Research

Machine learning prediction of premature death from multimorbidity among people with inflammatory bowel disease: a population-based retrospective cohort study

Gemma Postill BMScH, Ijeoma Uchenna Itanyi MD MPH, M. Ellen Kuenzig PhD, Furong Tang MSc, Vinyas Harish PhD, Emma Buajitti MPH, Laura Rosella PhD, Eric I. Benchimol MD PhD

Cite as: *CMAJ* 2025 March 24;197:E286-97. doi: 10.1503/cmaj.241117

Abstract

Background: Multimorbidity, the co-occurrence of 2 or more chronic conditions, is important in patients with inflammatory bowel disease (IBD) given its association with complex care plans, poor health outcomes, and excess mortality. Our objectives were to describe premature death (age < 75 yr) among people with IBD and to identify patterns between multimorbidity and premature death among decedents with IBD.

Methods: Using the administrative health data of people with IBD who died between 2010 and 2020 in Ontario, Canada, we conducted a population-based, retrospective cohort study. We described the proportion of premature deaths among people with IBD. We developed statistical and machine learning models to predict premature death from the presence of 17 chronic conditions and the patients' age at diagnosis. We evaluated models using accuracy, positive predictive value, sensitivity, F_1 scores, area under the receiver operating curve (AUC), calibration plots, and explainability plots.

Results: All models showed strong performance (AUC 0.81–0.95). The best performing was the model that incorporated age at diagnosis for each chronic condition developed at or before age 60 years (AUC 0.95, 95% confidence interval 0.94–0.96). Salient features for predicting premature death were young ages of diagnosis for mood disorder, osteo- and other arthritis types, other mental health disorders, and hypertension, as well as male sex.

Interpretation: By comparing results from multiple approaches modelling the impact of chronic conditions on premature death among people with IBD, we showed that conditions developed early in life (age ≤ 60 yr) and their age of onset were important for predicting their health trajectory. Clinically, our findings emphasize the need for models of care that ensure people with IBD have access to high-quality, multidisciplinary health care.

Inflammatory bowel disease (IBD) is a group of chronic inflammatory disorders that primarily affect the gastrointestinal tract, including Crohn disease and ulcerative colitis. Canada has among the highest incidence and prevalence of IBD in the world.[1] About 470 000 (equivalent to 1 in 91) people in Canada are forecasted to be living with IBD by 2035.[2] People with IBD have a higher mortality rate (17 v. 12 per 1000 person-years), with rate differences greatest in mid-life (ages 20–59 years, depending on IBD subtype),[3] and shorter life expectancy (differences of up to 8 years among females and 6 years among males) than people without IBD.[4] Premature death (< 75 yr)[5–7] is a robust marker of population health and health system performance, as many premature deaths are considered avoidable through appropriate prevention or early and effective treatment.[6–11] Understanding the predictors of premature death can guide improvement of health systems.[5–8]

Multimorbidity, the co-occurrence of 2 or more chronic conditions,[11] is an emerging concept of study in IBD.[13] People with IBD are more likely to develop chronic health conditions than those without IBD.[14,15] Multimorbidity can alter the natural history of IBD.[16] In the general population, multimorbidity is associated with complex care plans,[17] poor health outcomes,[18] and excess deaths.[19] Although multidisciplinary care is an established IBD care goal, structural barriers inhibit the implementation of multidisciplinary care in practice, including the funding and organization of clinical care.[20–22] Understanding whether (and which)

 © 2025 CMA Impact Inc. or its licensors

comorbidities predict premature death among people with IBD can ascertain priorities for improving the multidisciplinary management of IBD.

Machine learning models have been used to predict premature death in the general population.[23–25] Although such predictive models cannot determine causal relationships,[26,27] they have identified data-driven patterns to inform further causal research. Our objectives were thus to describe the proportion of premature deaths among people with IBD in Ontario and to identify patterns between non-IBD chronic conditions and premature death among decedents with IBD.

Methods

Study design and data sources

We conducted a predictive, population-based, retrospective cohort study using Ontario administrative health data. In Ontario, all residents are eligible for publicly funded health coverage from the Ontario Health Insurance Plan (OHIP), so health administrative data include more than 99% of the Ontario population (around 14.5 million people). However, the health system covers medications only for those aged 65 years or older and those on social assistance. We used data housed at ICES, a prescribed entity under section 45 of Ontario's *Personal Health Information Protection Act*, allowing it to collect and analyze health care and demographic data, without consent, for health system evaluation and improvement.

Study population

Eligible participants were residents of Ontario who had received a diagnosis of IBD and died between Jan. 1, 2010, and Jan. 31, 2020. Decedents with missing ages or dates of death were excluded. We further excluded people with invalid health cards or those who were non-Ontario residents at their date of death. To identify those with IBD, we used the Ontario Crohn's and Colitis Cohort, which identifies incident IBD cases among residents of Ontario using validated age-specific algorithms applied to inpatient hospitalization records, same-day surgery records, physician billing claims, and medications data (among those aged ≥ 65 years only).[28,29] We calculated the minimum sample size — including a 95% confidence interval, a 5% margin of error, and 18 parameters — with pmsampsize R package (version 1.1.3)[30] using the approach by Riley and colleagues (Appendix 1, Supplemental Methods, available at www.cmaj.ca/lookup/doi/10.1503/cmaj.241117/tab-related-content).[31]

Chronic conditions

Using validated algorithms for health administrative data, we identified people with a history of certain chronic conditions, including asthma,[32] congestive heart failure,[33] chronic obstructive pulmonary disease,[34] diabetes,[35] rheumatoid arthritis,[36] hypertension,[37] and dementia.[38,39] Cancer was identified from the Ontario Cancer Registry.[40] For chronic conditions without validated algorithms, we required a relevant diagnostic code present on any hospital discharge or on 2 physician visits within a 2-year period; these included acute myocardial infarction, osteoporosis, cardiac arrhythmia, chronic coronary syndrome, stroke, renal failure, osteo- and other arthritis types (nonrheumatoid), mood disorders, and other mental health disorders (Appendix 1, Supplementary Table S1).[41–45]

Sociodemographic characteristics

We identified age of death and sex from the Registered Persons Database. We identified material resources and rurality of all decedents based on their postal code on their date of death. Material resources is a dimension of the Ontario Marginalization Index that is closely connected to poverty and refers to the ability for individuals and communities to attain basic material needs (i.e., housing, food, clothing, and education). Material resources is measured at the level of the dissemination area (the smallest standard census geographic area) and categorized into quintiles.[46] Rurality was defined using the Rurality Index for Ontario scores, with scores of 40 or higher indicating rural locations.[47]

Outcome

The primary outcome of interest was premature death, defined as any death before 75 years of age. The age of 75 years is commonly employed in Canada as a standard benchmark for premature deaths[5–7] and used widely in the literature.[6,8] In Ontario, 37% of all deaths between 2018 and 2022 occurred prematurely.[48]

Statistical and machine learning analyses

We descriptively summarized sociodemographic and clinical characteristics of the cohort. We used supervised machine learning approaches across 3 predictive tasks (Figure 1). Model features (chronic conditions) were determined using a priori clinical knowledge and preprocessed differently across each task. Task 1 predicted premature death from conditions present (yes or no) at death to capture entire health trajectories. Task 2 predicted premature death from conditions developed by age 60 years (early-life conditions). Task 3 predicted premature death from normalized age of diagnosis of early-life conditions. We chose the threshold of 60 years for early-life conditions because most people with IBD are diagnosed by age 60 years,[49] this age is when increases in multimorbidity across age cohorts move from gradual to substantial and graded,[42] and this threshold provides clinical lead time for premature death prevention. Choosing these tasks allowed us to explore different aspects of the relationship between non-IBD chronic conditions and premature death.

Model development and evaluation

We randomly split the data into training and testing cohorts (80:20), with the same cohorts used across tasks. For each task, we developed logistic regression,[50] random forest,[51] and Extreme Gradient Boosting (XGBoost)[52] models to determine which model best characterized the relationship between non-IBD chronic conditions and premature death. For task 3, we trained only an XGBoost model (XGB3) as it alone could handle missing values. The logistic regression and random forest models were developed using scikit-learn version 1.3.0.[53] The XGBoost model was developed using the XGBoost package version 1.7.6.[52] The hyperparameters of all models (Appendix 1, Supplemental Table S3)

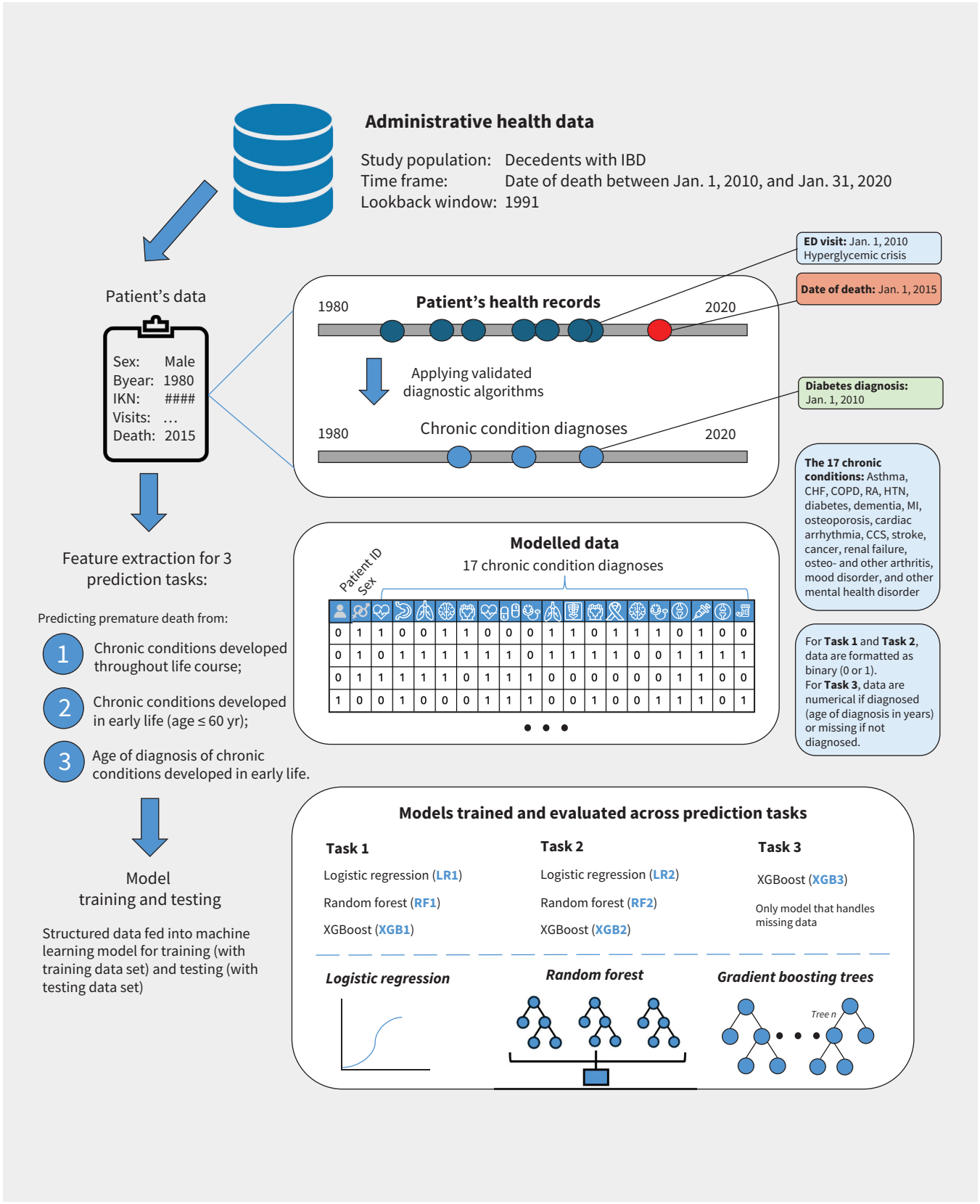

Figure 1: Overview of prediction pipeline and models used across prediction tasks. We specified 3 modelling tasks for predicting premature death among people with inflammatory bowel disease (IBD). We then used 3 types of models, namely logistic regression, random forest, and Extreme Gradient Boosting (XGBoost); XGBoost was the only model used for task 3 as it enabled direct modelling of missing data (those without conditions would have missing data). See Related Content for accessible version. Note: CCS = chronic coronary syndrome, CHF = congestive heart failure, COPD = chronic obstructive pulmonary disease, ED = emergency department, HTN = hypertension, MI = myocardial infarction, RA = rheumatoid arthritis.

were optimized on accuracy using 5-fold cross-validation and GridSearchCV.[53] For logistic regression, we used a classification threshold of 0.5.

We evaluated model performance by assessing the confusion matrix, accuracy (percentage of correct predictions), positive predictive value (PPV), sensitivity (recall), F_1 scores (balance of precision and recall), area under the receiver operating curves (AUC), and calibration plots. As the goal was to determine whether and which chronic conditions were predictive of premature death, AUC and precision were the most clinically important metrics. We derived confidence intervals from bootstrapping, with 1000 random resamples of the testing data. In the best performing model, we evaluated model performance stratified by sex and IBD subtype (Crohn disease, ulcerative colitis, and unspecified) and conducted a manual analysis of predictive errors (decedents misclassified with a probability threshold of 0.5).

We assessed feature importance using coefficients (for logistic regression) and the Gini importance (for random forest and XGBoost).[54,55] For the XGBoost models, we employed Shapley Additive Explanations analysis to interpret the contribution of feature values to the prediction.[56] All analyses (except sample size calculation) were performed using SAS version 8.3 and Python version 3.7. Appendix 1, Supplemental Methods, provides further details on model development and evaluation.

Sensitivity analyses

For the best performing model, we modified the outcome variable by using age of death instead of premature death and employed XGBoost regression predictive modelling. The objective was to determine the specificity of our findings to the definition of premature death. Second, we included all conditions diagnosed up to age 75 years and repeated both the classification (premature death) and regression (age of death) predictive modelling. The objective was to determine the specificity of our findings to the definition of early onset conditions.

Study reporting

For study reporting, we used the Reporting of Studies Conducted using Observational Routinely Collected Health Data (RECORD) reporting guidelines,[57] the 2024 Transparent Reporting of Multivariable Prediction Model for Individual Prognosis or Diagnosis plus Artificial Intelligence (TRIPOD+AI) checklist,[58] and the APPRAISE-AI framework.[59]

Ethics approval

Data use was authorized under section 45 of Ontario's *Personal Health Information Protection Act* and did not require review by institutional research ethics boards. All data were deidentified before access. Data are available in uncleaned, unedited format to ICES researchers and data analysts.

Results

Cohort characteristics

The cohort included 9278 decedents (49.3% female) with IBD (Table 1 and Appendix 1, Supplementary Figure S1). Of these, 47.2% were premature deaths (44% for females and 50% for males; Appendix 1, Supplemental Table S4). Conditions most prevalent at age 60 years were osteo- and other arthritis types (39.0%), mood disorders (38.3%), and hypertension (29.5%). At death, conditions with the highest prevalence were osteo- and other arthritis types (76.8%), hypertension (72.8%), mood disorders (69.0%), renal failure (49.6%), and cancer (46.1%).

Prediction of premature death

All 7 models demonstrated strong performance and calibration on testing data (n = 1856) (Table 2). Model performance improved when features included only conditions diagnosed before age 60 years (Task 2 and 3), and further improved when age of diagnosis was included (Task 3).

The strongest feature for predicting premature death varied across and within tasks, demonstrating that although models had similar performance metrics, they may have drawn on different relationships within the data (Figure 2 and Appendix 1, Supplemental Figure S3). For task 1, the absence of chronic conditions, particularly those commonly developed later in life (e.g., dementia, chronic coronary syndrome, congestive heart failure), was leveraged by our model to predict premature death. However, models in task 2 and 3 captured a relationship between the presence of non-IBD chronic conditions and premature death. For task 3, important features were young age at diagnosis for mood disorders, osteo- and other arthritis types, mental health disorders, and hypertension, as well as male sex (Figure 2).

The best performing model in task 3 had few prediction errors (11%), most of which were for patients with conditions diagnosed close to age 60 years (Appendix 1, Supplemental Figure S5). The false positive predictions commonly involved osteo- and other arthritis types (58%), hypertension (56%), and mood disorders (53%); false negative errors largely occurred among those with few comorbidities (Appendix 1, Supplemental Table S5). Similar performance was observed across sexes (Table 2 and Appendix 1, Supplemental Figure S4) and IBD subtypes (Appendix 1, Supplemental Table S8), with slight improvements among those with Crohn disease or unspecified IBD, potentially owing to the increased prevalence of premature deaths among people with those subtypes (Table 3), as indicated by the difference in model performance for the prediction of premature and nonpremature deaths (Appendix 1, Supplemental Table S8).

Sensitivity analyses

In the first sensitivity analysis predicting age of death, we observed strong performance and similar feature importance as the main analyses, with subtle differences in the relative order of features (Appendix 1, Supplemental Table S9 and Supplemental Figure S6). In the second sensitivity analysis predicting from chronic conditions developed by age 75 years, we also observed similar performance and feature importance. However, unlike the original analysis, these models identified conditions developed near 75 years as protective of premature death (Figure 2C).

Table 1 (part 1 of 2): Characteristics of decedents with inflammatory bowel disease (IBD) at time of death

Characteristic	No. (%) of decedents*		
	All decedents with IBD *n* = 9278	Training set *n* = 7422	Testing set *n* = 1856
Age of death, yr, median (IQR)	76 (64–85)	76 (64–85)	75 (64–85)
Premature death			
Yes	4380 (47.2)	3478 (46.9)	902 (48.6)
No	4898 (52.8)	3944 (53.1)	954 (51.4)
Sex			
Female	4577 (49.3)	3679 (49.6)	898 (48.4)
Male	4701 (50.7)	3743 (50.4)	958 (51.6)
IBD subtype			
Crohn disease	4109 (44.3)	3325 (44.8)	784 (42.2)
Ulcerative colitis	4769 (51.4)	3789 (51.1)	980 (52.8)
Unspecified IBD	400 (4.3)	308 (4.1)	92 (4.9)
Rurality Index for Ontario score			
Rural (≥ 40)	944 (10.2)	759 (10.2)	185 (10.0)
Urban (< 40)	8234 (88.7)	6580 (88.7)	1654 (89.1)
Missing	100 (1.1)	83 (1.1)	17 (0.9)
Area-level material resources			
Q1 (most resources)	1710 (18.4)	1347 (18.1)	363 (19.6)
Q2	1713 (18.5)	1360 (18.3)	353 (19.0)
Q3	1835 (19.8)	1483 (20.0)	352 (19.0)
Q4	1871 (20.2)	1521 (20.5)	350 (19.0)
Q5 (least resources)	2069 (22.3)	1648 (22.2)	421 (22.7)
Missing	80 (0.9)	63 (0.8)	17 (1.0)
Chronic conditions diagnosed by death			
Hypertension	6755 (72.8)	5438 (73.2)	1317 (71.0)
Cancer	4274 (46.1)	3406 (45.9)	868 (46.8)
Renal failure	4606 (49.6)	3685 (49.6)	921 (49.6)
Chronic coronary syndrome	4194 (45.2)	3368 (45.4)	826 (44.5)
Cardiac arrhythmia	3065 (33.0)	2474 (33.3)	591 (31.8)
Congestive heart failure	3042 (32.8)	2445 (32.9)	597 (32.2)
Myocardial infarction	2949 (31.8)	2370 (31.9)	579 (31.2)
Diabetes	3233 (34.8)	2607 (35.1)	626 (33.7)
Stroke	3309 (35.7)	2653 (35.7)	656 (35.3)
COPD	4011 (43.2)	3208 (43.2)	803 (43.3)
Asthma	1982 (21.4)	1563 (21.1)	419 (22.6)
Dementia	2109 (22.7)	1667 (22.5)	442 (23.8)
Mood disorder	6402 (69.0)	5128 (69.1)	1274 (68.6)
Other mental health disorder	4485 (48.3)	3625 (48.4)	860 (46.3)
Rheumatoid arthritis	530 (5.7)	410 (5.5)	120 (6.5)
Osteo- and other arthritis types	7122 (76.8)	5713 (77.0)	1409 (75.9)
Osteoporosis	1965 (21.2)	1572 (21.2)	393 (21.2)

Table 1 (part 2 of 2): Characteristics of decedents with inflammatory bowel disease (IBD) at time of death

Characteristic	No. (%) of decedents*		
	All decedents with IBD *n* = 9278	Training set *n* = 7422	Testing set *n* = 1856
Chronic conditions diagnosed by age 60 years			
Hypertension	2739 (29.5)	2211 (29.8)	528 (28.4)
Cancer	1376 (14.8)	1072 (14.4)	304 (16.4)
Renal failure	1247 (13.4)	985 (13.3)	262 (14.1)
Chronic coronary syndrome	1202 (13.0)	968 (13.0)	234 (12.6)
Cardiac arrhythmia	615 (6.6)	482 (6.5)	133 (7.2)
Congestive heart failure	420 (4.5)	323 (4.4)	97 (5.2)
Myocardial infarction	715 (7.7)	568 (7.7)	147 (7.9)
Diabetes	1245 (13.4)	1008 (13.6)	237 (12.8)
Stroke	739 (8.0)	601 (8.1)	138 (7.4)
COPD	1399 (15.1)	1108 (14.9)	291 (15.7)
Asthma	1030 (11.1)	804 (10.8)	226 (12.2)
Dementia	79 (0.9)	62 (0.8)	17 (0.9)
Mood disorder	3558 (38.3)	2835 (38.2)	723 (39.0)
Other mental health disorder	2097 (22.6)	1676 (22.6)	421 (22.7)
Rheumatoid arthritis	208 (2.2)	161 (2.2)	47 (2.5)
Osteo- and other arthritis types	3618 (39.0)	2885 (38.9)	733 (39.5)
Osteoporosis	627 (6.8)	499 (6.7)	128 (6.9)
No. of chronic conditions at death			
0–1	198 (2.1)	166 (2.2)	32 (1.7)
2–4	1860 (20.0)	1460 (19.7)	400 (21.6)
5–7	3320 (35.8)	2671 (36.0)	649 (35.0)
≥ 8	3900 (42.0)	3125 (42.1)	775 (41.8)

Note: COPD = chronic obstructive pulmonary disease, IQR = interquartile range.
*Unless indicated otherwise.

Interpretation

Among Ontario residents with IBD who died between 2010 and 2020, almost half died prematurely. Non-IBD comorbidities showed a strong predictive relationship with premature death (AUC 0.81–0.95). Important features for predicting premature death were younger ages of diagnosis for mood disorders, osteo- and other arthritis types, mental health disorders, and hypertension, as well as male sex. Strengths of this study include its large sample size, population-based nature, and exploration of the relationship between non-IBD chronic conditions and premature death. Our findings emphasize the importance of further causal work on multimorbidity in IBD and integrated, multidisciplinary care of those with IBD across the lifespan.

We built on previous research by modelling the ability of individuals' non-IBD chronic conditions to predict premature death with machine learning and using premature death as our outcome. Using comorbidity scores (i.e., Charlson Comorbidity Index and Adjusted Clinical Groups) can mask interrelationships between conditions and impedes understanding the specific conditions associated with the outcome.[60–62] Indeed, although having 3 or more conditions was associated with an increased risk of death, there are discordant effects depending on the specific combinations of conditions.[63] Machine learning models can better reflect complex relationships between features without a priori specification. As such, machine learning models predicting death in other clinical areas have demonstrated higher discrimination than traditional classification and regression models. For example, machine learning performed better than traditional regression models for predicting in-hospital death after elective cardiac surgery[64] and premature death in the general population.[65,66] The use of premature death as the outcome more directly identifies opportunities for health system improvements, as premature deaths are considered avoidable through appropriate prevention or early and effective treatment.[6,9,10]

Early-life conditions important for predicting premature death among those with IBD align with previous literature. In the general population, premature mortality rates are higher among males.[8] Mental health and mood disorders have been associated with premature death owing to health behaviours

Table 2: Model performance for predicting premature death from chronic conditions among individuals (*n* = 1856) with inflammatory bowel disease

Task*	Prediction class evaluated†	Model performance (95% CI)		
		Positive predictive value‡	Sensitivity§	F_1 score¶
Task 1				
Logistic regression (accuracy: 0.75, 95% CI 0.73–0.77; AUC: 0.82, 95% CI 0.81–0.84)	Nonpremature death	0.75 (0.72–0.78)	0.77 (0.75–0.80)	0.76 (0.74–0.78)
	Premature death	0.75 (0.72–0.78)	0.73 (0.70–0.75)	0.74 (0.72–0.76)
	Weighted average	0.75 (0.73–0.77)	0.75 (0.73–0.77)	0.75 (0.73–0.77)
Random forest (accuracy: 0.74, 95% CI 0.72–0.76; AUC: 0.82, 95% CI 0.80–0.84)	Nonpremature death	0.72 (0.70–0.75)	0.81 (0.78–0.83)	0.76 (0.74–0.78)
	Premature death	0.77 (0.74–0.80)	0.67 (0.64–0.70)	0.72 (0.69–0.74)
	Weighted average	0.75 (0.72–0.76)	0.74 (0.73–0.77)	0.74 (0.72–0.76)
XGBoost (accuracy: 0.75, 95% CI 0.72–0.77; AUC: 0.82, 95% CI 0.81–0.84)	Nonpremature death	0.74 (0.71–0.77)	0.79 (0.76–0.82)	0.76 (0.74–0.78)
	Premature death	0.76 (0.73–0.79)	0.71 (0.67–0.74)	0.73 (0.71–0.76)
	Weighted average	0.75 (0.73–0.77)	0.75 (0.73–0.77)	0.75 (0.73–0.77)
Task 2				
Logistic regression (accuracy: 0.82, 95% CI 0.81–0.84; AUC: 0.91, 95% CI 0.90–0.92)	Nonpremature death	0.81 (0.78–0.83)	0.87 (0.85–0.89)	0.84 (0.82–0.85)
	Premature death	0.85 (0.83–0.87)	0.78 (0.75–0.80)	0.81 (0.79–0.83)
	Weighted average	0.83 (0.81–0.84)	0.83 (0.81–0.84)	0.82 (0.81–0.84)
Random forest (accuracy: 0.82, 95% CI 0.81–0.84; AUC: 0.91, 95% CI 0.90–0.93)	Nonpremature death	0.82 (0.80–0.84)	0.84 (0.82–0.86)	0.83 (0.81–0.85)
	Premature death	0.83 (0.80–0.85)	0.80 (0.78–0.83)	0.82 (0.80–0.83)
	Weighted average	0.82 (0.81–0.84)	0.82 (0.81–0.84)	0.82 (0.81–0.84)
XGBoost (accuracy: 0.82, 95% CI 0.81–0.84; AUC: 0.91, 95% CI 0.90–0.93)	Nonpremature death	0.82 (0.80–0.84)	0.85 (0.82–0.87)	0.83 (0.81–0.85)
	Premature death	0.83 (0.81–0.86)	0.80 (0.78–0.83)	0.82 (0.80–0.84)
	Weighted average	0.83 (0.81–0.84)	0.83 (0.81–0.84)	0.82 (0.81–0.84)
Task 3				
XGBoost (accuracy: 0.89, 95% CI 0.88–0.91; AUC: 0.95, 95% CI 0.94–0.96)	Nonpremature death	0.87 (0.85–0.89)	0.94 (0.92–0.95)	0.90 (0.89–0.92)
	Premature death	0.93 (0.91–0.94)	0.85 (0.83–0.87)	0.89 (0.87–0.90)
	Weighted average	0.89 (0.88–0.91)	0.89 (0.88–0.91)	0.89 (0.88–0.91)
XGBoost, males (accuracy: 0.88, 95% CI 0.86–0.90; AUC: 0.94, 95% CI 0.92–0.95)	Nonpremature death	0.85 (0.82–0.88)	0.92 (0.89–0.94)	0.88 (0.86–0.94)
	Premature death	0.91 (0.89–0.94)	0.84 (0.81–0.87)	0.87 (0.85–0.90)
	Weighted average	0.88 (0.86–0.90)	0.88 (0.86–0.90)	0.88 (0.86–0.90)
XGBoost, females (accuracy: 0.91, 95% CI 0.89–0.93; AUC: 0.96, 95% CI 0.95 – 0.97)	Nonpremature death	0.89 (0.86–0.92)	0.95 (0.93–0.97)	0.92 (0.91–0.97)
	Premature death	0.94 (0.91–0.96)	0.87 (0.84–0.90)	0.90 (0.88–0.92)
	Weighted average	0.91 (0.89–0.93)	0.91 (0.89–0.93)	0.91 (0.89–0.93)

Note: AUC = area under the curve, CI = confidence interval, XGBoost = Extreme Gradient Boosting.
*Accuracy refers to pooled premature and nonpremature prediction. AUC refers to premature death.
†We report metrics for each class (premature and nonpremature death) and as a weighted average across classes.
‡ Positive predictive value (also known as precision in the machine learning literature) refers to the proportion of true positive predictions out of all positive predictions.
§Sensitivity (also known as recall in the machine learning literature) is the proportion of true positive predictions from all actual positive samples in the data set.
¶The F_1 score considers both positive predictive value (precision) and sensitivity (recall).

and psychosocial and medication-related factors.[67–71] Hypertension, particularly when diagnosed young, has also been associated with all-cause death.[72–74] Osteoarthritis has been associated with increased risk of death, particularly among those younger than 65 years,[75] although this relationship is unclear.[76] In the context of IBD, the co-occurrence of these conditions with IBD can be related to pathophysiology or the psychosocial implications of living with IBD. For example, mood disorders are more common among people with IBD,[77] perhaps because of chronic pain, chronic fatigue, systemically circulating cytokines (e.g., tumour necrosis factor α), medications (e.g., corticosteroids),[78] or psychosocial reasons (e.g., social isolation).[79] Importantly, other conditions have well-known associations with IBD (e.g., cardiovascular disease,[80] renal failure,[81] osteoporosis[82,83]) but were less important in our final model.[13] This may have occurred because these conditions commonly occur after age 60 years. However,

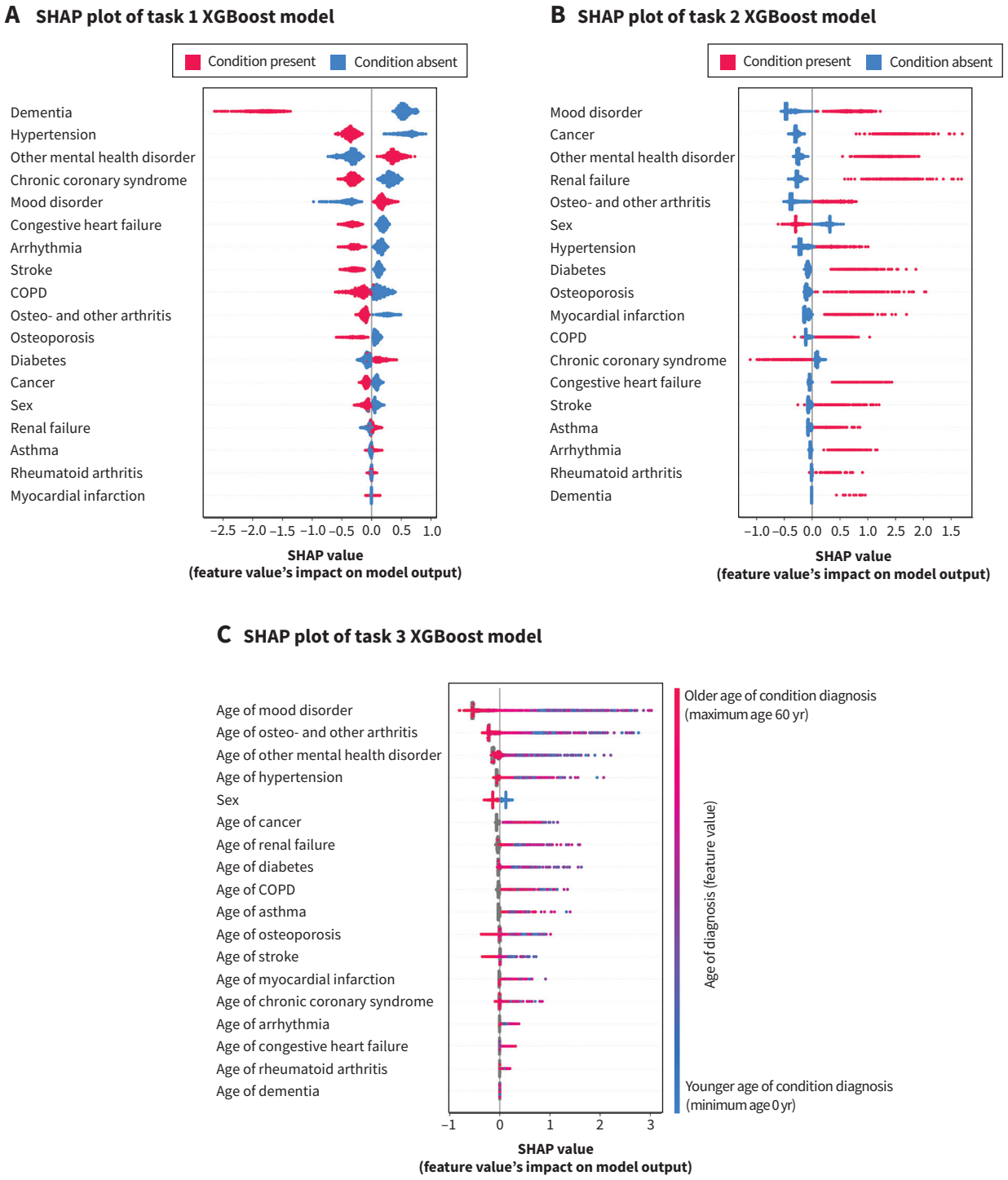

Figure 2: Feature importance for predicting premature death from (A) presence or absence of all chronic conditions (Extreme Gradient Boosting 1 [XGB1] model), (B) presence or absence of chronic conditions developed at or before age 60 years (XGB2 model), and (C) age of chronic conditions developed at or before age 60 years (XGB3 model), as denoted by the Shapley Additive Explanations (SHAP) analysis. Each dot represents a single patient in the data set. The X axis shows SHAP values, which quantify the impact of each predictor variable's value on the model's output. Positive SHAP values indicate a higher likelihood of premature death, while negative values indicate a lower likelihood. Colours represent the variable values for each condition across each patient in the data set. In subplots A and B, red values indicate the condition was present, while blue values indicate that the condition was absent. In subplot C, red corresponds to an older age of condition diagnosis (up to age 60 yr) and blue corresponds to younger age. In all panels, sex is distinguished such that red indicates males and blue indicates females. Grey values represent patients with missing values (e.g., no diagnosis of condition). The correlation between blue dots and positive SHAP values indicates that younger ages of diagnosis were primarily used by the model to predict premature death. The vertical dispersion of points for each feature reflects variability in its impact across patients. This visualization helps identify key predictors and their relative influence on model decisions. See Related Content for accessible version. Note: COPD: chronic obstructive pulmonary disease.

Table 3: Number of deaths among those with inflammatory bowel disease (IBD) by sex, age group, sex-stratified age group, and IBD subtype

Group	No. (%) of deaths	
	Premature deaths *n* = 4380	Nonpremature deaths *n* = 4898
Sex		
Male	2371 (54.1)	2330 (47.6)
Female	2009 (45.9)	2568 (56.1)
Age group (both sexes), yr		
< 34	178 (4.1)	–
35–49	540 (12.3)	–
50–74	3662 (63.6)	–
75–94	–	4605 (94.0)
≥ 95	–	293 (6.0)
Age group (males), yr		
< 34	110 (2.5)	–
35–49	286 (6.1)	–
50–74	1975 (42.0)	–
75–94	–	2245 (45.8)
≥ 95	–	85 (54.2)
Age group (females), yr		
< 34	68 (1.5)	–
35–49	254 (6.5)	–
50–74	1687 (38.5)	–
75–94	–	2360 (48.2)
≥ 95	–	208 (51.8)
IBD subtype		
Crohn disease	2319 (52.9)	1790 (36.5)
Ulcerative colitis	1887 (43.1)	2882 (58.8)
Unspecified IBD	174 (4.0)	226 (4.6)

our second sensitivity analysis (including conditions diagnosed up to age 75 years) demonstrated similar features.

The uniqueness of our findings is twofold. We predicted premature death among decedents with IBD without including details of IBD diagnosis or management in the model, and we used models trained on early-life chronic conditions, which improved performance relative to including conditions developed at any age. The clinical implication is that chronic conditions developed early in life may be more important in determining a patient's health trajectory, although further causal research is needed to elucidate this relationship. Although our insights are not causal insights, they identify patients potentially at higher risk of premature death. Our model helps dissect and capture patient heterogeneity, identifying areas where more targeted follow-up is needed to better understand their clinical importance and relation to IBD severity. Literature on multidisciplinary care in IBD focuses on increasing availability of IBD nurses, dietitians, and mental health professionals, with the occasional mention of rheumatologists or dermatologists who are specifically involved in treating IBD extraintestinal manifestations. Our work points to a need to expand availability to other medical specialties and to enhance care coordination among specialties. Surveillance, early intervention, and preventive care efforts led by multidisciplinary teams that include diverse medical specialties will be key in ensuring patients receive appropriate care for their co-occurring health conditions. Our results can be used to inform strategies for multidisciplinary and personalized approaches to care; however, these strategies should be evaluated to determine their effectiveness prior to widespread implementation.

Limitations

We used health administrative data to identify chronic conditions and premature death; all studies using health administrative data are at risk of misclassification bias and people who seek care may be systematically different from those who do not.[57] We minimized these biases by conducting our study in a single-payer health care system and using validated disease identification algorithms. The lookback period for incident cases extended only to the 1990s; conditions diagnosed earlier may not have been captured if no additional care was received after 1990. However, most conditions included require ongoing health care and would be captured. Feature importance for prediction of premature death does not signify clinical importance. We fit a prediction model for premature death; however, to say chronic conditions developed early in life cause premature death, alternative analytic approaches, such as target trial emulations, are needed. From our study, it was not possible to conclude whether conditions predictive of premature death indicate severe IBD or function independently. Lastly, the models developed were not intended for use as population risk prediction tools because the cohort includes only deceased individuals; rather, our findings provide insights on the relationship between multimorbidity and premature death. Given our strong model performance, future work could extend and validate these results by applying similar prediction methods to living individuals, developing time-to-death prediction models, and including data on IBD treatment regimes.

Conclusion

We demonstrated that, among decedents with IBD, machine learning models can accurately predict premature death associated with non-IBD comorbidities, with stronger performance for models trained on early-life conditions (age ≤ 60 yr), suggesting these may be more important in determining one's health trajectory. These findings provide scientific support for providing multidisciplinary and integrated health care across the lifespan (particularly during young and middle adulthood). Further research is needed to provide more nuanced understanding on how multimorbidity can cause premature death among people with IBD to inform effective preventive care at population scale.

References

1. Ng SC, Yun Shi H, Hamidi N, et al. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *Lancet* 2017;390:2769-78.

2. Coward S, Benchimol EI, Bernstein CN, et al. Forecasting the incidence and prevalence of inflammatory bowel disease: a Canadian nationwide analysis. *Am J Gastroenterol* 2024;119:1563-70.
3. Card T, Hubbard R, Logan RFA. Mortality in inflammatory bowel disease: a population-based cohort study. *Gastroenterology* 2003;125:1583-90.
4. Kuenzig ME, Manuel DG, Donelle J, et al. Life expectancy and health-adjusted life expectancy in people with inflammatory bowel disease. *CMAJ* 2020;192:E1394-402.
5. Subedi R, Greenberg TL, Roshanafshar S. Does geography matter in mortality? An analysis of potentially avoidable mortality by remoteness index in Canada. *Health Rep* 2019;30:3-15.
6. *Premature and potentially avoidable mortality, Canada, provinces and territories*. Ottawa: Statistics Canada; 2023. Available: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1310074401 (accessed 2025 Mar. 02).
7. *Avoidable deaths from preventable causes*. Ottawa: The Canadian Institute for Health Information; 2023. Available: https://www.cihi.ca/en/indicators/avoidable-deaths-from-preventable-causes (accessed 2025 Jan. 20).
8. Buajitti E, Watson T, Norwood T, et al. Regional variation of premature mortality in Ontario, Canada: a spatial analysis. *Popul Health Metr* 2019;17:9.
9. Norheim OF, Jha P, Admasu K, et al. Avoiding 40% of the premature deaths in each country, 2010–30: review of national mortality trends to help quantify the UN Sustainable Development Goal for health. *Lancet* 2015;385:239-52.
10. Shiels MS, Chernyavskiy P, Anderson WF, et al. Trends in premature mortality in the USA by sex, race, and ethnicity from 1999 to 2014: an analysis of death certificate data. *Lancet* 2017;389:1043-54.
11. Nolte E, McKee M. Measuring the health of nations: analysis of mortality amenable to health care. *BMJ* 2003;327:1129.
12. Mosli MH, Alsahafi M, Alsanea MN, et al. Multimorbidity among inflammatory bowel disease patients in a tertiary care center: a retrospective study. *BMC Gastroenterol* 2022;22:487.
13. Argollo M, Gilardi D, Peyrin-Biroulet C, et al. Comorbidities in inflammatory bowel disease: a call for action. *Lancet Gastroenterol Hepatol* 2019;4:643-54.
14. Bähler C, Schoepfer AM, Vavricka SR, et al. Chronic comorbidities associated with inflammatory bowel disease: prevalence and impact on healthcare costs in Switzerland. *Eur J Gastroenterol Hepatol* 2017;29:916-25.
15. Shaffer SR, Kuenzig ME, Windsor JW, et al. The 2023 impact of inflammatory bowel disease in Canada: special populations — IBD in seniors. *J Can Assoc Gastroenterol* 2023;6:S45-54.
16. Román ALS, Muñoz F. Comorbidity in inflammatory bowel disease. *World J Gastroenterol* 2011;17:2723-33.
17. Salisbury C, Johnson L, Purdy S, et al. Epidemiology and impact of multimorbidity in primary care: a retrospective cohort study. *Br J Gen Pract* 2011;61:e12-21.
18. Boyd CM, Fortin M. Future of multimorbidity research: How should understanding of multimorbidity inform health system design? *Public Health Rev* 2010;32:451-74.
19. Nunes BP, Flores TR, Mielke GI, et al. Multimorbidity and mortality in older adults: A systematic review and meta-analysis. *Arch Gerontol Geriatr* 2016;67:130-8.
20. Kuenzig ME, Fung SG, Marderfeld L, et al. Twenty-first century trends in the global epidemiology of pediatric-onset inflammatory bowel disease: systematic review. *Gastroenterology* 2022;162:1147-1159.e4.
21. Kuenzig ME, Im JHB, Coward S, 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* 2023;6:S16-22.
22. Mathias H, Rohatinsky N, Murthy SK, et al. The 2023 impact of inflammatory bowel disease in Canada: access to and models of care. *J Can Assoc Gastroenterol* 2023;6:S111-21.
23. Liu B, Fang L, Xiong Y, et al. A machine learning model based on genetic and traditional cardiovascular risk factors to predict premature coronary artery disease. *Front Biosci (Landmark Ed) 2022*;27:211.
24. Zhang D, Li Y, Kalbaugh CA, et al. Machine learning approach to predict in-hospital mortality in patients admitted for peripheral artery disease in the United States. *J Am Heart Assoc* 2022;11:e026987.
25. Hernández Guillamet G, Morancho Pallaruelo AN, Miró Mezquita L, et al. Machine learning model for predicting mortality risk in patients with complex chronic conditions: retrospective analysis. *Online J Public Health Inform* 2023;15:e52782.
26. van Diepen M, Ramspek CL, Jager KJ, et al. Prediction versus aetiology: common pitfalls and how to avoid them. *Nephrol Dial Transplant* 2017;32:ii1-5.
27. Ramspek CL, Steyerberg EW, Riley RD, et al. Prediction or causality? A scoping review of their conflation within current observational research. *Eur J Epidemiol* 2021;36:889-98.
28. Benchimol EI, Guttmann A, Griffiths AM, et al. Increasing incidence of paediatric inflammatory bowel disease in Ontario, Canada: evidence from health administrative data. *Gut* 2009;58:1490-7.
29. Benchimol EI, Guttmann A, Mack DR, et al. Validation of international algorithms to identify adults with inflammatory bowel disease in health administrative data from Ontario, Canada. *J Clin Epidemiol* 2014;67:887-96.
30. Ensor J. pmsampsize: Calculates the minimum sample size required for developing a multivariable prediction mode; 2023. Available: https://cran.r-project.org/web/packages/pmsampsize/pmsampsize.pdf (accessed 2024 Apr. 07).
31. Riley RD, Ensor J, Snell KIE, et al. Calculating the sample size required for developing a clinical prediction model. *BMJ* 2020;368:m441.
32. Gershon AS, Wang C, Guan J, et al. Identifying patients with physician-diagnosed asthma in health administrative databases. *Can Respir J* 2009;16:183-8.
33. Schultz SE, Rothwell DM, Chen Z, et al. Identifying cases of congestive heart failure from administrative data: a validation study using primary care patient records. *Chronic Dis Inj Can* 2013;33:160-6.
34. Gershon AS, Wang C, Guan J, et al. Identifying individuals with physcian diagnosed COPD in health administrative databases. *COPD* 2009;6:388-94.
35. Hux JE, Ivis F, Flintoft V, et al. Diabetes in Ontario. *Diabetes Care* 2002;25:512-6.
36. Widdifield J, Bernatsky S, Paterson JM, et al. Accuracy of Canadian health administrative databases in identifying patients with rheumatoid arthritis: a validation study using the medical records of rheumatologists. *Arthritis Care Res (Hoboken)* 2013;65:1582-91.
37. Tu K, Campbell NR, Chen Z-L, et al. Accuracy of administrative databases in identifying patients with hypertension. *Open Med* 2007;1:e18-26.
38. Jaakkimainen L, Duchen R, Lix L, et al. Identification of early onset dementia in population-based health administrative data: a validation study using primary care electronic medical records. *J Alzheimers Dis* 2022;89:1463-72.
39. Jaakkimainen RL, Bronskill SE, Tierney MC, et al. Identification of physician-diagnosed Alzheimer's disease and related dementias in population-based administrative data: a validation study using family physicians' electronic medical records. *J Alzheimers Dis* 2016;54:337-49.
40. Prodhan S, King MJ, De P, et al. Health services data: the Ontario Cancer Registry (a unique, linked, and automated population-based registry). *Health Serv Eval* 2019;363-9.
41. Mondor L, Maxwell CJ, Bronskill SE, et al. The relative impact of chronic conditions and multimorbidity on health-related quality of life in Ontario long-stay home care clients. *Qual Life Res* 2016;25:2619-32.
42. Ryan BL, Bray Jenkyn K, Shariff SZ, et al. Beyond the grey tsunami: a cross-sectional population-based study of multimorbidity in Ontario. *Can J Public Health* 2018;109:845-54.
43. Rosella L, Kornas K, Huang A, et al. Accumulation of chronic conditions at the time of death increased in Ontario from 1994 To 2013. *Health Aff (Millwood)* 2018;37:464-72.
44. Ryan B, Allen B, Zwarenstein M, et al. Multimorbidity and mortality in Ontario, Canada: a population-based retrospective cohort study. *J Comorb 2020* Aug 27;10:2235042X20950598.
45. Rosella LC, Negatu E, Kornas K, et al. Multimorbidity at time of death among persons with type 2 diabetes: a population-based study in Ontario, Canada. *BMC Endocr Disord* 2023;23:127.
46. van Ingen T, Matheson FI. The 2011 and 2016 iterations of the Ontario Marginalization Index: updates, consistency and a cross-sectional study of health outcome associations. *Can J Public Health* 2022;113:260-71.
47. Boris K. Measuring rurality — RIO 2008_BASIC: methodology and results. Toronto: Ontario Medical Association; 2009.
48. *Deaths, by age group and sex*. Ottawa: Statistics Canada; 2023. Available: https://www150.statcan.gc.ca/t1/tbl1/en/tv.
49. Baars JE, Kuipers EJ, van Haastert M, et al. Age at diagnosis of inflammatory bowel disease influences early development of colorectal cancer in inflammatory bowel disease patients: a nationwide, long-term survey. *J Gastroenterol* 2012;47:1308-22.
50. Scikit-Learn. sklearn.linear_model.LinearRegression. Available: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html (accessed 2025 Mar. 02).
51. Scikit-Learn. sklearn.ensemble.RandomForestClassifier.
52. xgboost developers. XGBoost Python Package; 2023.
53. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. Ithaca (NY): Cornell university; 2012. doi:10.48550/ARXIV.1201.0490.
54. Breiman L. Random forests. *Mach Learn* 2001;45:5-32.

55. Menze BH, Kelm BM, Masuch R, et al. A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC Bioinformatics* 2009;10:213.
56. Lundberg SM, Lee S-I. *A unified approach to interpreting model predictions. Proceedings from 31st Conference on Neural Information Processing Systems (NIPS 2017); 2017 Dec. 4–9; Long Beach (CA)*. Seattle (WA): University of Washington; 2017. Available: https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf (accessed 2025 Apr. 07).
57. Benchimol EI, Smeeth L, Guttmann A, et al.; RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Med* 2015;12(10):e1001885.
58. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* 2024;Apr 16:385:e078378.
59. Kwong JCC, Khondker A, Lajkosz K, et al. APPRAISE-AI tool for quantitative evaluation of AI studies for clinical decision support. *JAMA Netw Open* 2023;6:e2335377.
60. Austin PC, Van Walraven C, Wodchis WP, et al. Using the Johns Hopkins Aggregated Diagnosis Groups (ADGs) to predict mortality in a general adult population cohort in Ontario, Canada. *Med Care* 2011;49:932-9.
61. Austin SR, Wong Y-N, Uzzo RG, et al. Why summary comorbidity measures such as the Charlson Comorbidity Index and Elixhauser Score work. *Med Care* 2015;53:e65-72.
62. Drosdowsky A, Gough K. The Charlson Comorbidity Index: problems with use in epidemiological research. *J Clin Epidemiol* 2022;148:174-7.
63. Caughey GE, Ramsay EN, Vitry AI, et al. Comorbid chronic diseases, discordant impact on mortality in older people: a 14-year longitudinal population study. *J Epidemiol Community Health* 2010;64:1036-42.
64. Allyn J, Allou N, Augustin P, et al. A comparison of a machine learning model with EuroSCORE II in predicting mortality after elective cardiac surgery: a decision curve analysis. *PLoS One* 2017;12:e0169772.
65. Weng SF, Vaz L, Qureshi N, et al. Prediction of premature all-cause mortality: a prospective general population cohort study comparing machine-learning and standard epidemiological approaches. *PLoS One* 2019;14:e0214365.
66. Yu Q, Zhang L, Ma Q, et al. Predicting all-cause mortality and premature death using interpretable machine learning among a middle-aged and elderly Chinese population. *Heliyon* 2024;10:e36878.
67. Walker ER, McGee RE, Druss BG. Mortality in mental disorders and global disease burden implications: a systematic review and meta-analysis. *JAMA Psychiatry* 2015;72:334.
68. Weye N, Momen NC, Christensen MK, et al. Association of specific mental disorders with premature mortality in the danish population using alternative measurement methods. *JAMA Netw Open* 2020;3:e206646.
69. Launders N, Hayes JF, Price G, et al. Clustering of physical health multimorbidity in people with severe mental illness: an accumulated prevalence analysis of United Kingdom primary care data. *PLoS Med* 2022;19:e1003976.
70. de Mooij LD, Kikkert M, Theunissen J, et al. Dying too soon: excess mortality in severe mental illness. *Front Psychiatry* 2019;10:855.
71. O'Connor RC, Worthman CM, Abanga M, et al. Gone too soon: priorities for action to prevent premature mortality associated with mental illness and mental distress. *Lancet Psychiatry* 2023;10:452-64.
72. Wang C, Yuan Y, Zheng M, et al. Association of age of onset of hypertension with cardiovascular diseases and mortality. *J Am Coll Cardiol* 2020;75:2921-30.
73. Blom JW, de Ruijter W, Witteman JCM, et al. Changing prediction of mortality by systolic blood pressure with increasing age: the Rotterdam study. *Age (Dordr)* 2013;35:431-8.
74. Martín-Fernández J, Alonso-Safont T, Polentinos-Castro E, et al. Impact of hypertension diagnosis on morbidity and mortality: a retrospective cohort study in primary care. *BMC Prim Care* 2023;24:79.
75. Cleveland RJ, Alvarez C, Nelson AE, et al. Hip symptoms are associated with premature mortality: the Johnston County Osteoarthritis Project. *Osteoarthritis Cartilage* 2020;28:1330-40.
76. Veronese N, Cereda E, Maggi S, et al. Osteoarthritis and mortality: a prospective cohort study and systematic review with meta-analysis. *Semin Arthritis Rheum* 2016;46:160-7.
77. Neuendorf R, Harding A, Stello N, et al. Depression and anxiety in patients with Inflammatory Bowel Disease: A systematic review. *J Psychosom Res* 2016;87:70-80.
78. Moulton CD, Pavlidis P, Norton C, et al. Depressive symptoms in inflammatory bowel disease: an extraintestinal manifestation of inflammation? *Clin Exp Immunol* 2019;197:308-18.
79. Chen J, Geng J, Wang J, et al. Associations between inflammatory bowel disease, social isolation, and mortality: evidence from a longitudinal cohort study. *Therap Adv Gastroenterol* 2022;15:17562848221127474.
80. Hudson M, Chitolie A, Hutton RA, et al. Thrombotic vascular risk factors in inflammatory bowel disease. *Gut* 1996;38:733-7.
81. Corica D, Romano C. Renal involvement in inflammatory bowel diseases. *J Crohns Colitis* 2016;10:226-35.
82. Sobh MM, Abdalbary M, Elnagar S, et al. Secondary osteoporosis and metabolic bone diseases. *J Clin Med* 2022;11:2382.
83. Gregson CL, Armstrong DJ, Bowden J, et al. UK clinical guideline for the prevention and treatment of osteoporosis [published erratum in *Arch Osteoporos* 2022;17:80]. *Arch Osteoporos* 2022;17:58.

Competing interests: Eric Benchimol reports funding from the Canadian Association of Gastroenterology, the Canadian Institutes of Health Research (CIHR), Crohn's Colitis Canada, and Helmsley Charitable Trust; travel support from Samsung Bioepis; and stipends from Catrile & Associates, Cleveland Clinic, University of Calgary, HMP Global, the College of Physicians and Surgeons of Ontario, European Horizon Grant Program, and the Asian Pan-Pacific Society for Pediatric Gastroenterology, Hepatology and Nutrition (APPSPGHAN). He has acted as a consultant for McKesson Canada and the Dairy Farmers of Ontario for matters unrelated to medications used to treat inflammatory bowel disease. He has also acted as a consultant for the Canadian Drug Agency. He reports membership and board roles with Crohn's and Colitis Canada, Crohn's and Colitis Young Adult Network, CIHR, Empowering Next-Generation Researchers in Perinatal and Child Health, Health Canada, and the Canadian Children Inflammatory Bowel Disease Network. No other competing interests were declared.

This article has been peer reviewed.

Affiliations: Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public Health (Postill, Harish, Rosella, Benchimol); and Temerty Faculty of Medicine (Postill, Harish, Benchimol), University of Toronto; ICES (Postill, Tang, Buajitti, Rosella, Benchimol); Department of Public Health Sciences, Dalla Lana School of Public Health (Uchenna Itanyi, Buajitti, Rosella), University of Toronto; SickKids Inflammatory Bowel Disease Centre, Division of Gastroenterology, Hepatology and Nutrition (Kuenzig, Tang, Benchimol), and Child Health Evaluative Sciences, SickKids Research Institute (Kuenzig, Tang, Benchimol), The Hospital for Sick Children; Toronto, Ont.; Department of Epidemiology, Biostatistics and Occupational Health (Buajitti), McGill University, Montréal, Que.; Institute for Better Health (Rosella), Trillium Health Partners, Mississauga, Ont.; Department of Paediatrics, Temerty Faculty of Medicine (Benchimol), University of Toronto, Toronto, Ont.

Contributors: Gemma Postill, Eric Benchimol, and Laura Rosella conceived and designed the study, with substantial input from all authors. All authors contributed to the design of the study cohort and creation of the data collection plan. Furong Tang was the data analyst who created the cohort, with initial cohort ideation support from Emma Buajitti and Ellen Kuenzig, and provided support for the data analysis including review of the statistical code. Gemma Postill performed all data analyses. Ijeoma Uchenna Itanyi and Vinyas Harish provided machine learning methodological guidance. All authors contributed to interpretation of the results. Gemma Postill drafted the manuscript. All authors revised the manuscript critically for important intellectual content, gave final approval of the version to be published, and agreed to be accountable for all aspects of the work.

Content licence: This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY-NC-ND

4.0) licence, which permits use, distribution and reproduction in any medium, provided that the original publication is properly cited, the use is noncommercial (i.e., research or educational use), and no modifications or adaptations are made. See: https://creativecommons.org/licenses/by-nc-nd/4.0/

Funding: This study received funding from the Temerty Centre for Artificial Intelligence Research and Education in Medicine (T-CAIREM) 2022 summer studentship and the 2023 American College of Gastroenterology Medical Student Research Award (both received by Gemma Postill). Laura Rosella is funded by a Canada Research Chair in Population Health Analytics (72051628) and Stephen Family Chair in Community Health from Trillium Health Partners. Eric Benchimol holds the Northbridge Financial Corporation Chair in Inflammatory Bowel Disease, a joint Hospital–University Chair between the University of Toronto, The Hospital for Sick Children, and the SickKids Foundation. The funders had no role in the design or conduct of the research project.

Data sharing: The data set from this study is held securely in coded form at ICES. While legal data sharing agreements between ICES and data providers (e.g., health care organizations and government) prohibit ICES from making the data set publicly available, access may be granted to those who meet prespecified criteria for confidential access, available at www.ices.on.ca/DAS (email: das@ices.on.ca). The full data set creation plan and underlying analytic code are available from the authors upon request, understanding that the computer programs may rely upon coding templates or macros that are unique to ICES and are therefore either inaccessible or may require modification. The code and pre-trained model are publicly available: https://github.com/PopHealthAnalytics/Multimorbidity-IBD.

Disclaimer: This study was supported by ICES, which is funded by an annual grant from the Ontario Ministry of Health (MOH) and the Ministry of Long-Term Care (MLTC). The opinions, results and conclusions reported in this paper are those of the authors and are independent from the funding sources. No endorsement by ICES or the Ontario MOH or MLTC is intended or should be inferred. This document used data adapted from the Statistics Canada Postal Code^{OM} Conversion File, which is based on data licensed from Canada Post Corporation, and/or data adapted from the Ontario Ministry of Health Postal Code Conversion File, which contains data copied under license from Canada Post Corporation and Statistics Canada. The authors thank the Toronto Community Health Profiles Partnership for providing access to the Ontario Marginalization Index.

Accepted: Feb. 18, 2025

Correspondence to: Eric Benchimol, eric.benchimol@sickkids.ca