nature portfolio

Corresponding author(s): Gilaad G. Kaplan, Siew C. Ng

Last updated by author(s): Mar 24, 2025

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 include 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 = 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
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	N/A
Novel plant genotypes	N/A
Authentication	N/A