

Global evolution of inflammatory bowel disease across epidemiologic stages

<https://doi.org/10.1038/s41586-025-08940-0>

Received: 8 January 2024

Accepted: 26 March 2025

Published online: 30 April 2025

Open access

 Check for updates

Lindsay Hracs^{1,70}, Joseph W. Windsor^{1,70}, Julia Gorospe¹, Michael Cummings², Stephanie Coward¹, Michael J. Buie¹, Joshua Quan¹, Quinn Goddard¹, Léa Caplan¹, Ante Markovicinović¹, Tyler Williamson², Yvonne Abbey², Murdani Abdullah⁴, Maria T. Abreu^{5,6}, Vineet Ahuja^{6,7}, Raja Affendi Raja Ali⁸, Mansour Altuwaijri⁹, Domingo Balderramo¹⁰, Rupa Banerjee^{6,11}, Eric I. Benchimol^{12,13,14,15}, Charles N. Bernstein^{6,16}, Eduard Brunet-Mas^{17,18,19}, Johan Burisch^{6,20,21,22}, Vui Heng Chong²³, Iris Dotan^{6,24}, Usha Dutta²⁵, Sara El Ouali^{26,27}, Angela Forbes²⁸, Anders Forss²⁹, Richard Gearry^{6,28}, Viet Hang Dao³⁰, Juanda Leo Hartono^{31,32}, Ida Hilm³³, Phoebe Hodges^{34,35}, Gareth-Rhys Jones³⁶, Fabián Juliao-Baños³⁷, Jamilya Kaibullayeva^{38,39}, Paul Kelly^{34,35}, Taku Kobayashi^{40,41}, Paulo Gustavo Kotze^{6,42}, Peter L. Lakatos^{6,43,44}, Charlie W. Lees^{45,46}, Julajak Limsrivilai⁴⁷, Bobby Lo^{20,21}, Edward V. Loftus Jr^{6,48}, Jonas F. Ludvigsson^{29,49}, Joyce W. Y. Mak⁵⁰, YingLei Miao^{51,52}, Ka Kei Ng⁵³, Shinji Okabayashi⁵⁴, Ola Olén⁵⁵, Remo Panaccione^{1,6}, Mukesh Sharma Paudel⁵⁶, Abel Botelho Quaresma^{57,58}, David T. Rubin^{6,59}, Marcellus Simadibrata⁴, Yang Sun^{51,52}, Hidekazu Suzuki⁶⁰, Martin Toro⁶¹, Dan Turner^{6,62}, Beatriz Iade⁶³, Shu Chen Wei⁶⁴, Jesus K. Yamamoto-Furusho⁶⁵, Suk-Kyun Yang⁶⁶, Siew C. Ng^{6,67,68,69,71}  Gilaad G. Kaplan^{1,6,71}  & Global IBD Visualization of Epidemiology Studies in the 21st Century (GIVES-21) Research Group*

During the twentieth century, inflammatory bowel disease (IBD) was considered a disease of early industrialized regions in North America, Europe and Oceania¹. At the turn of the twenty-first century, IBD incidence increased in newly industrialized and emerging regions in Africa, Asia and Latin America, while the prevalence in early industrialized regions continued to grow steadily^{2–4}. Changes in the incidence and prevalence denote the evolution of IBD across four epidemiologic stages: stage 1 (emergence), characterized by low incidence and prevalence; stage 2 (acceleration in incidence), marked by rapidly rising incidence and low prevalence; and stage 3 (compounding prevalence), where the incidence decelerates, plateaus or declines while the prevalence steadily increases. A fourth stage (prevalence equilibrium) has been proposed in which the prevalence slope plateaus due to demographic shifts in an ageing IBD population, but it has not yet been evidenced. To date, these stages have remained theoretical, lacking specific numerical indicators to define transition points. Here, using real-world data from 522 population-based studies encompassing 82 global regions and spanning more than a century (1920–2024), we show spatiotemporal transitions across stages 1–3 and model stage 4 progression. Understanding the evolution of IBD across epidemiologic stages enables healthcare systems to better anticipate the future worldwide burden of IBD.

IBD, consisting of Crohn's disease (CD) and ulcerative colitis (UC), was first recognized in the 1800s⁵. In the early twentieth century, IBD was considered a rare disease among the descendants of Europeans who colonized North America and Oceania (hereafter referred to as early industrialized regions)³. The changing epidemiology of IBD is characterized in terms of incidence (new diagnoses reported per 100,000 person-years) and prevalence (total affected individuals per 100,000 persons at a given time); for brevity, the units 'per 100,000' are omitted in the following text. After the Second World War, the incidence of IBD in early industrialized regions increased rapidly³. Although the reasons for this increase remain incompletely understood, evidence suggests

environmental factors associated with Westernization of society—for example, increased smoking, Western diet and improved hygiene—may have substantially contributed by altering mucosal immune responses to the intestinal microbiome in genetically susceptible individuals^{6,7}.

During the latter half of the twentieth century, IBD was infrequently diagnosed in regions of Africa, Asia and Latin America that predominantly began industrialization after the Second World War (hereafter, newly industrialized regions) and low-income, developing areas (hereafter, emerging regions)³. By the twenty-first century, the incidence stabilized in many early industrialized regions, except in children, in whom it continues to rise, whereas the prevalence of IBD continued to

A list of affiliations appears at the end of the paper. *A list of authors and their affiliations appears at the end of the paper.

climb steadily across all age groups^{4,8}. Although IBD cases in emerging regions remain sporadic, since 2000, newly industrialized regions have reported a sharp increase in the incidence of UC, followed by CD¹. Today, IBD affects millions of people worldwide⁴.

The globalization of IBD has invalidated the historical notion that IBD is specific to the 'Western world'¹. Here, we advance the theory that IBD evolves temporally and spatially across four distinct epidemiologic stages. Stage 1 (emergence) is characterized by low incidence and prevalence; stage 2 (acceleration in incidence) involves rapidly rising incidence year-over-year, while the prevalence remains low; stage 3 (compounding prevalence) is marked by the slowing, stabilization or decrease in the incidence, with the prevalence continuing to accumulate due to decades of rising incidence outpacing mortality; and stage 4 (prevalence equilibrium) occurs when the prevalence plateaus due to mortality approximating incidence as the IBD population advances in age². By clearly defining these epidemiologic strata with specific benchmarks for transition across stages, regions can better prepare their healthcare systems to manage the stage-specific burden of IBD.

Trends in incidence and prevalence of IBD

We identified real-world data of 522 population-based studies reporting the incidence ($n = 463$) and/or prevalence ($n = 243$) of CD and/or UC, encompassing 82 countries, nations or territories (henceforth referred to as regions) and spanning the years 1920–2024 (Supplementary Fig. 1 and Supplementary Tables 1 and 2). Over the past century, the epidemiologic trends in the incidence (Fig. 1a,b) and prevalence (Fig. 1c,d) of IBD follow distinct geographical and temporal patterns.

UC was first recognized in the nineteenth century, while Crohn, Ginzburg and Oppenheimer's seminal 1932 paper on regional ileitis (later, CD) solidified CD as a distinct condition⁹. Consequently, data on the incidence of IBD during the early decades of the twentieth century are sparse, and prevalence data are non-existent (Fig. 1 and Supplementary Fig. 2). Today, with over a century of epidemiologic data, we can analyse distributions by region and decade. We calculated coalescing ranges (CR) for incidence (CR-I) and prevalence (CR-P), as defined by the 25th–75th percentiles within these strata. By the 1940s, the diagnosis of IBD was established in early industrialized regions (Fig. 1a,b). For example, in the 1940s, the CR-I for CD and UC in the United States was 1.15–2.30 and 1.02–2.41, respectively, while in the 1950s, Europe showed a rising IBD incidence, with UC more commonly diagnosed than CD, such as Sweden's CR-I of 1.88–7.50 for UC and 0.97–2.18 for CD (Fig. 1a,b and Supplementary Figs. 3a and 4a).

Subsequently, early industrialized regions in North America, Europe and Oceania experienced rapidly rising incidence, which stabilized in many regions around the turn of the twenty-first century (Fig. 1a,b). Longitudinal data from Cardiff, Wales (1931–2008) demonstrated a steady increase in CD incidence, with rates nearly doubling each decade from 1.17 (1946–1955) to 7.09 (1976–1985), before levelling off at 5.88 (1986–1995) and 6.64 (1996–2005)¹⁰ (Supplementary Fig. 5a).

The highest incidence of IBD, particularly UC, has been reported in Scandinavia (Fig. 1a,b). In Denmark, incidence steadily increased from the 1970s (CR-I: 2.05–3.55 for CD; 6.50–9.12 for UC) to 1990s (CR-I: 7.30–10.68 for CD; 13.18–20.99 for UC) and to 2010s (CR-I: 12.62–15.84 for CD; 22.21–30.64 for UC) (Fig. 1a,b and Supplementary Fig. 6a). The highest ever reported UC incidence was 73.7 in the Faroe Islands in 2011 (ref. 11) (Fig. 1a,b and Supplementary Fig. 7a). Only a small number of early industrialized regions report population-based incidence exceeding 40 per 100,000 for either UC or CD separately (Extended Data Fig. 1). Thus, 40 per 100,000 serves as a ceiling threshold for the incidence of CD or UC. Regions with methodologically reliable data that exceed this threshold should be prioritized for further study to identify environmental determinants of IBD.

After decades of rising incidence in early industrialized regions, the prevalence of IBD has steadily climbed (Fig. 1c,d). In Olmsted County,

the prevalence of IBD increased from 0.12% in 1965 to 0.35% in 1991, 0.56% in 2011 and 0.63% in 2019 (Supplementary Fig. 3b). A separate US study estimated the national prevalence to be 0.72% in 2018¹². In Lothian, Scotland, IBD prevalence increased from 0.57% to 0.78% between 2008 and 2018 (ref. 13) (Supplementary Fig. 8b), with forecasts from Canada and Scotland predicting that 1% of the population will be living with IBD by 2030^{13–15}.

During the twentieth century, epidemiologic data from newly industrialized and emerging regions in Africa, Asia and Latin America were sparse, with any available data reporting incidence and prevalence far lower than those in early industrialized regions (Fig. 1). By the turn of the twenty-first century, epidemiologic data began to indicate that newly industrialized regions were entering a stage of rapidly increasing incidence. This increase in reported incidence can be attributed to both improved identification of cases through advanced diagnostic capabilities and a true increase in incidence driven by environmental determinants^{5,7}.

Japan provides some of the earliest data from a newly industrialized region, spanning 1955 to 2000 (Fig. 1a,b and Supplementary Fig. 9). The incidence of IBD in Japan before the 1970s was less than 0.25, increasing to over 0.4 by 1980. By 2000, the incidence had increased tenfold reaching 4.77 and 1.27 for UC and CD, respectively (Supplementary Fig. 9a). South Korea showed similar patterns, with low incidence in the 1980s (CR-I: 0–0.03 for CD; 0.21–0.33 for UC) increasing steadily into the 2010s (CR-I: 2.20–3.20 for CD; 4.11–6.27 for UC) (Fig. 1a,b and Supplementary Fig. 10a). Regions with slower economic development experienced a delayed onset of rapidly rising incidence, as seen in China and Malaysia, where incidence substantially increased after 2000 (Supplementary Figs. 11a and 12a).

Brazil demonstrates a clear case of rising incidence in Latin America, starting with low rates in the 1980s (CR-I: 0.08–0.40 for CD; 0–0.46 for UC), and increasing in the 2000s (CR-I: 0.34–0.98 for CD; 0.53–1.04 for UC) and 2010s (CR-I: 1.21–3.22 for CD; 2.42–5.66 for UC) (Fig. 1a,b and Supplementary Fig. 13a). Heterogeneity within Brazil highlights higher IBD incidence in more-urbanized, developed areas; for example, the prevalence in the more densely populated and economically advanced São Paulo (182.81 in 2020) was three times that of Piauí (59.94 in 2020)¹⁶.

The highest prevalence of IBD in newly industrialized regions was observed in areas where incidence increased earlier. Japan's prevalence increased from 0.067% in 2000 to 0.165% in 2016 (Supplementary Fig. 9b). Brazil showed a similar trend, with prevalence rising from 0.014% in 2000 to 0.1% by 2020 (Supplementary Fig. 13b). By contrast, regions in which the incidence surged after 2000 report much lower prevalence. For example, Colombia's prevalence reached 0.067% in 2017 (Supplementary Fig. 14b), matching Japan's 2000 level. Similarly, China (Supplementary Fig. 11b), Malaysia (Supplementary Fig. 12b) and Taiwan (Supplementary Fig. 15b) all had an IBD prevalence below 0.03% in the most recent year of data (China, 2016; Malaysia, 2018; Taiwan, 2023).

The first three epidemiologic stages of IBD

The observed differences in IBD incidence and prevalence across various geographical areas over the past century suggest that epidemiologic patterns shift through time. To further explore these trends and characterize the epidemiology of IBD independently of geography and time, we developed a machine-learning classifier to determine the epidemiologic stage of global regions. This methodology not only automates the classification of stages across a large, heterogeneous dataset but also establishes benchmarks for incidence and prevalence that can be applied to new data as they become available.

Derived from the systematic review, the observed incidence and prevalence data, along with their change over time (Supplementary Fig. 16), informed an iterative labelling process that resulted in a training dataset with a subset of regions labelled as epidemiologic stage 1,

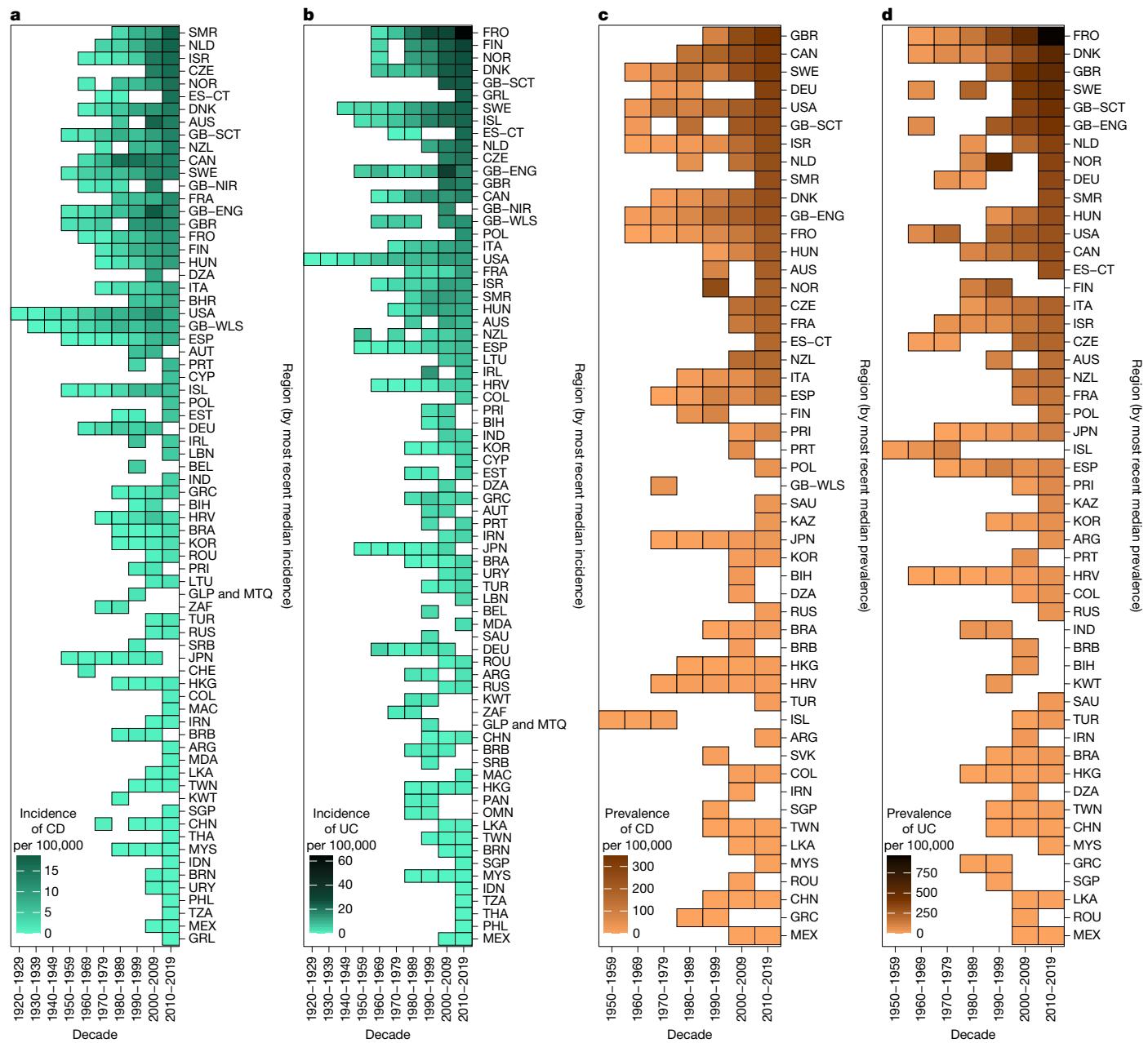


Fig. 1 | The incidence (per 100,000 person-years) and prevalence (per 100,000) of IBD by decade and region. **a**, The incidence of CD by decade, with regions ranked from the highest (top) to lowest (bottom) most recent median incidence value available. **b**, The incidence of UC by decade, with regions ranked from the highest (top) to lowest (bottom) most recent median incidence value available. **c**, The prevalence of CD by decade, with regions ranked from the highest (top) to lowest (bottom) most recent median prevalence value available. **d**, The prevalence of UC by decade, with regions ranked from the highest (top) to lowest (bottom) most recent median prevalence value available. Colour saturation represents median incidence/prevalence, calculated from all studies within a given region for that decade. CR values (25th to 75th percentiles) of incidence/prevalence for the corresponding region-decade pairing are provided in Supplementary Fig. 2. All data are available at Figshare (<https://doi.org/10.6084/m9.figshare.24952557>). An interactive map depicting year-over-year changes in incidence and prevalence is available online (<https://kaplan-gi.shinyapps.io/GIVES21/>). Regions are labelled with their International Organization for Standardization (ISO) 3166-1 alpha-3 codes: Algeria (DZA), Argentina (ARG),

Australia (AUS), Austria (AUT), Bahrain (BHR), Barbados (BRB), Belgium (BEL), Bosnia and Herzegovina (BIH), Brazil (BRA), Brunei (BRN), Canada (CAN), China (CHN), Colombia (COL), Croatia (HRV), Cyprus (CYP), Czechia (CZE), Denmark (DNK), Estonia (EST), Faroe Islands (FRO), Finland (FIN), France (FRA), Germany (DEU), Greece (GRC), Greenland (GRL), Guadeloupe and Martinique (GLP and MTQ), Hong Kong (HKG), Hungary (HUN), Iceland (ISL), India (IND), Indonesia (IDN), Iran (IRN), Ireland (IRL), Israel (ISR), Italy (ITA), Japan (JPN), Kazakhstan (KAZ), Kuwait (KWT), Lebanon (LBN), Lithuania (LTU), Macao (MAC), Malaysia (MYS), Mexico (MEX), Moldova (MDA), Netherlands (NLD), New Zealand (NZL), Norway (NOR), Oman (OMN), Panama (PAN), Philippines (PHL), Poland (POL), Portugal (PRT), Puerto Rico (PRI), Romania (ROU), Russia (RUS), San Marino (SMR), Saudi Arabia (SAU), Serbia (SRB), Singapore (SGP), Slovakia (SVK), South Africa (ZAF), South Korea (KOR), Spain (ESP), Sri Lanka (LKA), Sweden (SWE), Switzerland (CHE), Taiwan (TWN), Tanzania (TZA), Thailand (THA), Türkiye (TUR), United Kingdom (GBR), United States of America (USA) and Uruguay (URY). Regions with ISO 3166-2 codes are as follows: Catalonia (ES-CT), England (GB-ENG), Northern Ireland (GB-NIR), Scotland (GB-SCT) and Wales (GB-WLS).

2 or 3. We began with the assumption that many early industrialized regions are currently in stage 3. From there, we visually inspected historical trends in these regions to define preliminary stage assignments

on the basis of observable changes in the epidemiology of stages 1 and 2. These observed trends were visually compared to trends in emerging and newly industrialized regions suspected to be in stage 1 or 2

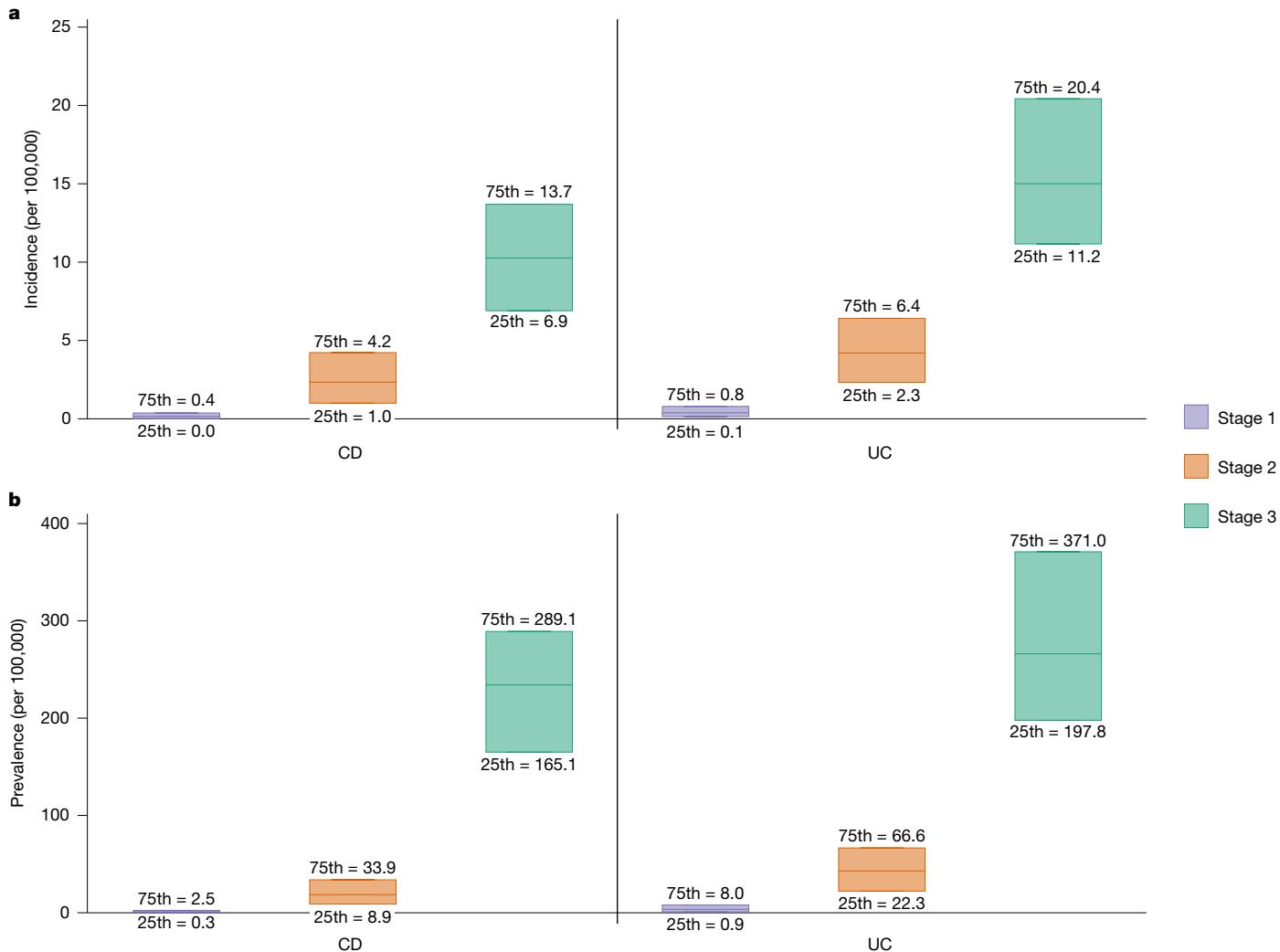


Fig. 2 | CRs of CD and UC across epidemiologic stages 1, 2 and 3. **a**, The CR-I of CD (left) and UC (right) across the three epidemiologic stages. Number of observations: $n = 263$ (CD stage 1), $n = 1,011$ (CD stage 2), $n = 796$ (CD stage 3), $n = 277$ (UC stage 1), $n = 847$ (UC stage 2), $n = 760$ (UC stage 3). **b**, The CR-P of CD (left) and UC (right) across the three epidemiologic stages. Number of observations: $n = 86$ (CD stage 1), $n = 247$ (CD stage 2), $n = 443$ (CD stage 3), $n = 118$ (UC stage 1), $n = 238$ (UC stage 2), $n = 435$ (UC stage 3). Data are categorized by type (incidence/prevalence), disease type (CD/UC) and epidemiologic stage

(stage 1, stage 2, stage 3), as determined by the random-forest classifier. For the box plots, the centre line shows the median, the lower hinge shows the 25th percentile (that is, first quartile) and the upper hinge shows the 75th percentile (that is, third quartile). The 25th and 75th percentiles are labelled and correspond to the CRs. Statistical analysis was performed using negative binomial regression with post hoc comparisons of estimated marginal means with Tukey adjustment for multiple comparisons, showing significant differences between all stages for the incidence and prevalence of CD and UC ($P < 0.001$ for all comparisons).

to corroborate classifications. This method enabled us to effectively compare current low-incidence regions with historical periods of lower incidence in current high-incidence regions, establishing equivalencies across varying times and geographies.

A k -nearest neighbours (k -NN)¹⁷ model supported the manual labelling process, facilitating an iterative refinement of classifications until stage labels for a subset of data were agreed upon by three independent reviewers. The labelled dataset was used as the training input for a random-forest classifier¹⁸, which used the magnitude of incidence and prevalence, along with the rates of change, to inform the model and define the three epidemiologic stages (Extended Data Fig. 2). After classification using a supervised random-forest model, benchmarked CR-I and CR-P values corresponding to stages 1–3 were calculated (Fig. 2). Although the number of epidemiologic stages was predefined, the supervised machine-learning classifier accurately classified unseen validation data into the three stages with a high accuracy of 95.15% (95% confidence interval (CI) = 92.60–97.01). Classification errors on validation data mainly occurred in regions with datapoints between

the CRs, which probably represent regions in transition toward the next epidemiologic stage.

Negative binomial regression models with post hoc comparisons of estimated marginal means with Tukey adjustment revealed significant differences between all stages for the incidence and prevalence of CD and UC (all values, $P < 0.001$). The clear distinctions across stages define the CRs for the incidence and prevalence of IBD—stage 1: CR-I = 0.1–1.2, CR-P = 1.2–10.5; stage 2: CR-I = 3.3–10.6, CR-P = 31.2–100.5; stage 3: CR-I = 18.1–34.1, CR-P = 362.9–660.1 (Fig. 2). Values that fall between CRs indicate regions transitioning between stages. For example, Taiwan was classified as stage 1 in 2010–2019, showing increases in incidence from 1.32 and prevalence from 11.24 in 2010 to an incidence of 2.51 and a prevalence of 21.16 in 2019. These rising trends suggest that Taiwan was transitioning towards stage 2; with newly available data, the model subsequently assigned a stage 2 classification in the 2020s based on prevalence of 29.2 in 2023 (Supplementary Fig. 15b). These ranges may help to define the burden of IBD on healthcare systems by providing estimated numbers of incident and prevalent IBD cases year over year

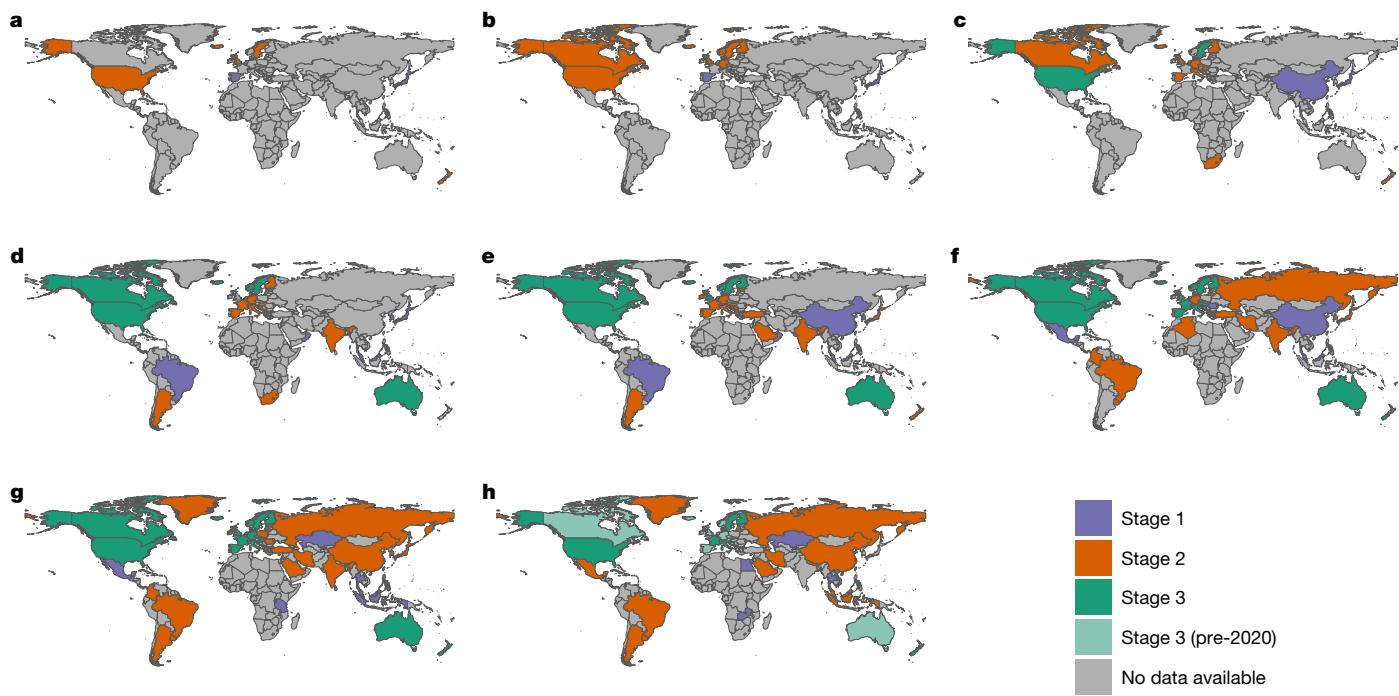


Fig. 3 | Global maps depicting epidemiologic stages of IBD evolution from 1950 to 2024. **a**, Epidemiologic stages from 1950 to 1959. **b**, Epidemiologic stages from 1960 to 1969. **c**, Epidemiologic stages from 1970 to 1979. **d**, Epidemiologic stages from 1980 to 1989. **e**, Epidemiologic stages from 1990 to 1999. **f**, Epidemiologic stages from 2000 to 2009. **g**, Epidemiologic stages from 2010 to 2019. **h**, Epidemiologic stages from 2020 to 2024; because regions cannot

regress in stage, regions without data in 2020–2024 but with a previous stage 3 classification are shaded in a lighter green than regions in stage 3 that do have data during this period. Each region is coloured according to its epidemiologic stage as predicted by the random-forest classifier. Interactive maps are available online (<https://kaplan-gi.shinyapps.io/GIVES21/>).

and linking the estimated rates to their respective costs and resource needs^{2,19}.

An overall increase in incidence rates was observed over time across stages 1, 2 and 3, with several regions transitioning to stage 3 after 1990 (Extended Data Fig. 2). Prevalence remains consistently low in regions classified as stage 1. Prevalence noticeably increases in stage 2 and then rises rapidly in stage 3 (Extended Data Fig. 2). Our findings demonstrate regional transitions across stages and highlight the increasing number of regions included in population-based studies over time (Fig. 3). For example, data spanning a century from the United States display a transition from stage 1 to stage 2 in the 1950s, followed by a shift to stage 3 in the 1970s. Today, most early industrialized regions in Europe, North America and Oceania are classified as stage 3, while many newly industrialized regions in Latin America, East Asia and the Middle East are in stage 2 (Fig. 3). Data from emerging regions in stage 1 (such as many regions in Africa) remain limited, as data scarcity is a typical characteristic of this stage.

The underlying drivers of transition across stages remain unclear. The shift from stage 1 to 2 can be partially attributed to the unmasking of incidence. As IBD begins to emerge in a region, the local medical infrastructure (such as access to colonoscopy) must be sufficient to diagnose CD. In low-income regions where access to colonoscopy is more limited than sigmoidoscopy, cases of CD may be misdiagnosed as UC or missed entirely. We observed a significant difference ($P < 0.001$) in the UC:CD ratio across the three epidemiologic stages: a median ratio of 3.24:1 in stage 1, decreasing to 1.87:1 in stage 2 and further to 1.54:1 in stage 3 (Supplementary Fig. 17).

Beyond the unmasking of incidence, environmental factors have also contributed to a true increase in incidence. Industrialization, urbanization and Westernization have been linked to increasing incidence of IBD^{4–5}. To quantify societal changes over time, we examined five time- and region-specific indicators: the Augmented Human Development Index (AHD)²⁰, obesity²¹, percentage urbanization²²,

the Universal Health Coverage Service Index²³ and the Western Diet Index (WDI)²⁴ (Methods). Each of these societal indicators showed a significant difference when analysed by stage (Extended Data Fig. 3); for example, on a 0–1 scale, the median AHD increased from 0.39 in stage 1 to 0.53 in stage 2 and to 0.70 in stage 3 ($P < 0.001$ between epidemiologic stages).

Mathematically modelling the transition to stage 4

Partial differential equations (PDEs) were developed to model time-dependent prevalence²⁵ for three stage 3 regions: Canada, Denmark and Scotland. Prevalence was modelled out to 2043, the last common year of projected population data available^{26–28} (Fig. 4a). On the basis of the assumption of stable incidence over time, calculated as the mean incidence for each age group over the most recent 8-year period, the models indicate a rising prevalence in each region: Canada (0.65% in 2014 to 0.83% in 2025 to 0.96% in 2035 to 1.05% in 2043); Denmark (0.86% in 2014 to 1.19% in 2025 to 1.44% in 2035 to 1.59% in 2043); and Scotland (0.74% in 2014 to 1.04% in 2025 to 1.32% in 2035 to 1.51% in 2043) (Fig. 4a). Time-dependent prevalence stratified by IBD type shows a similar distribution of CD and UC prevalence in Canada over the next 20 years, while UC is projected to be more prevalent than CD in Scotland and Denmark (Supplementary Fig. 18a,b).

Modelled prevalence continues to climb for each region; however, the rate of prevalence growth decreases, signalling a transition towards stage 4. This transition occurs as mortality begins to approximate incidence, leading to a slowing of prevalence growth—a state driven by an ageing IBD population. For example, in Canada (observed data, 2007–2014), we observe a greater growth in prevalence among older adults compared with in the paediatric and young-adult age groups (Supplementary Fig. 19a); similar trends are seen in Denmark (observed data, 2010–2017) and Scotland (observed data, 2010–2017) (Supplementary Fig. 19b,c). In each region, PDE-modelled prevalence indicates

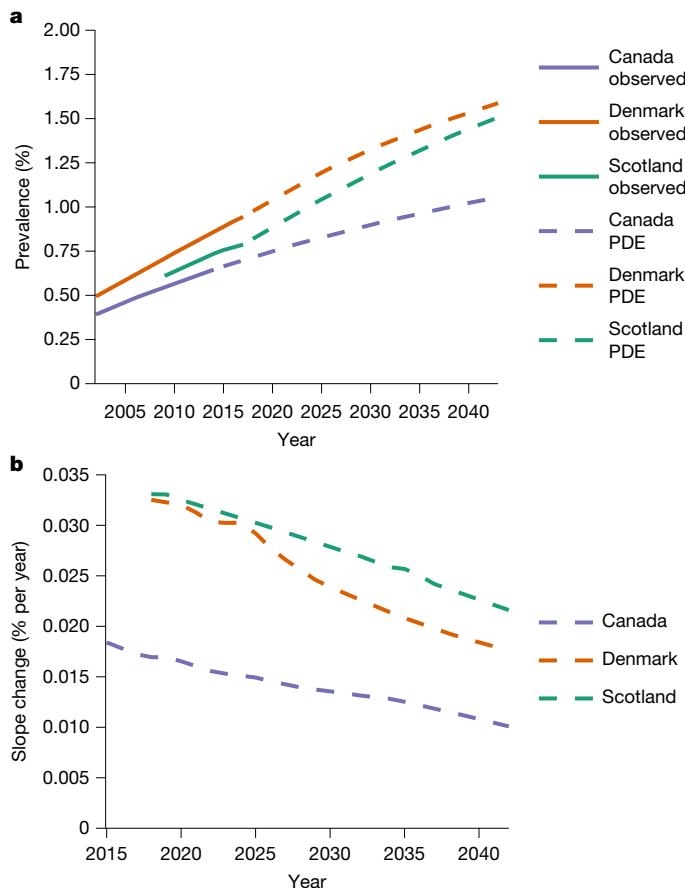


Fig. 4 | PDE-modelled time-dependent prevalence. **a**, Modelled time-dependent prevalence of IBD in Canada (observed data, 2007–2014), Denmark (observed data, 2010–2017) and Scotland (observed data, 2010–2017). **b**, Central difference approximations of the slopes of time-dependent IBD prevalence.

that the prevalence among older adults and older people with IBD will steadily increase across 2023–2043 (Supplementary Fig. 19a–c).

The prevalence growth rate (that is, slope) decreases in each analysed region, as calculated using central difference approximations²⁹ (Fig. 4b). In Canada, the slope decreases from 0.018% per year in 2015 to 0.010% per year in 2042. In Denmark and Scotland, the slope decreases from 0.033% per year in 2018 to 0.018% and 0.022% per year in 2042, respectively. When examining a 20-year period of overlapping data (2022–2042), all three regions exhibit similar reductions in prevalence growth rates ($P = 0.947$). A slowing of prevalence growth rates was also observed in central difference approximations for both CD and UC (Supplementary Fig. 18c,d).

In addition to modelling prevalence with stable incidence rates, we modelled four additional incidence rate scenarios: 2% and 1% decreases in incidence per year, and 1% and 2% increases in incidence per year. These incidence scenarios were modelled for Canada (Fig. 5a), Denmark (Fig. 5b) and Scotland (Fig. 5c). The no-change scenario corresponds to model predictions based on stable average incidence for each region (Fig. 4a). In 2043, the modelled prevalence across incidence scenarios (−2% to 2%) ranges between 0.97% and 1.16% for Canada, 1.45% and 1.76% for Denmark, and 1.36% and 1.69% for Scotland.

We propose that stage 4—prevalence equilibrium—occurs when the change in prevalence slope reduces to an average of 0% ($\pm 0.01\%$) over a 5-year period. Under the model assumption of stable yearly incidence, prevalence slope changes near 2040 are approaching zero; however, by 2042, only Canada (among analysed regions) reaches this threshold (Extended Data Table 1). Thus, if incidence remains stable, the models suggest that it will take decades for the examined stage 3 regions to

reach prevalence equilibrium. If incidence is decreased by 2% per year through preventative medicine and policy interventions targeting modifiable behavioural and environmental exposures, our models suggest that prevalence equilibrium may be achievable in the 2040s for all three regions³⁰ (Fig. 5d–f and Extended Data Table 2). Prevalence equilibrium is a desired state in the evolution of IBD for healthcare administrators and policymakers, as it would allow for appropriate allocation of resources to ensure timely, high-quality healthcare for those affected with IBD.

Discussion

IBD serves as a case study on the evolution of chronic inflammatory diseases that have predominantly manifested in the last two centuries following the industrial revolution⁵. For centuries, the occurrence of IBD was confined to sporadic incident cases with persistently low prevalence (stage 1: emergence)². Economic advancement, industrialization, urbanization, improved healthcare access and delivery, and shifting environmental exposures trigger a rapid increase in incidence, whereas prevalence levels remain low (stage 2: acceleration in incidence)². IBD is typically diagnosed in young adults and, as mortality remains low, prevalence steadily climbs (stage 3: compounding prevalence)². During stage 3, incidence rates tend to show less growth, stabilize or even decline; this, combined with an aging IBD population, is proposed to slow prevalence growth, and eventually plateau (stage 4: prevalence equilibrium)². Understanding IBD epidemiology prepares healthcare systems for the increasing burden of IBD as regions transition through these stages³¹.

Several low-income regions in Africa, Asia and Latin America are likely in stage 1. The exact drivers and timing of transition to stage 2 are unclear. However, in the absence of robust epidemiologic data, societal-level indicators such as the AHDI may provide insights into an impending increase in IBD cases as economies develop, healthcare infrastructure improves and lifestyles shift towards environmental factors that trigger IBD. As these regions transition to stage 2, raising awareness of IBD, differentiating from infectious diseases, and enhancing training and resources, including access to colonoscopy, are crucial³². Population-based studies in regions with sparse data, particularly Africa, are essential. Moreover, studying regions as they progress into stage 2 offers opportunities to investigate genetic, environmental and microbial determinants of IBD³³.

Many newly industrialized regions in Asia and Latin America are currently in stage 2, with robust population-based data indicating rising incidence throughout the past two decades. Japan and South Korea, which underwent rapid industrialization in the latter half of the twentieth century, have reported prevalence exceeding 0.1%, placing them on the brink of transitioning towards stage 3^{34,35}. By contrast, regions like China and Malaysia, which industrialized later, entered stage 2 more recently, and report fivefold lower prevalence^{36,37}. Even if newly industrialized regions never reach the incidence and prevalence levels of IBD seen in Europe, achieving a prevalence of 0.1% over the next few decades will still impose a substantial burden on societies with large populations like India and China. Moreover, the age distribution of IBD cases in these regions is likely to be skewed towards the young³⁸, exacerbating the economic burden for working-age adults with IBD or caregivers of children with IBD. As managing IBD is expensive, equitable access to treatments is essential to avoid the indirect burden of a chronic illness in a young population, such as loss of work productivity^{19,39}.

Many early industrialized regions are in stage 3. Variations in IBD-susceptibility genes and environmental exposures influence the magnitude and change of incidence over time⁴⁰. Methodological differences in surveillance systems may also contribute to heterogeneity, potentially leading to outliers in epidemiologic data^{41,42}. Stage 3 regions face the rising burden of IBD, with increasing numbers of individuals living with IBD and a growing proportion of older people.

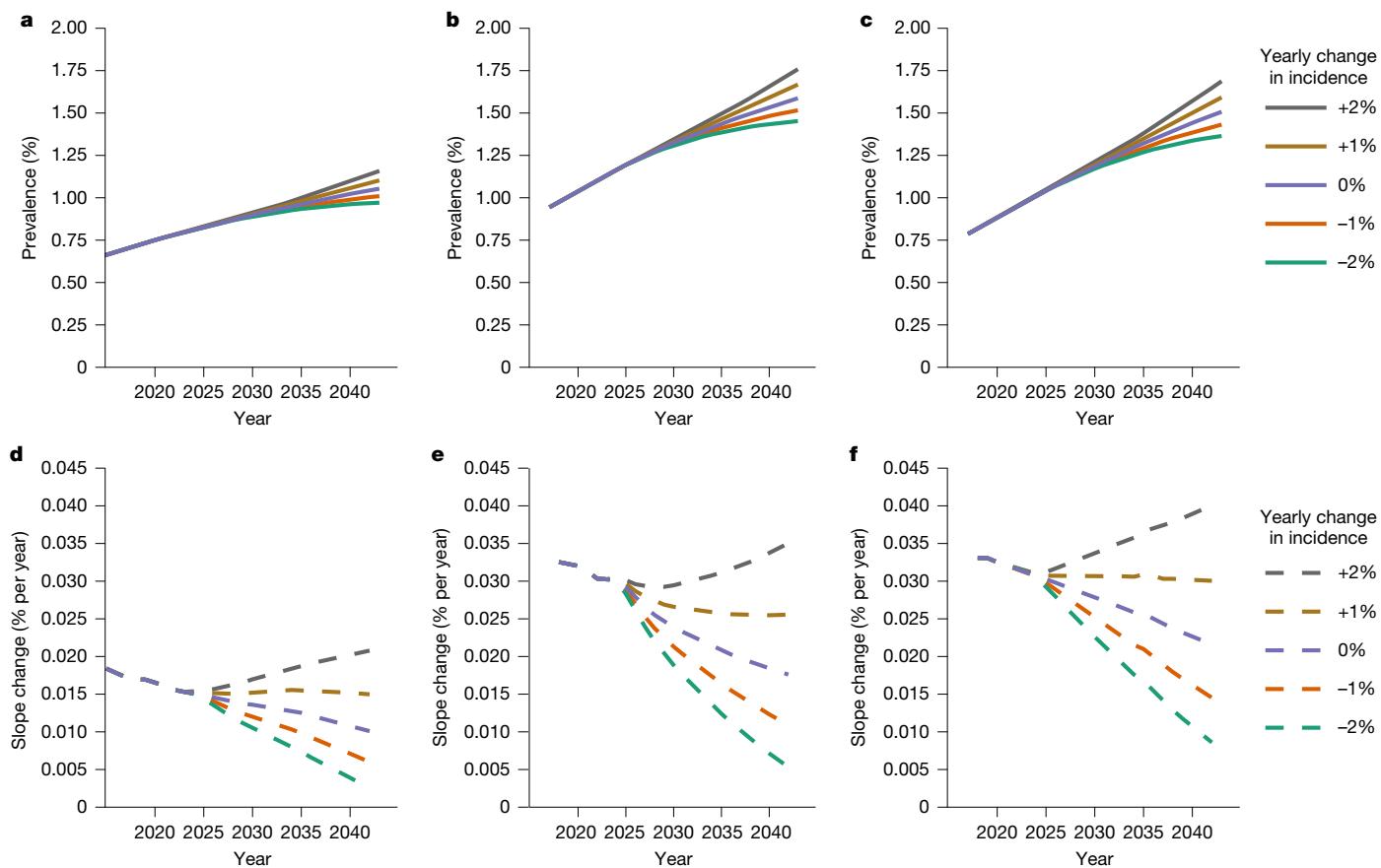


Fig. 5 | Calculations of IBD prevalence over time under scenarios of incidence rate change. **a**, The prevalence in Canada under a 2%, 1%, 0%, -1% and -2% change in incidence per year, with the base incidence rate set equal to the average incidence from 2007 to 2014. **b**, The prevalence in Denmark under a 2%, 1%, 0%, -1% and -2% change in incidence per year, with the base incidence rate set equal to the average incidence from 2010 to 2017. **c**, The prevalence in Scotland under a 2%, 1%, 0%, -1% and -2% change in incidence per year, with the base incidence

rate set equal to the average incidence from 2010 to 2017. **d**, Central difference approximations of the slopes of the time-dependent prevalence in Canada for each of the five yearly incidence rate change scenarios. **e**, Central difference approximations of the slopes of the time-dependent prevalence in Denmark for each of the five yearly incidence rate change scenarios. **f**, Central difference approximations of the slopes of the time-dependent prevalence in Scotland for each of the five yearly incidence rate change scenarios.

These regions must balance the demands of new diagnoses in young individuals with the intricacies of caring for an ageing IBD population. The challenges are particularly pronounced in adult gastroenterology clinics, as managing IBD with immunosuppressive therapies and/or intestinal resections becomes more complex in older people with age-related comorbidities such as diabetes, cancer or dementia⁴³. Proactive healthcare planning is essential, as multidisciplinary care teams will be needed to manage ageing IBD populations⁴⁴.

Prevalence will continue to steadily climb while mortality is low². However, the shifting demographics of an ageing IBD population over the next few decades is projected to slow the rate of prevalence growth: although the prevalence will still increase, it will do so at a slower pace. Assuming a stable incidence, the prevalence is estimated to range between 1.02% and 1.59% in the 2040s in Canada, Denmark and Scotland, and between 1.10% and 1.76% if the incidence increases by 2% per year. Future epidemiologic studies in stage 3 regions that monitor the slowing of prevalence growth are essential for determining the timing of the transition to stage 4. Our models underscore that interventions aimed at reducing IBD incidence have the potential to stabilize and possibly decrease IBD prevalence.

Prevention of some IBD cases is an ambitious yet potentially achievable goal^{40,45}. Pre-disease cohorts have linked biological markers (such as genes, serology, microorganisms) to future IBD development, indicating the potential to identify those at moderate risk who might benefit from targeted interventions, such as modifying their

microbiome^{46–48}. For those at higher risk, pharmaceutical interventions could be explored to delay IBD onset, similar to strategies used in type 1 diabetes with teplizumab⁴⁹ and rheumatoid arthritis with abatacept⁵⁰. Observational research also indicates that lifestyle and dietary changes may reduce CD and UC incidence³⁰. Thus, prioritizing research on preventative strategies to reduce the risk of IBD is crucial.

This study represents the most comprehensive analysis of population-based data on IBD incidence and prevalence, but several limitations must be considered (see the ‘Strengths and limitations’ section in the Methods). The quality of data sources varied, particularly for historical data from the twentieth century and data from regions without robust healthcare surveillance systems. While our machine-learning classifier demonstrated high accuracy (>95%), classification errors were more common in regions with limited data and in those transitioning between stages. Investigating distinct transition phases between predefined stages or further subdividing these stages may become feasible as more longitudinal data from diverse regions become available. Moreover, the PDE model did not account for differential mortality between IBD and non-IBD populations, potentially overestimating future prevalence.

Conclusion

Incidence and prevalence data spanning the past century were analysed to explore epidemiologic trends in the evolution of IBD across the

world. Machine-learning algorithms were applied to the data to create a classification system across three epidemiologic stages: stage 1 (emergence), low incidence and prevalence; stage 2 (acceleration in incidence), rapidly increasing incidence and low prevalence; and stage 3 (compounding prevalence), reduced incidence growth while prevalence rapidly increases. Our classification system was used to define benchmarks for transition across these stages for regions worldwide. Modelling population demographics enabled us to predict prevalence over the next two decades, while also demonstrating the plateauing of prevalence that characterizes stage 4 (prevalence equilibrium). These data can be used by healthcare systems and society to address the rising global burden of IBD. Furthermore, the testable predictions made by the epidemiologic transition theory that we advance here should serve as a model for future researchers investigating analogous diseases with global, longitudinal epidemiologic data.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-025-08940-0>.

1. Kaplan, G. G. The global burden of IBD: from 2015 to 2025. *Nat. Rev. Gastroenterol. Hepatol.* **12**, 720–727 (2015).
2. Kaplan, G. G. & Windsor, J. W. The four epidemiological stages in the global evolution of inflammatory bowel disease. *Nat. Rev. Gastroenterol. Hepatol.* **18**, 56–66 (2021).
3. Molodecky, N. A. et al. Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. *Gastroenterology* **142**, 46–54 (2012).
4. Ng, S. C. et al. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *Lancet* **390**, 2769–2778 (2017).
5. Kaplan, G. G. & Ng, S. C. Globalisation of inflammatory bowel disease: perspectives from the evolution of inflammatory bowel disease in the UK and China. *Lancet Gastroenterol. Hepatol.* **1**, 307–316 (2016).
6. Xue, M. et al. Environmental factors associated with risk of Crohn's disease development in the Crohn's and Colitis Canada—Genetic, Environmental, Microbial Project. *Clin. Gastroenterol. Hepatol.* **22**, 1889–1897 (2024).
7. Kaplan, G. G. & Ng, S. C. Understanding and preventing the global increase of inflammatory bowel disease. *Gastroenterology* **152**, 313–321 (2017).
8. Kuenzig, M. E. et al. Twenty-first century trends in the global epidemiology of pediatric-onset inflammatory bowel disease: systematic review. *Gastroenterology* **162**, 1147–1159 (2022).
9. Mulder, D. J., Noble, A. J., Justinich, C. J. & Duffin, J. M. A tale of two diseases: the history of inflammatory bowel disease. *J. Crohns Colitis* **8**, 341–348 (2014).
10. Gunesh, S., Thomas, G. A., Williams, G. T., Roberts, A. & Hawthorne, A. B. The incidence of Crohn's disease in Cardiff over the last 75 years: an update for 1996–2005. *Aliment Pharmacol. Ther.* **27**, 211–219 (2008).
11. Vegh, Z. et al. Incidence and initial disease course of inflammatory bowel diseases in 2011 in Europe and Australia: results of the 2011 ECCO-EpiCom inception cohort. *J. Crohns Colitis* **8**, 1506–1515 (2014).
12. Lewis, J. D. et al. Incidence, prevalence, and racial and ethnic distribution of inflammatory bowel disease in the United States. *Gastroenterology* **165**, 1197–1205 (2023).
13. Jones, G. R. et al. IBD prevalence in Lothian, Scotland, derived by capture-recapture methodology. *Gut* **68**, 1953–1960 (2019).
14. Coward, S. et al. Past and future burden of inflammatory bowel diseases based on modeling of population-based data. *Gastroenterology* **156**, 1345–1353 (2019).
15. Coward, S. et al. Forecasting the incidence and prevalence of inflammatory bowel disease: a Canadian nationwide analysis. *Am. J. Gastroenterol.* **119**, 1563–1570 (2024).
16. Quaresma, A. B. et al. Temporal trends in the epidemiology of inflammatory bowel diseases in the public healthcare system in Brazil: a large population-based study. *Lancet Reg. Health Am.* **13**, 100298 (2022).
17. Wu, X. et al. Top 10 algorithms in data mining. *Knowl. Inform. Syst.* **14**, 1–37 (2008).
18. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
19. Burisch, J. et al. The cost of inflammatory bowel disease in high-income settings: a Lancet Gastroenterology & Hepatology Commission. *Lancet Gastroenterol. Hepatol.* **8**, 458–492 (2023).
20. Prados de la Escosura, L. Augmented human development in the age of globalization. *Econ. Hist. Rev.* **74**, 946–975 (2021).
21. Global Health Observatory. *Prevalence of Obesity Among Adults, BMI≥30, Age-Standardized Estimates by Country* (World Health Organization, 2017). apps.who.int/gho/data/node/main.A900A?lang=en.
22. Department of Economic and Social Affairs-Population Division. *World Urbanization Prospects: The 2018 Revision* (United Nations, 2018). population.un.org/wup/Download/.
23. Tracking Universal Health Coverage: 2023 Global Monitoring Report (World Health Organization and International Bank for Reconstruction and Development/The World Bank, 2023).
24. Azzam, A. Is the world converging to a 'Western diet'? *Publ. Health Nutr.* **24**, 309–317 (2021).
25. Brinks, R. & Landwehr, S. Age- and time-dependent model of the prevalence of non-communicable diseases and application to dementia in Germany. *Theor. Popul. Biol.* **92**, 62–68 (2014).
26. *Table 17-10-0057-01 Projected Population, by Projection Scenario, Age and Sex, as of July 1 (x 1,000)* (Statistics Canada, 2022). www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1710005701.
27. *Population Projection (FRDK123)* (Statbank Denmark, 2023). extranet.dst.dk/pyramide/pyramide.htm#&lang=en.
28. *2018-Based Principal Population Projections For 2018-2043, By Sex, NHS Board Area and Single Year of Age* (National Records of Scotland, 2020). www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/population/population-projections/sub-national-population-projections/2018-based/detailed-datasets.
29. Cheney, W. & Kincaid, D. *Numerical Mathematics and Computing* 5th edn, 633 (Brooks/Cole-Thompson, 2004).
30. Lopes, E. W. et al. Lifestyle factors for the prevention of inflammatory bowel disease. *Gut* **72**, 1093–1100 (2023).
31. Burisch, J., Clayton, J., Hernandez, I., Hou, J. K. & Kaplan, G. G. The cost of IBD care—how to make it sustainable. *Clin. Gastroenterol. Hepatol.* **23**, 386–395 (2024).
32. Ananthakrishnan, A. N., Kaplan, G. G. & Ng, S. C. Changing global epidemiology of inflammatory bowel diseases: sustaining health care delivery into the 21st century. *Clin. Gastroenterol. Hepatol.* **18**, 1252–1260 (2020).
33. Mak, J. W. Y. et al. Development of the global inflammatory bowel disease visualization of epidemiology studies in the 21(st) century (GIVES-21). *BMC Med. Res. Methodol.* **23**, 129 (2023).
34. Murakami, Y. et al. Estimated prevalence of ulcerative colitis and Crohn's disease in Japan in 2014: an analysis of a nationwide survey. *J. Gastroenterol.* **54**, 1070–1077 (2019).
35. Park, S. H. et al. A 30-year trend analysis in the epidemiology of inflammatory bowel disease in the Songpa-Kangdong district of Seoul, Korea in 1986–2015. *J. Crohns Colitis* **13**, 1410–1417 (2019).
36. Yang, H., Zhou, R., Bai, X., Guo, M., Ruan, G., Wang, L. & Qian, J. Trend and geographic variation in incidence and prevalence of inflammatory bowel disease in regions across China: a nationwide employee study between 2013 and 2016. *Front. Med.* **9**, 900251 (2022).
37. Mokhtar, N. M. et al. A four-decade analysis of the incidence trends, sociodemographic and clinical characteristics of inflammatory bowel disease patients at single tertiary centre, Kuala Lumpur, Malaysia. *BMC Publ. Health* **19**, 550 (2019).
38. Shah, S. C. et al. Sex-based differences in the incidence of inflammatory bowel diseases—pooled analysis of population-based studies from the Asia-Pacific region. *Aliment Pharmacol. Ther.* **49**, 904–911 (2019).
39. Kuenzig, M. E. et al. The 2023 impact of inflammatory bowel disease in Canada: indirect (individual and societal) and direct out-of-pocket costs. *J. Can. Assoc. Gastroenterol.* **6**, S16–S22 (2023).
40. Herauf, M. et al. Commentary on the epidemiology of inflammatory bowel disease in compounding prevalence nations: toward sustaining healthcare delivery. *Gastroenterology* **166**, 949–956 (2024).
41. Long, M. D. et al. Challenges in designing a national surveillance program for inflammatory bowel disease in the United States. *Inflamm. Bowel Dis.* **20**, 398–415 (2014).
42. Molodecky, N. A. et al. Challenges associated with identifying the environmental determinants of the inflammatory bowel diseases. *Inflamm. Bowel Dis.* **17**, 1792–1799 (2011).
43. Shaffer, S. R. et al. The 2023 impact of inflammatory bowel disease in Canada: special populations-IBD in seniors. *J. Can. Assoc. Gastroenterol.* **6**, S45–S54 (2023).
44. Mathias, H. et al. The 2023 impact of inflammatory bowel disease in Canada: access to and models of care. *J. Can. Assoc. Gastroenterol.* **6**, S11–S21 (2023).
45. Chhibba, T. et al. Environmental risk factors of inflammatory bowel disease: towards a strategy of preventative health. *J. Crohns Colitis* <https://doi.org/10.1093/ecco-jcc/jiac042> (2025).
46. Raygoza Garay, J. A. et al. Gut microbiome composition is associated with future onset of Crohn's disease in healthy first-degree relatives. *Gastroenterology* **165**, 670–681 (2023).
47. Torres, J. et al. Serum biomarkers identify patients who will develop inflammatory bowel diseases up to 5 years before diagnosis. *Gastroenterology* **159**, 96–104 (2020).
48. Livanos, A. E. et al. Anti-integrin avß6 autoantibodies are a novel biomarker that antedate ulcerative colitis. *Gastroenterology* **164**, 619–629 (2023).
49. Herold, K. C. et al. An anti-CD3 antibody, teplizumab, in relatives at risk for type 1 diabetes. *N. Engl. J. Med.* **381**, 603–613 (2019).
50. Cope, A. P. et al. Abatacept in individuals at high risk of rheumatoid arthritis (APIPPRA): a randomised, double-blind, multicentre, parallel, placebo-controlled, phase 2b clinical trial. *Lancet* **403**, 838–849 (2024).

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



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

- ¹Division of Gastroenterology and Hepatology, Department of Medicine, University of Calgary, Calgary, Alberta, Canada. ²Centre for Health Informatics and Department of Community Health Sciences, University of Calgary, Calgary, Alberta, Canada. ³Maidstone and Tunbridge Wells NHS Trust, Kent, UK. ⁴Division of Gastroenterology, Department of Internal Medicine, HCRC IMERI, Faculty of Medicine, Universitas Indonesia, Dr. Cipto Mangunkusumo Hospital, Jakarta, Indonesia. ⁵F. Widjaja Inflammatory Bowel Disease Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA. ⁶International Organization for the study of Inflammatory Bowel Disease (IOIBD), Stockholm, Sweden. ⁷Department of Gastroenterology, All India Institute of Medical Sciences, New Delhi, India. ⁸Sir Jeffrey Cheah Sunway Medical School, Faculty of Medical and Life Sciences, Sunway University, Selangor, Malaysia. ⁹Division of Gastroenterology, Department of Medicine, College of Medicine, King Saud University, Riyadh, Saudi Arabia. ¹⁰IBD Unit, Gastroenterology Department, Hospital Privado Universitario de Córdoba, Instituto Universitario de Ciencias Biomédicas de Córdoba, Córdoba, Argentina. ¹¹IBD Center, Asian Institute of Gastroenterology, Hyderabad, India. ¹²SickKids Inflammatory Bowel Disease Centre, Division of Gastroenterology, Hepatology and Nutrition, The Hospital for Sick Children (SickKids), Toronto, Ontario, Canada. ¹³Department of Paediatrics and Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, Ontario, Canada. ¹⁴Child Health Evaluative Sciences, SickKids Research Institute, Toronto, Ontario, Canada. ¹⁵ICES, Toronto, Ontario, Canada. ¹⁶Department of Medicine, and the University of Manitoba IBD Clinical and Research Centre, Max Rady College of Medicine, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, Manitoba, Canada. ¹⁷Gastroenterology Department, Parc Taulí Hospital Universitari, Institut d'Investigació i Innovació Parc Taulí (I3PT-CERCA), Sabadell, Spain. ¹⁸CIBERehd, Instituto de Salud Carlos III, Madrid, Spain. ¹⁹Departament de Medicina, Universitat Autònoma de Barcelona, Barcelona, Spain. ²⁰Gastrourine, Medical Division, Copenhagen University Hospital—Amager and Hvidovre, Hvidovre, Denmark. ²¹Copenhagen Center for Inflammatory Bowel Disease in Children, Adolescents and Adults, Copenhagen University Hospital—Amager and Hvidovre, Denmark. ²²Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. ²³Division of Gastroenterology and Hepatology, Department of Medicine, Raja Isteri Pengiran Anak Saleha (RIPAS) Hospital, Bandar Seri Begawan, Brunei Darussalam. ²⁴Division of Gastroenterology, Rabin Medical Center, Petah-Tikva, Israel, Affiliated to the Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel. ²⁵Department of Gastroenterology, Postgraduate Institute of Medical Education and Research, Chandigarh, India. ²⁶Digestive Disease Institute, Cleveland Clinic Abu Dhabi, Abu Dhabi, UAE. ²⁷Digestive Disease and Surgery Institute, Cleveland Clinic Foundation, Cleveland, OH, USA. ²⁸Department of Medicine, University of Otago, Christchurch, New Zealand. ²⁹Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. ³⁰Internal Medicine Faculty, Hanoi Medical University, Hanoi, Vietnam. ³¹Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. ³²Division of Gastroenterology & Hepatology, National University Hospital, Singapore, Singapore. ³³Division of Gastroenterology, Department of Medicine, Faculty of Medicine, University Malaya, Kuala Lumpur, Malaysia. ³⁴Blizard Institute, Barts & The London School of Medicine, Queen Mary University of London, London, UK. ³⁵Tropical Gastroenterology & Nutrition Group, University of Zambia School of Medicine, Lusaka, Zambia. ³⁶University of Edinburgh, Institute for Regeneration and Repair, Edinburgh, UK. ³⁷Department of Gastroenterology, Pablo Tobon Uribe Hospital, Medellín, Colombia. ³⁸JSC Research Institute of Cardiology and Internal Diseases of the Ministry of Health of the Republic of Kazakhstan, Astana, Republic of Kazakhstan. ³⁹Asfendiyarov Kazakh National Medical University, Almaty, Kazakhstan. ⁴⁰Center for Advanced IBD Research and Treatment, Kitasato University Kitasato Institute Hospital, Tokyo, Japan. ⁴¹Department of Gastroenterology, Kitasato University School of Medicine, Sagamihara, Japan. ⁴²Health Sciences Postgraduate Program, Pontifícia Universidade Católica do Paraná, Curitiba, Brazil. ⁴³Division of Gastroenterology, McGill University, Montreal, Quebec, Canada. ⁴⁴Department of Internal Medicine and Oncology, Semmelweis University, Budapest, Hungary. ⁴⁵Centre for Genomics and Experimental Medicine (CGEM), Institute of Genetics & Cancer, University of Edinburgh, Edinburgh, UK. ⁴⁶Edinburgh IBD Unit, Western General Hospital, NHS Lothian, Edinburgh, UK. ⁴⁷Division of Gastroenterology, Siriraj Hospital, Mahidol University, Bangkok, Thailand. ⁴⁸Division of Gastroenterology and Hepatology, Mayo Clinic College of Medicine and Science, Rochester, MN, USA. ⁴⁹Department of Pediatrics, Örebro University Hospital, Stockholm, Sweden. ⁵⁰Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Hong Kong, China. ⁵¹Gastroenterology, The First Affiliated Hospital of Kunming Medical University, Kunming, China. ⁵²Yunnan Province Clinical Research Center for Digestive Diseases, Kunming, China. ⁵³Conde S. Januário Hospital, Macao SAR, China. ⁵⁴Department of Healthcare Epidemiology, School of Public Health in the Graduate School of Medicine, Kyoto University, Kyoto, Japan. ⁵⁵Division of Clinical Epidemiology, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden. ⁵⁶Department of Gastroenterology, National Academy of Medical Sciences, Kathmandu, Nepal. ⁵⁷UNOESC Curso de Medicina: Universidade do Oeste de Santa Catarina, Joaçaba, Brazil. ⁵⁸Health Sciences Postgraduate Program, Pontifícia Universidade Católica do Paraná, Curitiba, Brazil. ⁵⁹Inflammatory Bowel Disease Center, University of Chicago Medicine, Chicago, IL, USA. ⁶⁰Division of Gastroenterology and Hepatology, Department of Internal Medicine, Tokai University School of Medicine, Isehara, Japan. ⁶¹Head of the Inflammatory Bowel Diseases Unit, HIGEA, Mendoza, Argentina. ⁶²The Juliet Keidan Institute of Pediatric Gastroenterology & Nutrition, The Hebrew University of Jerusalem, Shaare Zedek Medical Center, Jerusalem, Israel. ⁶³Cooperativa de Servicios Médicos (COSEM), Montevideo, Uruguay. ⁶⁴Department of Internal Medicine, National Taiwan University Hospital and College of Medicine, Taipei, Taiwan. ⁶⁵Inflammatory Bowel Disease Clinic, Department of Gastroenterology, National Institute of Medical Sciences and Nutrition and National Autonomous University of Mexico (UNAM), Mexico City, Mexico. ⁶⁶Department of Gastroenterology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea. ⁶⁷Department of Medicine and Therapeutics, Institute of Digestive Disease, State Key Laboratory of Digestive Diseases, Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Hong Kong, China. ⁶⁸Microbiota I-Center (MagIC), Hong Kong, China. ⁶⁹New Cornerstone Science Laboratory, The Chinese University of Hong Kong, Hong Kong, China. ⁷⁰These authors contributed equally: Lindsay Hracs, Joseph W. Windsor. ⁷¹These authors jointly supervised this work: Siew C. Ng, Gilad G. Kaplan. ⁷²e-mail: siewchiennng@cuhk.edu.hk; ggkaplan@ucalgary.ca

Global IBD Visualization of Epidemiology Studies in the 21st Century (GIVES-21) Research Group

- Lindsay Hracs^{1,70}, Joseph W. Windsor^{1,70}, Stephanie Coward¹, Yvonne Abbey³, Murdani Abdullah⁴, Maria T. Abreu^{5,6}, Vineet Ahuja^{6,7}, Raja Affendi Raja Ali⁸, Mansour Attuwajiri⁹, Domingo Balderramo¹⁰, Rupa Banerjee^{6,11}, Eric I. Benchimol^{12,13,14,15}, Charles N. Bernstein^{6,16}, Eduard Brunet-Mas^{17,18,19}, Johan Burisch^{6,20,21,22}, Vui Heng Chong²³, Iris Dotan^{6,24}, Usha Dutta²⁵, Sara El Ouali^{26,27}, Angela Forbes²⁸, Anders Forss²⁹, Richard Gearry²⁸, Viet Hang Dao³⁰, Juanda Leo Hartono^{31,32}, Ida Hilmi³³, Phoebe Hodges^{34,35}, Gareth-Rhys Jones³⁶, Fabián Juliao-Baños³⁷, Jamilia Kaibullayeva^{38,39}, Paul Kelly^{34,35}, Taku Kobayashi^{40,41}, Paulo Gustavo Kotze^{6,42}, Peter L. Lakatos^{6,43,44}, Charlie W. Lees^{45,46}, Julajak Limsrivilai⁴⁷, Bobby Lo^{20,21}, Edward V. Loftus Jr^{6,48}, Jonas F. Ludvigsson^{29,49}, Joyce W. Y. Mak⁵⁰, YingLei Miao^{51,52}, Ka Kei Ng⁵³, Shinji Okabayashi⁵⁴, Ola Olén⁵⁵, Remo Panaccione¹⁶, Mukesh Sharma Paudel⁵⁶, Abel Botelho Quaresma^{57,58}, David T. Rubin^{6,59}, Marcellus Simadibrata⁴, Yang Sun^{51,52}, Hidekazu Suzuki⁶⁰, Martin Toro⁶¹, Dan Turner^{6,62}, Beatriz Iade⁶³, Shu Chen Wei⁶⁴, Jesus K. Yamamoto-Furusho⁶⁵, Suk-Kyun Yang⁶⁶, Siew C. Ng^{6,67,68,69,71} & Gilad G. Kaplan^{1,6,71}

Methods

Systematic review

We conducted a systematic review of population-based studies to investigate changes in incidence and prevalence of CD and UC across global regions over time. The systematic review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement⁵¹.

Search strategy. We reassessed population-based studies reporting the incidence and prevalence of IBD from our team's two previous systematic reviews^{3,4}, covering the period before 2017. We updated the previous systematic reviews by performing a search of Embase, MEDLINE, PubMed and Web of Science for the period covering 1 January 2017 to 8 May 2024 with no language restrictions (the search strategy is shown in Supplementary Table 3).

Study selection. All studies underwent independent title and abstract screening by at least two reviewers. The reference lists were also inspected for additional citations. We included population-based studies published in article form at any time or abstracts from 2020 or later, which reported the incidence and/or prevalence of CD and/or UC separately or provided sufficient information to calculate the corresponding incidence or prevalence. Population-based studies were defined as studies deriving incidence and/or prevalence for the entire population of a specified geographical region or a representative sample. Exclusion criteria were as follows: no defined geographical region, less than 1 year of data, study population limited to paediatric cases only, self-reported cases identified by survey, reviews and clinical trials. Studies published in languages other than English were translated using Google Translate⁵². All included articles were re-evaluated by authors with local expertise and local language proficiency (see the 'Data verification' section below).

Data extraction. Author, publication date, geographical area (region and/or subregion), study period, ages, year (or midpoint of range), total population, data type (incidence or prevalence), rate type (crude or standardized), rate of UC, CD and/or IBD-unclassified (IBD-u) and case counts were extracted from each selected study. Aggregate rates were removed if an annual rate for the mid-period year was available. Missing population values were pulled from publicly available official statistical records^{53–61}. For incidence data, we calculated the summed population for the time period; for prevalence data, we recorded the population at a given timepoint or used the mid-period population in cases of period prevalence. Data extraction and verification were conducted by at least two reviewers. In cases of disagreement between reviewers, consensus was reached through discussion.

Quality assessment. The quality of each study was independently assessed by two reviewers using a modified Joanna Briggs Institute Checklist for Prevalence Studies⁶². An additional quality measure was included to identify studies that were strictly population-based (that is, those that identified the entire population sample). Studies that used a representative sample or may have missed some cases due to the sampling method were of lower quality. The results of the quality assessment are presented in Supplementary Table 4.

Data verification. All steps of the systematic review (abstract review, full-text review, data extraction, quality assessment) were conducted by at least two members of a trained centralized team at the University of Calgary to ensure methodological consistency. Additional data verification occurred through international partners from the International Organization for the study of IBD (IOIBD) and the Global IBD Visualization of Epidemiology Studies in the 21st Century (GIVES-21) consortium. Experts from IOIBD and GIVES-21 confirmed the inclusion

of studies, verified data accuracy and suggested additional studies that may have been missed (Supplementary Fig. 1). They confirmed the population-based status of the studies, addressed conflicts in language translation and provided local context (for example, study quality), including explanations for outlier results.

Rate calculation. Crude incidence or prevalence was calculated from case counts and the population of the catchment area. If case counts were not provided or the population of the catchment area was not available, we used the rates reported in the papers, whether presented in text, tables or extracted from plots using OriginPro⁶³ or juicr⁶⁴.

Temporal data analysis. When multiple studies overlapped in region and time, we averaged rates to create a single value for that region and year. Data preprocessing for analysis and visualization of decade-level rates involved weighting the mean rates calculated by region for each year according to the population of that area for that year. Weighted rates were aggregated by decade. This weighting method enabled us to combine regional and subregional data to generate a single value for a region in a particular decade. Aggregated subregional data were treated as representative for a region when no other regional data were available. Missing case counts were back-calculated from provided per 100,000 values and populations. In rare cases in which the catchment area population values were not available, non-weighted means were used in the aggregation. When IBD-u was reported separately, it was combined with UC.

Temporal trends for studies and regions that reported at least three datapoints within a 5-year period were established using Poisson regression (or, if overdispersed, negative binomial regression) models built in R⁶⁵, with the year as the sole predictor variable and the incidence rate of either CD or UC as the outcome. We determined the average annual percentage change with associated 95% CI for incidence by exponentiating the β coefficients from the regression models (Supplementary Table 1).

Machine-learning classification

Exploratory data analysis. We visually inspected scatterplots of historical trends in incidence and prevalence for early industrialized, newly industrialized and emerging regions to determine epidemiologic stages for a subset of regions with data extracted during the systematic review. On the basis of the assumption that Canada, the United States, most of western Europe, Australia and New Zealand are currently in stage 3, we assigned stage classifications to approximately 65% of our dataset (Supplementary Fig. 16).

k-nearest-neighbors-assisted labelled dataset creation. As regions and subregions do not transition between stages simultaneously, we built a k -NN classification algorithm to support manual data labelling. The k -NN algorithm facilitated an iterative labelling process by classifying regions with robust historical incidence or prevalence data into one of three classifications: stage 1, stage 2 or stage 3. Owing to skewed class proportions (that is, scarcity of stage 1 data) and missing values within classes, we developed four separate models: incidence of CD, incidence of UC, prevalence of CD and prevalence of UC. Each model used two features: incidence or prevalence and the absolute difference in incidence or prevalence from the previously available year of data.

A 75/25 train/test split ($n = 1,581/n = 527$) was applied. The value of k -neighbours was set to the square root of the number of instances in the training data and adjusted accordingly for each model. A Monte Carlo simulation with 1,000 sampling loops was run for each model. The model accuracy (Supplementary Table 5) was determined using the following formula:

$$\text{acc}(\%) = 100 \times \left(\frac{M}{L} \right) \quad (1)$$

Article

where M equals the number of misclassifications (that is, instances in which the model's classification output differed from our manual classification) and L equals the number of Monte Carlo sampling loops. Three iterations were performed in which the k -NN model output was inspected with successive relabelling of the input data, achieved through consensus among three analysts. At each iteration, additional regions and subregions were added to the labelled dataset until approximately 80% of our total dataset was labelled, with the intention of using it for training and validating a random-forest classifier.

Random-forest classifier. A random-forest classifier was built in R⁶⁵ using the randomForest⁶⁶ package. Random forests use an ensemble learning method, in which a specified number of decision trees is generated, and the results are aggregated. The class selected by the model most frequently becomes the resulting classification¹⁸. Each individual decision tree determines a class prediction based on a random subset of features, which reduces the likelihood of overfitting and improves the accuracy of class prediction⁶⁷. Random-forest classifiers do not require data scaling and are robust to outliers and noise, which is essential given the heterogeneity and imbalance in our dataset and the complexity of analysing systematic review data.

Features and data imputation. Classifying a region's stage in a particular year was based on 16 possible features: (1) CD incidence; (2) UC incidence; (3) CD prevalence; (4) UC prevalence; (5) rates of change for incidence of CD; (6) rates of change for incidence of UC; (7) rates of change for prevalence of CD; (8) rates of change for prevalence of UC; and (9–16) indicators of imputed values for each of the preceding eight variables. The structure of the classifier required a value to be specified for each of the above features for each year of available data by region.

Imputation was used to ensure that a value was specified for each of the primary eight features. Imputation was conducted within a single region across all available datapoints for that region. When at least two datapoints were available for a region, linear interpolation was used, using the next observation carried backward and the last observation carried forward to extrapolate missing values outside the available interval. In cases in which only a single datapoint was available for a region, that value was extrapolated to all missing data for that region. Zero imputation was applied to the remaining missing datapoints. Features 5–8 were included to account for the potential increase or decrease in incidence and prevalence over time. As classifications were made by the random-forest model at the level of the region and year, and many regions lacked incidence or prevalence values for certain years, features 9–16 were included so the random-forest classifier could account for unavailable data that were imputed for classification purposes.

Random-forest models provide a measure of feature importance, allowing for an examination of which data types (incidence or prevalence) and which disease types (CD or UC) contributed most to the classifications across the three stages (Supplementary Fig. 20).

Training, validation and model architecture. The random-forest model was trained using a subset of data from the labelled dataset ($n = 1,647$). Training and validation sets were created using a 75/25 split ($n = 1,235/n = 412$) of the labelled dataset, with the out-of-bag (OOB) error estimate used to tune model hyperparameters: ntree (number of trees aggregated) and mtry (number of features used at each split in the tree). The model with the smallest OOB error estimate (OOB = 4.86%) used ntree = 1,000 and mtry = 5 (out of 16 possible features). The random-forest's classification accuracy on the unseen validation data was 95.15% (95% CI = 92.60–97.01); this means that the random forest correctly classified a region-year as stage 1, stage 2 or stage 3 approximately 95% of the time, indicating an appropriate model fit and performance (Supplementary Table 6).

Random-forest output. The output from the random-forest model was used to assign an epidemiologic stage to regions with limited incidence and prevalence data ($n = 842$), resulting in a complete dataset with stage classifications for all regions in the dataset across time ($n = 2,489$). As classifications were based on a single year of data, a decade-level stage classification for a region was calculated by identifying the mode class label from the random-forest model output for years within a decade. In cases in which the model provided an even split of stage classifications for a region within a decade ($n = 2$: Poland, 2010s; China, 2020s), four expert reviewers assessed available incidence and prevalence data and manually assigned a stage classification for that decade. When no regional data were available, the stage classification for a region was determined based solely on subregional data.

Coalescing ranges. The distribution of incidence and prevalence data for each stage was derived from the machine-learning models and used to calculate CRs, defined by the 25th and 75th percentiles. Negative binomial regression models followed by post hoc comparisons of estimated marginal means using the emmeans⁶⁸ package in R with Tukey adjustment for multiple comparisons were used to evaluate significant differences in CR between stage 1, stage 2 and stage 3 in each of the following: CD incidence, UC incidence, CD prevalence and UC prevalence.

Ulcerative colitis/Crohn's disease ratio. To calculate the UC:CD ratio, the annual population-weighted mean UC incidence was divided by the annual population-weighted mean CD incidence and assigned an epidemiologic stage on the basis of the output of the machine-learning model. The association between UC:CD ratio and epidemiologic stage was modelled using negative binomial regression. Pairwise comparisons between the stages were performed using the emmeans⁶⁸ library in R, with estimated marginal means using Tukey adjustment for multiple comparisons.

Societal indicators. Five societal indicators were examined in relationship to epidemiologic stages: the AHDI, obesity rate, percentage urbanization, the Universal Health Coverage Service Coverage Index (UHC) and the WDI. AHDI is an index score (0–1) that captures the geometric mean of normalized life expectancy, mean years of education, gross domestic product (GDP) per capita and Varieties of Democracy's Liberal Democracy Index²⁰. AHDI is available in 5-year increments, for which we performed inner-linear interpolation to achieve annual measures. Obesity data were extracted from the WHO database, providing a measure of the percentage of an adult population with a BMI > 30 kg m⁻² (ref. 69). Percentage urbanization data were extracted from a United Nations database, providing a measure of the mid-year percentage of a population living in an urban setting⁷⁰. UHC is an index score (0–100) that quantifies various aspects of healthcare, including reproductive health (for example, the percentage of pregnant people with ≥4 prenatal care visits), prevention of communicable diseases (for example, the percentage of 1-year-old children with adequate diphtheria, tetanus and pertussis vaccination), non-communicable diseases (for example, the percentage of cervical cancer screening in women aged 30–49 years) and healthcare access (for example, the number of hospital beds per capita)²². The UHC is also available in 5-year increments, for which we performed an inner-linear interpolation to achieve annual measures. WDI is an index score (0–1) calculated by dividing the available calories per person per day from animal oils and fats, milk, eggs, plant oils and fats, and sugars by the total available calories per person per day, similar to the methodology provided by Azzam in 2021, extended to all regions within our dataset²⁴. Regional calorie availability was extracted from the Food Balance sheets published by the Food and Agriculture Organization⁷¹.

Time- and region-specific values from each of these five indicators were stratified by the three epidemiologic stages (as derived by the machine-learning classifier) and statistically compared across

stages using Kruskal–Wallis nonparametric tests with post hoc comparisons using Wilcoxon rank-sum tests with adjustment for multiple comparisons.

Modelling stage 4 using PDEs

To explore the potential growth characteristics of IBD prevalence in stage 3 regions, we modelled the time-dependent prevalence using historic prevalence and incidence data, plus population projections from Canada, Denmark and Scotland (Lothian). Calculations were completed using Mathematica (v.13.1)⁷².

Data sources. Incidence and prevalence of IBD were calculated for Canada, Denmark and Scotland (Lothian) from administrative data provided by each region (Supplementary Table 7).

For Canada, population-based provincial administrative healthcare data were combined from Alberta, British Columbia, Manitoba, Newfoundland and Labrador, Nova Scotia, Ontario, Quebec and Saskatchewan to capture data for 2007–2014¹⁵. In Denmark, nationwide individual-level healthcare information was obtained from the Danish National Patient Register for 2010–2017 (ref. 73). For Scotland, data for the Lothian region (Edinburgh and surrounding area) were sourced from TrakCare (InterSystems) electronic health records for 2010–2017¹³. Projections of the population age-distribution for 2018–2043 were gathered from Statistics Canada^{26,53}, Denmark^{27,74} and Scotland (Lothian)^{28,70} (Supplementary Fig. 21).

Region-specific data transformations. Raw historic prevalence, incidence and population data for Canada, Denmark and Scotland (Lothian) were transformed. All age categories (<10 years, 10–17, 18–24, 25–34, 35–44, 45–54, 55–64, 64–79, 80+) were closed, with an upper and lower bound, except for the highest age category (80+), for which the centre of the age bin was set to the estimated population-averaged mean value (86 years for 80+, 93 years for 90+ and 106 years for 105+). The prevalence up to age 110 was estimated using linear extrapolation, constrained to have a zero or negative slope, and to be equal to or greater than zero, while the historic incidence and population projections were set to zero at age 110.

Deriving the equation. The change in prevalence over time was modelled using a PDE^{75,76} that has previously been used to estimate the future prevalence of chronic diseases, including diabetes mellitus^{77,78} and dementia²⁵. The PDE is derived from a compartment model^{75,76} (Supplementary Fig. 22) and has widespread applicability to diseases where the incidence rate is known or can be modelled.

The prevalence of a disease p depends on both time t and the age distribution of the disease cohort a . The change in prevalence as a function of both time and age is modelled by equation (2)²⁵:

$$\frac{\partial p}{\partial t} + \frac{\partial p}{\partial a} = (1-p) \left[i - m \frac{p(R-1)}{p(R-1)+1} \right] - r p + \mu \quad (2)$$

Here, $\frac{\partial p}{\partial t}$ and $\frac{\partial p}{\partial a}$ are the partial derivatives of prevalence with respect to time and age. The incidence rate i , full population mortality rate m , relative mortality ratio R (ratio of those with the disease to those without) and recovery rate r are all functions of both time and age. The migration term μ is a function of time, age and prevalence.

Simplifying the equation. For parsimony, we neglect terms in equation (2) that are not applicable or are small enough to be ignored. IBD is a chronic, incurable disease and we therefore ignored the recovery rate, r . The mortality term, $m p(R-1)/[p(R-1)+1]$ was set to zero owing to the relatively small difference in life expectancy between those living with IBD and the general population⁷⁹. Similarly, the migration term, μ , was set to zero because immigration to the regions being examined greatly exceeds the emigration^{80–82}, and the prevalence of IBD is lower

in new immigrants than in the existing population^{83,84}. Omitting these three terms yields the simplified equation

$$\frac{\partial p}{\partial t} + \frac{\partial p}{\partial a} = (1-p)i(t, a) \quad (3)$$

which makes explicit the age and time dependence of the incidence.

Solving the equation. Given the age distribution of prevalence at an initial time, $p_0(a)$, and assuming that the prevalence must be zero at age zero, equation (3) is solved using the method of characteristics⁸⁵ to yield

$$p(t, a) = \begin{cases} 1 - [1 - p_0(a-t)] e^{- \int_0^t i(x, x+t-a) dx}, & 0 \leq t < a \\ 1 - e^{- \int_0^a i(x+a-t, x) dx}, & 0 \leq a < t \end{cases} \quad (4)$$

To calculate prevalence as a function of time, we multiply by the normalized population age-distribution $\sigma(a)$ and integrate over age, such that

$$p(t) = \int_0^\infty p(t, a) \sigma(a) da \quad (5)$$

Equation (5) allows us to extrapolate the prevalence of IBD over the coming decades. To do so, three additional pieces of information are required: the initial prevalence of IBD as a function of age, the incidence rate of IBD as a function of age and time, and projections of the population age distribution.

Age-dependent IBD incidence and prevalence. A linear interpolation of the age-dependent IBD prevalence data for Canada (2014), Denmark (2017) and Scotland (2017) served as the initial prevalence, $p_0(a)$, for modelling (Supplementary Fig. 23). Data for IBD incidence during 2007–2014 for Canada and 2010–2017 for Denmark and Scotland were compiled (Supplementary Fig. 24), with the average incidence from these 8 years serving as the age-dependent component of the model incidence rate, $f(a)$. A time-dependent component was added to the incidence, modelled as an exponential growth process with growth rate g :

$$i(t, a) = f(a) e^{t \times \ln(g)} \quad (6)$$

Model verification. To verify that our model describes the prevalence of IBD, we compared the model's output to historical data, specifically the age-structured prevalence from 2014 for Canada and 2017 for Denmark and Scotland. Starting with equation (4), $p_0(a)$ was taken as the age-structured prevalence from 2002 (Canada), 2010 (Denmark) and 2009 (Scotland), and a time-independent incidence rate $f(a)$ as the average over the years 2007–2014 (Canada) or 2010–2017 (Denmark and Scotland) (Supplementary Figs. 23 and 24). For Canada, the model produced a 2014 prevalence of 0.65%, compared with the observed 0.65% (Supplementary Fig. 25a). For Denmark, the model produced a 2017 prevalence of 0.98%, compared with the observed 0.94% (Supplementary Fig. 25b). For Scotland, the model produced a 2017 prevalence of 0.88%, compared with the observed 0.79% (Supplementary Fig. 25c). The high degree of concordance suggests that the simplified model is reasonably accurate for longer-term modelling of general prevalence trends.

Slopes and central difference approximations. Central difference approximations to the slope of the time-dependent IBD prevalences in Canada, Denmark, and Scotland were calculated to determine the percent change in prevalence for each year of data. The central difference approximations is calculated by:

$$\delta p(x) = \frac{p(t+h) - p(t-h)}{2h} \quad (7)$$

Article

where $p(t)$ is the prevalence at year t , and h is the time-step, set equal to 1 year (that is, the frequency with which the time-dependent prevalence was calculated). Central difference approximations were averaged over 5-year periods to determine whether Canada, Denmark or Scotland had reached prevalence equilibrium by 2043.

Yearly incidence change scenarios. We selected five values of the growth constant g in equation (6): 0.98, 0.99, 1.00, 1.01 and 1.02, corresponding to 2% and 1% incidence rate decreases per year, no change, and 1% and 2% incidence growth per year, respectively. To model current trends, we assume that the incidence rate is constant, as per the 2007–2014 (Canada) and 2010–2017 (Denmark and Scotland) observed data, until 2024 at which time the exponential increases or decreases in incidence begin.

Strengths and limitations

This study represents, to our knowledge, the most comprehensive summation of population-based data on the incidence and prevalence of IBD, spanning a century of historical data that was used to explore spatial and temporal epidemiologic patterns across the world. We used a unique machine-learning approach to classify regions into three epidemiologic stages over time and established benchmarks to define incidence and prevalence ranges for each stage. Furthermore, to our knowledge, this is the first study to model the transition to a theoretical fourth stage, where prevalence growth plateaus due to an ageing IBD population and stable incidence rates. However, the interpretation of our findings should be evaluated in the context of inherent limitations.

We relied on incidence and prevalence data that varied in quality. Historical data from the twentieth century, regions lacking healthcare surveillance systems and fractionated healthcare systems that impair population-based case capture were of lower quality, leading to heterogeneity in the reported data. To address this, we conducted a quality assessment of each paper through a centralized evaluation process, which was further complemented by assessment from regional experts. Regions with administrative healthcare databases can electronically capture IBD cases using coding algorithms (for example, ICD coding); however, validation studies have demonstrated misclassification errors that lead to the inclusion of false positives, potentially inflating incidence and prevalence. By contrast, regions without population-based electronic healthcare surveillance systems relied on medical registries to identify IBD cases. While this approach results in highly accurate diagnoses, it may miss cases (for example, milder cases not followed by gastroenterologists), potentially leading to underestimation of incidence and prevalence.

The primary source of inconsistency in our random-forest model arises from the availability epidemiologic data, leading to an imbalance across stage classes. The scarcity of data from emerging regions and historical data from the earlier part of the twentieth century resulted in fewer studies available to assess stage 1. Moreover, data from stage 1 regions, which are predominantly low-income, often lacked electronic surveillance systems to support case identification. Consequently, we allowed case-ascertainment approaches that may have missed IBD cases such as surveying gastroenterology offices or hospital-based capture systems, provided that the hospitals serve a defined catchment area. However, in our quality assessment, we differentiated population-based studies that met our strict definition (that is, complete capture of cases in a defined region) from those that did not (that is, a representative sample). Our repository (<https://kaplan-gi.shinyapps.io/GIVES21/>) allows users to subset data by these two categories of population based.

Regional data are missing or limited in many highly populated areas in Africa, Asia and Latin America. Our machine-learning classifier was trained on robust data, enabling us to reliably classify regions during different time periods when data are scarce. Moreover, our GIVES-21 consortium is currently conducting high-quality, population-based epidemiologic research in over 30 regions with limited IBD data from newly industrialized and emerging regions³³. Our repository (<https://kaplan-gi.shinyapps.io/GIVES21/>) is being continuously updated to allow for integration and reanalyses of global epidemiologic data as new information is available. To capture the most recent population-based data available, data from 2020–2024 were included as a partial decade in our analyses. Although this period does not provide a complete decade for stratifying stage classification, its inclusion highlights the best estimate of the current burden of IBD in regions where these years of data are available.

While the model was highly accurate (>95%) in classifying regions into one of three stages, a few regions exhibited unexpected classifications: Ireland was classified as stage 2 during 2010–2019, which contrasts with other contemporary regions in Western Europe. Deviations between the model's output and expected classifications were primarily observed in regions transitioning between stages, especially in cases in which there was a lack of comprehensive data or when subregional data were used in place of regional values. For example, Ireland's classification for 2010–2019 was based on data from County Meath in 2010⁸⁶. These discrepancies highlight the need for more recent, population-based studies in such regions to improve the classification accuracy and refine estimates.

Our PDE models were developed with model parsimony in mind, and so did not include the differential mortality between the IBD and non-IBD populations; this implies that the projections serve as a likely upper bound on the future IBD prevalence. Furthermore, our classifier and PDE models did not directly account for immigration. Research shows that individuals immigrating from stage 1 or 2 regions to a stage 3 region eventually assume the IBD risk of their host region, particularly among the first-degree offspring^{83,84}. Finally, we acknowledge that unexpected future events may influence future projections. For example, we varied incidence growth in the model within a range of +2% to -2% over a 20-year horizon; however, events that could result in a more substantial change in incidence (such as the discovery of an IBD cure, or a high-mortality pandemic) are not accounted for in our model. Moreover, we limited our analyses to three regions currently in stage 3 to predict the transition towards stage 4. Future research that includes data from other stage 3 regions are needed to ensure generalizability of these predictions.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The raw data used for systematic review data analysis and input to our models are available at Figshare⁸⁷ (<https://doi.org/10.6084/m9.figshare.24952557>) and Code Ocean (<https://codeocean.com/capsule/8524418/tree/v1>). Data identified in our systematic review are also available for download or to view and visualize in our interactive data repository (<https://kaplan-gi.shinyapps.io/GIVES21/>). The static maps in Fig. 3 were created in R using the ggplot2 package.

Code availability

The code for the systematic review analysis, the machine-learning algorithms and the mathematical modelling, as well as code used to generate figures, is available at Code Ocean (<https://codeocean.com/capsule/8524418/tree/v1>).

51. Page, M. J. et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Brit. Med. J.* **372**, n71 (2021).
52. Google Translate (Google, 2022); translate.google.com.
53. *Table 17-10-0005-01 Population Estimates on July 1st, By Age and Sex* (Statistics Canada, 2021); www150.statcan.gc.ca/t1/tbl1/en/cv.action?pid=1710000501.
54. *Maddison Style Estimates of the Evolution of the World Economy: A New 2020 Update* (Maddison Project Database, 2020); www.rug.nl/ggdc/historicaldevelopment/maddison/releases/maddison-project-database-2020?lang=en.

55. *Population estimates for UK, England and Wales, Scotland and Northern Ireland: mid-2020-april-2021-geography* (Office for National Statistics, 2022); www.ons.gov.uk/datasets/mid-year-pop-est-editions/mid-2020-april-2021-geography/versions/1.
56. *FOLK1A: Population at the First Day of the Quarter by Region, Sex, Age and Marital Status* (Statistics Denmark, 2022); www.statbank.dk/statbank5a/selectvarval/define.asp?PLangage=1&subword=tabsel&MainTable=FOLK1A&PXSId=199113&tablestyle=&ST=SD&button=0
57. *Population by Region, Marital Status, Age and Sex: Year 1968–2021* (Statistics Sweden, 2022); www.statistikdatabasen.scb.se/pxweb/en/ssd/START_BE_BE0101_BE0101A/BefolkningsNy/.
58. Brazilian Institute of Geography and Statistics. *Cidades* (2022); cidades.ibge.gov.br/brazil/panorama.
59. *World Population Prospects 2019: Data Portal* (United Nations Department of Economic and Social Affairs—Population Division, 2019); population.un.org/dataportal/home.
60. *Estimates of Resident Population Change and Rankings for Counties in Minnesota* (United States Census Bureau, 2022); www2.census.gov/programs-surveys/popest/tables.
61. *Number of Inhabitants: Minnesota* (United States Census Bureau, 1952); www2.census.gov/library/publications/decennial/1950/population-volume-1/vol-01-26.pdf.
62. Munn, Z., Moola, S., Lisy, K., Riihano, D. & Tufanaru, C. Methodological guidance for systematic reviews of observational epidemiological studies reporting prevalence and cumulative incidence data. *Int. J. Evid. Based Healthc.* **13**, 147–153 (2015).
63. OriginPro v.b9.5.5.409 (OriginLab, 2018).
64. Lajeunesse, M. J. juicr: automated and manual extraction of numerical data from scientific images v.0.1 (2021).
65. R: a language and environment for statistical computing v.4.2.2 (R Foundation for Statistical Computing, 2022).
66. Liaw, A. & Wiener, M. Classification and regression by randomForest. *R News* **2**, 18–22 (2002).
67. Rokach, L. Ensemble-based classifiers. *Artific. Intel. Rev.* **33**, 1–39 (2010).
68. Lenth, R. V. et al. emmeans: estimated marginal means, aka Least-Squares Means v.1.10.3.
69. Global Health Observatory. *Prevalence of Obesity Among Adults, BMI ≥ 30, Age-Standardized Estimates by Country* (World Health Organization, 2017); apps.who.int/gho/data/node.main.A900A?lang=en.
70. National Records of Scotland. *Mid-year Population Estimates: Time Series Data* (National Records of Scotland, 2022); www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/population/population-estimates/mid-year-population-estimates/population-estimates-time-series-data.
71. Morotz, G. M. et al. The PTPPIP51 coiled-coil domain is important in VAPB binding, formation of ER-mitochondria contacts and IP3 receptor delivery of Ca^{2+} to mitochondria. *Front. Cell Dev. Biol.* **10**, 920947 (2022).
72. Mathematica v13.1 (Wolfram Research, 2022).
73. Dorn-Rasmussen, M. L. et al. The incidence and prevalence of paediatric- and adult-onset inflammatory bowel disease in Denmark during a 37-year period—a nationwide cohort study (1980–2017). *J. Crohns Colitis* **17**, 259–268 (2023).
74. *Population 1. January (FOLK2)* (Statbank Denmark, 2023); extranet.dst.dk/pyramide/pyramide.htm#!&l=en.
75. Brinks, R. & Landwehr, S. A new relation between prevalence and incidence of a chronic disease. *Math. Med. Biol.* **32**, 425–435 (2015).
76. Keiding, N. Age-specific incidence and prevalence: a statistical perspective. *J. R. Stat. Soc.* **154**, 371–412 (1991).
77. Tönnies, T. et al. Projected number of people with diagnosed Type 2 diabetes in Germany in 2040. *Diabet. Med.* **36**, 1217–1225 (2019).
78. Voeltz, D., Tönnies, T., Brinks, R. & Hoyer, A. Future prevalence of type 2 diabetes—a comparative analysis of chronic disease projection methods. *PLoS ONE* **17**, e0264739 (2022).
79. Kuenzig, M. E., Manuel, D. G., Donelle, J. & Benchimol, E. I. Life expectancy and health-adjusted life expectancy in people with inflammatory bowel disease. *Can. Med. Assoc. J.* **192**, E1394–E1402 (2020).
80. *Table 17-0040-01 Estimates of the Components of International Migration, Quarterly* (Statistics Canada, 2023); www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1710004001.
81. *Immigration and emigration* (Statbank Denmark, 2023); www.dst.dk/en/Statistik/emner/borgere/flytninger/ind-og-udvandring.
82. *Total Migration To or From Scotland* (National Records of Scotland, 2023); www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/migration/migration-statistics/migration-flows/total-migration-to-or-from-scotland.
83. Benchimol, E. I. et al. Inflammatory bowel disease in immigrants to Canada and their children: a population-based cohort study. *Am. J. Gastroenterol.* **110**, 553–563 (2015).
84. Agrawal, M. et al. Inflammatory bowel diseases among first-generation and second-generation immigrants in Denmark: a population-based cohort study. *Gut* **70**, 1037–1043 (2021).
85. Polyanin, A. D., Zaitsev, V. F. & Moussiaux, A. *Handbook of First Order Partial Differential Equations* Vol. 1 (Taylor & Francis, 2002).
86. Burisch, J. et al. East–west gradient in the incidence of inflammatory bowel disease in Europe: the ECCO-EpiCom inception cohort. *Gut* **63**, 588–597 (2014).
87. Hracs, L. et al. Data for ‘Global evolution of inflammatory bowel disease across epidemiological stages’. *Figshare* <https://doi.org/10.6084/m9.figshare.24952557> (2025).

Author contributions L.H., J.W.W., S.C.N. and G.G.K. led the study. L.H., J.W.W., J.G., M.C., S.C., S.C.N. and G.G.K. conceived and designed the study, conducted analyses, interpreted the results and drafted the manuscript. L.H. led the machine-learning analyses. M.C. led the PDE model analyses. J.G. led the societal indicator and ratio analyses. J.W.W. led the systematic review. J.W.W., J.G., M.J.B., J.Q., A.M., S.C.N. and G.G.K. were responsible for primary data acquisition through the systematic review. E.I.B., C.N.B., M.C., R.G., Q.G., R.P. and T.W. interpreted the analysis and critically reviewed the work for important intellectual content. Y.A., M. Abdullah, M.T.A., V.A., R.A.R.A., M. Altuwaijri, D.B., R.B., E.B.-M., J.B., V.H.C., I.D., U.D., S.E.O., A. Forbes, A. Forss, R.G., V.H.D., J.L.H., I.H., P.H., G.-R.J., F.J.-B., J.K., P.K., T.K., P.G.K., P.L.L., C.W.L., J.L., B.L., E.V.L., J.F.L., J.W.Y.M., Y.M., K.K.N., S.O., O.O., M.S.P., A.B.Q., D.T.R., M.S., Y.S., H.S., M.T., D.T., B.I., S.C.W., J.K.Y.-F. and S.-K.Y. formed an international group of collaborators who verified data, provided additional data with local expertise and critically reviewed the work. S.C.N. and G.G.K. provided funding for this study. All of the authors edited the manuscript and provided final approval. L.H., J.W.W., S.C.N. and G.G.K. contributed equally to this study.

Competing interests M.T.A. has received research funding from the National Institute of Health, Department of Defense, The Leona M. and Harry B. Helmsley Charitable Trust (through the University of Miami), Crohn’s and Colitis Foundation and Kenneth Rainin Foundation; she is a consultant or served on advisory boards for AbbVie, Arena Pharmaceuticals (now Pfizer), Bristol Myers Squibb, Celsius Therapeutics, Eli Lilly and Company, Gilead Sciences, Janssen Pharmaceuticals, Janssen Global Services, Pfizer Pharmaceutical, Prometheus Biosciences, UCB Biopharma; she has received fees for lecturing from Alimentiv, Janssen Pharmaceuticals, Prime CME and WebMD Global LLC. D.B. has received speaking fees from AbbVie, Takeda and Janssen, and consulting fees from AbbVie, Takeda, Janssen and Amgen. R.B. has received grants/research support from the Asian Healthcare Foundation, Pfizer Global, Indian council of Medical Research (ICMR) and the Leona M and Harry B Helmsley Charitable Trust; and advisory board fees from Abbott, AstraZeneca, Cadila, Cipla, Celtrion, Dr Reddy Labs, Emcure, Ferring Pharmaceuticals, Hetero Drugs, Janssen, La Renon, MSN Labs, Mankind Pharma, Menarini, Micro Labs, Nestle, Pfizer, RPG Lifesciences, Sun Pharmaceuticals, Systopic, Takeda Pharmaceuticals, Torrent, Waterley and Zydus. E.I.B. has acted as a consultant for McKesson Canada and the Dairy Farmers of Ontario for matters unrelated to medications used to treat IBD; he has also acted as a consultant for the Canadian Agency for Drugs and Technology in Health (CADTH). C.N.B. is supported by the Bingham Chair in Gastroenterology; has served on advisory boards for AbbVie Canada, Amgen Canada, Bristol Myers Squibb Canada, Eli Lilly Canada, JAMP Pharmaceuticals, Janssen Canada, Pendopharm Canada, Sandoz Canada, Takeda Canada and Pfizer Canada; Consultant for Takeda; Educational grants from AbbVie Canada, Bristol Myers Squibb Canada, Ferring Canada, Pfizer Canada, Takeda Canada, Janssen Canada, Organon Canada, Eli Lilly Canada and Amgen Canada; speaker’s panel for AbbVie Canada, Janssen Canada, Pfizer Canada and Takeda Canada. Received research funding from AbbVie Canada, Amgen Canada, Sandoz Canada, Takeda Canada and Pfizer Canada. E.B.-M. has served as a speaker and consultant for Janssen, Chiesi, Takeda and Kern. J.B. reports grants and personal fees from AbbVie, grants and personal fees from Janssen-Cilag, personal fees from Celgene, grants and personal fees from MSD, personal fees from Pfizer, grants and personal fees from Takeda, grants and personal fees from Tillotts Pharma, personal fees from Samsung Bioepis, grants and personal fees from Bristol Myers Squibb, grants from Novo Nordisk, personal fees from Pharmacosmos, personal fees from Ferring and personal fees from Galapagos; outside the submitted work. I.D. received consultation fees/served in advisory boards/speakers’ bureaus for AbbVie, Athos, Arena, Celtrion, Celgene/BMS, Eli-Lilly, Ferring, Food Industries Organization, Gilead, Galapagos, Iterative Scopes, Janssen, Neopharm, Pfizer, Roche/Genentech, Sangamo, Sublimity, Sandoz, Takeda, Wildbio and Prometheus. Jill Roberts IBD Center, shareholder: Gutreat, Harp diagnostics. S.E.O. has received honoraria for speaking from AbbVie and Janssen. A. Forss has served as a speaker and advisory board member for Janssen and Tillotts Pharma. R.G. has received research funding from AbbVie, Janssen, Zespir, Comvita, Goodman Fielder and Takeda; he has served on advisory boards for AbbVie, Janssen, Zespir, Comvita and Takeda; and has received honoraria from AbbVie, Janssen, Zespir and Takeda. J.L.H. has received consultation and speaking fee from Janssen, Ferring, AbbVie and Takeda. I.H. has received consulting and speaking honoraria from AbbVie, Janssen, Ferring, LF Asia, Pfizer and Takeda. B.I. has been member of advisory boards for AbbVie, Janssen and Takeda. G.-R.J. is funded by a Wellcome Trust Clinical Research Career Development Fellowship and has received speaker fees from AbbVie, Ferring Pharmaceuticals, Janssen, Takeda, Fresenius Kabi and Pfizer. F.J.-B. has received consulting and speaking fees from AbbVie, Janssen, Pfizer, Celtrion and Takeda. J.K. has served on advisory boards for Janssen Kazakhstan and has been a consultant for Takeda and Ferring; has been on speaker panels for Janssen, Takeda, Ferring, Thermo Fisher Scientific; and has received research funding from Ferring Kazakhstan. G.G.K. has received honoraria for speaking or consultancy from AbbVie, Amgen, Janssen, Pfizer, Sandoz and Pendopharm; received grants for research from Ferring and for educational activities from AbbVie, Bristol Myers Squibb, Ferring, Fresenius-Kabi, Janssen, Pfizer, Takeda; he shares ownership of a patent: Treatment of inflammatory disorders, autoimmune disease and PBC; UTI Limited Partnership, assignee; patent WO2019046959A1, PCT/CA2018/051098; 7 September 2018. T.K. has served as an advisory board member, consultant or speaker for AbbVie, Activaid, Ajinomoto Bio-Pharma, Alfresa Pharma, Alimentiv, Astellas Pharma, Bristol Myers Squibb, Celtrion, Coviden, EA Pharma, Eisai, Eli Lilly, Ferring Pharmaceuticals, Gilead Sciences, Janssen Pharmaceuticals, JIMRO, JMD, Kissei Pharmaceutical, Kyorin Pharmaceutical, Mitsubishi Tanabe Pharma, Mochida Pharmaceutical, Nippon Kayaku, Pfizer, Takeda, Thermo Fisher Scientific and Zeria Pharmaceutical, and has received research funding from AbbVie, Alfresa Pharma, EA Pharma, Kyorin Pharmaceutical, Mochida Pharmaceutical, Nippon Kayaku, Otsuka Holdings, Sekisui Medical, Takeda, Thermo Fisher Scientific and Zeria Pharmaceutical. P.G.K. has received consulting and speaking honorarium from AbbVie, Janssen, Pfizer and Takeda. He also has received scientific grants from Pfizer and Takeda. P.L.L. has been a speaker and/or advisory board member for AbbVie, Amgen, BioJamp, Bristol Myers Squibb, Fresenius Kabi, Genetech, Gilead, Janssen, Merck, Mylan, Organon, Pendopharm, Pfizer, Roche, Sandoz, Takeda, Tillotts and Viatris; and has received unrestricted research grants from AbbVie, Gilead, Takeda and Pfizer. C.W.L. is funded by a UK Research and Innovation Future Leaders Fellowship and has acted as a speaker and/or consultant to AbbVie, Janssen, Takeda, Pfizer, Galapagos, GSK, Gilead, Vifor Pharma, Ferring, Dr Falk, BMS, Boehringer Ingelheim, Novartis, Sandoz, Celtrion, Cellgene, Amgen, Samsung

Acknowledgements This work was conducted by the Global Cluster of IBD. The work was funded by The Leona M. and Harry B. Helmsley Charitable Trust (grant no. G-2108-04777); the Canadian Institutes of Health Research, Project Scheme Operating Grant (reference no. PJT-162393); and the International Organization for the study Inflammatory Bowel Disease (IOIBD). S.C.N. has received research funding from the Research Grants Council–Research Impact Fund (RGC-RIF, grant no. R4030-22), New Cornerstone Science Foundation (grant no. NCI202346), InnoHK, The Government of Hong Kong, Special Administrative Region of the People’s Republic of China, and the Leona M. and Harry B. Helmsley Charitable Trust (grant no. 2017PG-IBD003).

Article

Bioepis, Fresenius Kabi, Tillotts, Trellus Health and Iterative Health. B.L. has received speaking fees from Tillotts Pharma and Janssen Cilag; consultancy fees from Tillotts Pharma and Bristol Myers Squibb; and has received unrestricted research grants from Janssen Cilag, Tillotts Pharma; outside the submitted work. E.V.L. has consulted for AbbVie, Alvotech, Amgen, Arena, Astellas, Avalo, Boehringer Ingelheim, Bristol-Myers Squibb, Celtrion, Eli Lilly, Fresenius Kabi, Genentech, Gilead, GlaxoSmithKline, Gossamer Bio, Iota Biosciences, Iterative Health, Janssen, Morphic, Ono, Protagonist, Sun, Surrozen, Takeda, TR1X and UCB; has received research support from AbbVie, AstraZeneca, Bristol-Myers Squibb, Celgene/Receptos, Genentech, Gilead, Gossamer Bio, Janssen, Takeda, Theravance and UCB; and is a shareowner of Exact Sciences. S.C.N. has served as an advisory board member for Pfizer, Ferring, Janssen and AbbVie and received honoraria as a speaker for Ferring, Tillotts, Menarini, Janssen, AbbVie and Takeda; has received research grants through her affiliated institutions from Olympus, Ferring and AbbVie; is a founder member, non-executive director, non-executive scientific advisor and shareholder of GenieBiome Ltd; and receives patent royalties through her affiliated institutions. S.O. has received speaking fees from Mitsubishi Tanabe Pharma and consulting fees from EA Pharma. O.O. has been PI on projects at Karolinska Institutet partly financed by investigator-initiated grants from Janssen, Takeda, AbbVie, Pfizer, Bristol Myers Squibb and Ferring, and also report grants from Pfizer, AbbVie, Janssen and Galapagos in the context of a national safety monitoring programs; none of these studies have any relation to the present study. The Karolinska Institutet has received fees for lectures and participation on advisory boards by O.O. from Janssen, Ferring, Takeda and Pfizer on topics not related to the present study. R.P. has received consulting fees from Abbott, AbbVie, Alimentiv (formerly Robarts), Amgen, Arena Pharmaceuticals, AstraZeneca, Biogen, Boehringer Ingelheim, Bristol-Myers Squibb, Celgene, Celtrion, Cosmos Pharmaceuticals, Eisai, Elan, Eli Lilly, Ferring, Galapagos, Fresenius Kabi, Genentech, Gilead Sciences, Glaxo-Smith Kline, JAMP Bio, Janssen, Merck, Mylan, Novartis, Oppidan Pharma, Organon, Pandion Pharma, Pendopharm, Pfizer, Progenity, Protagonist Therapeutics, Roche, Sandoz, Satisfi Health, Shire, Sublimity Therapeutics and Takeda. A.B.Q. has received speaking honorarium from AbbVie, Astra Zeneca, Ferring, Pfizer, Janssen and Takeda. D.T.R. has received grant support from Takeda; and has served as a consultant for AbbVie, Altrubio, Abivax SA, Altrubio, Avalo Therapeutics, Bausch Health, Bristol-Myers Squibb, Buhlmann Diagnostics, ClostraBio, Connect BioPharma, Douglas

Pharmaceuticals, Eli Lilly, Genentech, InDex Pharmaceuticals, Iterative Health, Janssen Pharmaceuticals, Odyssey Therapeutics, Pfizer, Sanofi and Takeda Pharmaceuticals; he serves on the board of trustees for the Crohn's & Colitis Foundation and is on the board of directors for Cornerstones Health. H.S. has received honoraria for speaking from Astellas, AbbVie, Biofermin, EA Pharma, Janssen, Kissei, Miyarisan, Otsuka, Tanabe-Mitsubishi, Takeda, Tsumura and Viatris; and received grants for research from Toh-so and Biofermin. M.T. has received consulting and speaking fees from AbbVie, Janssen and Takeda. D.T. received consultation fees, research grants, royalties or honoraria from Janssen, Pfizer, Shaare Zedek Medical Center, Hospital for Sick Children, Ferring, AbbVie, Takeda, Prometheus Biosciences, Lilly, Roche, Thermo Fisher Scientific, BMS and SorrisoPharma. S.C.W. has received consultancy fees from AbbVie, Bristol-Myers Squibb, Celtrion, Ferring Pharmaceuticals, Janssen, Pfizer, Takeda and Tanabe, and speaker fees from AbbVie, Bristol-Myers Squibb, Celtrion, Excelsior Biopharma, Ferring Pharmaceuticals, Pfizer, Janssen, Sanofi, Takeda and Tanabe. T.W. received funding for educational activities from AbbVie and Pfizer. J.K.Y.-F. is a speaker bureau member for AbbVie, Alfasigma, Asofarma, Bristol Myers Squibb, Carnot, Celtrion, Chinoin, Farmasa, Ferring, Janssen, Siegfried Rhein and Takeda; is a clinical research investigator Bristol Myers Squibb and Takeda; has been member of advisory boards for AbbVie, Celtrion, Ferring, Janssen and Takeda; has received research grants from Takeda; received payments for lectures from AbbVie, Alfasigma, Asofarma, Bristol Myers Squibb, Carnot, Celtrion, Chinoin, Farmasa, Ferring, Janssen, Siegfried Rhein and Takeda; and is an advisory board member of the AbbVie, Celtrion, Ferring, Janssen and Takeda. The other authors declare no competing interests.

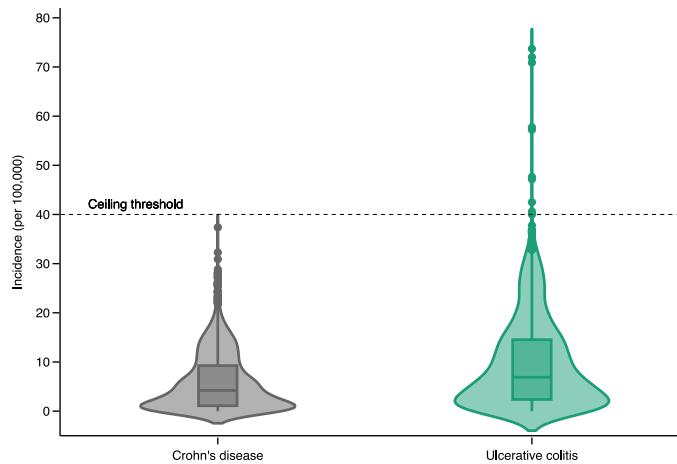
Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41586-025-08940-0>.

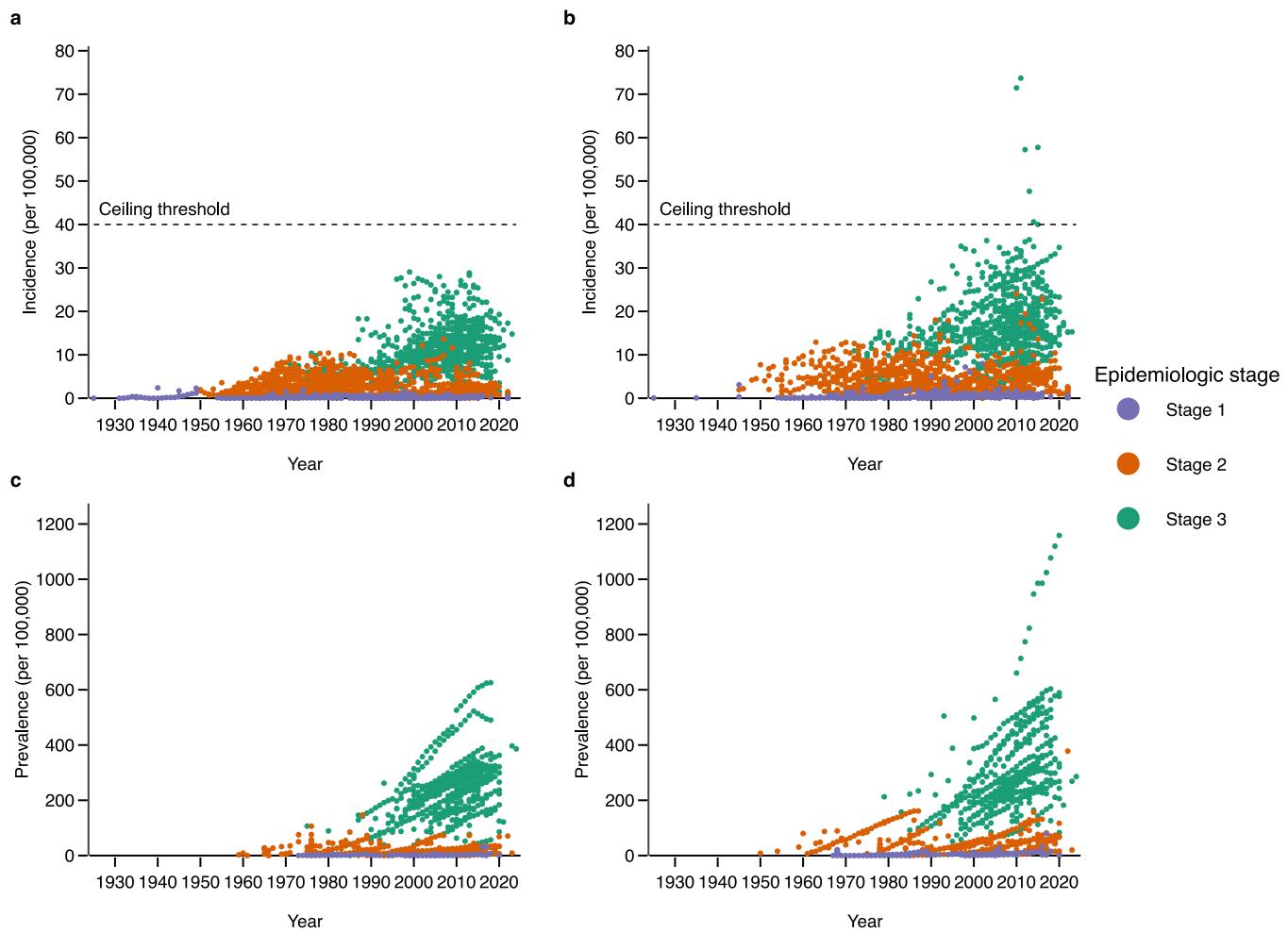
Correspondence and requests for materials should be addressed to Siew C. Ng or Gilaad G. Kaplan.

Peer review information *Nature* thanks Ashwin Ananthakrishnan, Andrew Mertens, Siddharth Singh and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

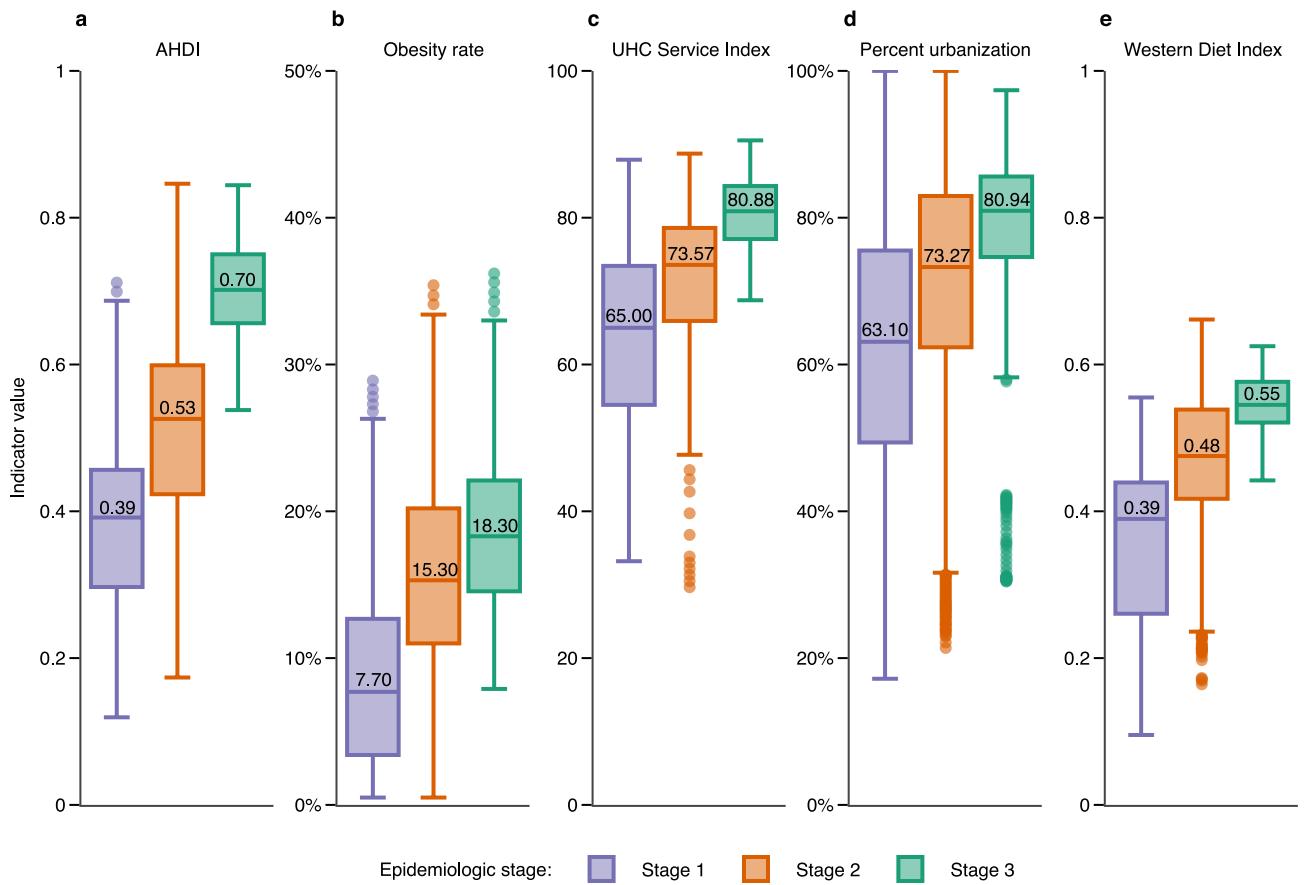


Extended Data Fig. 1 | Density of all incidence data included in the systematic review for Crohn's disease and ulcerative colitis across regions and time. All data are represented in violin plots with boxplot overlays where the middle line on the boxplot is the median, the lower hinge on the boxplot is the 25th percentile (i.e., first quartile), the upper hinge on the boxplot is the 75th percentile (i.e., third quartile); the violin is the data density at the associated value on the y-axis, and points are individual data points outside of the density distribution. Incidence rates of Crohn's disease or ulcerative colitis that exceed the ceiling threshold of 40 per 100,000 are considered outlier data.



Extended Data Fig. 2 | Annual incidence and prevalence of Crohn's disease and ulcerative across epidemiologic stages of IBD evolution. **a**, Annual incidence of Crohn's disease categorized by three epidemiologic stages of IBD evolution as predicted by the random forest classifier. **b**, Annual incidence of ulcerative colitis categorized by three epidemiologic stages of IBD evolution as predicted by the random forest classifier. **c**, Annual prevalence of Crohn's

disease categorized by three epidemiologic stages of IBD evolution as predicted by the random forest classifier. **d**, Annual prevalence of ulcerative colitis categorized by three epidemiologic stages of IBD evolution as predicted by the random forest classifier. Each point corresponds to an annual aggregate mean of CD or UC incidence/prevalence for all regions with available data.



Extended Data Fig. 3 | Comparison of societal indicators across three epidemiologic stages. **a**, Augmented Human Development Index (AHDI) stratified by epidemiologic stage, as predicted by the random forest classifier (number of observations: stage 1 $n = 470$; stage 2 $n = 1,090$; stage 3 $n = 540$). **b**, Obesity rate stratified by epidemiologic stage, as predicted by the random forest classifier (number of observations: stage 1 $n = 295$; stage 2 $n = 791$; stage 3 $n = 476$). **c**, Universal Health Coverage (UHC) Service Index stratified by epidemiologic stage, as predicted by the random forest classifier (number of observations: stage 1 $n = 160$; stage 2 $n = 400$; stage 3 $n = 380$). **d**, Percent urbanization stratified by epidemiologic stage, as predicted by the random forest classifier (number of observations: stage 1 $n = 460$; stage 2 $n = 1,180$; stage 3 $n = 610$). **e**, Western Diet Index (WDI) stratified by epidemiologic stage,

as predicted by the random forest classifier (number of observations: stage 1 $n = 348$; stage 2 $n = 980$; stage 3 $n = 540$). All data are represented in boxplots where the middle line is the median, the lower hinge is the 25th percentile (i.e., first quartile), the upper hinge is the 75th percentile (i.e., third quartile), the lower whisker extends to $1.5 \times$ the interquartile range (IQR) from the first quartile, the upper whisker extends to $1.5 \times$ IQR from the third quartile, and the data points beyond the end of the whiskers are individually plotted outliers. A Kruskal-Wallis non-parametric test with *post hoc* comparisons using Wilcoxon rank-sum adjusted for multiple comparisons showed significant differences between epidemiologic stages for AHDI, obesity rate, UHC Service Index, percent urbanization, and Western Diet Index ($p < 0.001$ for all comparisons).

Article

Extended Data Table 1 | Average percent change in prevalence slope

Period	Canada (% change)	Denmark (% change)	Scotland (% change)
2010–2014	0.019	0.030	0.025
2015–2019	0.017	0.030	0.025
2020–2024	0.016	0.031	0.032
2025–2029	0.014	0.027	0.029
2030–2034	0.013	0.023	0.027
2035–2039	0.012	0.020	0.024
2040–2042	0.010	0.018	0.022

Slopes calculated using central difference approximation, for observed and modelled data from 2010–2042 (historical data in bold).

Extended Data Table 2 | Percent change in prevalence slope in 2042 by yearly incidence rate change

Region	2% decrease	1% decrease	no change	1% increase	2% increase
Canada	0.003	0.006	0.010	0.015	0.021
Denmark	0.005	0.011	0.018	0.026	0.035
Scotland	0.009	0.015	0.022	0.030	0.040

Slopes calculated using central difference approximation, for modelled yearly incidence rate changes of 2% decrease, 1% decrease, no change, 1% increase, and 2% increase. We propose that stage 4—Prevalence Equilibrium—occurs when the change in the slope of prevalence decelerates to an average of 0% ($\pm 0.01\%$).

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	Data extraction from published graphs was conducted using OriginPro2018 (v. b9.5.5.409) or juicer (v. 0.1). No additional software was used for data collection.
Data analysis	Statistical analyses and machine learning were performed in R (v.4.2.2) and included packages randomForest (v. 4.7-1.1) and emmeans (v. 1.10.3). Mathematical modelling was performed in Mathematica (v.13.1). Maps in Figure 3 were created with ggplot2 (v. 3.5.1).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The raw data used for systematic review data analysis and input to our models are available at <https://doi.org/10.6084/m9.figshare.24952557> and in the Code Ocean capsule linked to this manuscript. Data identified in our systematic review are also available for download or to view and visualize in our interactive data

repository at <https://kaplan-gi.shinyapps.io/GIVES21/>.

The systematic review was performed through Medline (https://www.nlm.nih.gov/medline/medline_overview.html), Embase (<https://www.elsevier.com/products/embase>), PubMed (<https://pubmed.ncbi.nlm.nih.gov/>), and Web of Science (<https://clarivate.com/academia-government/scientific-and-academic-research/research-discovery-and-referencing/web-of-science/>).

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	Sex- and gender-based analyses were not performed. Although sex and gender differences can be observed in the epidemiology of inflammatory bowel disease, the systematic review data gathered and subsequently analyzed were not stratified by sex or gender.
Reporting on race, ethnicity, or other socially relevant groupings	Analyses based on race and ethnicity were not performed. The systematic review data gathered and subsequently analyzed corresponded to global geography, but stratifications by race, ethnicity, or other socially relevant groupings were not collected as non-general populations was an exclusion criteria.
Population characteristics	Global populations: all ages, genders, and races.
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	A total of 522 population-based studies reporting on the incidence and prevalence of inflammatory bowel disease were included in our analyses. Machine learning analyses were performed on a total of 5,521 data points (Crohn's disease incidence $n = 2,070$; ulcerative colitis incidence $n = 1,884$; Crohn's disease prevalence $n = 776$; ulcerative colitis prevalence $n = 791$). Incidence and prevalence data for Crohn's disease and ulcerative colitis across each global region for each year of data were merged to create a total of 2,489 random forest classifications. Mathematical modelling of age-stratified incidence and prevalence data from 3 regions, i.e., Canada, Denmark, and Scotland, was completed (Supplementary Table 7).
Data exclusions	Our exclusion criteria were as follows: No defined geographic region, less than one year of data, study population is pediatric only (i.e., did not include adult and/or senior populations), self-reported cases identified by survey, reviews, and clinical trials. Exclusion criteria were pre-established at the outset of the systematic review and align with previous systematic reviews that the current work builds on.
Replication	We identified and reassessed population-based studies included in two previous systematic reviews on the incidence and prevalence of inflammatory bowel disease. Our systematic review was conducted in accordance with Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA 2020). Data and code are made available for reproducibility.
Randomization	Training and validation data were randomly sampled.
Blinding	A validation set was withheld at model fitting. All test data (approximately 35% of our dataset) were unseen by the model.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	Antibodies
<input checked="" type="checkbox"/>	Eukaryotic cell lines
<input checked="" type="checkbox"/>	Palaeontology and archaeology
<input checked="" type="checkbox"/>	Animals and other organisms
<input checked="" type="checkbox"/>	Clinical data
<input checked="" type="checkbox"/>	Dual use research of concern
<input checked="" type="checkbox"/>	Plants

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	ChIP-seq
<input checked="" type="checkbox"/>	Flow cytometry
<input checked="" type="checkbox"/>	MRI-based neuroimaging

Plants

Seed stocks

N/A

Novel plant genotypes

N/A

Authentication

N/A