

Regarding surgical outcomes for patients with CD, most US databases (with the exception of AMB-EMR) demonstrated stable or increasing rates of partial colonic resection within one year from diagnosis, over the past two decades (e.g., 7.9% and 8.9% in the CCAE database and 11.2% to 14% in the MDCR database, during 2005-2010 and 2015-2020, respectively; Supplementary Table 20). Notably, rates in IMRD-UK decreased considerably from 6.9% in 2000–2005 to 4.1% in 2015–2020. A similar trend is observed when examining the colonic resection rates 3 years after diagnosis (CCAE: 10.3% to 12.8%, MDCD: 10% to 13.7%, and, conversely, IMRD-UK: 11% to 6.7%, in 2005 and after 2015, respectively; Supplementary Table 21).

Small bowel resection rates had demonstrated similar course over time among patients with CD: stable or increasing in most US databases (e.g., 3.3% and 4.1% at the early 2000s, 5.3% and 4.4% in 2015–2020 in the CCAE and MDCR data sets, respectively; Supplementary Table 22); but decreasing in AMB-EMR (1% in 2005–2010 to 0.4% in 2015–2020), IMRD-UK (8% in 2000–2005 to 4.7% in 2015–2020) and JMDC (4.8% in 2010–2015 to 2.4% after 2020). These same trends were kept when examining small

bowel resection rate 3 years after diagnosis (Supplementary Table 23).

Partial colectomy rates for patients with ulcerative colitis varied over time and across cohorts. Within one year from diagnosis (Supplementary Table 24), resection rates increased between 2000 and 2020 (e.g., MDCR database from 8.8% in 2000 to 14.1% in 2020). When examining colectomy rates within three years following diagnosis (Supplementary Table 25), we observed a similar trend: In the CUIMC database, colectomy rate was 5.8% in 2000–2005 and 7.2% in 2015–2020; in the Optum-CDM cohort, the rates were 12.8% and 16.9% for 2000 and 2020, respectively. The most significant increase in colectomy rates was evident in the CCAE and MDCR cohorts, where rates were 10.1% and 15.6% in 2000, respectively, and then increased to 14% and 20.5% by 2020.

Discussion

This study builds on the meticulous harmonization process of multiple data sources to a common data model, carried out by numerous members of the OHDSI community. These efforts allowed us to characterize IBD patient cohorts identified in numerous databases from around the globe, along different axes of the disease trajectory.

Male predominance in Asian IBD cohorts has been reported before [10, 11] and its underlying cause is still unclear. For the Australian cohorts, 59% of the entire DB population had no designated sex, hence sex composition of the IBD population is unknown. The average age at diagnosis across most cohorts in our study is higher than previously reported [12, 13], likely due to the distinctive patient demographics under examination and, specifically, the small proportion of pediatric populations within certain cohorts (e.g., MDCR and, to a lesser extent, CCAE). Indeed, our findings show that the average age at diagnosis is highly correlated with the average age of the covered population in each database (Supplemental Table 1; $R^2 = 0.86$ and 0.8, P-value < 0.001 for Crohn's disease and ulcerative colitis, respectively). Additionally, in some DBs, incident events may be missed due to lags in coverage. These distinctions should be borne in mind when interpreting our findings. Nonetheless, and consistent with prior investigations [13], our data reveals a younger age at diagnosis among patients with CD in comparison to those with UC, reinforcing the reliability of our findings.

The present study demonstrated a trend of a declining mean age at diagnosis for both UC and CD, a pattern consistent with several [14], but not all [15] studies. In contrast to numerous studies that predominantly scrutinized patient populations in Europe and North America, the current study reveals a consistent age-of-diagnosis trend in South-East Asia and Australia. This broader geographical scope underscores the global nature of the observed phenomenon. In addition to a possible actual decline in onset age, improved diagnostic capabilities and heightened clinician awareness represent alternative potential explanations for this observation.

Over the past decade, numerous studies reported a decline in the prevalence of small bowel resection among patients with Crohn's disease [16]. However, several recent investigations presented conflicting evidence, challenging this established observation [17, 18]. Our data demonstrate that, overall, there is no universally consistent trend of decreasing resection rates. In most US databases, we found no significant decrease in the prevalence of small bowel resections within the first year after diagnosis for patients diagnosed in the early 2000s compared to those diagnosed in 2020 [19]. It is essential to acknowledge that we specifically examined surgical events within the first year and three years after diagnosis. This focus potentially minimizes the impact of drug treatment on disease progression. Furthermore, determining the exact indication for surgery (whether it is due to complications, treatment failure, or other factors) is, in general, challenging (and impossible using our aggregated data). Additionally, it is worth noting that recent advancements in surgical techniques and lower complication rates have led to a growing tendency to consider surgical treatment earlier in the disease course, before the use of advanced medical options [20]. These novel approaches may have influenced the outcomes, especially for specific patient subgroups.

Interestingly, some of our datasets have revealed a notable uptick in the frequency of colectomy procedures among patients with UC over the past two decades. This finding contrasts with several recent studies [21]. A number of factors could contribute to this discrepancy. Firstly, this disparity may be contingent upon the demographic makeup of the studied populations, with Asian datasets notably exhibiting a higher escalation. Moreover, disparities in accessing timely and appropriate medical care could disproportionately steer certain patient cohorts towards surgical intervention. Lastly, various biases, including encompassing coding discrepancies (e.g., more two or three staged procedures) and limitations in data collection methodologies, may have influenced our findings. It is also important to acknowledge that, owing to the inherent characteristics of our cohorts, the definitive surgical procedures undergone by patients remain unspecified. Nonetheless, considering the established standards of care in UC surgical management, the term 'partial colectomy' likely denotes the initial step of the definitive surgical intervention for this particular patient population.

Individuals with IBD exhibit a heightened vulnerability to mental health disorders compared to the general population [22–24]. Specifically, anxiety and depression manifest at notably elevated rates among patients with IBD and can

predict disease complications [25]. Our study unveiled a pronounced sex disparity, with females exhibiting higher prevalence rates of both anxiety and depression compared to their male counterparts. This observation echoes findings in the general population, where anxiety disorders are more prevalent among women [26]. Moreover, there was an escalation in the prevalence of anxiety and depression prior to IBD diagnosis in the last two decades. This might be attributed to various factors. Primarily, this trend might mirror a global surge in anxiety disorders [27]. Additionally, heightened awareness among both patients and healthcare providers may influence diagnostic practices and documentation among patients seeking care for other symptoms. We note that diagnostic criteria and assessment methods for depression and anxiety may vary considerably across healthcare systems (and physicians), affecting the observed prevalence of these conditions, particularly in IBD cohorts. Whether these factors account for the upward prevalence trajectory or if there have been specific shifts in clinical presentation warrants further research.

We observed similar epidemiological trends—i.e., disease incidence, clinical presentation, and outcomes—across diverse populations in Southeast Asia, Europe, and the US, despite substantial genetic and dietary differences [28]. Global shifts toward westernized diets, urbanization, and changes in microbiota may contribute to these common trends, as well as higher worldwide detection rates, owing to improved diagnostics and healthcare access. Furthermore, common pathways in immune dysregulation and inflammation, along with similar therapeutic responses, may play a more significant role in IBD epidemiology than previously appreciated. This convergence highlights the importance of investigating the universal mechanisms underlying IBD pathogenesis and treatment.

Our study may have several limitations. First, interpretation of the reported characteristics requires careful consideration of the broader context underlying each data source—e.g., in terms of its demographic composition, data type (insurance claims versus EHRs), and healthcare setting (primary care versus hospital)—to identify potential biases and data gaps; we provided such information in Supplementary Methods 1, but additional aspects may be considered. Second, to identify a disease onset event, we required a minimum look-back period of 365 days with no IBD-related diagnoses or prescriptions; longer periods could have increased sensitivity of the incidence cohorts [4]. We note, however, that for most databases, the median IBD-free history length is much longer (Tables 1 and 2). Third, we extracted characteristics for a set of pre-defined strata (e.g., elderly individuals or males) and leave more specific sub-group (e.g., elderly females) analyses for future work. Finally, the results of endoscopic evaluations are unavailable

in the DBs we analyzed, and it is, therefore, impossible to characterize or stratify patients by disease severity.

Notwithstanding these limitations, our study provides a comprehensive characterization of IBD cohorts, including over one million patients, from multiple countries and healthcare settings, in great detail. We presented several analyses that use these data to study various aspects of the disease and made all characterization data publicly available. The IBD community can utilize this open resource to guide and accompany their research avenues and, specifically, apply similar demographic or temporal trend analyses to other attributes or subpopulations; run time-to-event analyses, e.g., to hospitalization, using the follow-up time window strata; or readily obtain answers to certain questions, e.g., regarding medication usage or malignancy prevalence, in specific geography and healthcare system. In the future, we hope to update and expand it to include additional data sources and attributes.

Data availability IBD characterization results can be found at https://data.ohdsi.org/IbdTable1/ and https://data.ohdsi.org/IbdCharacterization/.

This study was performed as a federated network study and patient-level data remained with the data partner. Some of the datasets used within this study are available via license. The data that support the findings of this study are available to license from Merative™ (CCAE, MDCD, MDCR), Optum (OPTUM-EHR, OPTUM-CDM), JDMC, IQVIA™ (FRANCE, GERMANY, AUSTRALIA, AMB-EHR, PharMetrics+, IMRD-UK). Data are available from Merative™ at https://www.merative.com/documents/brief/marketscan-explainer-general, from Optum™ at https://www.optum.com/business/solutions/life-sciences/real-world-data.html, from JMDC at https://www.jmdc.co.jp/en/jmdc-claims-database/, and from IQVIA™ at https://www.iqvia.com/solutions/real-world-evidence/real-world-data-and-insights.

The AUSOM data generated and analyzed in the study can be available from Prof Rae Woong Park upon reasonable request. The CUIMC patient-level data that were used in this study are accessible only to IRB approved study authors. Other organizations would need to be contacted in order to gain access to their data assets.

Declarations

Conflict of interest EAV, JS, AS, and NH are employees of Janssen Research and Development LLC and shareholders of Johnson & John-

son (J&J) stock; SM previously consulted for Surescripts and now consults for First Databank; YC is employee of CytoReason.

Ethical approval The New England Institutional Review Board determined that studies conducted in CCAE, MDCR, MDCD, OPTUM-EHR, and OPTUM-CDM were exempt from study-specific Institutional Review Board review, as these studies do not qualify as human subject research. For use of IQVIA™'s FRANCE, GERMANY, AUSTRALIA, AMB-EMR, and PharMetrics+, no patient permission was necessary because all patient data are deidentified for research purposes. Based on Ethical Guidelines for Epidemiological Research issued by the Japanese Ministry of Health, Labor and Welfare, ethics approval and informed consent were not applicable for use of JMDC. The use of AUSOM data in this study was approved by the Institutional Review Board of Ajou University Hospital (IRB number: AJOUIRB-MDB-2022–263) and the use of KDH data—by the Institutional Review Board of Kangdong Sacred Heart Hospital (IRB number: 2022–09-021); both committees waived the requirement for informed consent. IQVIA™ Scientific Review Committee approved the use of IMRD-UK data for this study (SRC Reference Number: 21SRC066). The research was approved by the Columbia University Institutional Review Board, as an OHDSI network study, and by the Johns Hopkins institutional review board.

References

1. Benchimol EI, Manuel DG, Guttmann A, Nguyen GC, Mojaverian N, Quach P et al. Changing Age Demographics of Inflammatory Bowel Disease in Ontario, Canada: A Population-based Cohort Study of Epidemiology Trends. *Inflamm Bowel Dis.* 2014;20:1761–1769.

2. Windsor JW, Kaplan GG. Evolving Epidemiology of IBD. *Curr Gastroenterol Rep.* 2019;21:40.

3. Freeman K, Ryan R, Parsons N, Taylor-Phillips S, Willis BH, Clarke A. The incidence and prevalence of inflammatory bowel disease in UK primary care: a retrospective cohort study of the IQVIA Medical Research Database. *BMC Gastroenterol.* 2021;21:139.

4. Friedman MY, Leventer-Roberts M, Rosenblum J, Zigman N, Goren I, Mourad V et al. Development and validation of novel algorithms to identify patients with inflammatory bowel diseases in Israel: an epi-IIRN group study. *Clin Epidemiol.* 2018;10:671–681.

5. Lee CK, Ha HJ, Oh SJ, Kim J-W, Lee JK, Kim H-S et al. Nationwide validation study of diagnostic algorithms for inflammatory bowel disease in Korean National Health Insurance Service database. *J Gastroenterol Hepatol.* 2020;35:760–768.

6. OHDSI. The Book of OHDSI: Observational Health Data Sciences and Informatics. OHDSI; 2019.Available from: https://books.google.co.il/books?id=JxpnzQEACAAJ

7. Li X, Ostropolets A, Makadia R, Shoaibi A, Rao G, Sena AG et al. Characterising the background incidence rates of adverse events of special interest for covid-19 vaccines in eight countries: multinational network cohort study. *BMJ.* 2021;373:n1435.

8. Suchard MA, Schuemie MJ, Krumholz HM, You SC, Chen R, Pratt N et al. Comprehensive comparative effectiveness and safety of first-line antihypertensive drug classes: a systematic, multinational, large-scale analysis. *The Lancet.* 2019;394:1816–1826.

9. Haas L, Nagy P, Bowring M, Cook M, Magen-Rimon R, Weisshof R et al. Characterizing the Johns Hopkins Inflammatory Bowel Disease Cohort using OMOP CDM. *Gastroenterology.* 2023;164:S43–S44.

10. Kim HJ, Hann HJ, Hong SN, Kim KH, Ahn IM, Song JY et al. Incidence and Natural Course of Inflammatory Bowel Disease in Korea, 2006–2012: A Nationwide Population-based Study. *Inflamm Bowel Dis.* 2015;21:623–630.

11. Park SH, Kim Y-J, Rhee KH, Kim Y-H, Hong SN, Kim KH et al. A 30-year Trend Analysis in the Epidemiology of Inflammatory Bowel Disease in the Songpa-Kangdong District of Seoul, Korea in 1986–2015. *J Crohns Colitis.* 2019;13:1410–1417.

12. Matsuoka K, Fujii T, Okamoto R, Yamada A, Kunisaki R, Matsuura M et al. Characteristics of adult patients newly diagnosed with Crohn's disease: interim analysis of the nation-wide inception cohort registry study of patients with Crohn's disease in Japan (iCREST-CD). *J Gastroenterol.* 2022;57:867–878.

13. Loftus CG, Loftus EV Jr, Harmsen SW, Zinsmeister AR, Tremaine WJ, Melton JL III et al. Update on the incidence and prevalence of Crohn's disease and ulcerative colitis in Olmsted County, Minnesota, 1940–2000. *Inflamm Bowel Dis.* 2007;13:254–261.

14. Ghersin I, Khteeb N, Katz LH, Daher S, Shamir R, Assa A. Trends in the epidemiology of inflammatory bowel disease among Jewish Israeli adolescents: a population-based study. *Aliment Pharmacol Ther.* 2019;49:556–563.

15. Braegger CP, Ballabeni P, Rogler D, Vavricka SR, Friedt M, Pittet V et al. Epidemiology of Inflammatory Bowel Disease. *J Pediatr Gastroenterol Nutr.* 2011;53:141–144.

16. Ma C, Moran GW, Benchimol EI, Targownik LE, Heitman SJ, Hubbard JN et al. Surgical Rates for Crohn's Disease are Decreasing: A Population-Based Time Trend Analysis and Validation Study. *Am J Gastroenterol.* 2017;112:1840–1848.

17. Stöss C, Berlet M, Reischl S, Nitsche U, Weber M-C, Friess H et al. Crohn's disease: a population-based study of surgery in the age of biological therapy. *Int J Colorectal Dis.* 2021;36:2419–2426.

18. Valvano M, Vinci A, Cesaro N, Frassino S, Ingravalle F, Ameli M et al. The long-term effect on surgery-free survival of biological compared to conventional therapy in Crohn's disease in real world-data: a retrospective study. *BMC Gastroenterol.* 2023;23:438.

19. Burisch J, Kiudelis G, Kupcinskas L, Kievit HAL, Andersen KW, Andersen V et al. Natural disease course of Crohn's disease during the first 5 years after diagnosis in a European population-based inception cohort: an Epi-IBD study. *Gut.* 2019;68:423–433.

20. Stevens TW, Haasnoot ML, D'Haens GR, Buskens CJ, de Groot EJ, Eshuis EJ et al. Laparoscopic ileocaecal resection versus infliximab for terminal ileitis in Crohn's disease: retrospective long-term follow-up of the LIR!C trial. *Lancet Gastroenterol Hepatol.* 2020;5:900–907.

21. Dai N, Haidar O, Askari A, Segal JP. Colectomy rates in ulcerative colitis: A systematic review and meta-analysis. *Dig Liver Dis.* 2023;55:13–20.

22. Barberio B, Zamani M, Black CJ, Savarino EV, Ford AC. Prevalence of symptoms of anxiety and depression in patients with

inflammatory bowel disease: a systematic review and meta-analysis. *Lancet Gastroenterol Hepatol.* 2021;6:359–370.

23. Fousekis FS, Katsanos AH, Kourtis G, Saridi M, Albani E, Katsanos KH et al. Inflammatory Bowel Disease and Patients With Mental Disorders: What Do We Know? *J Clin Med Res.* 2021;13:466–473.

24. Frolkis AD, Vallerand IA, Shaheen A-A, Lowerison MW, Swain MG, Barnabe C et al. Depression increases the risk of inflammatory bowel disease, which may be mitigated by the use of antidepressants in the treatment of depression. *Gut.* 2019;68:1606–1612.

25. Ananthakrishnan AN, Gainer VS, Perez RG, Cai T, Cheng S-C, Savova G et al. Psychiatric co-morbidity is associated with increased risk of surgery in Crohn's disease. *Aliment Pharmacol Ther.* 2013;37:445–454.

26. Javaid SF, Hashim IJ, Hashim MJ, Stip E, Samad MA, Ahbabi AA. Epidemiology of anxiety disorders: global burden and sociodemographic associations. *Middle East Curr Psychiatry.* 2023;30:44.

27. Goodwin RD, Weinberger AH, Kim JH, Wu M, Galea S. Trends in anxiety among adults in the United States, 2008–2018: Rapid increases among young adults. *J Psychiatr Res.* 2020;130:441–446.

28. Zhao M, Feng R, Ben-Horin S, Zhuang X, Tian Z, Li X et al. Systematic review with meta-analysis: environmental and dietary differences of inflammatory bowel disease in Eastern and Western populations. *Aliment Pharmacol Ther.* 2022;55:266–276.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.