

Applying Machine Learning Models Derived From Administrative Claims Data to Predict Medication Nonadherence in Patients Self-Administering Biologic Medications for Inflammatory Bowel Disease

Christian Rhudy, PharmD, MBA,^{*}  Courtney Perry, DO, MS,[†]  Michael Wesley, PhD,[‡]
David Fardo, PhD,[§] Cody Bumgardner, PhD,[¶] Syed Hassan, MBBS/MD,[†] Terrence Barrett, MD,[†] and
Jeffery Talbert, PhD,[¶] 

^{*}Department of Pharmacy Services, University of Kentucky Healthcare, Lexington, KY, USA

[†]Division of Digestive Diseases and Nutrition, Department of Medicine, University of Kentucky College of Medicine, Lexington, KY, USA

[‡]Department of Behavioral Science, Psychiatry and Psychology, University of Kentucky College of Medicine, Lexington, KY, USA

[§]Department of Biostatistics, University of Kentucky, College of Public Health, Lexington, KY, USA

[¶]Department of Pathology and Laboratory Medicine, University of Kentucky College of Medicine, Lexington, KY, USA

[¶]Division of Biomedical Informatics, University of Kentucky College of Medicine, Lexington, KY, USA

Address correspondence to: Christian Rhudy, PharmD, MBA, Department of Pharmacy Services, University of Kentucky Healthcare, 531 Wellington Way, Office 311, Lexington, KY 40503, USA (Office: 859.218.0727; 502.657.8075; christian.rhudy@uky.edu).

Background: Adherence to self-administered biologic therapies is important to induce remission and prevent adverse clinical outcomes in Inflammatory bowel disease (IBD). This study aimed to use administrative claims data and machine learning methods to predict nonadherence in an academic medical center test population.

Methods: A model-training dataset of beneficiaries with IBD and the first unique dispense of a self-administered biologic between June 30, 2016 and June 30, 2019 was extracted from the Commercial Claims and Encounters and Medicare Supplemental Administrative Claims Database. Known correlates of medication nonadherence were identified in the dataset. Nonadherence to biologic therapies was defined as a proportion of days covered ratio <80% at 1 year. A similar dataset was obtained from a tertiary academic medical center's electronic medical record data for use in model testing. A total of 48 machine learning models were trained and assessed utilizing the area under the receiver operating characteristic curve as the primary measure of predictive validity.

Results: The training dataset included 6998 beneficiaries ($n = 2680$ nonadherent, 38.3%) while the testing dataset included 285 patients ($n = 134$ nonadherent, 47.0%). When applied to test data, the highest performing models had an area under the receiver operating characteristic curve of 0.55, indicating poor predictive performance. The majority of models trained had low sensitivity and high specificity.

Conclusions: Administrative claims-trained models were unable to predict biologic medication nonadherence in patients with IBD. Future research may benefit from datasets with enriched demographic and clinical data in training predictive models.

Lay Summary

This study used insurance data and machine learning to predict if patients with inflammatory bowel disease would obtain their medications on time. The models did not make accurate predictions, suggesting more data is needed for better predictions.

Key Words: medication adherence, biologics, machine learning

Background

Inflammatory bowel disease (IBD), comprised of Crohn disease and ulcerative colitis, affects approximately 6.8 million patients globally, with a prevalence rate of 464.5 cases per 100 000 in the United States.¹ Moderate to severe IBD is often medically managed with biologic therapies, several of which are self-administered in the outpatient setting, including adalimumab, certolizumab, golimumab, ustekinumab, and risankizumab.^{2,3} Though biologics are considered cost-effective therapies, particularly when compared to costs associated with poorly controlled IBD, they still contribute

significantly to the overall cost of IBD care, averaging \$36 051 per patient per year in 2015.⁴

Poor adherence to biological therapy contributes to worse IBD outcomes and higher care-associated costs. Current estimates of nonadherence to biologic therapy range from 17.4% to 45%.⁵⁻⁸ Several risk factors for biologic nonadherence have been identified, including younger age, female gender, tobacco use, payor type, Crohn's disease diagnosis, and comorbid diagnoses such as anxiety and depression.⁵⁻¹² Patient medication utilization patterns are also associated with biologic nonadherence, such as nonadherence

Received for publication: May 4, 2024. Editorial Decision: June 26, 2024

© The Author(s) 2024. Published by Oxford University Press on behalf of Crohn's & Colitis Foundation.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

Key Messages

What is already known? Despite efficacy, nonadherence to biologic therapy in inflammatory bowel disease remains high and previous work has characterized risk factors for medication nonadherence.

What is new here? Established risk factors for biologic nonadherence were identified in a large administrative claims dataset to train predictive machine learning models. Models were tested for validity in an academic medical center patient population.

How can this study help patient care? While models in this study were unsuccessful in reliably predicting biologic nonadherence, the results described can provide a basis for future investigations utilizing alternate training datasets and methodologies.

to prior IBD therapies, concurrent dual therapy with a biologic and an immunomodulator, and chronic opioid use.^{5,9,12-14}

The development of predictive models to identify patients at high risk of nonadherence and proactively address barriers to adherence is appealing, as risk factors are often present at baseline. Machine learning models have previously been developed to predict medication nonadherence in other disease states, as well as non-biologic immunomodulator therapy in IBD.¹⁵⁻¹⁸ Furthermore, many of the risk factors for nonadherence can be identified from the data contained in administrative claims databases, providing a sufficiently large dataset for training such a model.^{18,19} However, the usefulness of this model hinges on the predictive validity when applied in a clinical environment. This study will examine the utility of several competing machine learning models trained on administrative claims data to predict IBD biologic nonadherence in a tertiary academic medical center patient population. Investigators hypothesized training predictive models on large-scale administrative claims datasets including variables previously associated with nonadherence would produce models to accurately identify nonadherence in individual patient populations.

Materials and Methods

Data Sources and Study Design

Model-training datasets were derived from the Meritvet MarketScan Commercial Claims and Encounters and Medicare Supplemental Administrative Claims Database, hereafter referred to as MarketScan. This database contains demographics, records of inpatient and outpatient healthcare encounters, and prescription medication transactions for more than 273 million unique beneficiaries covered by employer health plans.²⁰ This dataset includes *International Classification of Diseases, Tenth Revision, Clinical Modification* (ICD-10-CM) and *Healthcare Common Procedure Coding System/Current Procedure Terminology* (HCPCS/CPT) codes associated with encounter diagnoses and procedures performed, as well as the *Generic Product Identifier* (GPI) associated with dispensed medications.²¹⁻²³

MarketScan beneficiaries with a first unique medication dispense between June 30, 2016 and June 30, 2019 of any self-administered biologic medication with a current

FDA-approved indication for IBD (adalimumab, certolizumab, golimumab, and ustekinumab) were considered for inclusion. The first dispense date was considered the index date for study purposes. Otherwise eligible MarketScan beneficiaries were excluded if: (1) they did not have at least 2 outpatient encounter claims or 1 inpatient encounter claim associated with an ICD-10-CM code in a 180-day preindex period indicating IBD (K50*, K51*); (2) they were not continuously enrolled for 180 days prior to and 365 days post-index date, with prescription benefits and no gap in enrollment greater than 30 days; or (3) were <18 years of age at the index date. From this cohort, beneficiary demographics and all encounter-level data (including dates of service and associated ICD-10-CM and HCPCS/CPT) were extracted from the preindex period, while all medication dispensing records were obtained from the pre- and postindex periods.

In order to provide a dataset to test model validity outside of MarketScan beneficiaries, a separate cohort was identified from patients treated at a tertiary academic medical center that serves a rural and disproportionate share population. Eligible patient records were identified as having a first unique biologic dispense date between June 30, 2016 and June 30, 2022. Similar to the training cohort, patients were excluded if (1) they were <18 years of age at index date, or (2) had no diagnosis of IBD during the preindex period. While additional relevant clinical and demographic data was available in electronic medical records, data extraction was performed to mirror only the same data elements available from the MarketScan beneficiary records. Only medication dispensing records from pharmacies affiliated with the health system were available for analysis, in contrast to the training dataset which contained all dispenses where insurance was billed across any location. This study was reviewed and approved by the Medical Institutional Review Board of the academic medical center.

Data Preprocessing

After extraction, raw data was used to generate 2 dataset types for model training and assessment. In the first dataset type (investigator-only), only investigator-selected features of interest were included. Features were identified from a review of the relevant literature and clinical experience of the investigators. Demographic features included index IBD diagnosis, sex, age, Medicare payor, and index biologic. Encounter-level features included preindex diagnoses of anxiety/depression or smoking, number of inpatient and outpatient encounters within the 180-day preindex period, and the Charlson Comorbidity Index (CCI) assessed at the index date.²⁴ CCI was calculated through the use of ICD-10-CM codes via previously published methods.²⁵ Medication-related features included prior IBD therapies utilized in the preindex period, identified nonadherence to prior IBD therapies, chronic opioid use, dual therapy with an immunomodulator in addition to the index biologic, number of unique medications dispensed, and prednisone milligram equivalents of corticosteroids dispensed. Chronic opioid use was defined as greater than or equal to 90-day supply of outpatient opioid dispenses in the preindex period without any 30-day gaps in supply.¹⁴ All features were binary indicators, with the exception of age (categorical), number of inpatient and outpatient encounters, number of unique medications dispensed, and prednisone milligram equivalents (continuous). In the

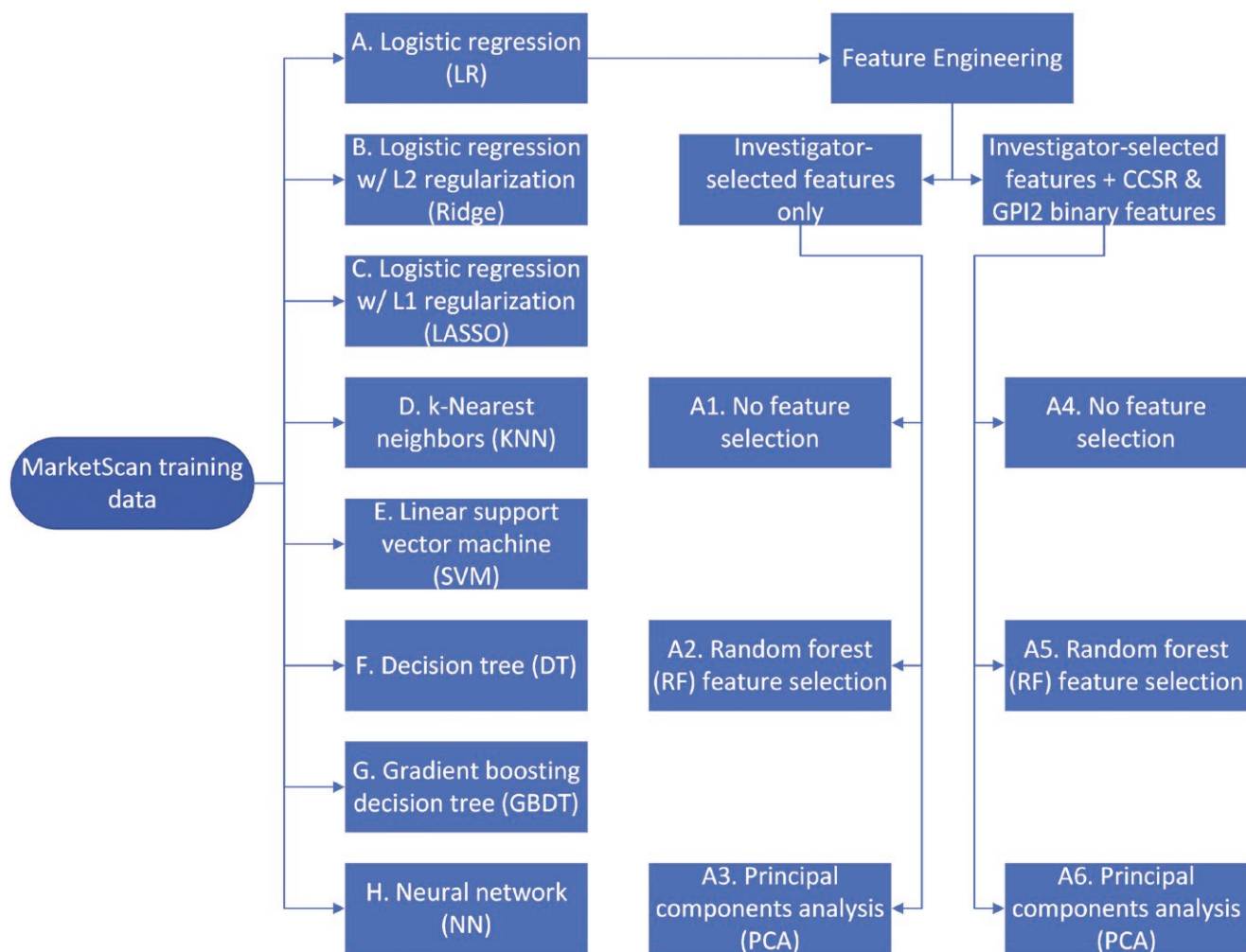


Figure 1. Summary of approach to model creation.

event of missing or incomplete data, the binary indicator was considered negative or zero was substituted for continuous data. Full study definitions of administrative claims codes used to generate investigator-selected features are available in [Supplementary Table 1](#).

In the second dataset type the same investigator-selected features were generated, but additional binary indicators were created utilizing all ICD-10-CM codes and GPI codes from preindex encounters and medication dispenses. Binary indicator features were created for each category of Clinical Classification Software Refined (CCSR), a tool which groups ICD-10-CM codes into clinically relevant categories.²⁶ In addition, binary features for the dispense of a particular medication group (GPI 2-digit) in the preindex period were also created. These additional features were intended to aid in identification of any associations with medication nonadherence not previously identified in the literature. This dataset is henceforward referred to as the investigator + CCSR + GPI2 dataset. In both dataset types, features were standardized to a mean of zero and a standard deviation of one.

Data Labeling

Training and testing datasets were labeled with a binary feature indicating nonadherence derived from a proportion of days covered (PDC) calculation. PDC is an indirect measure

of outpatient medication adherence that assesses the degree of adherence in a ratio from 0 to 1 based upon the intervals between medication dispenses and the supply dispensed in a given time period. PDC for this study was calculated as follows:

$$PDC = \frac{\text{Days covered by biologic supply in post-index period}}{365}$$

Where days covered are the number of days in which a subject would have supply under the assumption that they are taking the medication as prescribed and only receiving medication from observed dispenses.^{27,28} Dispensed supply of any biologic qualifying for inclusion in this study was considered in the numerator. PDC values were dichotomized to create the nonadherence feature, with a PDC < 0.8 corresponding to nonadherence (a threshold commonly used in prior literature).²⁹

Model Training

The approach to model training is outlined in [Figure 1](#). Eight machine learning classification algorithms (labeled A-H) were used to develop six candidate models (#1-6) each for a total of 48 models (A1–H6). Classifier algorithms chosen included logistic regression, ridge regression, lasso regression, k-nearest neighbors, linear support vector machines, decision

trees, gradient boosting decision trees, and neural networks. All classifier algorithms used were obtained from the publicly available Python package scikit-learn, version 1.3.0.³⁰ These algorithms were chosen based on availability within the scikit-learn package, as well as to investigate several competing approaches with distinct advantages in handling the study datasets. In specific, logistic regression was selected as a base model for comparison to other approaches. Given the number of considered features, algorithms with built-in feature selection (ridge, lasso regression) or the ability to capture complex decision boundaries in high-dimensional space (k-nearest neighbors, linear support vector machines) were considered. Additionally, investigators suspected non-linear relationships between features for which decision trees and neural network classifiers may be more suited.

For each classifier algorithm, three models were trained using the investigator-only dataset (models 1–3). In model 1, no feature selection or transformation beyond zero standardization was performed. In model 2, a random forest algorithm was first applied to the dataset to estimate feature importance and only features with importance greater than the mean were included in model training. This preprocessing step was conducted to reduce potential errors introduced by possibly irrelevant features. In model 3, data was transformed the data with principal components analysis prior to model training. This step was taken to reduce dimensionality while preserving variance, in hopes of reducing overfitting and enhancing the generalization of the model. Models 4, 5, and 6 were trained using the same methods as 1, 2, and 3, respectively, but used the investigator + CCSR + GPI2 dataset to incorporate additional and potentially relevant unknown features.

All models were trained using ten-fold cross-validation on the training dataset to obtain the optimal model for each algorithm. The optimal model was defined as having the highest F_β score on the training data, where:

$$F_\beta = \frac{(1 + \beta^2) * Precision * Recall}{(\beta^2 * Precision + Recall)}$$

$$Precision = \frac{True\ Positives}{(True\ Positives + False\ Positives)}$$

$$Recall = \frac{True\ Positives}{(True\ Positives + False\ Negatives)}$$

A $\beta = 2$ value was chosen for this study in order to minimize false negatives in the model. This was chosen given an assumption that the preference would be to overtreat false positive risks for nonadherence rather than undertreat false negatives. For models in which there were multiple hyperparameters or multiple valid settings for a single hyperparameter, a grid search of all combinations of hyperparameters specified was conducted to obtain the combination that maximized F_β score (Supplementary Table 2).

Model Performance Evaluation and Statistical Analysis

Descriptive statistics were reported for all variables, including the number and percentage of beneficiaries with a given condition for categorical variables, as well as mean, median, standard deviation, and interquartile range for continuous variables, as observed in training and test datasets. To examine relationships between individual features and

biological nonadherence at baseline in each dataset, 2 logistic regression models including all investigator-only variables were constructed using training and test datasets. Odds ratios from each of these logistic regression models were reported.

Model predictive validity was assessed via several metrics, including accuracy, area under the receiver operating characteristic curve (AUC), Brier score, F1 & F_β score, negative predictive value, precision, sensitivity, and specificity.³¹ AUC was used as the primary metric of model predictive validity, and used to select the best model within a classifier algorithm for comparison against other classifier algorithms. Metrics were obtained for model performance on both training and testing datasets; however, only metrics derived from testing datasets were utilized in assessing model predictive validity.

Results

Description of Training and Test Datasets

After the application of inclusion and exclusion criteria (Figure 2), 6,998 eligible beneficiaries were included in the training dataset, and 285 patients were included in the test dataset. Rates of nonadherence were higher in the test dataset ($n = 134$, 47.02%) as compared to the training dataset ($n = 2680$, 38.3%; Table 1). Observed distributions of the PDC variable were left-skewed (Supplementary Figure 1). In the training dataset, factors associated with significantly lower odds of nonadherence included age groups of 45–54 (odds ratio [OR] 0.863; 95% confidence interval [95%CI] 0.756, 0.986) and 55–64 (OR 0.712; 95% CI: 0.599, 0.846) referent to ≤ 44 , diagnosis of Crohn's disease (OR 0.685; 95% CI: 0.586, 0.801), and prior use of mercaptopurines (OR 0.756; 95% CI: 0.62, 0.92) or azathioprine (OR 0.763; 95% CI: 0.653, 0.891). Features associated with higher odds of nonadherence included female sex (OR 1.131; 95% CI: 1.022, 1.252), index biologic of certolizumab (OR 2.288; 95% CI: 1.612, 3.248) or ustekinumab (OR 1.484; 95% CI: 1.261, 1.746) compared to index adalimumab, diagnosis of tobacco use (OR 1.262; 95% CI: 1.019, 1.563), CCI (OR 1.061; 95% CI: 1.001, 1.125), preindex vedolizumab administration (OR 1.324; 95% CI: 1.011, 1.734), or any inpatient admission (OR 1.149; 95% CI: 1.014, 1.302).

In the test dataset, a single variable was associated with significantly lower odds of nonadherence including Medicare coverage (OR 0.419; 95% CI: 0.203, 0.865). Conversely, several factors were associated with greater odds of nonadherence, including female sex (OR 1.885; 95% CI: 1.116, 3.183), age group 45–54 (OR 2.345; 95% CI: 1.097, 5.017) referent to ≤ 44 , and index biologic of certolizumab (OR 4.827; 95% CI: 1.195, 19.488).

Model Performance

Receiver operating characteristic curves assessing classification performance for all models on training and test datasets are available in Supplementary Figures 2 and 3. Based upon AUC when applied to testing data, the highest-performing models from each algorithm type were selected for comparison (models A3, B3, C3, D6, E3, F3, G1, and H6). Among the selected models, F3 and G1 tied for the highest AUC at 0.55 each, however, the difference from other compared models was minimal (Figure 3).

Model performance metrics on training and test datasets for the selected models is displayed in Table 2. On the training

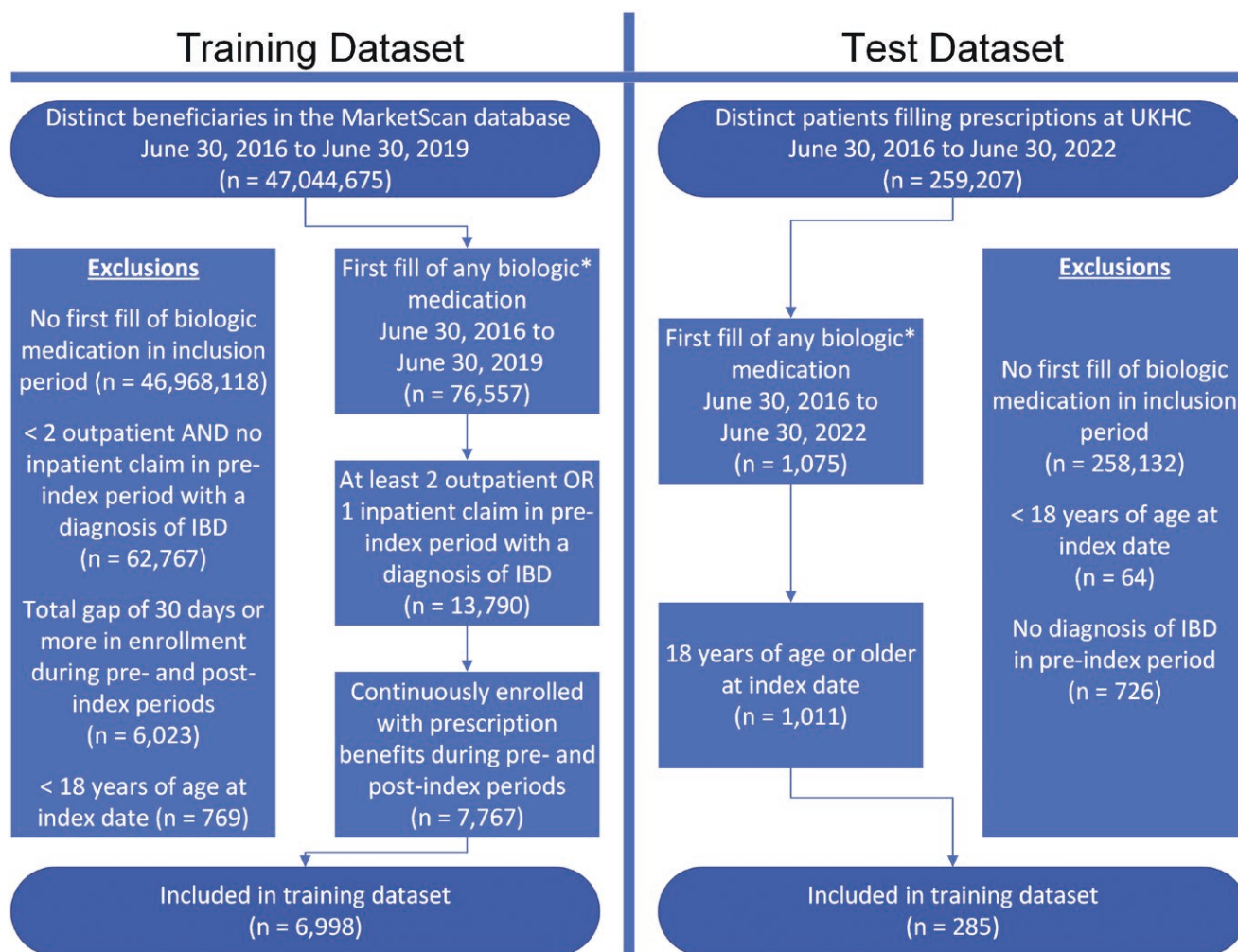


Figure 2. Application of inclusion and exclusion criteria to generate training and test datasets.

set, high accuracy was observed for models D6 (71.42%) and H6 (99.94%); however, these models had similar accuracy to other candidate models when applied to the test dataset. Confusion matrices for all selected models are provided for visual assessment of model accuracy (Figure 4).

In general, models trended towards low sensitivity and high specificity on the test dataset, with the exception of model G1 (gradient-boosted decision tree; sensitivity = 0.71; specificity = 0.36). Model G1 also featured the highest F1, F_{β} , and negative predictive value values as compared to other models (0.58, 0.65, and 0.59, respectively). This may be attributable due to the nature of the learning algorithm, which sequentially builds a series of decision trees that improve upon prior errors. This can result in overfitting to the minority/positive case (eg, nonadherent) and result in an increase in true and false positives. In addition to the performance metrics for selected models, assessment of metrics for all candidate models on training and test datasets are provided in [Supplementary Tables 3 and 4](#).

Discussion

The goal of this study was to train machine learning models using administrative claims data to accurately predict nonadherence to self-administered biologic therapies in

patients with IBD at an academic health system. While numerous associations with nonadherence to self-administered biologic medications were observed within the initial analysis, contrary to our main hypothesis, machine learning models trained on data available in administrative claims databases failed to reliably predict nonadherence in a test dataset derived from an academic medical center's patient population. Examining the primary metric of model predictive value (AUC) for this analysis, values of 0.5 to a model that performs in accordance with chance, with values between 0.7 and 1 considered a model with moderate to perfect accuracy.³² The highest-performing models generated by this study had an AUC of 0.55 when applied to the test dataset, suggesting limited usefulness in the prediction of nonadherence. While more acceptable AUC values were observed on the training data for some candidate models, they did not translate to higher performance on the testing dataset, suggesting that those models overfit training data and did not have sufficient bias to generalize to new data.³³

There are several possible explanations for why model predictive performance on unseen data was low. Medication nonadherence is a multifactorial issue, and administrative claims data does not contain data on several known correlates with self-administered biologic nonadherence. Datasets with extended demographics and information on patient-specific

Table 1. Logistic regression analysis of investigator-selected features in training and test datasets.

Variable	Training dataset			Test dataset		
	Beneficiaries (n = 6998)	Odds ratio	95% CI	Patients (n = 285)	Odds ratio	95% CI
Adherence						
Nonadherent	2680 (38.3%)	-	-	134 (47.02%)	-	-
<i>Index diagnosis</i>						
Crohns disease	4483 (64.06%)	0.685	(0.586, 0.801)	246 (86.32%)	1.512	(0.458, 4.990)
Ulcerative colitis	3543 (50.62%)	1.123	(0.967, 1.303)	63 (22.11%)	2.085	(0.799, 5.442)
<i>Sex</i>						
Female	3665 (52.37%)	1.131	(1.022, 1.252)	160 (56.14%)	1.885	(1.116, 3.183)
<i>Age group</i>						
≤ 44 (referent)	3967 (56.69%)	-	-	187 (65.61%)	-	-
45 - 54	1494 (21.35%)	0.863	(0.756, 0.986)	44 (15.44%)	2.345	(1.097, 5.017)
55 - 64	1292 (18.46%)	0.712	(0.599, 0.846)	25 (8.77%)	2.088	(0.716, 6.084)
≥ 65	245 (3.5%)	0.99	(0.417, 2.353)	29 (10.18%)	1.385	(0.365, 5.262)
<i>Coverage</i>						
Medicare	238 (3.4%)	0.838	(0.354, 1.986)	75 (26.32%)	0.419	(0.203, 0.865)
<i>Index biologic</i>						
Adalimumab (referent)	5940 (84.88%)	-	-	108 (37.89%)	-	-
Golimumab	115 (1.64%)	1.003	(0.686, 1.467)	3 (1.05%)	2.169	(0.156, 30.256)
Certolizumab	136 (1.94%)	2.288	(1.612, 3.248)	18 (6.32%)	4.827	(1.195, 19.488)
Ustekinumab	807 (11.53%)	1.484	(1.261, 1.746)	156 (54.74%)	0.774	(0.435, 1.378)
<i>Diagnoses</i>						
Anxiety/Depression	1347 (19.25%)	1.138	(0.998, 1.298)	71 (24.91%)	0.729	(0.377, 1.411)
Smoking	390 (5.57%)	1.262	(1.019, 1.563)	57 (20%)	0.993	(0.494, 1.994)
<i>Charlson comorbidity index</i>						
Mean (SD)	0.91 (1.3)	1.061	(1.001, 1.125)	0.95 (1.5)	1.031	(0.797, 1.334)
Median (IQR)	0 (1)			0 (1)		
<i>Prior IBD therapies</i>						
Infliximab	706 (10.09%)	0.995	(0.841, 1.175)	32 (11.23%)	1.224	(0.529, 2.831)
Vedolizumab	244 (3.49%)	1.324	(1.011, 1.734)	24 (8.42%)	1.18	(0.444, 3.137)
Aminosalicylates	3134 (44.78%)	0.919	(0.820, 1.031)	3 (1.05%)	-*	-*
Mercaptopurines	515 (7.36%)	0.756	(0.62, 0.92)	1 (0.35%)	-*	-*
Azathioprine	899 (12.85%)	0.763	(0.653, 0.891)	5 (1.75%)	0.211	(0.017, 2.6)
Budesonide	1649 (23.56%)	0.941	(0.835, 1.060)	9 (3.16%)	0.466	(0.067, 3.26)
Methotrexate	342 (4.89%)	0.939	(0.742, 1.189)	5 (1.75%)	1.515	(0.099, 23.222)
<i>Inpatient admissions</i>						
Any inpatient admission	1469 (20.99%)	1.149	(1.014, 1.302)	100 (35.09%)	1.479	(0.773, 2.828)
<i>Outpatient encounters</i>						
Mean (SD)	12.1 (9.5)	0.998	(0.992, 1.004)	14.5 (9.5)	0.993	(0.968, 1.019)
Median (IQR)	10 (9)			10 (12)		
<i>Medication utilization</i>						
Nonadherent to prior therapy	673 (9.62%)	1.033	(0.874, 1.220)	1 (0.35%)	-*	-*
Chronic opioid use	289 (4.13%)	1.211	(0.942, 1.557)	1 (0.35%)	-*	-*
Dual therapy with anti-inflammatory	2067 (29.54%)	0.889	(0.789, 1.001)	13 (4.56%)	2.63	(0.478, 14.477)
<i>Number of medications</i>						
Mean (SD)	6.7 (4.5)	1.012	(0.998, 1.027)	1.5 (2.8)	1.053	(0.916, 1.212)
Median (IQR)	6 (6)			0 (2)		
<i>Prednisone milligram equivalents dispensed</i>						
Mean (SD)	1465 (18 217)	1	(1,1)	54 (264)	0.999	(0.998, 1.001)
Median (IQR)	160 (1260)			0 (0)		

CI, confidence interval; IQR, interquartile range; SD, standard deviation. Odds ratios and 95% CI obtained a logistic regression model predicting nonadherence and incorporating all examined investigator-specified variables with no additional feature engineering. * indicates the 95% confidence interval was outside of interpretable range.

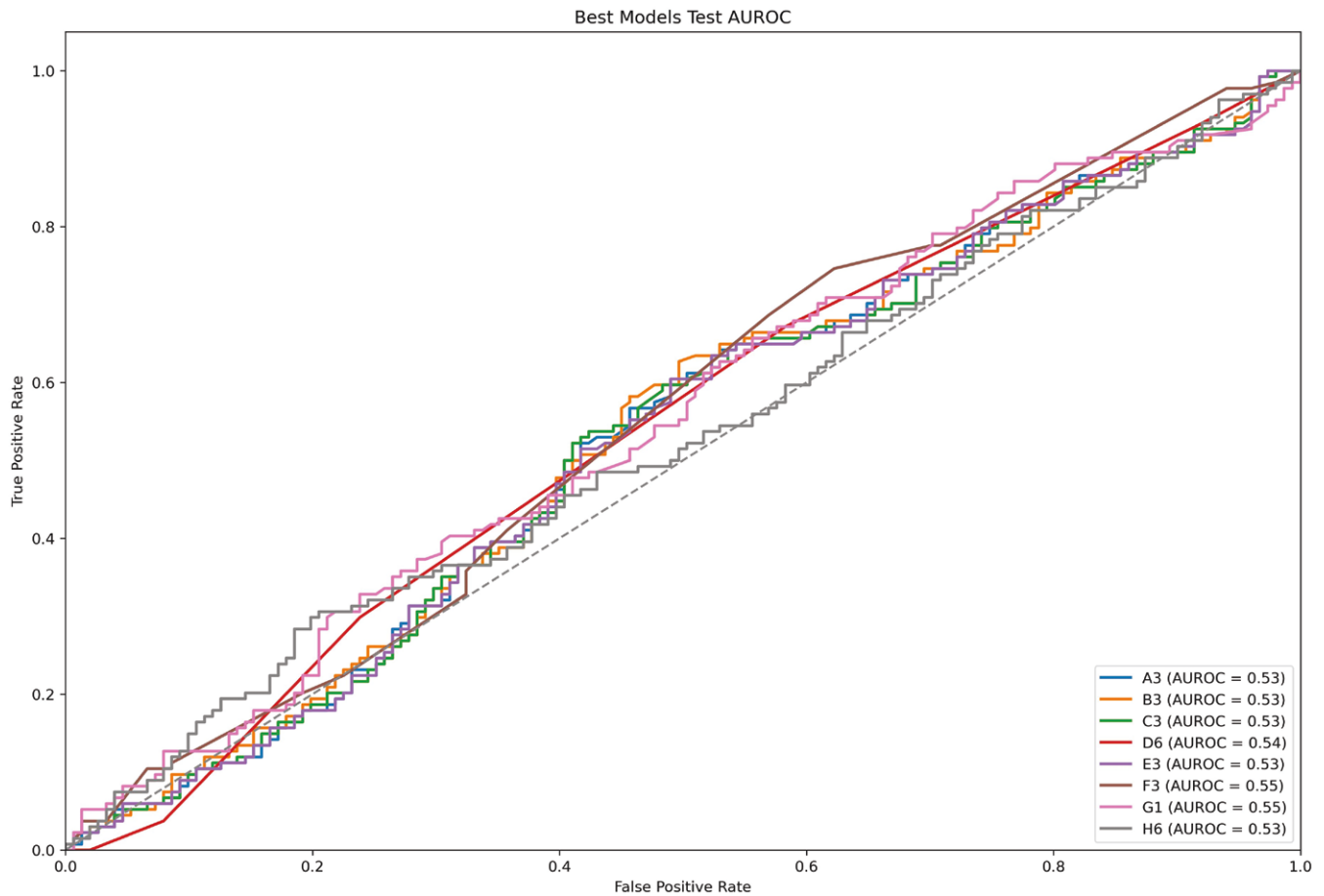


Figure 3. Comparison of receiver operating characteristic curves for selected high-performing models.

Table 2. Predictive performance measures for highest-performing models.

Model performance on training dataset								
Measure	A3	B3	C3	D6	E3	F3	G1	H6
Accuracy	55.94%	55.56%	55.72%	71.42%	55.99%	60.77%	61.66%	99.94%
F1 score	0.50	0.50	0.50	0.55	0.50	0.54	0.55	1.00
F_β score ($\beta = 2$)	0.54	0.54	0.54	0.48	0.54	0.57	0.58	1.00
Brier score	0.24	0.24	0.24	0.18	0.23	0.23	0.23	0.00
Sensitivity	0.57	0.58	0.58	0.45	0.57	0.59	0.61	1.00
Specificity	0.55	0.54	0.54	0.88	0.55	0.62	0.62	1.00
Precision	0.44	0.44	0.44	0.70	0.44	0.49	0.50	1.00
Negative predictive value	0.68	0.67	0.67	0.72	0.68	0.71	0.72	1.00
Model performance on test dataset								
Measure	A3	B3	C3	D6	E3	F3	G1	H6
Accuracy	52.98%	53.33%	52.98%	54.39%	52.98%	52.63%	52.63%	54.39%
F1 score	0.47	0.48	0.48	0.38	0.47	0.42	0.58	0.41
F_β score ($\beta = 2$)	0.46	0.46	0.46	0.33	0.46	0.38	0.65	0.36
Brier score	0.25	0.25	0.25	0.29	0.25	0.25	0.25	0.41
Sensitivity	0.45	0.46	0.46	0.30	0.45	0.36	0.71	0.34
Specificity	0.60	0.60	0.60	0.76	0.60	0.68	0.36	0.73
Precision	0.50	0.50	0.50	0.53	0.50	0.49	0.50	0.52
Negative predictive value	0.55	0.55	0.55	0.55	0.55	0.54	0.59	0.55

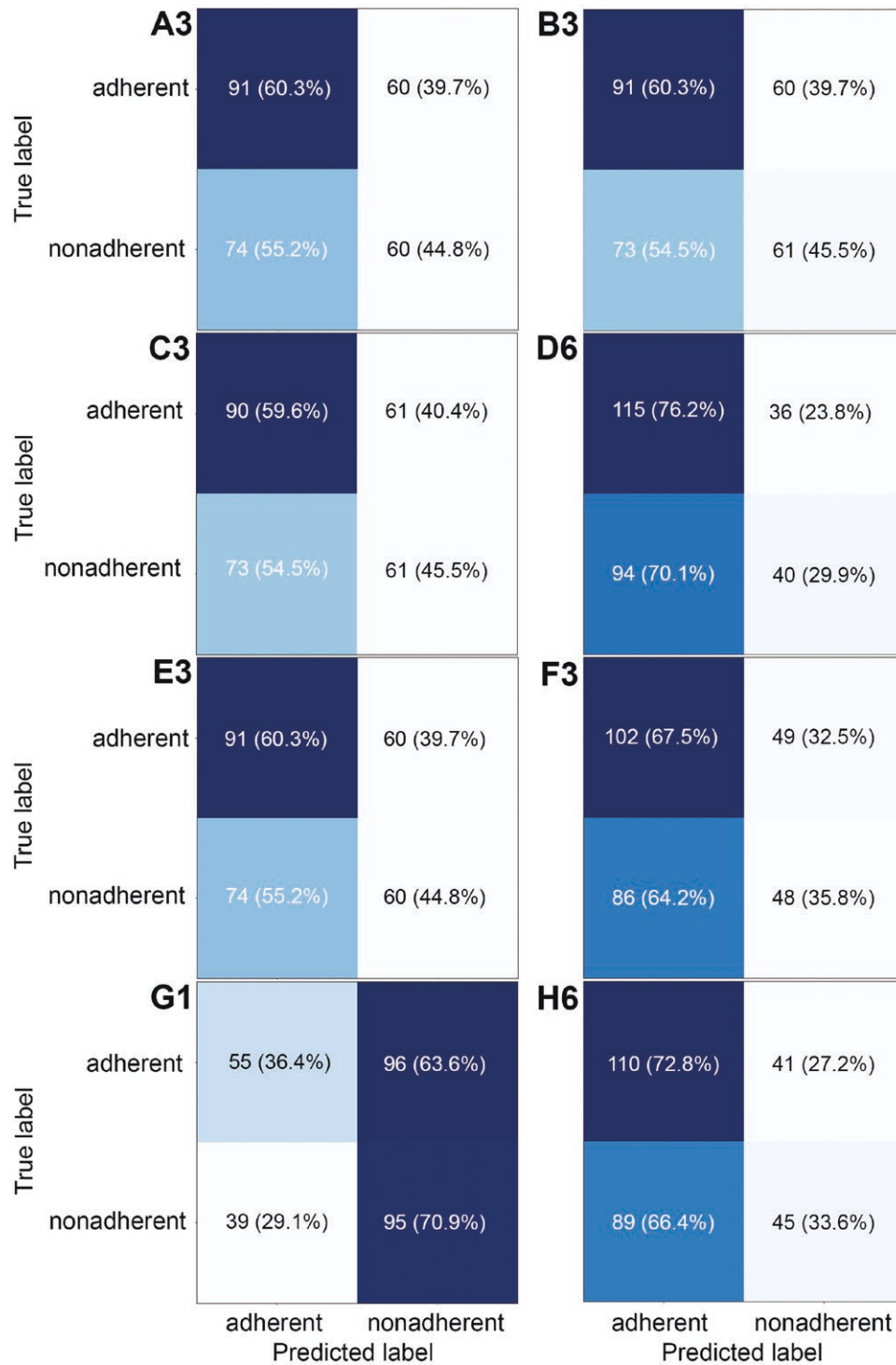


Figure 4. Confusion matrices for selected models.

social determinants of health, including race/ethnicity, income, and education level might provide additional predictive value to train more accurate models.^{5,34–37} In addition, patient health literacy and disease-related knowledge are known to have a significant correlation with medication adherence, although this factor is much more difficult to accurately assess and collect at a large scale.^{13,16,38,39} Furthermore, no clinical context is available in administrative claims data, such as provider assessments, imaging and procedure studies,

laboratory values, and others that have been used in successfully training models to predict adherence in IBD previously.¹⁶ Other studies using machine learning methodology trained on MarketScan data to predict adherence had similar difficulties identifying nonadherence.¹⁸ Future investigations should consider model training in enriched data sources, including electronic medical records or administrative claims data sources with additional socioeconomic and demographic variables. Additionally, assessments of health literacy or IBD-specific

knowledge or beliefs may be necessary for predictive models to accurately classify the risk of nonadherence, although the collection of such data is not widespread.

Upon closer examination of the additional metrics of model fit, in general, models trended towards low sensitivity and high specificity (ie, a tendency towards false positives rather than false negatives). Several of the candidate models in series F and G (decision trees and gradient-boosted decision trees) eschewed this trend, most notably candidate model G1 with higher sensitivity and low specificity. As we ultimately aim to identify patients at risk for nonadherence to biological medications, a model with higher sensitivity is preferred, as potential interventions (additional education/behavioral interventions) are likely to be inexpensive and the risk of positive misidentification is unlikely to pose a significant risk to the patient.^{39,40}

While model performance in predicting nonadherence in this study was low, potential insights for future studies can be gleaned. In contrast to prior literature suggesting increased risk, a diagnosis of Crohn's disease was observed to be associated with a significantly lower risk of nonadherence.¹² This suggests that the relationship between IBD diagnosis and biological nonadherence may involve relationships with other relevant correlates and should be more thoroughly explored in additional literature. Additionally, the highest-performing models in this analysis were trained from the investigator-only dataset, and utilized primary components analysis (PCA) dimensionality reduction as part of data preprocessing. The greater performance of the investigator-specified feature set suggests that the additional CCSR and GPI columns added no net information gain at best, and at worst, generated additional noise in the models. Omission of these columns, or a more targeted approach to selection in the future is likely appropriate. Furthermore, use of PCA to reduce training dataset noise and improve future models appears appropriate. In consideration of the numerous causes of nonadherence, dimensionality reduction with PCA or another mechanism is likely necessary.

Strengths of this study include the use of a large training dataset, which typically increases the likelihood of creating more generalizable models. In addition, utilizing multiple machine learning algorithms and feature engineering approaches to detect nonadherence ensures that conclusions are not based upon the failure of a single approach.

Weaknesses of this study include the inability to train models on extended demographic and clinical data due to the limitations of administrative claims data, which may have prevented a suitable model from being constructed. This study was also unable to include an evaluation of dosing interval as a possible predictor of nonadherence.

Additionally, it is possible the use of 2 distinct data sources contributed in part to poor model performance on test datasets. The training and test datasets substantially differed in the distribution of demographic and clinical data. As well, missing external dispense data may have limited identification of medication-related features in the testing dataset. While potentially representative of available data in many practice locations, this limits the predictive potential of a model trained with more comprehensive dispensing histories. Training models on a dataset more similar to (or drawn directly from) the target population in the future may improve predictive performance in that population.

In conclusion, machine learning models trained on administrative claims data were unable to accurately predict medication nonadherence in patients self-administering biologics for IBD in a tertiary academic medical center patient population. Future research into training models should consider training in datasets with additional demographic and disease-state relevant variables, while omitting excessive information on unrelated diagnoses or medication dispenses.

Supplementary Data

Supplementary data are available at *Crohn's & Colitis* 360 online.

Funding

The project described was supported by the NIH National Center for Advancing Translational Sciences through grant number UL1TR001998. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Conflict of Interest

The authors have no relevant competing or financial interests to declare.

Data Availability

Data used in this study was made available to the authors by a third-party license from Meritve™. The Meritve MarketScan Research Databases are available for researchers who purchase access to the data and complete the required data use agreement processes.

References

1. Collaborators GBDIBD. The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet Gastroenterol Hepatol*. 2020;5(1):17-30. doi: [10.1016/S2468-1253\(19\)30333-4](https://doi.org/10.1016/S2468-1253(19)30333-4)
2. Feuerstein JD, Isaacs KL, Schneider Y, Siddique SM, Falck-Ytter Y, Singh S; AGA Institute Clinical Guidelines Committee. AGA clinical practice guidelines on the management of moderate to severe ulcerative colitis. *Gastroenterology*. 2020;158(5):1450-1461. doi: [10.1053/j.gastro.2020.01.006](https://doi.org/10.1053/j.gastro.2020.01.006)
3. Feuerstein JD, Ho EY, Shmidt E, et al.; American Gastroenterological Association Institute Clinical Guidelines Committee. AGA clinical practice guidelines on the medical management of moderate to severe luminal and perianal fistulizing crohn's disease. *Gastroenterology*. 2021;160(7):2496-2508. doi: [10.1053/j.gastro.2021.04.022](https://doi.org/10.1053/j.gastro.2021.04.022)
4. Yu H, MacIsaac D, Wong JJ, et al. Market share and costs of biologic therapies for inflammatory bowel disease in the USA. *Aliment Pharmacol Ther*. 2018;47(3):364-370. doi: [10.1111/apt.14430](https://doi.org/10.1111/apt.14430)
5. Lopez A, Billioud V, Peyrin-Biroulet C, Peyrin-Biroulet L. Adherence to anti-TNF therapy in inflammatory bowel diseases: a systematic review. *Inflamm Bowel Dis*. 2013;19(7):1528-1533. doi: [10.1097/MIB.0b013e31828132cb](https://doi.org/10.1097/MIB.0b013e31828132cb)
6. Aluzaita K, Braund R, Seeley L, Amiesimaka OI, Schultz M. Adherence to inflammatory bowel disease medications in Southern New Zealand. *Crohn's Colitis* 360. 2021;3(3):otab056. doi: [10.1093/crocol/otab056](https://doi.org/10.1093/crocol/otab056)
7. Coenen S, Weyts E, Ballet V, et al. Identifying predictors of low adherence in patients with inflammatory bowel disease. *Eur*

- J Gastroenterol Hepatol.* 2016;28(5):503-507. doi: [10.1097/MEG.0000000000000570](https://doi.org/10.1097/MEG.0000000000000570)
8. Fidler HH, Singendonk MM, van der Have M, Oldenburg B, van Oijen MG. Low rates of adherence for tumor necrosis factor-alpha inhibitors in Crohn's disease and rheumatoid arthritis: results of a systematic review. *World J Gastroenterol.* 2013;19(27):4344-4350. doi: [10.3748/wjg.v19.i27.4344](https://doi.org/10.3748/wjg.v19.i27.4344)
 9. Severs M, Mangen MJ, Fidler HH, et al. Clinical predictors of future nonadherence in inflammatory bowel disease. *Inflamm Bowel Dis.* 2017;23(9):1568-1576. doi: [10.1097/MIB.0000000000001201](https://doi.org/10.1097/MIB.0000000000001201)
 10. Nahon S, Lahmek P, Saas C, et al. Socioeconomic and psychological factors associated with nonadherence to treatment in inflammatory bowel disease patients: results of the ISSEO survey. *Inflamm Bowel Dis.* 2011;17(6):1270-1276. doi: [10.1002/ibd.21482](https://doi.org/10.1002/ibd.21482)
 11. Bruna-Barranco I, Lue A, Gargallo-Puyuelo CJ, et al. Young age and tobacco use are predictors of lower medication adherence in inflammatory bowel disease. *Eur J Gastroenterol Hepatol.* 2019;31(8):948-953. doi: [10.1097/MEG.0000000000001436](https://doi.org/10.1097/MEG.0000000000001436)
 12. Shah NB, Haydek J, Slaughter J, et al. Risk factors for medication nonadherence to self-injectable biologic therapy in adult patients with inflammatory bowel disease. *Inflamm Bowel Dis.* 2020;26(2):314-320. doi: [10.1093/ibd/izz253](https://doi.org/10.1093/ibd/izz253)
 13. Selinger CP, Eaden J, Jones DB, et al. Modifiable factors associated with nonadherence to maintenance medication for inflammatory bowel disease. *Inflamm Bowel Dis.* 2013;19(10):2199-2206. doi: [10.1097/MIB.0b013e31829ed8a6](https://doi.org/10.1097/MIB.0b013e31829ed8a6)
 14. Rhudy C, Perry CL, Singleton M, Talbert J, Barrett TA. Chronic opioid use is associated with early biologic discontinuation in inflammatory bowel disease. *Aliment Pharmacol Ther.* 2021;53(6):704-711. doi: [10.1111/apt.16269](https://doi.org/10.1111/apt.16269)
 15. Koesmahargyo V, Abbas A, Zhang L, et al. Accuracy of machine learning-based prediction of medication adherence in clinical research. *Psychiatry Res.* 2020;294:113558. doi: [10.1016/j.psychres.2020.113558](https://doi.org/10.1016/j.psychres.2020.113558)
 16. Wang L, Fan R, Zhang C, et al. Applying machine learning models to predict medication nonadherence in crohn's disease maintenance therapy. *Patient Prefer Adherence.* 2020;14:917-926. doi: [10.2147/PPA.S253732](https://doi.org/10.2147/PPA.S253732)
 17. Desai RJ, Wang SV, Vaduganathan M, Evers T, Schneeweiss S. Comparison of machine learning methods with traditional models for use of administrative claims with electronic medical records to predict heart failure outcomes. *JAMA Netw Open.* 2020;3(1):e1918962. doi: [10.1001/jamanetworkopen.2019.18962](https://doi.org/10.1001/jamanetworkopen.2019.18962)
 18. Yerrapragada G, Siadimas A, Babaeian A, Sharma V, O'Neill TJ. Machine learning to predict tamoxifen nonadherence among u commercially insured patients with metastatic breast cancer. *JCO Clin Cancer Inform.* 2021;5:814-825. doi: [10.1200/CCI.20.00102](https://doi.org/10.1200/CCI.20.00102)
 19. MacKay EJ, Stubna MD, Chivers C, et al. Application of machine learning approaches to administrative claims data to predict clinical outcomes in medical and surgical patient populations. *PLoS One.* 2021;16(6):e0252585. doi: [10.1371/journal.pone.0252585](https://doi.org/10.1371/journal.pone.0252585)
 20. *Meritve MarketScan Research Databases.* Ann Arbor, MI: Meritve; 2022.
 21. *International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM).* Hyattsville, MD: National Center for Health Statistics; 2023.
 22. *HCPCS - General Information.* Baltimore, MD: U.S. Centers for Medicare & Medicaid Services; 2024.
 23. *Medi-Span Generic Product Identifier (GPI).* Alphen aan den Rijn, The Netherlands: Wolters Kluwer N.V.; 2024
 24. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis.* 1987;40(5):373-383. doi: [10.1016/0021-9681\(87\)90171-8](https://doi.org/10.1016/0021-9681(87)90171-8)
 25. Glasheen WP, Cordier T, Gumpina R, Haugh G, Davis J, Renda A. Charlson comorbidity index: ICD-9 update and ICD-10 Translation. *Am Health Drug Benefits.* 2019;12(4):188-197.
 26. (HCUP) HCUP. *Clinical Classifications Software Refined (CCSR).* Vol. 2023: Agency for Healthcare Research and Quality; 2022.
 27. Loucks J, Zuckerman AD, Berni A, Saulles A, Thomas G, Alonzo A. Proportion of days covered as a measure of medication adherence. *Am J Health Syst Pharm.* 2021;79(6):492-496. doi: [10.1093/ajhp/zxab392](https://doi.org/10.1093/ajhp/zxab392)
 28. Cramer JA, Roy A, Burrell A, et al. Medication compliance and persistence: terminology and definitions. *Value Health.* 2008;11(1):44-47. doi: [10.1111/j.1524-4733.2007.00213.x](https://doi.org/10.1111/j.1524-4733.2007.00213.x)
 29. *Pharmacy Quality Alliance (PQA): Adherence: PQA Adherence Measures.* Alexandria, VA: Pharmacy Quality Alliance; 2022.
 30. Pedregosa F, Varoquaux G, Gramfort A, et al. Machine learning in python. *J Mach Learn Res.* 2011;12:2825-2830.
 31. Jiang T, Gradus JL, Rosellini AJ. Supervised machine learning: a brief primer. *Behav Ther.* 2020;51(5):675-687. doi: [10.1016/j.beth.2020.05.002](https://doi.org/10.1016/j.beth.2020.05.002)
 32. Fischer JE, Bachmann LM, Jaeschke R. A readers' guide to the interpretation of diagnostic test properties: clinical example of sepsis. *Intensive Care Med.* 2003;29(7):1043-1051. doi: [10.1007/s00134-003-1761-8](https://doi.org/10.1007/s00134-003-1761-8)
 33. Belkin M, Hsu D, Ma S, Mandal S. Reconciling modern machine-learning practice and the classical bias-variance trade-off. *Proc Natl Acad Sci U S A.* 2019;116(32):15849-15854. doi: [10.1073/pnas.1903070116](https://doi.org/10.1073/pnas.1903070116)
 34. Vadhariya A, Fleming ML, Johnson ML, et al. Group-Based trajectory models to identify sociodemographic and clinical predictors of adherence patterns to statin therapy among older adults. *Am Health Drug Benefits.* 2019;12(4):202-211.
 35. Cai Q, Ding Z, Fu AZ, Patel AA. Racial or ethnic differences on treatment adherence and persistence among patients with inflammatory bowel diseases initiated with biologic therapies. *BMC Gastroenterol.* 2022;22(1):545. doi: [10.1186/s12876-022-02560-y](https://doi.org/10.1186/s12876-022-02560-y)
 36. Bernstein CN, Walld R, Marrie RA. Social Determinants of outcomes in inflammatory bowel disease. *Am J Gastroenterol.* 2020;115(12):2036-2046. doi: [10.14309/ajg.0000000000000794](https://doi.org/10.14309/ajg.0000000000000794)
 37. Wilder ME, Kulie P, Jensen C, et al. The impact of social determinants of health on medication adherence: a systematic review and meta-analysis. *J Gen Intern Med.* 2021;36(5):1359-1370. doi: [10.1007/s11606-020-06447-0](https://doi.org/10.1007/s11606-020-06447-0)
 38. Zhang NJ, Terry A, McHorney CA. Impact of health literacy on medication adherence: a systematic review and meta-analysis. *Ann Pharmacother.* 2014;48(6):741-751. doi: [10.1177/1060028014526562](https://doi.org/10.1177/1060028014526562)
 39. Gohil S, Majd Z, Sheneman JC, Abughosh SM. Interventions to improve medication adherence in inflammatory bowel disease: a systematic review. *Patient Educ Couns.* 2022;105(7):1731-1742. doi: [10.1016/j.pec.2021.10.017](https://doi.org/10.1016/j.pec.2021.10.017)
 40. Rubin DT, Mittal M, Davis M, Johnson S, Chao J, Skup M. Impact of a patient support program on patient adherence to adalimumab and direct medical costs in crohn's disease, ulcerative colitis, rheumatoid arthritis, psoriasis, psoriatic arthritis, and ankylosing spondylitis. *J Manag Care Spec Pharm.* 2017;23(8):859-867. doi: [10.18553/jmcp.2017.16272](https://doi.org/10.18553/jmcp.2017.16272)