HHS Public Access

Author manuscript

Gastroenterology. Author manuscript; available in PMC 2024 November 01.

Published in final edited form as:

Gastroenterology. 2023 November ; 165(5): 1197–1205.e2. doi:10.1053/j.gastro.2023.07.003.

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], **Michael David Kappelman**[5]

[1.]Division of Gastroenterology and Hepatology, University of Pennsylvania, Philadelphia, PA.

[2.]Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

[3.]Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

[4.]HealthCore, Inc. Wilmington, DE

Contact information for corresponding author: James D Lewis, MD, MSCE, 720 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104 lewisjd@pennmedicine.upenn.edu.

Author contributions

Author	Study concept and design	Acquisition of data	Analysis and interpretation	Drafting manuscript	Critical revisions	Statistical analysis
Lewis	X	X	X	X	X	
Parlett	X	X	X		X	X
Jonsson-Funk	X	X	X		X	
Brensinger	X		X		X	X
Pate	X		X		X	X
Wu			X		X	X
Dawwas	X		X		X	
Weiss			X		X	
Constant	X		X		X	
McCauley		X			X	
Haynes	X		X		X	
Yang			X		X	
Schaubel	X		X		X	
Hurtado-Lorenzo			X		X	
Kappelman	X		X		X	

[5]University of North Carolina at Chapel Hill, Chapel Hill, NC

[6]Division of Gastroenterology, Hepatology, and Nutrition, The Children's Hospital of Philadelphia, Philadelphia, PA

[7]Janssen Research and Development, LLC, Titusville, NJ

[8]Crohn's & Colitis Foundation, New York, NY

Abstract

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 utilized four administrative claims datasets: 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 de-identified 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-specific insurance-weighted estimates and pooled estimates standardized to 2018 United States census estimates with 95% confidence intervals (CI).

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 4th to 8th 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 census, there is an estimated 2.39 million Americans diagnosed with IBD. The prevalence of IBD per 100,000 population was 812 (95% CI 802 – 823) in White, 504 (482–526) in Black, 403 (373 – 433) in Asian and 458 (440–476) in Hispanic Americans.

Conclusions: IBD is diagnosed in more than 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.

Lay summary

There are approximately 2.4 to 2.7 million Americans diagnosed with inflammatory bowel diseases. The prevalence of IBD was highest among non-Hispanic whites and residents of the Northeast.

Keywords

Crohn's disease; ulcerative colitis; epidemiology; Medicare; Medicaid; race

Introduction

Inflammatory bowel diseases (IBD) include 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 incidence and prevalence of IBD in the United States. The highest estimate comes from the National Health Interview Study, although this was based on patients' self-report.[4] Other population-based studies from Olmstead County, MN 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 United States.[7]

The paucity of rich data on the incidence, prevalence, and racial and ethnic distribution of IBD in the United States stems from the lack of a unified health system with a common medical record or central data repository. Rather, in the United States, 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 United States 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 and geographic distribution of IBD.

Methods

Data sources

The study utilized four administrative claims datasets. 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 age 65 or older with at least one 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 as recording of diagnoses is less complete in Medicaid managed care plans. We also used two 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 (i.e. < age 65) from fourteen states and members with Elevance Health-managed Medicare plans. We also used claims data from Optum's de-identfied Clinformatics Data Mart (CDM), a collection of anonymized patient-level insurance data from 2000 to 2017. There is no overlap of Medicare, Medicaid, HealthCore and CDM data used in this research. See supplemental methods for additional details.

Each data set contains billing data for physician encounters, including diagnoses recorded using International Classification of Disease 9th edition, Coding Manual (ICD9-CM) (prior to October 1, 2015) or 10th edition (ICD10). Prescription drug claims are coded using National Drug Codes (NDC) and include quantity dispensed and days supplied. Procedures, including infused medications, are classified using American Medical Association's Current Procedural Terminology (CPT) and CMS Common Procedure Coding System (HCPCS) 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. For this research, we assumed that the variable aligned with the recording of sex within the United States Census data.

Algorithms to identify incident and prevalent diagnosis of IBD

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 prior to the first diagnosis of IBD in the claims data by any provider, 2) no prior therapy with a medication used to treat IBD unless there was another indication, such as rheumatoid arthritis in a patient treated with an anti-TNF medication, 3) had a colonoscopy, sigmoidoscopy, capsule endoscopy or bowel resection surgery within 6 weeks prior to the first recorded diagnosis code, and 4) had a second diagnosis of 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 either: 1) a first prescription for steroids (oral or rectal) and/or mesalamine (oral or rectal), sulfasalazine, olsalazine, balsalazide, adalimumab, infliximab, golimumab, or certolizumab within 90 days following 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 one IBD diagnosis by any provider regardless of specialty combined with either one or more IBD diagnosis by a gastroenterologist or surgeon or a therapy with a medication for IBD in the absence of another indication. We previously established that the PPV of this algorithm was 94% for two or more diagnoses by a gastroenterologist or surgeon and receipt of IBD specific medications, 92% for two or more diagnoses by a gastroenterologist or surgeon without receipt of IBD medications and 78% for one 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 prior to 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 supplemental methods for details on identification of provider specialty.

We categorized patients as having Crohn's disease versus ulcerative colitis if the most recent diagnosis was the same as most frequent diagnosis on or before the data of meas[8]urement. 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 prior to 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 census. For patients over the age of 65, we pooled Medicare fee for service data obtained from CMS with Medicare Advantage data from CDM and Anthem claims. For patients over age 65, the insurance weights were derived from data published by the Kaiser Foundation on Medicare beneficiaries enrollment in fee for service versus 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 USA 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 (e.g., driver's license records) for approximately 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 as compared to 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.[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 four 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 and/or the secondary definition of prevalence that included patients with a single IBD diagnosis by a gastroenterologist or surgeon, or 1 or more by any provider other than a gastroenterologist or surgeon, and no prescribed therapy. Because the PPV of one 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.[11] 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 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 lifetables 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 IBD derived from the four claims data sets to determine the number of people newly diagnosed with 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 50^{th} year of life as in the 59^{th} 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 IBD from the cohort of 100,000 at each age. We then divided the number of people estimated to be alive with 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 United States. We qualitatively compared this to our pooled estimate of prevalence derived from the four 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 IBD 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 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 algorithm, the pooled incidence rate per 100,000 person-years was 15.9 (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 4th to 8th decades, and declined further beyond age 80. (Figure 1, Supplemental 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, Supplemental Table 1).

Prevalence of IBD in the United States

We analyzed data for 14,420,692 individuals with 4 or more years of continuous insurance of whom 115,715 met the primary definition of IBD requiring two or more diagnoses or one 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 per 100,000 (95% CI 375 – 382) for ulcerative colitis and 305 (95% CI 302 – 308) for Crohn's disease. Extrapolated to the 2020 census estimate of 331,449,281 US population, there is 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.

While 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 males among children and females 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 over 80 years of age, particularly among those with Crohn's disease (Figure 2, Supplemental Table 2).

The prevalence of IBD per 100,000 population was 812 (95% CI 802 – 823) in White, 504 (482–526) in Black, 403 (373 – 433) in Asian and 458 (440–476) in Hispanic Americans. The higher prevalence among White Americans was observed for both Crohn's disease and ulcerative colitis. (Figure 3A). 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 United States. However, the relative prevalence of ulcerative colitis and Crohn's disease were similar across regions (Figure 3B).

The prevalence of IBD in adults age 20 – 64 years was nearly identical in Medicaid, HealthCore and CDM; however, among children, the prevalence was approximately 40% lower in Medicaid relative to the commercial plans (Supplemental Figure 1). When stratified by race, this was more evident in Black children (prevalence ratio for Medicaid vs CDM 0.7 in females and 0.6 in males) than in White (females 0.9 and males 0.9) or Hispanic (females 1.0 and males 1.1). There were too few Asian children with 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 1–3 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 United States. 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 United States. The primary estimate of the prevalence of IBD in the United States of 721 per 100,000 population extrapolates to an estimate of 2.39 million Americans with IBD. Our secondary analysis was approximately 15% higher, extrapolating to 2.74 million Americans with IBD. The age- and sex-standardized estimates of incidence translate to approximately 39,000 to 56,000 new IBD diagnoses per year in the United States. 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 United States. 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, MN (10.7 for Crohn’s disease and 12.2 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 approximately 1.6 million Americans with IBD.[6] Similarly, a 2016 estimate from CDM and Truven, a second commercial data source, estimated that there were approximately 1.4 million Americans with IBD,[12] and in a 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 four sources, including Medicare. With the aging of the population, compounding prevalence may contribute to a rise in prevalence over time.[2,13] 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 National Health Interview Study (NHIS) estimated there were 3.1 million Americans previously diagnosed with IBD,[4] while the 2009–2010 National Health and Nutrition Examination Survey (NHANES) estimated that there were 2.3 million Americans diagnosed with IBD.[14] 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 and/or under-estimation of IBD in claims-based analyses as some patients may patients 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 U.S. are similar to recent estimates from other high prevalence regions in Europe and Canada.[2] For example, the estimated prevalence per 100,000 population in 2010 was 744 in Germany,[9] in 2017 was 770 in Norway[15] and in 2020 was 872 in Denmark.[16] In Canada, the prevalence of IBD in 2018 was estimated at 700 per 100,000 based on extrapolation of data from 2002–2008.[17]

The prevalence of IBD has been described as passing through four phases: emergence of IBD, acceleration of incidence, compounding prevalence and finally prevalence equilibrium.[18] We observed a gradual increase in prevalence across the last decade, suggesting that the United States had not reached prevalence equilibrium. This is consistent with a recent study in Medicaid that showed an increasing prevalence of CD during the last decade, although most of that occurred prior to 2016.[19]

There is a general paucity of data on the racial distribution of IBD in the United States. We estimated that the prevalence of IBD was nearly twice as high among non-Hispanic White Americans as compared to 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 approximately 40% lower than in White Americans.[19–21] In the 2009–2010 NHANES, self-reported diagnosis with IBD was approximately 0.8% in non-Hispanic Black Americans vs 1.4% in non-Hispanic White Americans and 1.6% in Mexican Americans.[14] The NHIS had very similar results, with 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 was imputed which may under- or over-detect 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 United States population. According to the 2020 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 approximately 331 million people, one can estimate that there are approximately 224,000 Black, 79,000 Asian and 281,000 Hispanic Americans with IBD as compared to 1.6 million White Americans with IBD. Thus, we estimate that relative to Black, Asian and Hispanic Americans, there are 7x, 21x and 6x more White Americans with IBD, respectively. However, these estimates need to be viewed with caution as 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 social construct rather than a biologic construct. Nonetheless, this study provides critically important and novel estimates of the racial distribution of IBD in the United States.

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 are dependent on their guardians to advocate for them. Among possible explanations for this observation would be that poverty may make it more difficult for the guardian to seek medical care for their children due to many different social determinants of health that are linked to poverty, thereby leading to under diagnosis. 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.[22,23] 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.[24]

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, since 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 United States Veterans Affairs health system documented nearly identical prevalence estimates as ours.[25] 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. While 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% or more 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 approximately 15%. The component of the primary prevalence algorithm with the lowest PPV was for patients having one IBD diagnosis and at least one 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 over-estimation 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 utilized in this study differ from the overall population. Medicaid beneficiaries are more likely to be Black or Hispanic while commercially insured populations have a higher proportion of White Americans than the general population. To address this, we used direct standardization to assure that the incidence and prevalence estimates accounted for age, sex, and type of insurance.

There are a small proportion of Medicare beneficiaries who 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.[26] Given the small proportions, less than 10% for each, and the contrasting nature of these two groups, this is not expected to meaningfully impact the results.

In summary, we believe this to be the most comprehensive assessment of physician-diagnosed IBD in the United States to date. IBD is a relatively common chronic condition, affecting more than 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, it is not possible from these data to ascertain whether this is due to detection bias or biologic differences. 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

Refer to Web version on PubMed Central for supplementary material.

Funding Support:

Center for Disease Control and Prevention U01-DP006369

NIH: P30-DK050306

NIH: UL1TR002489

Disclosures

GKD receives funding from the American Society of Hematology and National Institutes of Health.

LEP is an employee of Elevance Health. She has received funding from Sanofi for an unrelated study.

JDL consulted or served on an advisory board for Eli Lilly and Company, Samsung Bioepis, UCB, Bristol-Myers Squibb, Nestle Health Science, Merck, Celgene, Janssen Pharmaceuticals, Bridge Biotherapeutics, Entasis Therapeutics, AbbVie, Pfizer, Gilead, Galapagos, Sanofi, Arena Pharmaceuticals, Protagonist Therapeutics, Amgen, and Scipher Medicine. He has had research funding from Nestle Health Science, Takeda, Janssen Pharmaceuticals, and AbbVie. He has had educational grants from Takeda and Janssen. He has performed legal work on behalf of generic manufacturers of ranitidine, including L. Perrigo Company, Glenmark Pharmaceuticals Inc., Amneal Pharmaceuticals LLC, Aurobindo Pharma USA, Inc., Dr. Reddy's Laboratories, Inc., Novitium Pharma, Ranbaxy Inc. and Sun Pharmaceutical Industries, Inc., Strides Pharma, Inc., and Wockhardt USA LLC. He owns stock in Dark Canyon Labs.

MJF: AbbVie, Astellas, Boehringer Ingelheim, GlaxoSmithKline, Sarepta, Takeda, and UCB Biosciences have collaborative agreements with the Center for Pharmacoepidemiology, Department of Epidemiology, UNC Chapel Hill which provides salary support to MJF as Center Director.

MDK has consulted for Abbvie, Pfizer, and Lilly, is a shareholder in Johnson & Johnson, and has received research support from Janssen and Abbvie.

KH is an employee of Janssen.

All other authors report no potential conflict of interest

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(1):46–54 e42; quiz e30. [PubMed: 22001864]
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.
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. The Lancet Gastroenterology Hepatology. 2020;5(1):17–30. [PubMed: 31648971]
4. Dahlhamer JM, Zammitti EP, Ward BW, Wheaton AG, Croft JB. Prevalence of Inflammatory Bowel Disease Among Adults Aged >/=18 Years - United States, 2015. MMWR Morb Mortal Wkly Rep. 2016;65(42):1166–1169. [PubMed: 27787492]
5. Herrinton LJ, Liu L, Lewis JD, Griffin PM, Allison J. Incidence and prevalence of inflammatory bowel disease in a Northern California managed care organization, 1996–2002. Am J Gastroenterol. 2008;103(8):1998–2006. [PubMed: 18796097]
6. Shivashankar R, Tremaine WJ, Harmsen WS, Loftus EV, Jr. Incidence and Prevalence of Crohn's Disease and Ulcerative Colitis in Olmsted County, Minnesota From 1970 Through 2010. Clin Gastroenterol Hepatol. 2017;15(6):857–863. [PubMed: 27856364]
7. Barnes EL, Nowell WB, Venkatachalam S, Dobes A, Kappelman MD. Racial and Ethnic Distribution of Inflammatory Bowel Disease in the United States. Inflamm Bowel Dis. 2021.
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. 2023:doi: 10.1093/ibd/izad1053. Online ahead of print.
9. DeFrank JT, Bowling JM, Rimer BK, Gierisch JM, Skinner CS. Triangulating differential nonresponse by race in a telephone survey. Prev Chronic Dis. 2007;4(3):A60. [PubMed: 17572964]
10. Jarrin OF, Nyandege AN, Grafova IB, Dong X, Lin H. 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(1):e1–e8. [PubMed: 31688554]

11. Dawwas G, Weiss A, Constant B, et al. Development and Validation of Claims-Based Definitions to Identify Incident and Prevalent Inflammatory Bowel Disease in Administrative Healthcare Databases. Inflamm Bowel Dis. Under review.
12. Ye Y, Manne S, Treem WR, Bennett D. 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(4):619–625. [PubMed: 31504515]
13. 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(4):1147–1159 e1144. [PubMed: 34995526]
14. Weisman MH, Oleg S, Seok Kim H, Hou JK, Miller FW, Dillon CF. Inflammatory Bowel Disease Prevalence: Surveillance data from the U.S. National Health and Nutrition Examination Survey. Prev Med Rep. 2023;33:102173. [PubMed: 37223580]
15. Lirhus SS, Høivik ML, Moum B, Anisdahl K, Melberg HO. Incidence and Prevalence of Inflammatory Bowel Disease in Norway and the Impact of Different Case Definitions: A Nationwide Registry Study. Clinical epidemiology. 2021;Volume 13:287–294. [PubMed: 33935519]
16. 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(3):501–510. [PubMed: 36728238]
17. 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(5):1345–1353 e1344. [PubMed: 30639677]
18. Kaplan GG, Windsor JW. The four epidemiological stages in the global evolution of inflammatory bowel disease. Nat Rev Gastroenterol Hepatol. 2021;18(1):56–66. [PubMed: 33033392]
19. Hutfless S, Jasper RA, Chen PH, Joseph S, Miller S, Brant SR. Burden of Crohn's Disease in the United States Medicaid Population, 2010 – 2019. Clin Gastroenterol Hepatol. 2023.
20. Xu F, Carlson SA, Liu Y, Greenlund KJ. Prevalence of Inflammatory Bowel Disease Among Medicare Fee-For-Service Beneficiaries — United States, 2001−2018. MMWR Morb Mortal Wkly Rep. 2021;70(19):698–701. [PubMed: 33983913]
21. Barnes EL, Nowell WB, Venkatachalam S, Dobes A, Kappelman MD. Racial and Ethnic Distribution of Inflammatory Bowel Disease in the United States. Inflamm Bowel Dis. 2022;28(7):983–987. [PubMed: 34473272]
22. Duricova D, Pedersen N, Elkjaer M, Gamborg M, Munkholm P, Jess T. Overall and cause-specific mortality in Crohn's disease: a meta-analysis of population-based studies. Inflamm Bowel Dis. 2010;16(2):347–353. [PubMed: 19572377]
23. Bewtra M, Kaiser LM, TenHave T, Lewis JD. Crohn's disease and ulcerative colitis are associated with elevated standardized mortality ratios: a meta-analysis. Inflamm Bowel Dis. 2013;19(3):599–613. [PubMed: 23388544]
24. Zhulina Y, Udumyan R, Tysk C, Halfvarson J. Mortality in patients with Crohn's disease in Örebro, Sweden 1963–2010. Scand J Gastroenterol. 2022;57(2):158–164. [PubMed: 34693837]
25. Hou JK, Kramer JR, Richardson P, Mei M, El-Serag HB. The incidence and prevalence of inflammatory bowel disease among U.S. veterans: a national cohort study. Inflamm Bowel Dis. 2013;19(5):1059–1064. [PubMed: 23448789]
26. 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.

What you need to know

Background and Contest:

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 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 more than 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 IBD is an important goal for future investigation.

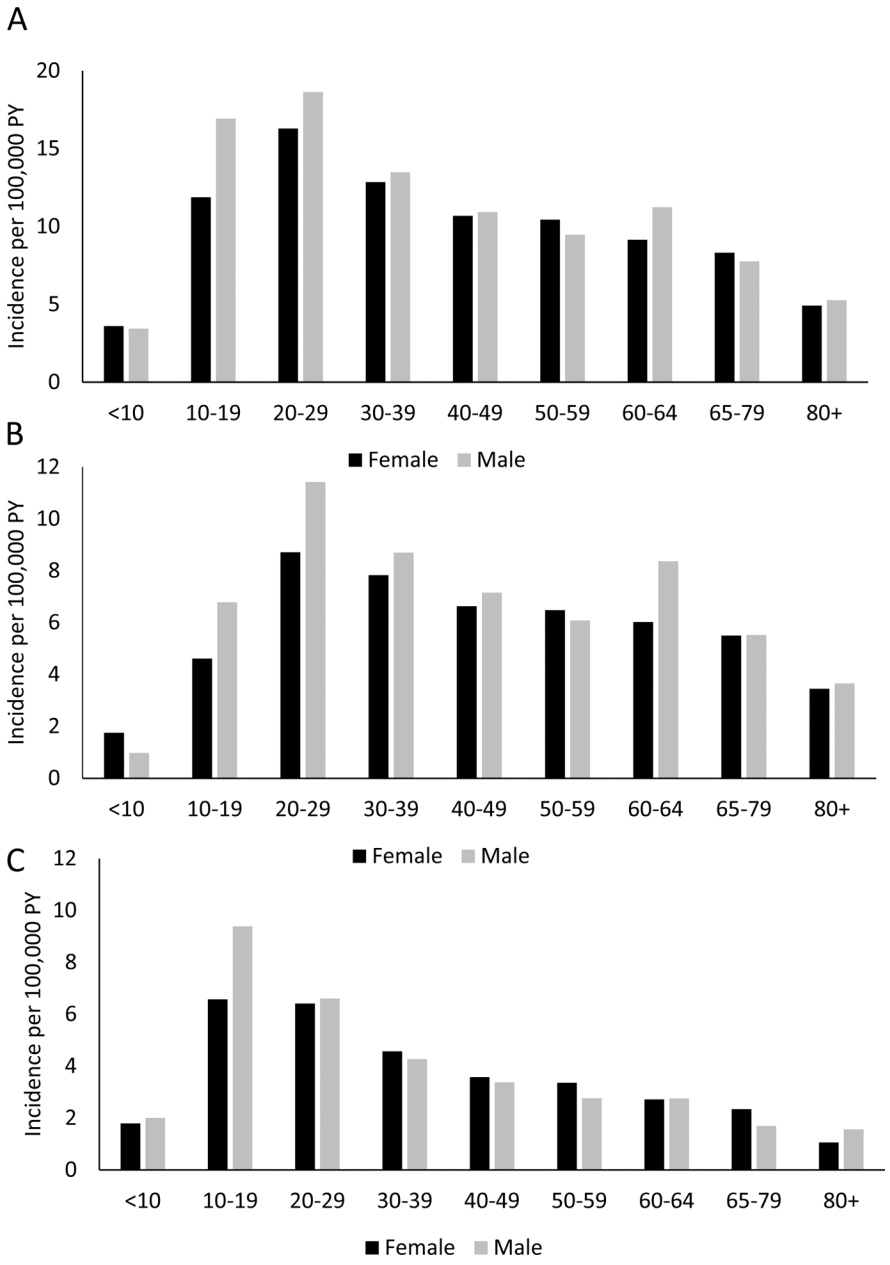

Figure 1.

Age- and sex-specific incidence per 100,000 person-years (A) IBD, (B) ulcerative colitis and (C) Crohn's disease in the United States

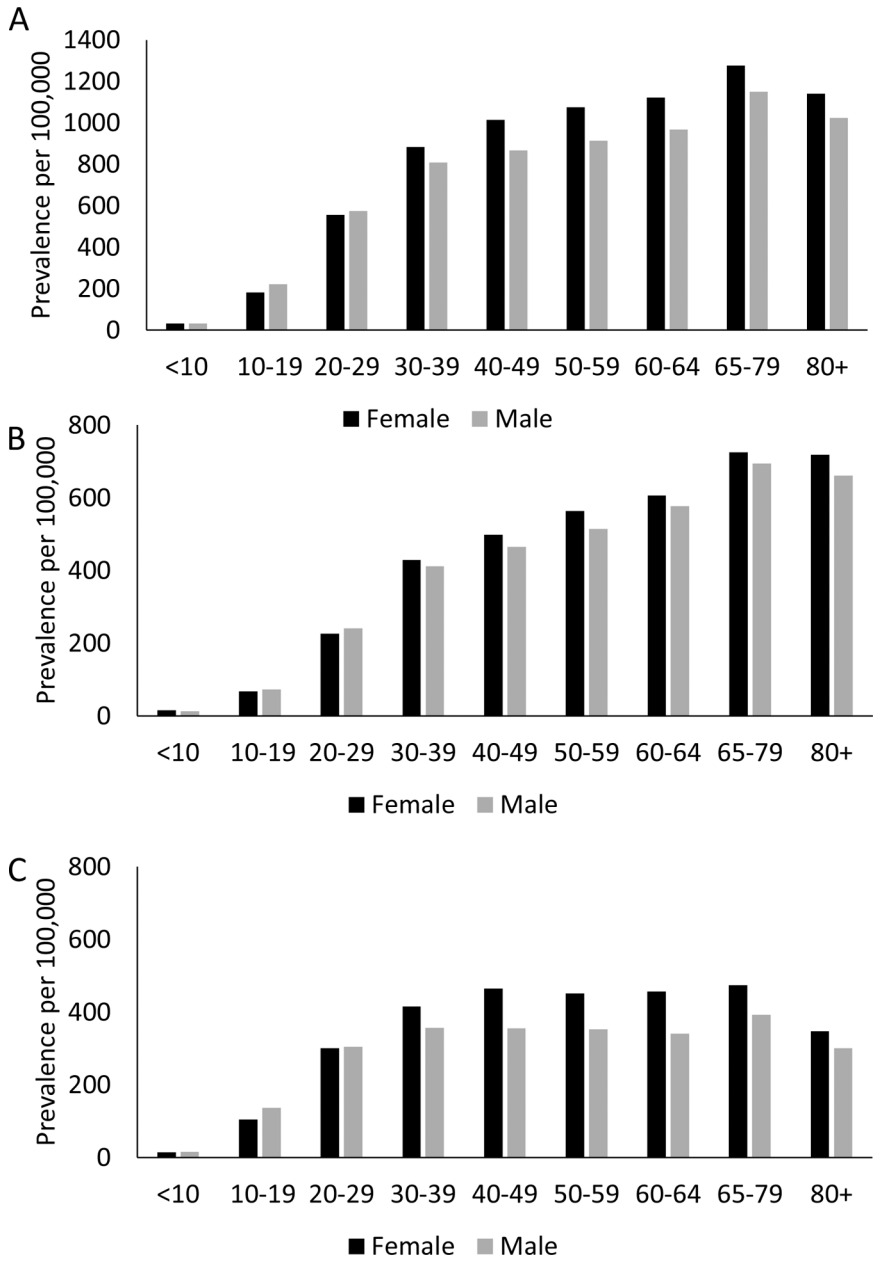

Figure 2.

Age- and sex-specific prevalence per 100,000 population of (A) IBD, (B) ulcerative colitis and (C) Crohn's disease in the United States

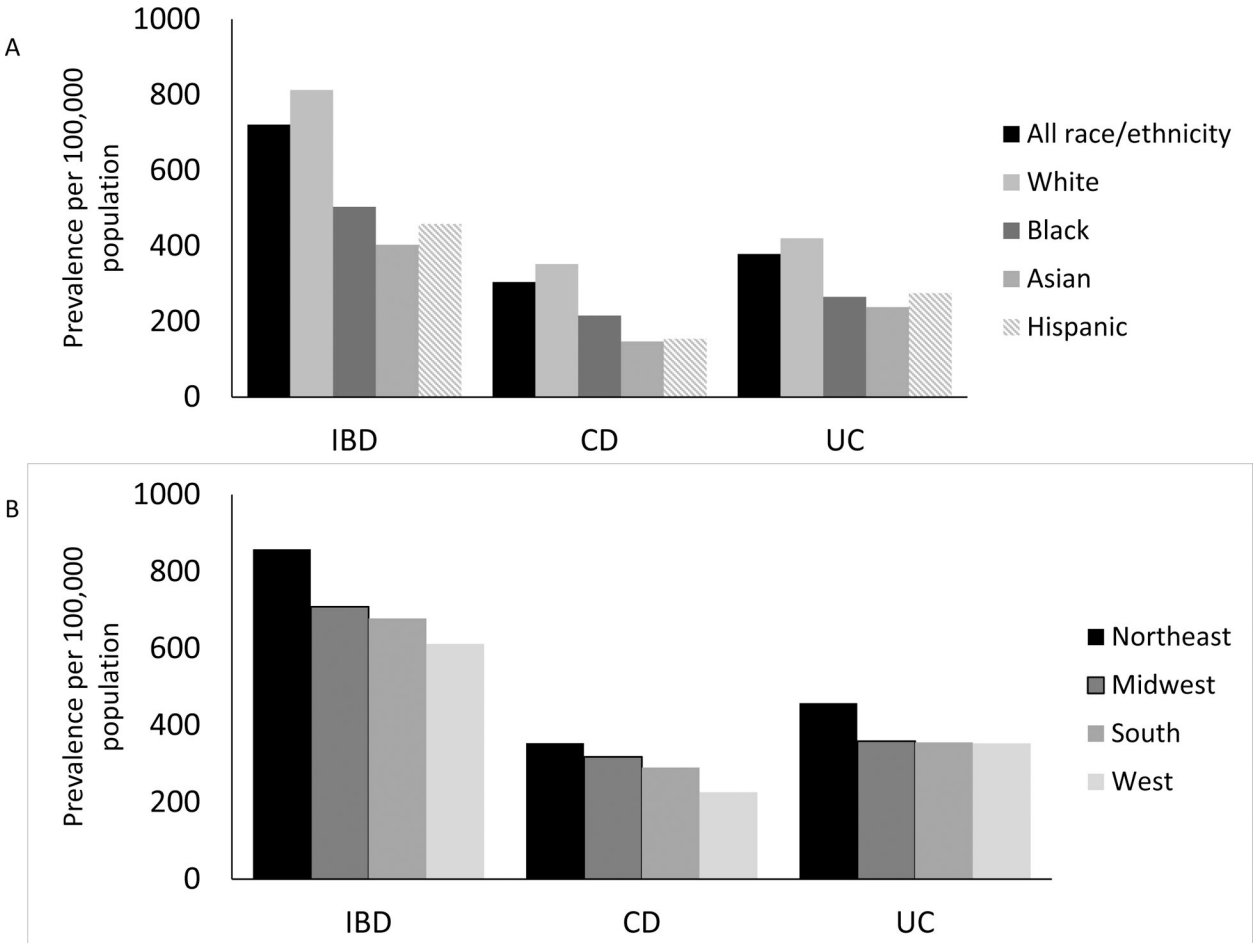

Figure 3.

Age- and sex-adjusted prevalence of IBD, ulcerative colitis and Crohn's disease by race and ethnicity (A) and region (B)

Table 1.

Sensitivity analysis using different criteria to determine prevalence of IBD, ulcerative colitis and Crohn's disease.

Diagnosis	Prevalence definition*	Minimum enrollment	Prevalence per 100,000 population
IBD overall	Primary	4 year	721 (717–726)
IBD overall	Secondary	4 year	826 (821–831)
IBD overall	Primary	1 year	600 (597–603)
IBD overall	Secondary	1 year	680 (677–683)
Crohn's disease	Primary	4 year	305 (302–308)
Crohn's disease	Secondary	4 year	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 year	378 (375–382)
Ulcerative colitis	Secondary	4 year	438 (435–441)
Ulcerative colitis	Primary	1 year	312 (310–314)
Ulcerative colitis	Secondary	1 year	358 (356–361)

*Secondary definition includes patients with a single diagnosis of IBD by a gastroenterologist or surgeon, or two or more diagnoses by providers other than gastroenterologists or surgeons, without any therapy, applying a weight of 0.22

Table 2.

Secular trends in prevalence of IBD

	Age- and sex-standardized prevalence per 100,000 population		
Year	**CDM***	**HealthCore***	**Medicare^**
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

*Includes under age 65 and Medicare advantage patients over age 65

^Includes only fee for service Medicare beneficiaries over age 65