Predicting Hospitalization and Outpatient Corticosteroid Use in Inflammatory Bowel Disease Patients Using Machine Learning

Akbar K. Waljee, MD, MS[*,†,‡], **Rachel Lipson, MSc**[*], **Wyndy L. Wiitala, PhD**[*], **Yiwei Zhang, PhD, MS**[§], **Boang Liu, MS**[§], **Ji Zhu, PhD**[§], **Beth Wallace, MD**[‡,¶], **Shail M. Govani, MD, MS**[†,‡], **Ryan W. Stidham, MD, MS**[†], **Rodney Hayward, MD**[*,‡,‖], and **Peter D. R. Higgins, MD, PhD, MSc**[†]

[*]VA Center for Clinical Management Research, VA Ann Arbor Health Care System, Ann Arbor, Michigan

[†]Division of Gastroenterology and Hepatology, University of Michigan Health System, Ann Arbor, Michigan

[¶]Division of Rheumatology, University of Michigan Health System, Ann Arbor, Michigan

[‖]Division of General Medicine, Department of Internal Medicine, University of Michigan Health System, Ann Arbor, Michigan

[‡]University of Michigan Medical School, Institute for Healthcare Policy and Innovation, Ann Arbor, Michigan

[§]Department of Statistics, University of Michigan, Ann Arbor, Michigan

Abstract

Background—Inflammatory bowel disease (IBD) is a chronic disease characterized by unpredictable episodes of flares and periods of remission. Tools that accurately predict disease course would substantially aid therapeutic decision-making. This study aims to construct a model that accurately predicts the combined end point of outpatient corticosteroid use and hospitalizations as a surrogate for IBD flare.

Methods—Predictors evaluated included age, sex, race, use of corticosteroid-sparing immunosuppressive medications (immunomodulators and/or anti-TNF), longitudinal laboratory data, and number of previous IBD-related hospitalizations and outpatient corticosteroid prescriptions. We constructed models using logistic regression and machine learning methods

Address correspondence to: Akbar K. Waljee, MD, MS, 2215 Fuller Road, Gastroenterology 111D, Ann Arbor, MI 48105 (awaljee@med.umich.edu).

(random forest [RF]) to predict the combined end point of hospitalization and/or corticosteroid use for IBD within 6 months.

Results—We identified 20,368 Veterans Health Administration patients with the first (index) IBD diagnosis between 2002 and 2009. Area under the receiver operating characteristic curve (AuROC) for the baseline logistic regression model was 0.68 (95% confidence interval [CI], 0.67–0.68). AuROC for the RF longitudinal model was 0.85 (95% CI, 0.84–0.85). AuROC for the RF longitudinal model using previous hospitalization or steroid use was 0.87 (95% CI, 0.87–0.88). The 5 leading independent risk factors for future hospitalization or steroid use were age, mean serum albumin, immunosuppressive medication use, and mean and highest platelet counts. Previous hospitalization and corticosteroid use were highly predictive when included in specified models.

Conclusions—A novel machine learning model substantially improved our ability to predict IBD-related hospitalization and outpatient steroid use. This model could be used at point of care to distinguish patients at high and low risk for disease flare, allowing individualized therapeutic management.

Keywords

inflammatory bowel disease; corticosteroids; complications

Inflammatory bowel disease (IBD) is a chronic, often debilitating idiopathic disease that affects more than 1.5 million people in the United States.[1–3] IBD is also a very costly disease. Data from 2008 suggest that the total cost of IBD treatment in the United States is $6.3 billion, the majority of which is associated with outpatient pharmaceutical costs and hospitalizations.[4] This pattern is also seen in other inflammatory diseases such as rheumatoid arthritis, lupus, and multiple sclerosis. These diseases are typically diagnosed relatively early in life and require lifelong treatment, with patients frequently suffering disease exacerbations resulting in substantial morbidity, premature mortality, and high lifetime costs.

The course of IBD is characterized by episodic acute flares with intervening periods of remission, making therapeutic decision-making quite complex. Without the ability to accurately predict future flares, many patients suffer disabling or even fatal disease exacerbations when earlier stepping up of treatment could have kept them in remission, while other patients undergo long periods of ineffective or unnecessary maintenance therapy when they might have benefited from an opportunity to step down therapy.[5, 6]

Tools that more accurately predict the disease course and offer advice on appropriate treatment could substantially improve the decision-making process. Such tools would allow targeted intervention in patients at the highest risk of disease exacerbations and complications, while simultaneously reducing treatment-related adverse events and improving cost-effectiveness by de-escalating treatment in patients at low risk of disease exacerbation or complications.[7] Recent studies of biomarkers for subclinical IBD activity have shown promise as predictors of disease exacerbations.[8–10] The most commonly used biomarkers are stool tests that assess intestinal inflammation. Some of these tests, such as

fecal leukocytes, are widely available and inexpensive but have limited accuracy.[11, 12] Fecal calprotectin has the most robust data supporting its use as a predictor of impending disease flare in IBD patients during the subsequent 3 months,[10] with stool test sensitivity of 77% and specificity of 71% on the basis of results of a recent meta-analysis.[13] However, it is expensive, not universally available, and its ability to predict 6-month risk of flares is limited.[10, 13] Serum markers as well as fecal excretions of various other serum proteins have also been evaluated, but they are limited by accuracy or cost barriers.[14]

This study aims to develop a model using longitudinal data routinely available in the electronic medical record to predict corticosteroid use and hospitalizations as surrogates for clinically meaningful flares among patients with IBD. Additional sensitivity analyses were conducted (1) examining prediction in ulcerative colitis (UC), Crohn's disease (CD), and indeterminate colitis (IC) patients separately; (2) excluding immunosuppressive (escalation) medications as a predictor; (3) using only outpatient corticosteroid use as the outcome; (4) using 12-month outcomes; and (5) a longitudinal logistic regression model.

METHODS

Overview

We used an institutional review board (IRB)–approved, national Veteran's Health Administration (VHA) electronic database to conduct a retrospective cohort study of patients with IBD. Patients were identified using previously validated algorithms based on a combination of inpatient and outpatient International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) codes for Crohn's disease (555.x), and ulcerative colitis (556.x).[15] Patients were selected for inclusion if they had 2 or more of these ICD-9 codes during at least 2 clinical encounters between 2002 and 2009, with at least 1 encounter being an outpatient visit. This approach has a positive predictive value for Crohn's disease of 0.84 and a positive predictive value for ulcerative colitis of 0.91 in the VHA.[15] Patients were classified as CD if all ICD-9 codes were 555.x, UC if all codes were 556.x, and IC otherwise. Patients identified as having IBD by this algorithm were monitored for IBD-related hospitalizations and outpatient corticosteroid use from 2002 to 2010.

Definition of Outcomes

The primary outcome was a composite measure capturing both use of outpatient corticosteroids prescribed for IBD and inpatient hospitalizations associated with a diagnosis of IBD. We extracted data on filled prescriptions for outpatient oral corticosteroids using generic names (Supplement 1) from the VHA Decision Support Systems (DSS) National Data Extraction data source. We determined the indication for corticosteroids by searching for ICD-9 diagnosis codes for a variety of common inflammatory comorbid conditions in the 7 days prior to the prescription fill date (Supplement 2) and excluded corticosteroid fills associated with these non-IBD diagnoses. Additional prescription exclusions included fills where the day supply was fewer than 7 days, including all Medrol dose packs (which are by definition a 5-day supply), and subjects with a positive *Clostridium difficile* diagnosis with vancomycin or metronidazole given 5–7 days after the stool test. Hospitalizations were coded based on an inpatient admission associated with an ICD-9 code of 555.x or 556.x and

a corticosteroid fill during the hospitalization. Outpatient corticosteroid use or hospitalization was assumed to be part of a previous treatment course if it occurred within 90 days of the previous hospitalization or corticosteroid prescription.

Predictor Variables

Predictor variables included patient age, sex, race, number of previous hospitalizations or corticosteroid prescriptions, use of immunosuppressive medication (immunomodulator and/or anti-TNF), and lab results derived from the CBC (with automated differential), chemistries, sedimentation rate (ESR), and C-reactive protein (CRP) serum values. Fecal calprotectin results were available for less than 1% of the study subjects, and therefore not used as a predictor (Table 1). Lab results were obtained from the VHA Corporate Data Warehouse LabChem tables. We used a combination of Logical Observation Identifiers Names and Codes (LOINC)[16–21] and test names to extract the relevant measures and excluded values that were out of the appropriate range.

No a priori variable selection was done for the Random Forests, as measuring variable importance is inherent to the methodology. A subselection of variables is chosen at each node, the most predictive variable of those is chosen, and then the process is repeated randomly for each node in each tree grown—in this case, 500. This prevents the strong predictors from dominating and allows the estimation of the effect of weak learners in an unbiased fashion. The same predictors were included in the logistic model for consistency between methods.

Missing Covariates

Missing lab covariate values were imputed based on the median value of the lab from all the previous visits. Patients missing more than 50% of lab data were excluded from analysis.

Statistical Analysis and Model Development

We developed 3 models to predict hospitalization and steroid use: (1) a logistic regression (LR) model using baseline data, (2) a random forest (RF) model using longitudinal summary variables, and (3) an RF model using longitudinal variables including the number of previous hospitalizations and corticosteroid prescriptions. The LR baseline model included data collected at each visit to predict outcomes in the 6 months following the visit (Fig. 1A).

In contrast to the baseline LR model, the longitudinal RF models incorporated predictor variables that included both longitudinal summaries of previous lab values and lab values for the current visit (Fig. 1B). Longitudinal summaries for lab values included the mean and maximum of all previously observed values, the mean of the differential (mean of the difference between sequential observed values divided by the sequential observation time, ie, the average slope), the maximum of the differential (maximum of the difference between sequential observed values divided by the sequential observation time), and the mean of acceleration (mean of the difference between sequential differential observed values divided by the difference between sequential differential observation time, ie, Dx/Dt) (Table 1). Longitudinal summaries were only calculated for patients with more than 1 visit, and acceleration was only calculated for patients with more than 2 visits. The outcome for each

model was defined on a per-visit level (any laboratory draw or clinical visit), with each patient contributing predictor variables from multiple visits. At the time of each visit, a prediction was made as to outcome status (a binary indicator of whether a hospitalization or steroid prescription occurred) within the next 6 months. A single event might be predicted by multiple visits, should they all fall within the same 6-month interval.

Development of a Logistic Regression Model

We first developed the predictive logistic regression model for the risk of hospitalization and steroid use during the 6-month window following every visit. This model was created using baseline visit predictor variables, as defined in Table 1. This model evaluates the predictive accuracy of using a single set of laboratory results in combination with age, sex, race, and use of immunosuppressive medications.

Development of Random Forest Machine Learning Models

Random forest is an ensemble method of prediction using decision trees.[22, 23] To classify a new observation, the observation is run through each of the trees in the forest. Each tree provides a classification (vote), and the forest combines the votes over all the trees to compute a predicted score of the outcome. Using this random forest method, baseline variables and summarized longitudinal values are put into the model as predictors. A primary model with these predictors was fit, and a "forest" of 500 trees was grown to produce the predictions. Each tree is grown using a sample of observations from the data with replacement, effectively a bootstrap sample. We also fit a second RF model that included the number of previous outcomes (ie, hospitalizations and/or outpatient corticosteroid prescriptions) as a predictor.

Variable importance—The relative importance of each predictor variable was determined by identifying nodes in the ensemble of trees in which the individual predictor variable appeared and summing the relative information content provided by all the nodes containing that variable. Predictor variables that provide the greatest combined discrimination have a higher importance.

Training and testing cohorts—To validate the predictive ability of the data, visits were split on a per-patient basis into training and testing sets, with 70% of a patient's visits in the training set and 30% reserved for the testing set. The training and testing split was based on the visit time; that is, different observations corresponding to the same individual were split into training and testing, such that earlier visits belong to the training set and later visits belong to the testing set. Patients with only 1 visit (n = 1625) were not included in the testing set; as a result, they were also not included in the sensitivity analyses.

Model performance—An optimal risk cutoff was identified to maximize the model sensitivity and specificity and its associated area under the receiver operating characteristic curve (AuROC). Brier scores, which capture both calibration and discrimination, are also reported as an overall measure of model performance. Brier scores can range from 0 to 1, with lower scores being consistent with increased accuracy and better model performance.

All statistical methods were performed using the statistical language R (version 3.3), with the packages *randomForest* and *gbm* (by Y.Z. and J.Z.)[22, 23] Two-sided *P* values <0.05 were considered statistically significant.

Sensitivity analysis—Additional sensitivity analyses were performed to evaluate the robustness of the models. These analyses evaluated: (1) the model in the subsets of UC, CD, and IC patients, (2) a model excluding the use of immunosuppressive medications as a predictor, (3) a model using only corticosteroid prescriptions without hospitalization as the outcome, (4) a model with a 12-month outcome window to determine the effect of a longer follow-up time on the predictive ability of the model, and (5) a logistic regression model using longitudinal data to predict the 6-month outcome.

RESULTS

Cohort and Demographics

Our initial cohort consisted of 30,456 patients with an index IBD diagnosis between 2002 and 2009. Patients were excluded if they did not have any visit data for at least 1 year after the index IBD diagnosis (n = 1794), if they were missing more than 50% of the predictor laboratory values before imputation (n = 8212), and if they were on a steroid or were an inpatient at the initiation of the study (n = 82). Our final cohort consisted of 20,368 patients and 351,112 visits. A majority of the patients had UC (52.8%) and were male (93.3%) and Caucasian (70.9%) (Table 2). Of patients in the final cohort, 4610 (22.6%) had at least 1 qualifying outpatient corticosteroid prescription or inpatient hospitalization between 2002 and 2010 and were deemed to have met the primary outcome. Of those 4610 patients, 3888 (19.1% of the cohort) had at least 1 visit within the 6 months prior to an outcome that was predicted by the model. A total of 8441 unique hospitalizations and outpatient steroid prescriptions were predicted by 38,112 visits. The average yearly outcome rate (defined as the number of hospitalizations or steroid prescriptions per year out of active patients in the cohort) was 11.5% for all outcomes. The yearly average rate of events predicted by the model was 6.5% as these outcomes were associated with a preceding visit. The median time from the first visit in a 6-month window to an event was 32 days (interquartile range [IQR], 8–64 days), suggesting an opportunity for therapeutic intervention. The median follow-up time was 67.48 months (IQR, 40.15–89.15 months).

Predicting Hospitalizations and Corticosteroid Prescriptions

The AuROC results for the 3 models on the testing cohort are displayed in Figure 2 and show the model accuracy in predicting IBD hospitalizations and steroid prescriptions using the baseline regression model compared with the longitudinal random forest models. For the baseline regression model, the AuROC was 0.68 (95% CI, 0.67–0.68). The AuROC for the primary RF longitudinal model was 0.85 (95% CI, 0.84–0.85), and the AuROC for the RF longitudinal model using previous hospitalization or steroid use as predictors was 0.87 (95% CI, 0.87–0.88). The variable importance graphs for the longitudinal models are shown in Supplement 3A and B. In the primary RF longitudinal model, the 5 strongest predictors of a future event were the patient's age, mean albumin, any prior use of an immunosuppressive medication, the mean platelet value, and the highest platelet count (Supplement 3A). When

included in the model, the number of previous hospitalizations or corticosteroid prescriptions was the strongest predictor of the primary outcome (Supplement 3B). Previous hospitalizations and steroid prescriptions were calculated as the number of previous outcomes at the time of a given visit. An alternative model including previous events as a binary variable demonstrated no difference in predictive ability, and so previous events was retained as a count variable. Table 3 shows the proportion of patient visits correctly classified by each model as high vs low risk of steroid prescription or hospitalization, and the associated Brier score. The RF longitudinal models predicted the primary outcome with 74%–80% sensitivity and 80%–82% specificity, while the baseline logistic regression model had only 64% sensitivity and specificity.

Sensitivity Analysis

UC, CD, and IC patients—RF model discrimination was similarly high when predicting outcomes for patients with ulcerative colitis, Crohn's disease, or indeterminate colitis separately. Adding the number of previous hospitalizations or steroid prescriptions ("previous events") as a predictor improved discrimination in all 3 models (Table 4). Specifically, in the RF model *without* the previous events predictor, the AuROC was 0.84 (95% CI, 0.83–0.85) for CD (n = 6448), 0.85 (95% CI, 0.84–0.86) for UC (n = 9863), and 0.82 (95% CI, 0.81– 0.83) for IC (n = 2432). Adding the previous outcome predictor to the models improved the AuROCs to 0.87 (95% CI, 0.87–0.88) for CD, to 0.88 (95% CI, 0.87– 0.88) for UC, and to 0.85 (95% CI, 0.84–0.86) for IC.

Excluding use of an immunosuppressive (escalation) medication as a predictor—Given the concern that use of an immunosuppressive medication may be a surrogate for a future hospitalization or steroid prescription, we performed an additional sensitivity analysis by constructing 2 models without this variable. One model excluded the number of previous hospitalization or steroid prescriptions, and the second model included the number of previous events. The AuROCs were still excellent at 0.84 (95% CI, 0.84– 0.85) and 0.87 (95% CI, 0.87–0.88), respectively (Table 4).

Predicting outpatient corticosteroid use only—Given that hospitalizations for IBD flares are infrequent (15.4% of outcomes were hospitalizations) and most patients are treated on an outpatient basis (84.6% of outcomes were treated with outpatient steroid fills), we evaluated whether the models would perform equally well for predicting an outcome of outpatient corticosteroid use only. This sensitivity analysis was performed with and without the immunosuppressive medication predictor and with and without the number of previous events predictor, producing 4 models. The AuROC results for the longitudinal models including the immunosuppressive medication predictor were (1) 0.90 (95% CI, 0.89–0.90) for the model with the previous event predictor and (2) 0.86 (95% CI, 0.86–0.87) for the model without the previous event predictor. The AuROC results for the longitudinal models without the immunosuppressive medication predictor were (1) 0.90 (95% CI, 0.89–0.90) for the model with the previous event predictor and (2) 0.86 (95% CI, 0.86–0.87) for the model without the previous event predictor (Table 4).

12-month outcome—In some cases, it would be clinically useful to have more durable predictions and to predict who might have a hospitalization or steroid fill within 12 months, so an additional sensitivity analysis was performed to evaluate this outcome. The longitudinal models performed equally well for 12-month outcomes. The AuROC results for the longitudinal model with the immunosuppressive medication predictor were (1) 0.90 (95% CI, 0.89–0.90) for the model with the previous event predictor and (2) 0.88 (95% CI, 0.88–0.89) for the model without the previous event predictor. The AuROC results for the longitudinal models without the immunosuppressive medication predictor were (1) 0.90 (95% CI, 0.89–0.90) for the model with the previous event predictor and (2) 0.88 (95% CI, 0.88–0.89) for the model without the previous event predictor (Table 4).

Longitudinal logistic regression model—In order to evaluate whether longitudinal prediction models perform equally well when traditional logistic regression is used, we evaluated the primary 6-month outcome of outpatient corticosteroid use and hospitalizations using a longitudinal logistic regression model. The AuROC results for the longitudinal logistic regression model with the immunosuppressive medication predictor were (1) 0.79 (95% CI, 0.79–0.80) for the model with the previous event predictor and (2) 0.70 (95% CI, 0.69–0.71) for the model without the previous event predictor. The AuROC results for the longitudinal logistic regression model without the immunosuppressive medication predictor were (1) 0.79 (95% CI, 0.79–0.80) for the model with the previous event predictor and (2) 0.68 (95% CI, 0.67–0.68) for the model without the previous event predictor (Table 4).

DISCUSSION

In this study, we demonstrated that a random forest prediction model incorporating longitudinal data readily available within the electronic medical record has excellent discrimination for risk of IBD-related hospitalization or steroid use in the next 6 months, outperforming fecal calprotectin. When integrated into a computerized decision aid, this model offers a simple, inexpensive tool that can be used at the point of care by clinicians to tailor IBD treatment, as well as by health systems or clinical coordinators for systems approaches to quality improvement. Such a tool has the potential to improve outcomes and reduce cost by both avoiding undertreatment of high-risk patients and reducing adverse events and waste related to overtreatment of low-risk patients.

Developing tools and decision support systems to guide clinicians in personalizing medical decision-making for patients with IBD has a particular application for integrated health care systems like the VHA, because having an IBD subspecialist provider at every facility is not feasible. However, our model is broadly applicable to any system working to provide a "targeted" or "tailored" prevention approach to risk stratifying individuals for disease exacerbation and treatment. Having risk stratification tools developed and validated within health care systems is an important first step toward realizing efficient patient-centered care in health care.

The model presented here is of great clinical interest because it outperforms traditional logistic regression models, yet requires only data readily available in the electronic medical record. While the techniques required are less conventional than logistic regression models

and may be less familiar to the clinician, prediction models based on RF methods can be implemented with no more difficulty than those based on any other statistical methodology. In either case, the mathematics of the prediction tool are not directly encountered by the clinician, but integrated into a computerized decision aid. In fact, because all data needed for the longitudinal RF model can be automatically extracted from the electronic medical record, using this model can reduce clinician time and effort at the point of care relative to a model requiring a provider to manually input clinical information. This approach is exemplified by our previously developed prediction models that have been incorporated at the University of Michigan to predict responses to thiopurine medications for IBD.[24, 25] Another clinical advantage of machine learning models is how easily they can be updated with new data and predictors, compared with models based on traditional regression techniques.

Our model also has obvious implications for the research setting. The ability to predict future IBD disease activity would assist greatly with selecting appropriate patients for clinical trials and other prospective studies. Our RF algorithm also uses information that is generally readily available from longitudinal administrative databases. If validated for use in this setting, our model could generate information about disease activity from such databases, addressing a major limitation facing most studies using these data sources.

There are several limitations to our study. As this study was performed using VHA data, its findings may not be generalizable to the overall US population due to demographic differences. In the US overall, the median age of diagnosis of IBD is approximately 30 years, with a slight male predominance,[26] whereas the VHA has an older population of veterans who are predominantly male. It is important to note that sex did not significantly influence the prediction model in this population, but this may change in a population with more gender diversity. Another consideration is that hospitalization and steroid usage rates may be lower in our cohort than they would be in a tertiary care population, as our population represents a more diverse community practice. While sensitivity analysis did show that our model performed well under a variety of different assumptions and subsets of patients in the VHA, further testing to validate it in a nationally representative patient cohort will therefore be necessary before it can be widely used. Additionally, it is possible that we are missing data on veterans who receive care that occurs outside of the VHA, that is, by veterans who purchase commercial insurance or are eligible for Medicare. Studies of patients who have access to VHA and civilian health care systems reveal that more than 60% utilize specialty care outside of the VHA.[27] While these patients may seek care outside of the VHA system, we postulate that this would only underestimate their event rates, biasing our results to the null. Furthermore, we believe many veterans with specialist care outside the VHA would nonetheless continue to obtain prescriptions through the VHA due to cheaper copayments for medications.[28] Another limitation is the lack of fecal calprotectin results in the VHA system. We were not able to include this laboratory value as a predictor in this study population; however, the high accuracy of predicting outcomes in our model using routine labs and information in the EMR shows the strength of using machine learning algorithms rather than individual predictors. Lastly, as in any predictive model, there is the potential for overfitting. While the usage of a large data set split into a training and a testing

set should allow for accurate assessment of model performance, it will again be important to assess the predictive value of this model in populations with different patient characteristics.

CONCLUSION

In summary, predicting the course of disease among patients with IBD is challenging, but tools to help target individual therapy by predicting flares have the potential to improve outcomes and reduce costs. We used novel machine learning techniques to create a predictive model that assimilates clinical and laboratory predictors readily available in the electronic medical record to accurately predict IBD-related hospitalizations and outpatient steroid use over both 6-month and 12-month time frames. Once validated for general use, such a model has far-reaching clinical and research implications; it would be easy to implement at the point of care to individualize and tailor therapeutic regimens, as well as having clear applications for enrollment in clinical trials, evaluating the impact of QI interventions, and evaluating disease activity in administrative databases. Additional prospective studies are needed to validate the utility of this model for general use.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Kappelman MD, Rifas-Shiman SL, Kleinman K, et al. The prevalence and geographic distribution of Crohn's disease and ulcerative colitis in the United States. Clin Gastroenterol Hepatol. 2007; 5:1424–9. [PubMed: 17904915]

2. Loftus EV. The burden of inflammatory bowel disease in the United States: a moving target? Clin Gastroenterol Hepatol. 2007; 5:1383–4. [PubMed: 18054749]

3. Loftus EV. Clinical epidemiology of inflammatory bowel disease: incidence, prevalence, and environmental influences. Gastroenterology. 2004; 126:1504–17. [PubMed: 15168363]

4. Kappelman MD, Rifas-Shiman SL, Porter CQ, et al. Direct health care costs of Crohn's disease and ulcerative colitis in US children and adults. Gastroenterology. 2008; 135:1907–13. [PubMed: 18854185]

5. Casellas F, Arenas JI, Baudet JS, et al. Impairment of health-related quality of life in patients with inflammatory bowel disease: a Spanish multicenter study. Inflamm Bowel Dis. 2005; 11:488–96. [PubMed: 15867589]

6. Ananthakrishnan AN, Weber LR, Knox JF, et al. Permanent work disability in Crohn's disease. Am J Gastroenterol. 2008; 103:154–61. [PubMed: 18076736]

7. Saini SD, Waljee AK, Higgins PDR. Cost utility of inflammation-targeted therapy for patients with ulcerative colitis. Clin Gastroenterol Hepatol. 2012; 10:1143–51. [PubMed: 22610010]

8. Consigny Y, Modigliani R, Colombel J-F, et al. A simple biological score for predicting low risk of short-term relapse in Crohn's disease. Inflamm Bowel Dis. 2006; 12:551–7. [PubMed: 16804391]

9. Kopylov U, Rosenfeld G, Bressler B, et al. Clinical utility of fecal biomarkers for the diagnosis and management of inflammatory bowel disease. Inflamm Bowel Dis. 2014; 20:742–56. [PubMed: 24562174]

10. Gisbert JP, Bermejo F, Pérez-Calle J-L, et al. Fecal calprotectin and lactoferrin for the prediction of inflammatory bowel disease relapse. Inflamm Bowel Dis. 2009; 15:1190–8. [PubMed: 19291780]

11. Schoepfer AM, Trummler M, Seeholzer P, et al. Accuracy of four fecal assays in the diagnosis of colitis. Dis Colon Rectum. 2007; 50:1697–706. [PubMed: 17762964]

12. Seva-Pereira A, Franco AO, de Magalhães AF. Diagnostic value of fecal leukocytes in chronic bowel diseases. Sao Paulo Med J. 1994; 112:504–6. [PubMed: 7871315]

13. Mao R, Xiao Y-L, Gao X, et al. Fecal calprotectin in predicting relapse of inflammatory bowel diseases: a meta-analysis of prospective studies. Inflamm Bowel Dis. 2012; 18:1894–9. [PubMed: 22238138]

14. Pardi DS, Sandborn WJ. Predicting relapse in patients with inflammatory bowel disease: what is the role of biomarkers? Gut. 2005; 54:321–2. [PubMed: 15710974]

15. Hou JK, Tan M, Stidham RW, et al. Accuracy of diagnostic codes for identifying patients with ulcerative colitis and Crohn's disease in the Veterans Affairs Health Care System. Dig Dis Sci. 2014; 59:2406–10. [PubMed: 24817338]

16. Kim H, El-Kareh R, Goel A, et al. An approach to improve LOINC mapping through augmentation of local test names. J Biomed Inform. 2012; 45:651–7. [PubMed: 22210167]

17. Lin MC, Vreeman DJ, McDonald CJ, et al. A characterization of local LOINC mapping for laboratory tests in three large institutions. Methods Inf Med. 2011; 50:105–14. [PubMed: 20725694]

18. Khan AN, Griffith SP, Moore C, et al. Standardizing laboratory data by mapping to LOINC. J Am Med Inform Assoc. 2006; 13:353–5. [PubMed: 16501183]

19. McDonald CJ, Huff SM, Suico JG, et al. LOINC, a universal standard for identifying laboratory observations: a 5-year update. Clin Chem. 2003; 49:624–33. [PubMed: 12651816]

20. Huff SM, Rocha RA, McDonald CJ, et al. Development of the Logical Observation Identifier Names and Codes (LOINC) vocabulary. J Am Med Inform Assoc. 1998; 5:276–92. [PubMed: 9609498]

21. Forrey AW, McDonald CJ, DeMoor G, et al. Logical observation identifier names and codes (LOINC) database: a public use set of codes and names for electronic reporting of clinical laboratory test results. Clin Chem. 1996; 42:81–90. [PubMed: 8565239]

22. Liaw A, Wiener M. Classification and regression by randomForest. R news. 2002

23. Breiman L. Random forests. Machine learning. 2001; 45:5–32.

24. Waljee AK, Joyce JC, Wang S, et al. Algorithms outperform metabolite tests in predicting response of patients with inflammatory bowel disease to thiopurines. Clin Gastroenterol Hepatol. 2010; 8:143–50. [PubMed: 19835986]

25. Waljee AK, Sauder K, Patel A, et al. Machine learning algorithms for objective remission and clinical outcomes with thiopurines. J Crohns Colitis. 2017; 11(7):801–10. [PubMed: 28333183]

26. Loftus CG, Loftus EVJ, Harmsen WS, et al. Update on the incidence and prevalence of Crohn's disease and ulcerative colitis in Olmsted County, Minnesota, 1940–2000. Inflamm Bowel Dis. 2007; 13:254–61. [PubMed: 17206702]

27. Liu CF, Chapko M, Bryson CL, et al. Use of outpatient care in Veterans Health Administration and Medicare among veterans receiving primary care in community-based and hospital outpatient clinics. Health Serv Res. 2010; 45:1268–86. [PubMed: 20831716]

28. Waljee AK, Wiitala WL, Govani S, et al. Corticosteroid use and complications in a US inflammatory bowel disease cohort. PLoS One. 2016; 11:e0158017. [PubMed: 27336296]

A From each visit look within 6 months for a hospitalization or corticosteroid fill

Month 0   6 Months   12 Months

◇ Visit

● Event

No outcome associated with visit

⌐‾‾‾‾‾‾‾‾‾‾‾¬ 6 Month window from visit

B Lab data summarized to predict 1st event

Lab data summarized to predict 2nd event

Month 0   6 Months   12 Months

◇ Visit

● Event

[#] Original lab data included in model to predict upcoming event

FIGURE 1.
A, Schematic of how baseline data predict IBD flares. Each visit looks forward 6 months to ascertain whether an outcome occurred. (A) demonstrates 3 visits whose lab data then predicts the same event. Three observations are added to the model with a positive outcome. One observation without an outcome is also included in the diagram. B, Schematic of how longitudinal data predicts IBD flares. All lab data for visits preceding an event are summarized and included in the model as predictor variables. The original lab data at the visit closest to the event are included in addition to the summarized measures.
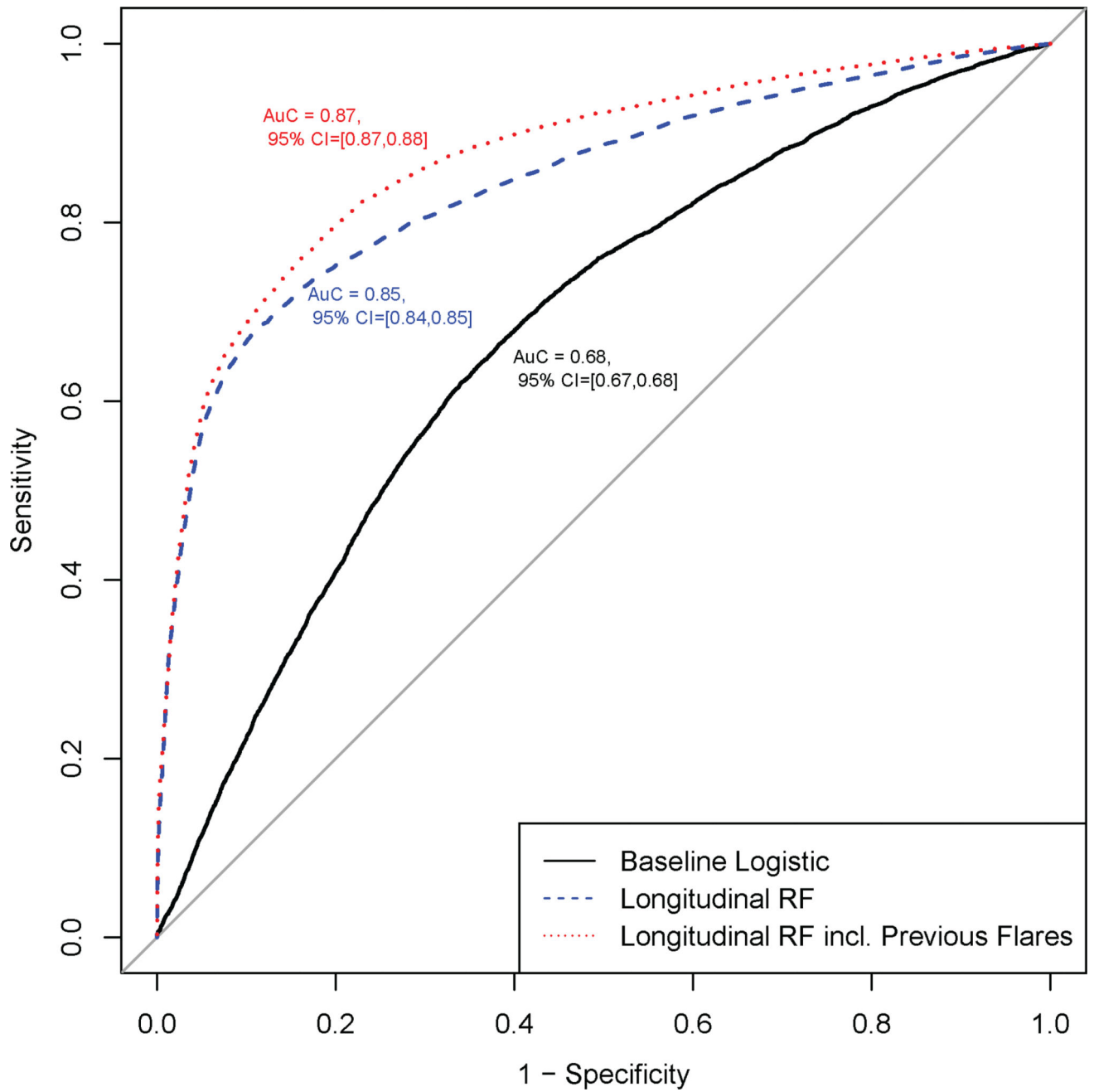
FIGURE 2.
Area under the receiver operating characteristic curve of models.

TABLE 1

Patient Variables

Variables included in both models	**Demographics:** age, sex, race.
	Labs: white blood cell count (WBC), hemoglobin, hematocrit (HCT), mean corpuscular volume (MCV), mean corpuscular hemoglobin concentrate (MCHC), platelets, sodium, potassium, glucose, blood urea nitrogen (BUN), serum creatinine, calcium, bicarbonate, chloride, albumin, aspartate aminotransferase (AST), alanine aminotransferase (ALT), total protein, alkaline phosphatase, bilirubin.
	Immunosuppressive (escalation) med: included any immunosuppressive medication such as thiopurine, methotrexate, anti-TNF, or combination therapy.
Additional variables for longitudinal model	**Previous hospitalization or steroid prescription:** number of previous outpatient corticosteroids or hospitalizations.
	Calculated: mean, maximum, mean of the differential, maximum of the differential, and mean of acceleration of labs.

TABLE 2

Patient Characteristics With IBD

	All	No Event (Hospitalization or Steroid Prescription)	Had Event
	N = 20,368	N = 16,480	N = 3888
Age, mean ± SD	59.1 ± 14.9	61.0 ± 14.1	53.2 ± 16.2
Male, No. (%)	19,003 (93.3)	15,417 (93.6)	3586 (92.2)
Race, No. (%)			
Caucasian	14,443 (70.9)	11,636 (70.6)	2807 (72.2)
African American	1637 (8.0)	1215 (7.4)	422 (10.9)
Other	356 (1.7)	271 (1.6)	85 (2.2)
Unknown or missing	3932 (19.3)	3358 (20.4)	574 (14.8)
Disease type, No. (%)			
Crohn's disease	7052 (34.6)	5667 (34.4)	1385 (35.6)
Ulcerative colitis	10,762 (52.8)	9178 (55.7)	1584 (40.7)
Indeterminate disease	2554 (12.5)	1635 (9.9)	919 (23.6)
Immunosuppressive medication use, No. (%)	3740 (18.4)	2036 (12.4)	1704 (43.8)
Nonevent visits per patient, median (IQR)	10 (2–21)	9 (4–18)	17 (7–35)
Follow-up time, median (IQR), mo	67.5 (40.1–89.1)	67.28 (40.2–89.1)	66.7 (39.7–89.1)

TABLE 3

Misclassification Table for the Models: Testing Cohort

Clinical Outcome		Visits With Event (%)		Nonevent Visits (%)		
Prediction Model	Cutoff	Predicted Event	Predicted Nonevent	Predicted Event	Predicted Nonevent	Brier Score
Logistic regression	0.111	5043 (64.1)	2830 (35.9)	35,259 (36.0)	62,586 (64.0)	0.36
Longitudinal random forest	0.143	5814 (73.8)	2059 (26.2)	17,395 (17.8)	80,450 (82.2)	0.18
Longitudinal random forest with previous events	0.143	6275 (79.7)	1598 (20.3)	19,555 (20.0)	78,290 (80.0)	0.20

TABLE 4

Sensitivity Ana

	With Predictor for Number of Previous Hospitalizations/ Steroid Prescriptions	Without Predictor for Number of Previous Hospitalizations/Steroid Prescriptions
	AuROC (95% CI)	AuROC (95% CI)
CD patients only (n = 6448)	0.87 (0.87–0.88)	0.84 (0.83–0.85)
UC patients only (n = 9863)	0.88 (0.87–0.88)	0.85 (0.84–0.86)
IC patients only (n = 2432)	0.85 (0.84–0.86)	0.82 (0.81–0.83)
Without predictor for immunosuppressive medication	0.87 (0.87–0.88)	0.84 (0.84–0.85)
Outcome: outpatient steroid prescriptions only (6 mo)		
With predictor for immunosuppressive medication	0.90 (0.89–0.90)	0.86 (0.86–0.87)
Without predictor for immunosuppressive medication	0.90 (0.89–0.90)	0.86 (0.86–0.87)
Outcome: any hospitalization or steroid prescription (12 mo)		
With predictor for immunosuppressive medication	0.90 (0.89–0.90)	0.88 (0.88–0.89)
Without predictor for immunosuppressive medication	0.90 (0.89–0.90)	0.88 (0.88–0.89)
Longitudinal logistic regression model		
With predictor for immunosuppressive medication	0.79 (0.79–0.80)	0.70 (0.69–0.71)
Without predictor for immunosuppressive medication	0.79 (0.79–0.80)	0.68 (0.67–0.68)