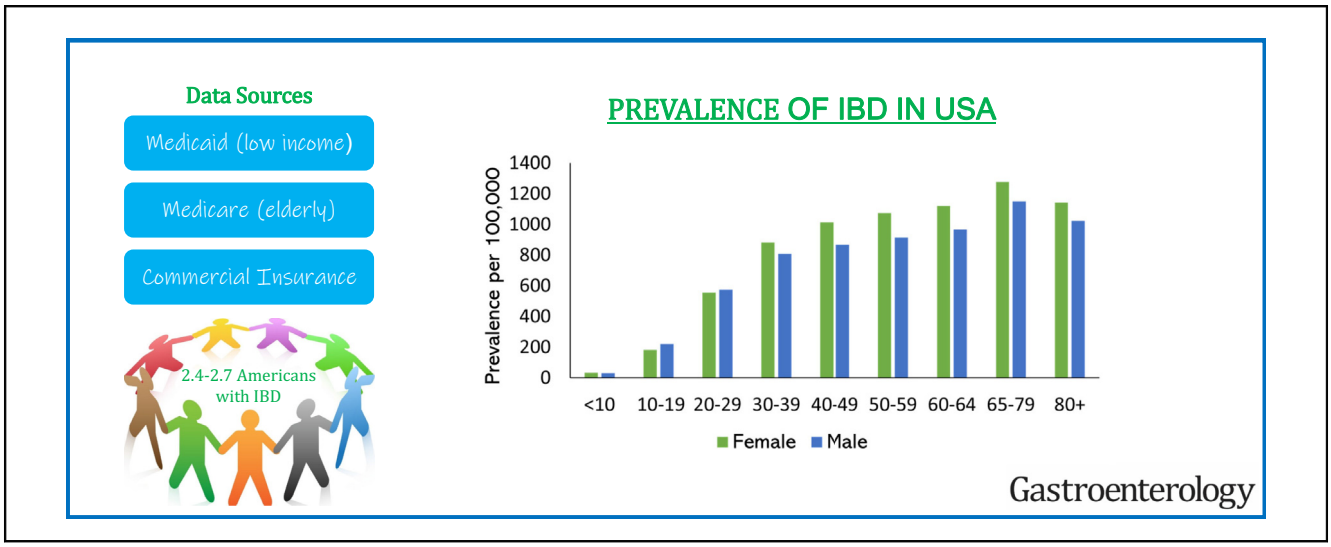
Gastroenterology 2023;165:1197–1205

Incidence, Prevalence, and Racial and Ethnic Distribution of Inflammatory Bowel Disease in the United States

James D. Lewis,[1,2,3] Lauren E. Parlett,[4] Michele L. Jonsson Funk,[5] Colleen Brensinger,[2] Virginia Pate,[5] Qufei Wu,[2] Ghadeer K. Dawwas,[2] Alexandra Weiss,[1] Brad D. Constant,[6] Maureen McCauley,[2] Kevin Haynes,[7] Jeff Yufeng Yang,[5] Douglas E. Schaubel,[3] Andres Hurtado-Lorenzo,[8] and Michael David Kappelman[9]

[1]Division of Gastroenterology and Hepatology, University of Pennsylvania, Philadelphia, Pennsylvania; [2]Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania; [3]Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania; [4]HealthCore, Inc, Wilmington, Delaware; [5]Center for Pharmacoepidemiology, Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina; [6]Division of Gastroenterology, Hepatology, and Nutrition, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania; [7]Janssen Research and Development, LLC, Titusville, New Jersey; [8]Crohn's & Colitis Foundation, New York, New York; and [9]Department of Pediatrics, Division of Pediatric Gastroenterology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina



BACKGROUND & AIMS: We sought to estimate the incidence, prevalence, and racial-ethnic distribution of physician-diagnosed inflammatory bowel disease (IBD) in the United States. **METHODS:** The study used 4 administrative claims data sets: a 20% random sample of national fee-for-service Medicare data (2007 to 2017); Medicaid data from Florida, New York, Pennsylvania, Ohio, and California (1999 to 2012); and commercial health insurance data from Anthem beneficiaries (2006 to 2018) and Optum's deidentified Clinformatics Data Mart (2000 to 2017). We used validated combinations of medical diagnoses, diagnostic procedures, and prescription medications to identify incident and prevalent diagnoses. We computed pooled age-, sex-, and race/ethnicity-specific insurance-weighted estimates and pooled estimates standardized to 2018 United States Census estimates with 95% confidence intervals (CIs). **RESULTS:** The age- and sex-standardized incidence of IBD per 100,000 person-years was 10.9 (95% CI, 10.6–11.2). The incidence of IBD peaked in the third decade of life, decreased to a relatively stable level across the fourth to eighth decades, and declined further. The age-, sex- and insurance-standardized prevalence of IBD was 721 per 100,000 population (95% CI, 717–726). Extrapolated to the 2020 United States Census, an estimated 2.39 million Americans are diagnosed with IBD. The prevalence of IBD per 100,000 population was 812 (95% CI, 802–823) in White, 504 (95% CI, 482–526) in Black, 403 (95% CI, 373–433) in Asian, and 458 (95% CI, 440–476) in Hispanic Americans. **CONCLUSIONS:** IBD is diagnosed in >0.7% of Americans. The incidence peaks in early adulthood and then plateaus at a lower rate. The disease is less commonly diagnosed in Black, Asian, and Hispanic Americans.

Keywords: Crohn's Disease; Ulcerative Colitis; Epidemiology; Medicare; Medicaid; Race.

Inflammatory bowel disease (IBD) includes ulcerative colitis, Crohn's disease, and IBD-unspecified. Throughout the world, the prevalence of people diagnosed with these diseases has increased throughout the last several decades.[1,2] North America is considered to have among the highest prevalence and incidence of IBD in the world,[2,3] yet there are few nationally representative data on the incidence and prevalence of IBD in the United States (US). The highest estimate comes from the National Health Interview Study (NHIS), although this was based on patients' self-report.[4] Other population-based studies from Olmstead County, Minnesota, and northern California that relied on provider diagnoses reported lower prevalence estimates.[5,6] Additionally, there are very limited data on the racial and geographic distribution of IBD in the US.[7]

The paucity of rich data on the incidence, prevalence, and racial and ethnic distribution of IBD in the US stems from the lack of a unified health system with a common medical record or central data repository. Rather, in the US, there are multiple different commercial health insurance plans as well as state and nationally administered health insurance plans for the poor and elderly or disabled, respectively. In this study, we sought to define the incidence and prevalence of physician-diagnosed IBD in the US in a nationally representative population by pooling data from multiple different health insurance plans, including commercial, Medicaid, and Medicare. Additionally, we sought to estimate the racial/ethnic and geographic distribution of IBD.

WHAT YOU NEED TO KNOW

BACKGROUND AND CONTEXT

There are few nationally representative data on incidence, prevalence, racial-ethnic composition, and regional variability of inflammatory bowel diseases in the United States.

NEW FINDINGS

The incidence of inflammatory bowel disease was 10.9 per 100,000 person-years. There are an estimated 2.39 million Americans with inflammatory bowel diseases. The prevalence is highest in White Americans and in the Northeastern United States.

LIMITATIONS

We measured incidence and prevalence by pooling Medicare, Medicaid, and commercial insurance claims data. Sensitivity and specificity may be imperfect. Race and ethnicity may be subject to misclassification.

CLINICAL RESEARCH RELEVANCE

Inflammatory bowel disease is a common chronic condition, affecting >0.7% of Americans. The prevalence varies by race, ethnicity, and geographic location. Future investigation is essential to understand the causes and consequences of these observed differences.

BASIC RESEARCH RELEVANCE

Identifying host factors and environmental exposures contributing to the incidence of inflammatory bowel disease is an important goal for future investigation.

Methods

Data Sources

The study used 4 administrative claims data sets. Medicare is a government run health insurance plan for older (age >65 years) and disabled Americans. There are fee-for-service and managed care Medicare plans. We used a 20% random sample of national fee-for-service Medicare data from 2007 to 2017 that included beneficiaries aged ≥65 with at least 1 month in which they were simultaneously enrolled in Parts A, B, and D fee-for-service coverage. Medicaid is a collection of state run health insurance plans. We used Medicaid data from 5 of the largest Medicaid plans (Florida, New York, Pennsylvania, Ohio, and California) from 1999 to 2012. We used only Medicaid fee-for-service because recording of diagnoses is less complete in Medicaid managed care plans.

We also used 2 sources of commercial health insurance data. One source was HealthCore, Inc, a wholly owned, independently operated subsidiary of Elevance Health, Inc (formerly Anthem, Inc). HealthCore provided aggregated data for beneficiaries from 2006 to 2018, including members with commercial plans (ie, age <65) from 14 states and members with Elevance Health-managed Medicare plans. We also used claims data from Optum's deidentified Clinformatics Data Mart (CDM), a collection of anonymized patient-level insurance data from 2000 to 2017. There was no overlap of Medicare, Medicaid, HealthCore, and CDM data used in this research. See the Supplementary Methods for additional details.

Each data set contains billing data for physician encounters, including diagnoses recorded using International Classification of Disease, 9th Edition, Clinical Modification (before October 1, 2015) or 10th Edition. Prescription drug claims are coded using National Drug Codes and include quantity dispensed and days supplied. Procedures, including infused medications, are classified using the American Medical Association's Current Procedural Terminology (CPT) and Centers for Medicare and Medicaid Services Healthcare Common Procedure Coding System codes.

Inclusion Criteria

For this research, we only included patients with inpatient, outpatient, and prescription medication benefits, with the exception of HealthCore, where only medical coverage was required for inclusion. Patients were also required to have their date of birth and sex recorded in the database. There is no distinction of sex assigned at birth and gender in the databases.

Abbreviations used in this paper: **CDM, Clinformatics Data Mart; CI, confidence interval; IBD, inflammatory bowel disease; NHANES, National Health and Nutrition Examination Survey; NHIS, National Health Interview Study; PPV, positive predictive value; US, United States.**

Most current article

© 2023 by the AGA Institute.
0016-5085/$36.00
https://doi.org/10.1053/j.gastro.2023.07.003

For this research, we assumed that the variable aligned with the recording of sex within the US Census data.

Algorithms to Identify Incident and Prevalent Diagnosis of Inflammatory Bowel Diseases

We used a combination of medical diagnoses, diagnostic procedures, and prescription medications to identify prevalent and incident diagnoses, as previously described.[8] Patients were considered to have an incident diagnosis if they had (1) a minimum of 4 years of follow-up before the first diagnosis of an IBD in the claims data by any provider, (2) no prior therapy with a medication used to treat an IBD, unless there was another indication, such as rheumatoid arthritis, in a patient treated with an anti-tumor necrosis factor medication, (3) had a colonoscopy, sigmoidoscopy, capsule endoscopy, or bowel resection surgery within 6 weeks before the first recorded diagnosis code, and (4) had a second diagnosis of an IBD within 12 months of the first diagnosis (the first and follow-up diagnoses were required to be from a gastroenterologist or surgeon). Incident cases were further subdivided into higher and lower probability based on medical therapy prescribed within 90 days of the diagnostic procedure that led to the diagnosis (the index date).

The high-probability group was defined based on having (1) a first prescription for steroids (oral or rectal) or mesalamine (oral or rectal), sulfasalazine, olsalazine, balsalazide, adalimumab, infliximab, golimumab, or certolizumab, or a combination of these, within 90 days after the index date or (2) no prescribed IBD medications if bowel resection surgery was used to define the index date. The lower-probability group was defined as those who did not meet the high probability therapy-based criteria but met the other criteria for an incident diagnosis. We previously established that the positive predictive value (PPV) to identify the diagnosis date within 90 days for the high-probability algorithm was 91% and for the lower-probability algorithm was 85%.[8]

The prevalence algorithm required 1 IBD diagnosis by any provider regardless of specialty combined with $\geq$1 IBD diagnoses by a gastroenterologist or surgeon or a therapy with a medication for an IBD in the absence of another indication. We previously established that the PPV of this algorithm was 94% for $\geq$2 diagnoses by a gastroenterologist or surgeon and receipt of IBD-specific medications, 92% for $\geq$2 diagnoses by a gastroenterologist or surgeon without receipt of IBD medications, and 78% for 1 diagnosis by any provider and receipt of IBD medications. A single diagnosis by any provider in the absence of prescribed medications had a PPV of 35%.[8] A secondary definition included patients with 1 diagnosis by any provider and no prescriptions for IBD-related therapies.

We required a minimum of 4 years of continuous benefits before December 31, 2017, for Medicare, HealthCore, and CDM data and December 31, 2012, for Medicaid beneficiaries to be included in the estimate of prevalence. Sensitivity analyses described below used a minimum of 1 year of benefits. See the Supplementary Methods for details on identification of provider specialty.

We categorized patients as having Crohn's disease vs ulcerative colitis if the most recent diagnosis was the same as the most frequent diagnosis on or before the data of measurement.[8] If the number of Crohn's disease diagnoses equaled the number of ulcerative colitis diagnosis or the most recent diagnosis was not the same as the most common, we categorized the patient as IBD not further specified.

Statistical Analysis

For incidence analyses, we excluded the last 6 months of data to avoid bias from delays in claims being filed. We identified incident diagnoses in the last 3 years of data before this 6-month cutoff.

To compute pooled estimates of incidence and prevalence, we first computed age- and sex-specific estimates within each of the data sets. For each age and sex stratum, we pooled the data from HealthCore and CDM using a fixed-effects meta-analytics method. Next, we computed age-, sex- and race-specific insurance-weighted estimates by pooling the data and applying weights proportional to the insurance coverage of Americans based on the 2018 US Census. For patients aged >65, we pooled Medicare fee-for-service data obtained from Centers for Medicare and Medicaid Services with Medicare Advantage data from CDM and Anthem claims. For patients aged >65, the insurance weights were derived from data published by the Kaiser Foundation on Medicare beneficiaries enrollment in fee-for-service vs managed Medicare plans (https://www.kff.org/medicare/issue-brief/medicare-advantage-in-2021-enrollment-update-and-key-trends/). Finally, we computed age- and sex-standardized estimates of the national prevalence of IBD using direct standardization to the 2018 US Census data.

To compute estimates by race and ethnicity, we relied on the race and ethnicity as recorded in CDM, Medicaid, and Medicare. In CDM, race and ethnicity has been collected from public records (eg, driver's license records) for ~30% of individuals and is imputed for the other members using an algorithm based on first and last names and US Census data ZIP Codes (ZIP + 4). This method is estimated to have 97% specificity, 48% sensitivity, and 71% PPV for predicting the race of Black individuals.[9] For Medicare, we used the Research Triangle Institute variable, which has improved accuracy compared with the data from beneficiary enrollment files.[10] Although Hispanic ethnicity is distinct from race, it is included in a single variable in these data sets and as such is reported together with race. Only Black, Asian, White, and Hispanic are included due to incomplete data and lower accuracy of the variables for other races/ethnicities.[10]

Estimates of incidence were computed using only the high-probability algorithm and combining the high- and low-probability algorithms. Estimates of prevalence were computed under 4 different assumptions based on the minimum enrollment period and the prevalence definitions used. The primary analysis required 4 years of minimum enrollment. Sensitivity analyses used a minimum enrollment period of 1 year or the secondary definition of prevalence that included patients with 1 IBD diagnosis by a gastroenterologist or surgeon, or $\geq$1 by any provider other than a gastroenterologist or surgeon, and no prescribed therapy. Because the PPV of 1 diagnosis without any medications was only 35% and the lowest PPV in any one data set was 22%, we applied this weight to the patients meeting only this definition.[8] Thus, in the sensitivity analysis, each patient meeting the primary definition had a weight of 1.0, and those meeting only the secondary definition had a weight of 0.22, the most conservative estimate based on our validation study results.

Stratum-specific variance estimates were computed in the same manner as the incidence and prevalence estimates. Weighted strata-specific variance estimates were summed to get the overall variance. Because the sample size was so large and variance estimates so small, essentially all comparisons would meet traditional definitions of statistical significance. As such, we report nominal values and 95% confidence intervals (CIs) for incidence and prevalence but did not compute *P* values for comparisons between groups.

To gauge whether our estimates of incidence and prevalence were concordant, we applied the principle that incidence multiplied by average duration of disease equals the prevalence. We created a theoretical population of 100,000 people born on the same day. We applied age-specific mortality rates from 2015 US life tables to determine the number of people who would still be alive at each age. Applying these mortality rates, >98% of the population would have died by age 100.

For each age up to 100, we multiplied the number of people alive by the pooled age- and sex-specific incidence rates of IBDs derived from the 4 claims data sets to determine the number of people newly diagnosed with an IBD at that age. For this step, we assumed that the age-specific estimates of incidence applied to all ages within a stratum. For example, we assumed that the incidence was the same in the 50th year of life as in the 59th year. We summed the number of cases from birth to each age and applied the age-specific mortality rates to determine the number of people alive with an IBD from the cohort of 100,000 at each age. We then divided the number of people estimated to be alive with an IBD by the overall number of people estimated to be alive to determine the prevalence of IBD at each age.

Next, we applied the US Census weights to each of the ages from 0 to 100 years and summed these weighted prevalence estimates to generate the expected age-standardized prevalence in the US. We qualitatively compared this to our pooled estimate of prevalence derived from the 4 claims data sets.

To assess for secular trends, we computed the prevalence of IBD in the CDM, HealthCore, and Medicare cohorts using our primary definition on December 31 of the years 2011, 2014, 2017, and 2020. Medicare data were not available for 2020. We used linear regression adjusted for the data source to test for linear trends in prevalence across time.

Results

Incidence of Inflammatory Bowel Disease in the United States

In the primary analysis of incidence, where we required 4 years of enrollment before the start of follow-up, the combined cohort contributed 42,964,750 person-years of follow-up, during which 4747 met the high-probability definition of a newly diagnosed IBD and 2221 met the lower-probability definition. The age- and sex-standardized incidence of IBD per 100,000 person-years in the United States was 10.9 (95% CI, 10.6–11.2). In a sensitivity analysis including both the high- and lower-probability algorithms, the pooled incidence rate per 100,000 person-years was 15.9 (95% CI, 15.5–16.3). The incidence of IBD, ulcerative colitis, and Crohn's disease peaked in the third decade of life, decreased to a stable relatively stable level across the fourth to eighth decades, and declined further beyond age 80 (Figure 1 and Supplementary Table 1). Overall, the incidence of ulcerative colitis (6.3; 95% CI, 6.1–6.6) was higher than that of Crohn's disease (4.1; 95% CI, 3.9–4.3). However, among children, the incidence of Crohn's disease was higher than ulcerative colitis (Figure 1 and Supplementary Table 1).

Prevalence of Inflammatory Bowel Diseases in the United States

We analyzed data for 14,420,692 individuals with ≥4 years of continuous insurance, of whom 115,715 met the primary definition of IBD requiring ≥2 diagnoses or 1 diagnosis and a prescription for an IBD medication. The age-, sex-, and insurance-standardized prevalence per 100,000 population was 721 (95% CI, 717–726) for IBD, 378 (95% CI, 375–382) for ulcerative colitis, and 305 (95% CI, 302–308) for Crohn's disease. Extrapolated to the 2020 US Census estimate of 331,449,281 US population, there are an estimated 2.390 million Americans with IBD, 1.253 million with ulcerative colitis, and 1.011 million with Crohn's disease. The combined estimates for ulcerative colitis and Crohn's disease do not total that of IBD due to patients identified with IBD but who could not be assigned specifically to either ulcerative colitis or Crohn's disease. Prevalence estimates using a range of other assumptions are included in Table 1.

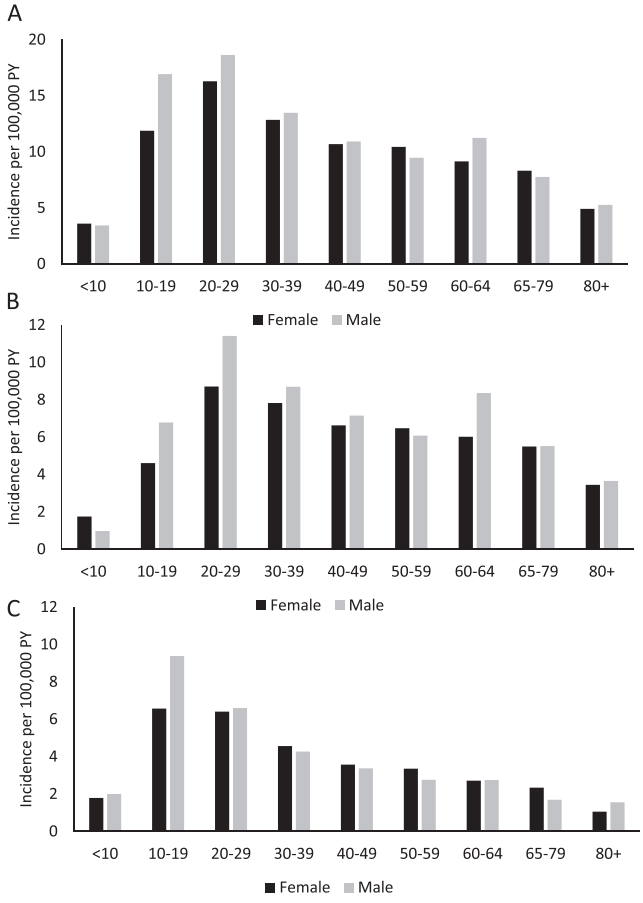


Figure 1. Age- and sex-specific incidence per 100,000 person-years (PY) for (*A*) IBD, (*B*) ulcerative colitis, and (*C*) Crohn's disease in the US.

Table 1. Sensitivity Analysis Using Different Criteria to Determine Prevalence of Inflammatory Bowel Disease, Ulcerative Colitis, and Crohn's Disease

Diagnosis	Prevalence definition[a]	Minimum enrollment	Prevalence per 100,000 population (95% CI)
IBD overall	Primary	4 years	721 (717–726)
IBD overall	Secondary	4 years	826 (821–831)
IBD overall	Primary	1 year	600 (597–603)
IBD overall	Secondary	1 year	680 (677–683)
Crohn's disease	Primary	4 years	305 (302–308)
Crohn's disease	Secondary	4 years	347 (344–350)
Crohn's disease	Primary	1 year	258 (256–260)
Crohn's disease	Secondary	1 year	290 (288–292)
Ulcerative colitis	Primary	4 years	378 (375–382)
Ulcerative colitis	Secondary	4 years	438 (435–441)
Ulcerative colitis	Primary	1 year	312 (310–314)
Ulcerative colitis	Secondary	1 year	358 (356–361)

[a]Secondary definition includes patients with a single diagnosis of IBD by a gastroenterologist or surgeon, or ≥2 diagnoses by providers other than gastroenterologists or surgeons, without any therapy, applying a weight of 0.22

Although the prevalence of ulcerative colitis was somewhat higher than that of Crohn's disease in most age-groups, this trend was reversed in the pediatric population. The prevalence of IBD was slightly higher in boys among children and women among adults. As expected, the prevalence of IBD overall, ulcerative colitis, and Crohn's disease generally increased with age, although there was a drop in the prevalence among those aged >80 years, particularly among those with Crohn's disease (Figure 2 and Supplementary Table 2).

The prevalence of IBD per 100,000 population was 812 (95%, CI 802–823) in White, 504 (95% CI, 482–526) in Black, 403 (95% CI, 373–433) in Asian, and 458 (95% CI, 440–476) in Hispanic Americans. The higher prevalence among White Americans was observed for both Crohn's disease and ulcerative colitis (Figure 3*A*). Of note, the ratio of ulcerative colitis to Crohn's disease was higher in Asian (1.6:1) and Hispanic (1.8:1) than in White (1.2:1) or Black (1.2:1) Americans.

The prevalence of IBD was the highest in the Northeast and lowest in the Western region of the US. However, the relative prevalence of ulcerative colitis and Crohn's disease was similar across regions (Figure 3*B*).

The prevalence of IBD in adults aged 20 to 64 years was nearly identical in Medicaid, HealthCore, and CDM; however, among children, the prevalence was ∼40% lower in Medicaid relative to the commercial plans (Supplementary Figure 1). When stratified by race, this was more evident in Black children (prevalence ratio for Medicaid vs CDM 0.7 in girls and 0.6 in boys) than in White (girls, 0.9; boys, 0.9) or Hispanic (girls, 1.0; boys, 1.1) children. There were too few Asian children with an IBD in Medicaid for reliable comparisons.

Concordance of Incidence and Prevalence

We assessed concordance of the pooled incidence and prevalence estimates by computing the expected age-standardized prevalence from the pooled age-specific incidence rates using the high probability algorithm and a 4-year minimum enrollment and comparing this to the pooled prevalence estimates using our algorithm with definitions 1 to 3 and a minimum 4-year enrollment. The estimated prevalence derived from our pooled high-probability incidence rates was 442 per 100,000 and from combined high- and low-probability incidence algorithms was 634 per 100,000. This latter estimate is close to the pooled prevalence estimate of 721 per 100,000.

Secular Trends

We examined change prevalence over time in the CDM, HealthCore, and Medicare populations from 2011 to 2020. The prevalence of IBD increased gradually during this time ($P = .04$) (Table 2).

Discussion

The prevalence of IBD in North America is among the highest in the world.[2,3] However, prior research in select populations has led to inconsistent estimates of the prevalence of IBD in the US. To overcome those limitations, in this study, we pooled data from commercial, Medicare, and Medicaid insurance plans to derive a population-based estimate of the incidence and prevalence of IBD throughout the US. The primary estimate of the prevalence of IBD in the US of 721 per 100,000 population extrapolates to an

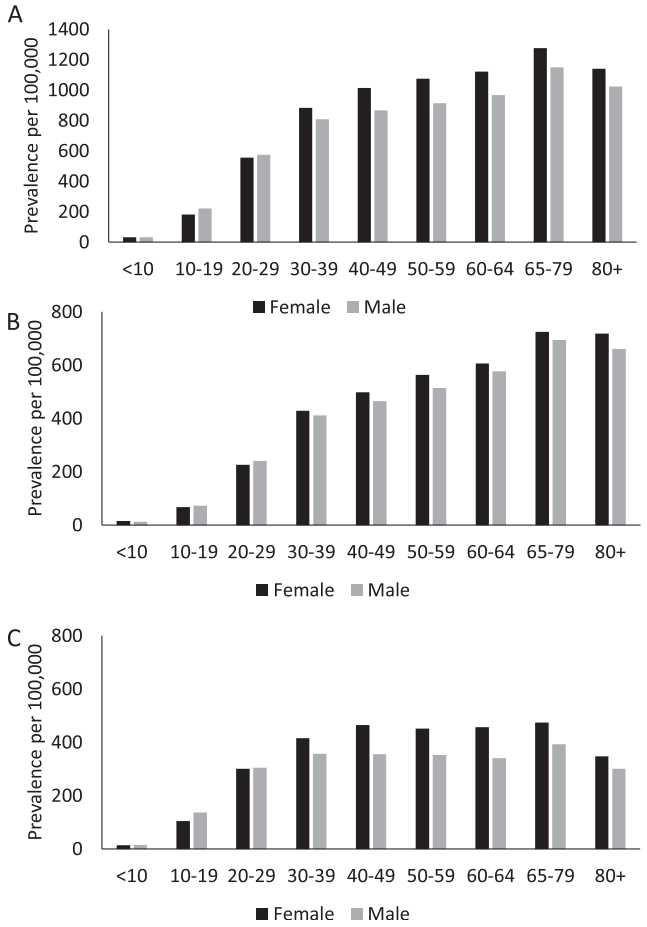


Figure 2. Age- and sex-specific prevalence per 100,000 population of (*A*) IBD, (*B*) ulcerative colitis, and (*C*) Crohn's disease in the US.

estimate of 2.39 million Americans with IBD. Our secondary analysis was ~15% higher, extrapolating to 2.74 million Americans with IBD. The age- and sex-standardized

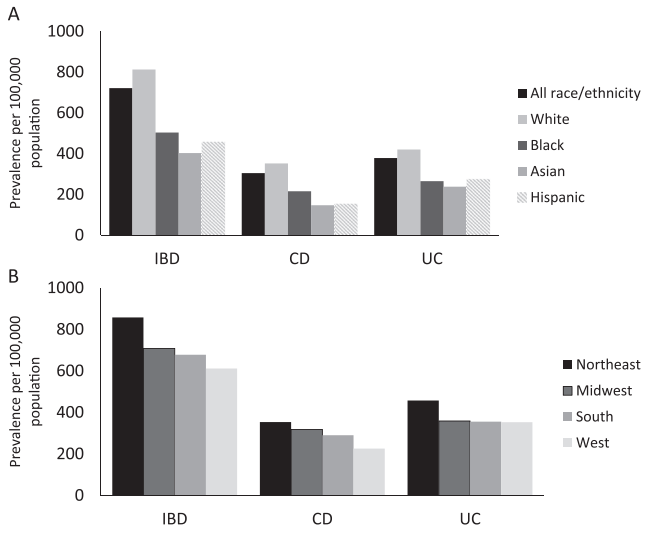


Figure 3. Age- and sex-adjusted prevalence of IBD, ulcerative colitis (UC), and Crohn's disease (CD) by (*A*) race and ethnicity and (*B*) region.

Table 2. Secular Trends in Prevalence of Inflammatory Bowel Disease

Year	Age- and sex-standardized prevalence per 100,000 population (95% CI)		
	CDM[a]	HealthCore[a]	Medicare[b]
2011	626 (618–633)	572 (567–577)	949 (932–966)
2014	644 (637–652)	650 (645–655)	1184 (1166–1202)
2017	659 (652–667)	695 (689–700)	1282 (1266–1298)
2020	654 (647–662)	725 (720–730)	Data not available

[a]Includes age <65 years and Medicare Advantage patients aged >65 years.
[b]Includes only fee-for-service Medicare beneficiaries aged >65 years.

estimates of incidence translate to ~39,000 to 56,000 new IBD diagnoses per year in the US. Thus, the burden of caring for these lifelong diseases is high and will likely increase as life expectancy increases.

The estimated incidence and prevalence of IBD from this study should be considered in the context of prior estimates from the US. We estimated the incidence of IBD to be between 11.8 and 17.0 per 100,000 person-years. This estimate is generally consistent with prior estimates from Olmstead County, Minnesota (10.7 per 100,000 for Crohn's disease and 12.2 per 100,000 for ulcerative colitis),[6] and Northern California (6.3 per 100,000 for Crohn's disease and 12.0 per 100,000 for ulcerative colitis).[5] In Olmstead County, the estimated prevalence of IBD in 2010 was 522.9 per 100,000 population, or ~1.6 million Americans with IBD.[6] Similarly, a 2016 estimate from CDM and Truven, a second commercial data source, estimated that there were ~1.4 million Americans with IBD,[11] and in the recent Global Burden of Disease Study that derived estimates from prior publications, the overall prevalence estimate was 1.8 million Americans.[3]

Our estimate was somewhat higher than those from prior administrative claims-based studies. A notable difference between our study and prior studies is that we included a more representative population by pooling data from 4 sources, including Medicare. With the aging of the population, compounding prevalence may contribute to a rise in prevalence over time.[2,12] In addition, by requiring patients to have 4 years of enrollment with their health insurance, we may have been better able to capture patients with less severe disease and who therefore have less frequent physician encounters for IBD.

In contrast to studies conducted using administrative data, the NHIS estimated there were 3.1 million Americans previously diagnosed with IBD,[4] whereas the 2009 to 2010 National Health and Nutrition Examination Survey (NHANES) estimated that there were 2.3 million Americans diagnosed with IBD.[13] The NHIS and NHANES studies relied on self-report of the IBD diagnosis The NHANES estimate is nearly identical to our primary estimate, and the NHIS estimates is rather close to our secondary estimate of 2.74 million Americans with IBD. Differences between our estimate and the prior NHIS estimate may reflect inaccuracy of

self-report in the NHIS study or underestimation of IBD in claims-based analyses, or both, because some patients may rarely see a physician or receive treatment for their IBD, such that they may not be detected by algorithms requiring multiple diagnoses of IBD or IBD prescriptions, even with a 4-year minimum enrollment period.

Our prevalence estimates from the US are similar to recent estimates from other high-prevalence regions in Europe and Canada.[2] For example, the estimated prevalence per 100,000 population was 744 in Germany in 2010,[9] 770 in Norway in 2017,[14] and 872 in Denmark in 2020.[15] In Canada, the prevalence of IBD in 2018 was estimated at 700 per 100,000, based on extrapolation of data from 2002 to 2008.[16]

The prevalence of IBD has been described as passing through 4 phases: emergence of IBD, acceleration of incidence, compounding prevalence, and finally, prevalence equilibrium.[17] We observed a gradual increase in prevalence across the last decade, suggesting that the US had not reached prevalence equilibrium. This is consistent with a recent study of Medicaid data that showed an increasing prevalence of CD during the last decade, although most of that occurred before 2016.[18]

There is a general paucity of data on the racial/ethnic distribution of IBD in the US. We estimated that the prevalence of IBD was nearly twice as high among non-Hispanic White Americans compared with Black, Hispanic, and Asian Americans. Our race- and ethnic-specific results are consistent with prior studies using Medicaid and Medicare claims and electronic health record data in which the prevalence in Black Americans was ~40% lower than in White Americans.[7,18,19] In the 2009 to 2010 NHANES, self-reported diagnosis with IBD was ~0.8% in non-Hispanic Black Americans vs 1.4% in non-Hispanic White Americans and 1.6% in Mexican Americans.[13] The NHIS had very similar results, with a self-reported IBD diagnosis of 0.5% in non-Hispanic Black participants, 1.2% in Hispanic participants, and 1.4% in non-Hispanic White particiapants.[4] Both of these studies benefited from self-reported race and ethnicity. In contrast, much of the race and ethnicity data in our study were imputed, which may underdetect or overdetect minority populations. Moreover, the uninsured population, which is not included in this study, are more likely to be Hispanic. It is reassuring that our estimates were very similar to the studies relying on self-reported race and ethnicity.

To put these data into context, it is important to consider the overall make-up of the US population. According to the 2020 US Census, 60.1% of Americans reported being non-Hispanic White, 13.4% Black alone, and 5.9% Asian alone; 18.5% of Americans reported being Hispanic. Based on a population of ~331 million people, one can estimate that there are ~224,000 Black, 79,000 Asian, and 281,000 Hispanic Americans with IBD compared with 1.6 million White Americans with IBD. Thus, we estimate that relative to Black, Asian, and Hispanic Americans, there are 7-times, 21-times, and 6-times more White Americans with IBD, respectively. However, these estimates need to be viewed with caution because the race and ethnicity data derived from Medicaid, Medicare, and CDM are not all from self-report. Rather, much of the race data are imputed based on statistical algorithms. Moreover, these must be interpreted as reflecting race as a social construct rather than as a biologic construct. Nonetheless, this study provides critically important and novel estimates of the racial distribution of IBD in the US.

We observed a lower prevalence of IBD in the children who were Medicaid beneficiaries, particularly Black children, but not in young adults. Because we only observed this finding in children, it is less likely that this is a bias in the design of the study. These hypothesis-generating data suggest the possibility that poverty has a greater impact on children with IBD than adults. Children depend on their guardians to advocate for them. Among possible explanations for this observation would be that poverty may make it more difficult for guardians to seek medical care for their children due to many different social determinants of health that are linked to poverty, thereby leading to underdiagnosis. An alternative hypothesis is that physicians are less likely to diagnose IBD in children from low-income families.

Prior studies have suggested slightly increased mortality among patients with IBD.[20,21] We observed a rise in prevalence of IBD, ulcerative colitis, and Crohn's disease with age, but a slight drop above age 80, particularly for Crohn's disease. This could reflect greater excess mortality in the oldest patients with IBD, and Crohn's disease in particular, as has been suggested in other populations.[22]

This study is unique in having pooled data that are representative of nearly the entire US population with health insurance. However, even with this complex design, we did not capture those with insurance through the Veterans Affairs system and the uninsured. We hypothesize that there are few patients diagnosed with IBD who lack health insurance, because people with a chronic disease would be less likely to forgo having insurance and may qualify for government-sponsored health plans. Under this hypothesis, missing data on the uninsured may result in an overestimate of the prevalence of IBD. An alternative hypothesis is that some people with IBD are unable to afford health insurance, which would mean that missing data on the uninsured population could bias to underestimating the prevalence of IBD. Prior studies within the US Veterans Affairs health system documented nearly identical prevalence estimates as ours.[23] As such, the lack of inclusion of the uninsured population in our estimate is unlikely to have significantly biased the results.

The sensitivity, specificity, and PPV of our claims-based algorithms are not 100%. To the extent that sensitivity is imperfect, we may have underestimated the prevalence and incidence. In contrast, to the extent that specificity and PPV are imperfect, we may have overestimated incidence and prevalence. Although it is impossible to perfectly balance underestimation due to imperfect sensitivity and overestimation based on imperfect specificity, we believe that our study provides a reliable range of estimates. For example, our primary definition of prevalence does not include patients with a single diagnosis of IBD in the claims data and no prescribed therapies. However, in our prior validation study, ≥22% of such patients were confirmed to have IBD, such as patients with a history of total colectomy for ulcerative colitis in the distant past. In our sensitivity

analysis, adding in such patients and applying a conservative weight of 0.22 to account for the low PPV increased the prevalence estimate by ∼15%.

The component of the primary prevalence algorithm with the lowest PPV was for patients who had 1 IBD diagnosis and at least 1 prescription, but most had multiple diagnoses with or without prescriptions. For example, within the CDM cohort, only 8.6% of the prevalent patients in the primary analysis had a single IBD diagnosis. Thus, the degree of overestimation that may have resulted from this is small. These sensitivity analyses allow for a more nuanced interpretation of the results, understanding that all prevalence estimates contain some degree of measurement error.

The racial and ethnic composition of the individual data sets used in this study differ from the overall population. Medicaid beneficiaries are more likely to be Black or Hispanic, whereas commercially insured populations have a higher proportion of White Americans than the general population. To address this, we used direct standardization to ensure that the incidence and prevalence estimates accounted for age, sex, and type of insurance.

A small proportion of Medicare beneficiaries do not have outpatient coverage (Part B) or prescription drug coverage (Part D). The former are more likely to be White and have higher income and the latter to have lower income and more medical conditions.[24] Given the small proportions, <10% for each, and the contrasting nature of these 2 groups, this is not expected to meaningfully impact the results.

Conclusions

In summary, we believe this to be the most comprehensive assessment of physician-diagnosed IBD in the US to date. IBD is a relatively common chronic condition, affecting >0.7% of Americans and is most prevalent in the Northeastern region. The incidence peaks in early adulthood and then plateaus at a lower rate. The disease is less commonly diagnosed in Black, Asian, and Hispanic Americans. However, ascertaining whether this is due to detection bias or biologic differences is not possible from these data. Future investigation is essential to understand the causes and consequences of these observed differences based on race and ethnicity. Finally, the lower prevalence in children with Medicaid insurance highlights the importance of additional research to understand the impact of social determinants of heath on the care of IBD.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at www.gastrojournal.org, and at https://doi.org/10.1053/j.gastro.2023.07.003.

References

1. Molodecky NA, Soon IS, Rabi DM, et al. Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. Gastroenterology 2012;142:46–54.e42; [quiz: e30].
2. Ng SC, Shi HY, Hamidi N, et al. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. Lancet 2017;390:2769–2778.
3. Alatab S, Sepanlou SG, Ikuta K, et al. The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet Gastroenterol Hepatol 2020;5:17–30.
4. Dahlhamer JM, Zammitti EP, Ward BW, et al. Prevalence of inflammatory bowel disease among adults aged ≥18 years—United States, 2015. MMWR Morb Mortal Wkly Rep 2016;65:1166–1169.
5. Herrinton LJ, Liu L, Lewis JD, et al. Incidence and prevalence of inflammatory bowel disease in a Northern California managed care organization, 1996-2002. Am J Gastroenterol 2008;103:1998–2006.
6. Shivashankar R, Tremaine WJ, Harmsen WS, et al. Incidence and prevalence of Crohn's disease and ulcerative colitis in Olmsted County, Minnesota from 1970 through 2010. Clin Gastroenterol Hepatol 2017;15:857–863.
7. Barnes EL, Nowell WB, Venkatachalam S, et al. Racial and ethnic distribution of inflammatory bowel disease in the United States. Inflamm Bowel Dis 2022;28:983–987.
8. Dawwas GK, Weiss A, Constant BD, et al. Development and validation of claims-based definitions to identify incident and prevalent inflammatory bowel disease in administrative healthcare databases. Inflamm Bowel Dis. Published online April 12, 2023. https://doi.org/10.1093/ibd/izad053.
9. DeFrank JT, Bowling JM, Rimer BK, et al. Triangulating differential nonresponse by race in a telephone survey. Prev Chronic Dis 2007;4:A60.
10. Jarrin OF, Nyandege AN, Grafova IB, et al. Validity of race and ethnicity codes in Medicare Administrative Data compared with gold-standard self-reported race collected during routine home health care visits. Med Care 2020;58:e1–e8.
11. Ye Y, Manne S, Treem WR, et al. Prevalence of inflammatory bowel disease in pediatric and adult populations: recent estimates from large national databases in the United States, 2007-2016. Inflamm Bowel Dis 2020;26:619–625.
12. Kuenzig ME, Fung SG, Marderfeld L, et al. Twenty-first century trends in the global epidemiology of pediatric-onset inflammatory bowel disease: systematic review. Gastroenterology 2022;162:1147–1159.e1144.
13. Weisman MH, Oleg S, Seok Kim H, et al. Inflammatory bowel disease prevalence: surveillance data from the U. S. National Health and Nutrition Examination Survey. Prev Med Rep 2023;33:102173.
14. Lirhus SS, Høivik ML, Moum B, et al. Incidence and prevalence of inflammatory bowel disease in Norway and the impact of different case definitions: a nationwide registry study. Clin Epidemiol 2021;13:287–294.
15. Larsen L, Karachalia Sandri A, Fallingborg J, et al. Has the incidence of inflammatory bowel disease peaked? Evidence from the population-based NorDIBD Cohort 1978-2020. Am J Gastroenterol 2023;118:501–510.

16. Coward S, Clement F, Benchimol EI, et al. Past and future burden of inflammatory bowel diseases based on modeling of population-based data. Gastroenterology 2019;156:1345–1353 e1344.
17. Kaplan GG, Windsor JW. The four epidemiological stages in the global evolution of inflammatory bowel disease. Nat Rev Gastroenterol Hepatol 2021;18:56–66.
18. Hutfless S, Jasper RA, Chen PH, et al. Burden of Crohn's disease in the United States Medicaid population, 2010–2019. Clin Gastroenterol Hepatol. Published online May 13, 2023. https://doi.org/10.1016/j.cgh.2023.05.002.
19. Xu F, Carlson SA, Liu Y, et al. Prevalence of inflammatory bowel disease among Medicare fee-for-service beneficiaries—United States, 2001−2018. MMWR Morb Mortal Wkly Rep 2021;70:698–701.
20. Duricova D, Pedersen N, Elkjaer M, et al. Overall and cause-specific mortality in Crohn's disease: a meta-analysis of population-based studies. Inflamm Bowel Dis 2010;16:347–353.
21. Bewtra M, Kaiser LM, TenHave T, et al. Crohn's disease and ulcerative colitis are associated with elevated standardized mortality ratios: a meta-analysis. Inflamm Bowel Dis 2013;19:599–613.
22. **Zhulina Y, Udumyan R**, Tysk C, et al. Mortality in patients with Crohn's disease in Örebro, Sweden 1963–2010. Scand J Gastroenterol 2022;57:158–164.
23. Hou JK, Kramer JR, Richardson P, et al. The incidence and prevalence of inflammatory bowel disease among U. S. veterans: a national cohort study. Inflamm Bowel Dis 2013;19:1059–1064.
24. Tarazi W, Welch W, Nguyen N, et al. Medicare Beneficiary Enrollment Trends and Demographic Characteristics. Issue Brief No. HP2022-08. Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services; March 2022.

Author names in bold designate shared co-first authorship.

Received February 9, 2023. Accepted July 7, 2023.

Correspondence
Address correspondence to: James D. Lewis, MD, MSCE, University of Pennsylvania, 720 Blockley Hall, 423 Guardian Drive, Philadelphia, Pennsylvania 19104. e-mail: lewisjd@pennmedicine.upenn.edu.

CRediT Authorship Contributions
James D. Lewis, MD, MSCE (Conceptualization: Lead; Funding acquisition: Equal; Methodology: Equal; Project administration: Equal; Supervision: Equal; Writing – original draft: Lead).
Lauren Parlett, PhD (Data curation: Equal; Formal analysis: Equal; Methodology: Supporting; Writing – review & editing: Equal).
Michele L. Jonsson Funk, PhD (Conceptualization: Supporting; Methodology: Equal; Writing – review & editing: Equal).
Colleen Brensinger, MS (Data curation: Equal; Formal analysis: Lead; Writing – review & editing: Equal).
Virginia Pate, MS (Data curation: Equal; Formal analysis: Equal; Writing – review & editing: Equal).
Qufei Wu, MS (Data curation: Equal; Formal analysis: Equal; Writing – review & editing: Equal).
Ghadeer Dawwas, PhD (Methodology: Equal; Writing – review & editing: Equal).
Alexandra Weiss, MD (Data curation: Supporting; Writing – review & editing: Equal).
Brad Constant, MD (Methodology: Equal; Writing – review & editing: Equal).
Maureen McCauley, BA (Data curation: Supporting; Project administration: Supporting; Writing – review & editing: Equal).
Kevin Haynes, PhD (Methodology: Equal; Writing – review & editing: Equal).
Jeff Yang, PhD (Investigation: Supporting; Writing – review & editing: Equal).
Douglas Schaubel, PhD (Methodology: Equal; Supervision: Supporting; Writing – review & editing: Equal).
Andres Hurtado-Lorenzo, PhD (Funding acquisition: Equal; Project administration: Equal; Writing – review & editing: Equal).
Michael Kappelman, MD, MPH (Conceptualization: Equal; Funding acquisition: Equal; Methodology: Equal; Project administration: Equal; Writing – review & editing: Equal).

Conflicts of interest
These authors disclose the following: Lauren E. Parlett is an employee of Elevance Health and has received funding from Sanofi for an unrelated study. James D. Lewis consulted or served on an advisory board for Eli Lilly and Company, Samsung Bioepis, UCB, Bristol-Myers Squibb, Nestlé Health Science, Merck, Celgene, Janssen Pharmaceuticals, Bridge Biotherapeutics, Entasis Therapeutics, AbbVie, Pfizer, Gilead, Galapagos, Sanofi, Arena Pharmaceuticals, Protagonist Therapeutics, Amgen, and Scipher Medicine, has had research funding from Nestlé Health Science, Takeda, Janssen Pharmaceuticals, and AbbVie, and has had educational grants from Takeda and Janssen. Michele L. Jonsson Funk receives salary support as Center Director of the Center for Pharmacoepidemiology, Department of Epidemiology, University of North Carolina Chapel Hill, which has collaborative agreements with AbbVie, Astellas, Boehringer Ingelheim, GlaxoSmithKline, Sarepta, Takeda, and UCB Biosciences. Michael David Kappelman has consulted for AbbVie, Pfizer, and Lilly, is a shareholder in Johnson & Johnson, and has received research support from Janssen and AbbVie. Kevin Haynes is an employee of Janssen. The remaining authors disclose no conflicts.

Funding
Ghadeer K. Dawwas receives funding from National Institutes of Health. This study received funding from the Centers for Disease Control and Prevention (U01-DP006369) and from the National Institutes of Health National Institute of Diabetes and Digestive and Kidney Diseases (P30-DK050306) And National Center for Advancing Translational Sciences (UL1TR002489).

Supplementary Methods and Results

Inclusion Criteria by Data Source

The inclusion criteria varied slightly by data source.

Medicare: Must have 4 years of continuous fee-for-service coverage with no gap greater than 1 month. Must have at least 1 month with Medicare Parts A, B, and D coverage.

Medicaid: Must have 4 years of continuous fee-for-service coverage. Must be aged <61 years at the start of the 4 years of continuous coverage

CDM: Must have 4 years of continuous coverage. All patients have pharmacy benefits coverage.

HealthCore: Must have 4 years of continuous coverage, with or without pharmacy benefits.

Time Period of Data Used in the Incidence and Prevalence Analyses

The time periods of the data used to compute incidence and prevalence varied slightly by the data source and the analysis. This is summarized in the following table.

Data source	Incidence	Prevalence
Medicare	January 2007–December 2017[a]	January 2007–December 2017
Medicaid	January 1999–December 2012[b]	January 1999–December 2012
CDM	May 2000–December 2017[c]	May 2000–December 2017
HealthCore	January 2006–June 2018[d]	January 2006–December 2017

[a]Could only be diagnosed as incident IBD between January 1, 2015, and December 31, 2017.
[b]Could only be diagnosed as incident IBD between July 1, 2009, and June 30, 2012.
[c]Could only be diagnosed as incident IBD between July 1, 2014, and June 30, 2017.
[d]Could only be diagnosed as incident IBD between July 1, 2014, and June 30, 2018.

Censoring Rules for Incidence Calculations

In each data set, patients were censored for death or loss of health insurance benefits. In Medicaid, patients were censored when they reached age 65 years or began a managed Medicaid plan.

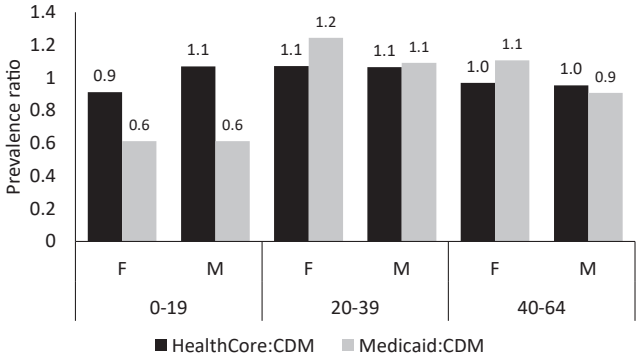
Avoiding Duplicate Counting of People

Although people may switch between insurance plans, duplicate counting of individuals was avoided by the following:

- People do not have 2 of these insurance plans as their primary insurance at the same time.
- People may have both Medicare and Medicaid benefits. Those aged <65 were included in the Medicaid population. Medicaid beneficiaries aged >65 were excluded. Thus, those aged >65 with both Medicare and Medicaid were included in the Medicare cohort.
- Medicare includes both fee-for-service and managed plans; however, a person may only have 1 of these benefits at a time. We used managed Medicare data from Anthem and CDM and only fee-for-service Medicare beneficiaries in the random sample data received from the Center for Medicare and Medicaid Services.
- There is a small chance that Medicaid beneficiaries could have been counted again in the commercial or Medicare data for the following reasons:
  - Because the Medicaid data ended in 2012, there is a small possibility that a patient could have been captured in Medicaid and later in commercial or Medicare data.
  - There are some states where children with IBD can have both Medicaid and commercial insurance.
  - Given the relatively small portion of Americans with Medicaid and the low probability that patients were double counted, this is expected to have a very small impact on the results.

Provider Specialty

Each data source has one or more variables that identify provider specialty. In CDM, we internally validated this code by examining the relationship between the code and performance of colonoscopy. In 2017, colonoscopies and sigmoidoscopies were far more likely to be performed by gastroenterologists (median, 44; interquartile range [IQR], 16–101) and colorectal surgeons (median, 16; IQR, 6–39) than by family practitioners (median, 2; IQR, 1–4), internists (median, 1; IQR, 1–8), or general surgeons (median, 1; IQR, 1–7). These data strongly support the validity of the provider specialty data in the commercial claims data. In Medicaid, however, physician specialty data were incomplete. Physician specialty can be identified in Medicaid using a specific provider taxonomy variable or by linking to the National Provider Identifier number of the provider. Unfortunately, there was a high level of missing data for lower endoscopy claims using either of these methods to identify provider specialty (nearly 50%). As such, we were not able to implement the coding requirement for physician specialty in the Medicaid data. The one other exception is that HealthCore's data did not distinguish colorectal surgeons from other surgeons, and as such, all surgeons were counted in the algorithms.



Supplementary Figure 1. Prevalence ratio comparing Medicaid, HealthCore, and CDM. F, female; M, male.