



Original Investigation | Gastroenterology and Hepatology

# Development and Validation of Machine Learning Models in Prediction of Remission in Patients With Moderate to Severe Crohn Disease

Akbar K. Waljee, MD, MSc; Beth I. Wallace, MD, MS; Shirley Cohen-Mekelburg, MD, MS; Yumu Liu, MS; Boang Liu, MS; Kay Sauder, BS; Ryan W. Stidham, MD; Ji Zhu, PhD; Peter D. R. Higgins, MD, PhD, MSc

## Abstract

**IMPORTANCE** Biological therapies have revolutionized inflammatory bowel disease management, but many patients do not respond to biological monotherapy. Identification of likely responders could reduce costs and delays in remission.

**OBJECTIVE** To identify patients with Crohn disease likely to be durable responders to ustekinumab before committing to long-term treatment.

**DESIGN, SETTING, AND PARTICIPANTS** This cohort study analyzed data from 3 phase 3 randomized clinical trials (UNITI-1, UNITI-2, and IM-UNITI) conducted from 2011 to 2015. Participants (n = 401) were individuals with active (C-reactive protein [CRP] measurement of  $\geq 5$  mg/L at enrollment) Crohn disease who received ustekinumab therapy. Data analysis was performed from November 1, 2017, to June 1, 2018.

**EXPOSURES** All included patients were exposed to 1 or more dose of ustekinumab for 8 weeks or more.

**MAIN OUTCOMES AND MEASURES** Random forest methods were used in building 2 models for predicting Crohn disease remission, with a CRP level lower than 5 mg/dL as a proxy for biological remission, beyond week 42 of ustekinumab treatment. The first model used only baseline data, and the second used data through week 8.

**RESULTS** In total, 401 participants, with a mean (SD) age of 36.3 (12.6) years and 170 male (42.4%), were included. The week-8 model had a mean area under the receiver operating characteristic curve (AUROC) of 0.78 (95% CI, 0.69-0.87). In the testing data set, 27 of 55 participants (49.1%) classified as likely to have treatment success achieved success with a CRP level lower than 5 mg/L after week 42, and 7 of 65 participants (10.8%) classified as likely to have treatment failure achieved this outcome. In the full cohort, 87 patients (21.7%) attained remission after week 42. A prediction model using the week-6 albumin to CRP ratio had an AUROC of 0.76 (95% CI, 0.71-0.82). Baseline ustekinumab serum levels did not improve the model's prediction performance.

**CONCLUSIONS AND RELEVANCE** In patients with active Crohn disease, demographic and laboratory data before week 8 of treatment appeared to allow the prompt identification of likely nonresponders to ustekinumab without the need for costly drug-level monitoring.

JAMA Network Open. 2019;2(5):e193721. doi:10.1001/jamanetworkopen.2019.3721

**Open Access.** This is an open access article distributed under the terms of the CC-BY License.

JAMA Network Open. 2019;2(5):e193721. doi:10.1001/jamanetworkopen.2019.3721

## Key Points

**Question** Can laboratory data be used to predict durable response to ustekinumab among patients with Crohn disease before they commit to long-term treatment?

**Findings** In this cohort study of 401 participants with active Crohn disease, a prediction model using demographic and laboratory data available after 1 dose of ustekinumab reliably identified patients with active Crohn disease who were likely to respond to treatment beyond week 42. The week-6 albumin to C-reactive protein ratio seemed to be a pragmatic predictor of future remission, but early ustekinumab serum levels appeared to add little to the predictive value of this model.

**Meaning** Long-term ustekinumab response may be predicted using demographic and laboratory data before week 8 of treatment, without the need for costly drug-level monitoring.

## + Supplemental content

Author affiliations and article information are listed at the end of this article.

## Introduction

Inflammatory bowel disease (IBD) has been estimated to affect more than 1.6 million people in the United States on the basis of the adjusted prevalence rates in the 2010 US white population.<sup>1,2</sup> Although the advent of expensive biological medications revolutionized the treatment of IBD,<sup>3</sup> the increasing use of these agents has revealed substantial heterogeneity of treatment effect across the population with IBD and likely overuse of these medications. Anti-tumor necrosis factor drugs, still the standard therapy for moderate to severe IBD, did not induce remission in 20% to 30% of patients and lost effectiveness within a year in another 30% to 40% of patients.<sup>3-6</sup> Novel biological agents emerging in recent years can help some patients who do not respond to anti-tumor necrosis factor therapy,<sup>6</sup> but these agents are ineffective in many others.<sup>7</sup> Initial efforts to improve remission rates with these agents have focused on optimizing serum drug levels and screening for antidrug antibodies, and these efforts have had some success.<sup>3,5</sup>

Interest has been growing in personalizing treatment strategies by using methods such as machine learning to match patients with IBD to the treatment most likely to work for them. These techniques using clinical and laboratory predictors that outperform drug levels alone may decrease avoidable drug costs, hospitalizations, surgical procedures, and complications while allowing patients with IBD to achieve symptomatic and biological remission much sooner than the typical 52-week clinical trial end point.<sup>8,9</sup> Such approaches have not yet been applied to ustekinumab. Previous studies of ustekinumab pharmacokinetics present mixed findings. Some suggest a dose-related association between serum drug concentrations of ustekinumab and treatment efficacy in Crohn disease similar to that seen with other novel agents<sup>8</sup>; others suggest that serum drug concentrations of ustekinumab are relatively poor predictors of clinical response in Crohn disease.<sup>10</sup>

The purpose of the present study was to (1) identify characteristics of patients with Crohn disease at baseline and at week 8 of ustekinumab therapy (before the start of subcutaneous maintenance therapy) that predict long-term remission on ustekinumab therapy using C-reactive protein (CRP) level as a biomarker of disease activity, and (2) evaluate the incremental advantage of using these predictors compared with drug level alone.

---

## Methods

We used publicly available data from 3 randomized clinical trials enrolling patients with active Crohn disease, UNITI-1 (ClinicalTrials.gov identifier [NCT01369329](#)), UNITI-2 (identifier [NCT01369342](#)), and IM-UNITI (identifier [NCT01369355](#)), conducted from 2011 to 2015. Trials UNITI-1 and UNITI-2 were conducted at 178 sites in 23 countries and 175 sites in 23 countries respectively. The IM-UNITI trial was conducted at 260 sites in 27 countries.<sup>7</sup> These data sets were obtained through the YODA (Yale Open Data Access) Project.<sup>11</sup> The University of Michigan Institutional Review Board confirmed that institutional review board approval for this study was not necessary because it evaluated previously collected and deidentified phase 3 clinical trial data. Informed consent was not obtained for this reason. This study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline and the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) reporting guideline. Data analysis was performed from November 1, 2017, to June 1, 2018.

The original clinical trial cohort had 1409 participants, 668 of whom were either randomized to placebo or lost to follow-up by week 8 of the trial and therefore excluded. The remaining 741 participants received ustekinumab at week 8. Of these, 462 participants had CRP measurements of 5 mg/L or higher at enrollment, suggesting active disease, and 61 participants were excluded for missing values. The final cohort comprised 401 participants, whose data were used for the week-8 model and sensitivity analyses (eTable in the [Supplement](#)). The baseline model excluded 30 additional patients for missing baseline laboratory tests and included 371 patients (eFigure 1 in the [Supplement](#)).

## Predictor Variables

The week-8 model included 401 patients, 5 demographic predictors (age, sex, race, ethnicity, and weight), and laboratory tests performed at week 8 (CRP level at weeks 0, 3, 6, and 8; ratio of serum ustekinumab level to CRP level at week 0 after the first ustekinumab dose, week 3, and week 6 [eTable in the Supplement]). The week-8 drug level was not used in modeling because of missing values. We also tracked CRP measurement in the long term and calculated mean, maximum, minimum, and the slope of CRP level from weeks 0 to 3, weeks 3 to 6, and weeks 6 to 8.

The baseline model at week 0 included 371 patients, and 30 patients were excluded from this cohort owing to missing baseline laboratory tests. This model included the 5 demographic predictors and 10 laboratory predictors (CRP level; platelet count; hemoglobin level; absolute counts of leukocytes, monocytes, lymphocytes, eosinophils, segmented neutrophils, and basophils; and serum ustekinumab concentration measured approximately 1 hour after the first drug dose).

## Definition of Outcomes

The primary outcome was the most recent available CRP measurement taken after week 42, which we used as a surrogate biological end point for Crohn disease inflammatory activity in patients with baseline CRP elevation. We dichotomized the outcome into 2 categories: CRP level lower than 5 mg/L as evidence of Crohn disease remission and CRP level of 5 mg/L or higher as evidence of continued Crohn disease activity and thus treatment failure (to convert CRP to nanomoles per liter, multiply by 9.524).<sup>12</sup> We chose this outcome because endoscopic data (the criterion standard for disease activity) were not available from the trial cohort of interest, and CRP level was the only objective disease activity measure available at multiple time points from the clinical trials.

No study visits or missing CRP measurements after week 42 were defined as treatment failures. We considered the utility of a subjective outcome measure, such as the Crohn Disease Activity Index (CDAI), but given the relatively low specificity of the CDAI, we instead chose an objective primary outcome variable. We also considered fecal calprotectin given its known high sensitivity over CRP for detection of active endoscopic disease activity.<sup>13</sup> However, we were limited by missing fecal calprotectin data, which might bias our results; instead, we chose to include fecal calprotectin as an outcome in a sensitivity analysis.

## Statistical Analysis and Model Development

Random forest is a machine learning method of prediction that uses an ensemble of decision trees<sup>14,15</sup> to classify observations. As previously described elsewhere,<sup>8,9</sup> information from a single patient (an observation) is evaluated by many decision trees (a forest) built on different bootstrap samples of the original data set. Each tree classifies the observation independently, which can be viewed as that tree's vote. The forest then combines the votes from the individual trees to generate a summary prediction. A cutoff value of the outcome variable is typically prespecified to provide the desired balance of sensitivity and specificity for the model. We used 500 trees to build each random forest model. Random forest modeling was chosen in this case because of its ability to model complex, nonlinear relationships and to easily accommodate complexities, such as interaction terms, which may not be captured by parametric models like logistic regression. This method has been used in previous studies.<sup>8,9</sup>

## Training and Testing Cohorts

As has been published,<sup>8,9</sup> the baseline and 8-week data sets were each split randomly into a 70% training set for model development and a 30% testing set. This splitting process was replicated 100 times, and each time the model was fit on the training set and used to predict the outcome in the testing set. This 100-times replication process was used to generate an estimate of the variation of the area under the receiver operating characteristic curve (AUROC) computed on the testing set. The mean AUROC was then obtained from the 100 replication test sets. Subsequently, the 1 split that produced a representative AUROC that was closest to the mean AUROC for the week-8 model was

selected as a representative split (training and testing cohorts) for both the week-8 and baseline models. This split was used to produce AUROC plots, cutoff point selection, and misclassification tables. These results were then compared across models.

The AUROC was used to evaluate the performance of each model. All statistical analyses were performed with the statistical language R, version 3.3, randomForest<sup>14</sup> and pROC packages (R Foundation for Statistical Computing). The optimal cutoff was defined as the point on the receiver operating characteristic plot with the highest sensitivity and specificity (ie, the point at which  $[1 - \text{sensitivity}]^2 + [1 - \text{specificity}]^2$  is minimized).

### Variable Importance and Partial Dependence Plots

We assessed predictor variable importance using a random forest model trained on the entire data set at baseline and at week 8.<sup>8,9</sup> We identified nodes in the group of trees in which each predictor appeared and then summed the information content of all nodes containing that predictor. Predictors providing greater combined discrimination were deemed to have higher importance. We built partial dependence plots on the entire cohort to demonstrate how each predictor affected the probability of success, defined by CRP level lower than 5 mg/L after week 42.<sup>8,9</sup> For each predictor, we used the random forest model on the entire data set at that time point to predict the probability of success by patient, with that predictor set to a given value and averaged over all patients to produce a mean probability of success. We repeated this procedure for each value of that predictor in the data, or for its 50 quantiles if more than 50 values were available.

### Simpler Pragmatic Models and Sensitivity Analyses

Given the complexity of random forest models, simpler models with comparable performance are typically preferred for clinical use. Because CRP<sup>16</sup> and albumin<sup>17</sup> are biomarkers used in the management of IBD, we constructed pragmatic single-variable prediction models with both baseline and weeks 6 to 8 data. This decision could also be justified by these 2 predictors' association with the variable importance plots. We evaluated these single-variable models on the full data set and compared the resulting predictions with those generated by the random forest models. We favored a simplified model that did not require week-8 data because a week-8 model would not allow clinicians to decide whether to continue ustekinumab until after the week-8 dose had been administered. To evaluate whether the outcome accurately represented Crohn disease activity, we performed a sensitivity analysis with the same variables used in the week-8 model to predict a dichotomized version of the patient's last CDAI score after week 42, with 150 as the cutoff point. Missing values were viewed as treatment failures. We included week-6 fecal calprotectin as a covariate in the week-8 model to evaluate prediction performance for a CRP-based outcome. We also created a model substituting dichotomized fecal calprotectin after week 42 as the outcome of interest, with thresholds of 100 µg/g or 400 µg/g.

## Results

In total, 401 participants were included in the final analysis. Among the participants, the mean (SD) age was 36.3 (12.6) years and 170 (42.4%) were male. The original clinical trial cohort comprised 741 participants who received ustekinumab at week 8, whereas the present study's week-8 model cohort had 401 participants and the baseline model cohort had 371. Additional demographic and clinical characteristics of all participants are presented in **Table 1**.

### Baseline Model and Predictors

The mean (SD) AUROC for the baseline model in the testing set was 0.56 (0.053), and the AUROC in the representative testing set was 0.59 (95% CI, 0.46-0.72) (**Figure 1A**). The best predictors of CRP at week 0 in the baseline model were baseline CRP level, weight, platelet count, ustekinumab serum level after the first dose of drug, and monocyte count (**Figure 1B**). Race, sex, and ethnicity were weak

predictors. Sensitivity for the baseline model, with 113 participants in the testing set, was 0.63, and specificity was 0.64. **Table 2** lists the best cutoff values, number of predicted successes and failures, and prediction accuracy rates for the baseline model.

Partial dependence plots were created to illustrate the association between each predictor and the outcome (eFigure 2 in the Supplement). Most predictors showed nonlinear associations. The likelihood of treatment success (ie, CRP level lower than 5 mg/L after week 42) increased steeply with lower baseline CRP measurements, with baseline CRP measurements of 14.65 mg/L or lower predicting treatment success. Weight and platelet count did not show monotonic trends.

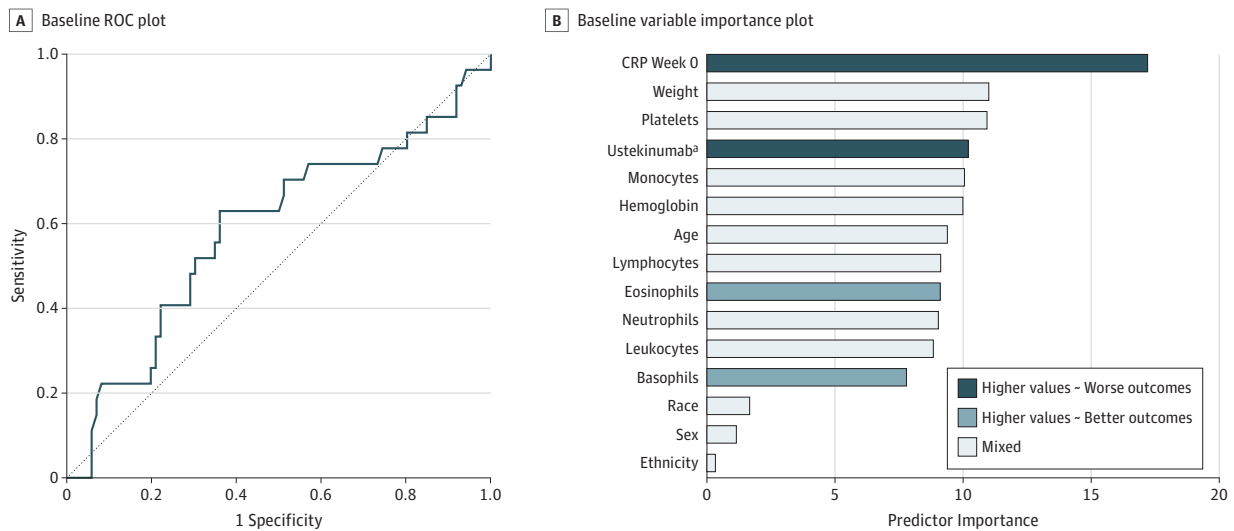
**Table 1. Participant Demographics**

Variable	Original Cohort (N = 741)	Modeling Cohort (N = 401)	Baseline Cohort (N = 371)
Age, mean (SD), y	37.7 (12.80)	36.3 (12.6)	36.3 (12.2)
Male sex, No. (%)	314 (42.)	170 (42.4)	158 (42.6)
White race, No. (%)	626 (84.5)	326 (81.3)	301 (81.1)
Weight, mean (SD), kg	71.0 (19.9)	70.2 (20.3)	70.3 (20.3)
Baseline CRP, median (IQR), mg/L	8.5 (3.4-22.2)	16.9 (9.2-32.3)	16.8 (9.3-32.8)
Concomitant medications, No. (%)			
Aminosalicylic acid	287 (38.7)	168 (41.9)	153 (41.2)
Immunosuppressants	308 (41.6)	170 (42.4)	152 (41.0)
Glucocorticoids	414 (55.9)	212 (52.9)	203 (54.7)
Location of disease, No. (%)			
Colon only	128 (17.3)	69 (17.2)	63 (17.0)
Ileum only	150 (20.2)	63 (15.7)	58 (15.6)
Colon and ileum	459 (61.9)	266 (66.3)	248 (66.8)
Perianal	273 (36.8)	176 (43.9)	167 (45.0)
Extraintestinal manifestation	387 (52.2)	214 (53.4)	203 (54.7)
Proximal small bowel	135 (18.2)	77 (19.2)	70 (18.9)

Abbreviations: CRP, C-reactive protein; IQR, interquartile range.

SI conversion factor: To convert CRP to nanomoles per liter, multiply by 9.524.

**Figure 1. Baseline Predictors**



A, Receiver operating characteristic (ROC) plot for baseline model. The area under the ROC curve in the representative testing set was 0.59 (95% CI, 0.46-0.72). B, Baseline variable importance plot. The importance of each predictor was computed by identifying nodes in the trees and summing up all information provided by the nodes containing the

predictor. The higher the importance, the larger the total discrimination provided by the predictor. CRP indicates C-reactive protein.

<sup>a</sup> Level 1 hour after first dose.

## Week-8 Model and Predictors

The mean (SD) AUROC for the week-8 model was 0.78 (0.042) (Figure 2A), and the AUROC on the representative testing set was 0.78 (95% CI, 0.69-0.87). The most important predictors in this model were CRP levels at weeks 3, 6, and 8 and serum ustekinumab to CRP ratio at weeks 3 and 6 (Figure 2B). Other than CRP, the most important laboratory variable included in this model was albumin measured at week 8. Race, sex, and ethnicity were weak predictors. Sensitivity for the week-8 model, with 120 participants in the testing set, was 0.79, and specificity was 0.67.

The week-8 model classified participants into 2 groups: those in whom treatment was predicted to succeed ( $n = 55$  [45.8%]), and those in whom treatment was predicted to fail ( $n = 65$  [54.2%]). Twenty-seven participants (49.1%) classified as likely to have treatment success achieved the goal of CRP level lower than 5 mg/L after week 42, and 7 of 65 participants (10.8%) classified as likely to have treatment failure achieved this outcome. For comparison, 87 participants (21.7%) of the full cohort achieved the specified outcome of remission after week 42.

## Predictors After a Single Dose of Ustekinumab

The probability of treatment success increased steeply with lower CRP levels at weeks 3, 6, and 8 (eFigure 3 in the Supplement). A higher value of the ustekinumab to CRP ratio at week 6 was associated with an increased treatment success rate, but the association was uncertain when the value was close to 0. The ustekinumab to CRP ratio at week 3 showed a nonmonotonic association with treatment success rate. A higher albumin level at week 6 was also associated with an increased treatment success rate. Other predictors showed weak or nonmonotonic associations with the outcome.

## Simpler Pragmatic Models

The predictive performance of the simpler pragmatic models evaluated on the full data set are shown in Table 2. We found that baseline CRP (prior to first ustekinumab dose) was a weak predictor of treatment success, with an AUROC of 0.67 (95% CI, 0.61-0.74). The week-6 albumin (grams per deciliter) to CRP (milligrams per liter) ratio was as accurate as week 8 CRP alone, with an AUROC of 0.76 (95% CI, 0.71-0.82) and a cutoff of 4.92. It was less accurate than the week-8 random forest model. A CRP value at week 6 or 8 also performed well, with an AUROC of 0.75 (95% CI, 0.70-0.81) at week 6 and 0.76 (95% CI, 0.71-0.82) at week 8. The CRP measurement at week 8 also predicted treatment success with reasonable accuracy in this cohort of patients with high baseline CRP level. However, it was inferior to the full model and to the week-6 albumin to CRP ratio by sensitivity, specificity, and accurate prediction rate. Adding the week-6 ustekinumab serum level to the week-6

Table 2. Prediction Performances of Models

Model Tested	Test Set Size	AUROC (95% CI)	Optimal Cutoff	Predicted Cases		Accurate Prediction Rate		Sensitivity	Specificity	Brier Score <sup>a</sup>
				Success (≥Cutoff)	Failure (<Cutoff)	Success (≥Cutoff)	Failure (<Cutoff)			
Baseline model	113	0.59 (0.46-0.72)	0.24	48	65	0.35	0.85	0.63	0.64	0.18
Baseline CRP alone	401	0.67 (0.61-0.74)	14.65 (mg/L)	182	219	0.32	0.87	0.67	0.61	NA
Week-8 model	120	0.78 (0.69-0.87)	0.26	55	65	0.49	0.89	0.79	0.67	0.17
Week-8 CRP alone	401	0.76 (0.71-0.82)	8.30	178	223	0.39	0.91	0.78	0.65	NA
Week-6 (albumin × ustekinumab-to-CRP ratio) model	398	0.71 (0.65-0.77)	24.71	177	221	0.33	0.88	0.69	0.62	NA
Week-6 CRP alone	401	0.75 (0.70-0.81)	6.77 (mg/L)	151	250	0.41	0.90	0.70	0.72	NA
Week-6 albumin to CRP ratio	398	0.76 (0.71-0.82)	4.92	166	232	0.40	0.91	0.77	0.68	NA

Abbreviations: AUROC, area under the receiver operating characteristic curve; CRP, C-reactive protein; NA, not applicable.

<sup>a</sup> Brier score, with a range of 0 to 1, is a measure of calibration.

SI conversion factor: To convert CRP to nanomoles per liter, multiply by 9.524.

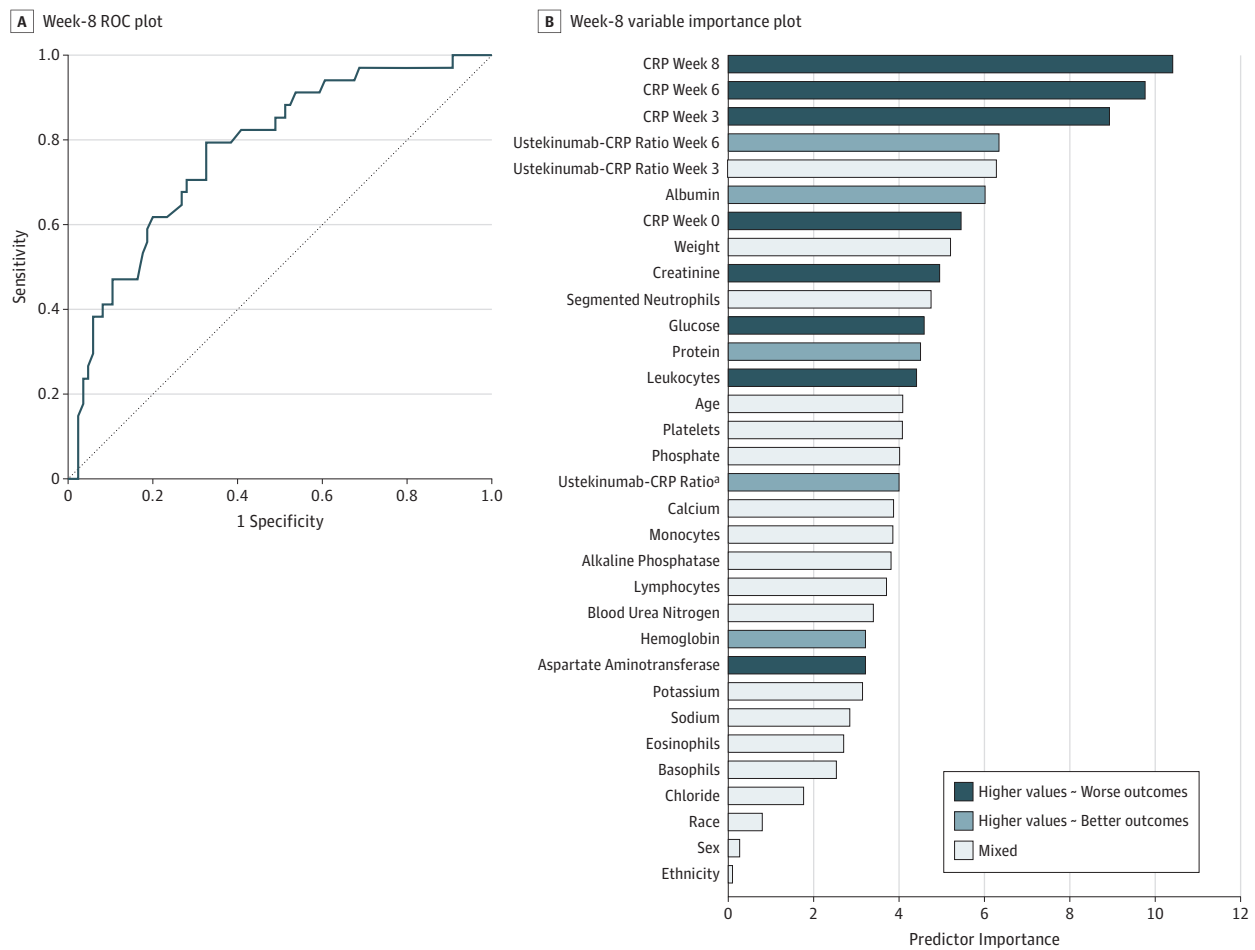
albumin to CRP ratio worsened the accuracy of this pragmatic model. These models are compared in Table 2.

**Sensitivity Analysis for Clinical End Point**

We fitted a random forest model with the same predictors used in the week-8 model, substituting the dichotomized CDAI score (<150 indicates clinical remission) after week 42 as the outcome of interest. Using this outcome definition, we identified 242 (60.3%) treatment failures and 159 (39.7%) treatment successes. Performance of the week-8 model was poor for this clinical end point, with an out-of-bag AUROC of 0.61 (95% CI, 0.55-0.66) for the CDAI end point.

Including week-6 fecal calprotectin value as a covariate in the week-8 model did not improve the prediction performance for a CRP-based outcome. The mean AUROC was 0.76 (95% CI, 0.71-0.82). Substituting the dichotomized fecal calprotectin value after week 42 as the outcome of interest with thresholds of 100 µg/g returned a mean (SD) AUROC of 0.64 (0.061) or 400 µg/g returned a mean (SD) AUROC of 0.63 (0.060).

**Figure 2. Week-8 Predictors**



A. Receiver operating characteristic (ROC) plot for week-8 model with C-reactive protein (CRP) and ustekinumab to CRP ratio. The area under the ROC curve in the representative testing set was 0.78 (95% CI, 0.69-0.87). B. Week-8 variable importance plot. The importance of each predictor was computed by identifying nodes in the trees and

summing all information provided by the nodes containing the predictor. The higher the importance, the larger the total discrimination provided by the predictor.

<sup>a</sup> One hour after infusion drug level with closest prior CRP.



## Discussion

Substantial heterogeneity in treatment response to costly biological therapies exists across the population with Crohn disease. For this reason, there is great interest in developing tools that can reliably identify patients who are unlikely to respond to costly medications after a short course of treatment. If effective, such tools could help clinicians identify patients who should be switched to an alternative treatment with a greater chance of success, reducing both the costs associated with an expensive but likely unsuccessful treatment and the harms associated with suboptimal disease control and avoidable steroid exposure.

We applied machine learning techniques to phase 3 trial data to generate predictive models of treatment response to ustekinumab at week 42 in patients with active Crohn disease, as defined by an elevated CRP level at enrollment using dichotomized CRP as a marker of biological remission. The model we constructed with baseline data was relatively inaccurate, performing worse than baseline CRP measurement alone (AUROC for CRP alone was 67%, with 67% sensitivity and 61% specificity). Thus, the baseline model should not be used clinically. This finding suggests, as do studies on vedolizumab,<sup>8,9</sup> that predicting an individual's response to a biological therapy before the first dose is difficult. The week-8 model provided the best performance characteristics for predicting long-term biological response with an AUROC of 0.78 (95% CI, 0.69-0.87). This model also proved able to distinguish long-term biological remission (as defined by CRP <5 mg/L beyond 42 weeks) from treatment failures. Simpler pragmatic models also showed good performance, with the week-6 albumin-to-CRP ratio showing an AUROC of 0.76 (95% CI, 0.71-0.82) at an albumin-to-CRP ratio cutoff of 4.92. Patients with higher values of this ratio at week 6 are increasingly likely to achieve biological remission.

Neither cross-sectional nor long-term measures of ustekinumab were associated with improvements in the performance of any of the models we tested, and they were associated with worse performance of week-6 pragmatic model. This finding supports previous studies demonstrating that serum ustekinumab level cannot adequately discriminate treatment responders from nonresponders in patients with Crohn disease enrolled in the UNITI clinical trials.<sup>10</sup> It seems unlikely that increasing the frequency of ustekinumab dosing will provide additional advantages in early nonresponders.

Knowing this finding might help clinicians decide to switch drug classes or to augment ustekinumab with an anti-tumor necrosis factor or a Janus kinase inhibitor if a substantial objective improvement in inflammation is not seen before the second dose rather than commit additional time and resources to costly monotherapy intensification, which appears unlikely to be effective. This finding allows clinicians to limit futile care with expensive biological agents such as ustekinumab 90 mg, which has a list price of approximately \$22 000 per dose.<sup>18</sup> In clinical practice, either the week-8 random forest model or the pragmatic model using the week-6 albumin to CRP ratio could predict outcomes, although the pragmatic model appears more likely to be used at the point of care.

## Limitations

This study has several limitations. First, we caution that the prediction performances of the pragmatic models should not be directly compared with those of the full models given that the pragmatic models were developed and tested on the entire cohort, whereas the random forest models were developed on a 70% random sample of the cohort and tested on the remaining 30%. Prior to clinical use, the pragmatic models would, ideally, be externally validated using a separate cohort. Second, we used patients with Crohn disease selected for enrollment in a clinical trial, which might not be generalizable to all patients with Crohn disease. Third, CRP is not a perfect proxy for biological remission given that up to 20% of patients with Crohn disease do not experience elevated CRP levels during a flare. Endoscopic data, which were not available for this cohort, would be ideal. To reduce the implications of this limitation, we used only the data from patients with elevated CRP levels at baseline and in whom CRP is more likely to be a reliable biomarker. Although this decision



improves the reliability of these results, it might limit generalizability to patients who received ustekinumab with an elevated baseline CRP levels. Fourth, before this model can be used in a real-world setting, long-term intervention studies are needed to evaluate the best course of action for patients deemed unlikely to achieve biological remission with ustekinumab. Fifth, this test, like any other diagnostic test, is imperfect, and clinical context should be considered when making decisions about its use and interpretation.

---

## Conclusions

In this study of clinical trial data, a model using demographic and laboratory data available after the first dose of ustekinumab reliably identified which patients with active Crohn disease at enrollment were likely to achieve treatment response beyond week 42. The week-6 albumin to CRP ratio is a potential predictor of future biological remission during ustekinumab therapy among patients with high baseline CRP levels. Unlike other biological agents for Crohn disease, ustekinumab serum did not appear to add substantial predictive value for future biological remission.

---

## ARTICLE INFORMATION

**Accepted for Publication:** March 24, 2019.

**Published:** May 10, 2019. doi:10.1001/jamanetworkopen.2019.3721

**Open Access:** This is an open access article distributed under the terms of the [CC-BY License](#). © 2019 Waljee AK et al. *JAMA Network Open*.

**Corresponding Author:** Akbar K. Waljee, MD, MSc, Division of Gastroenterology and Hepatology, Department of Internal Medicine, Michigan Medicine, Ann Arbor Veterans Affairs Medical Center, 2215 Fuller Rd, 111D, Ann Arbor, MI 48105 ([awaljee@med.umich.edu](mailto:awaljee@med.umich.edu)).

**Author Affiliations:** Michigan Integrated Center for Health Analytics and Medical Prediction, Ann Arbor (Waljee, Wallace); VA Center for Clinical Management Research, VA Ann Arbor Health Care System, Ann Arbor, Michigan (Waljee, Wallace, Cohen-Mekelburg); Division of Gastroenterology and Hepatology, Department of Internal Medicine, Michigan Medicine, Ann Arbor (Waljee, Cohen-Mekelburg, Sauder, Stidham, Higgins); Division of Rheumatology, Department of Internal Medicine, Michigan Medicine, Ann Arbor (Wallace); Department of Statistics, University of Michigan, Ann Arbor (Y. Liu, B. Liu, Zhu).

**Author Contributions:** Dr Waljee had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

*Concept and design:* Waljee, Stidham, Higgins.

*Acquisition, analysis, or interpretation of data:* All authors.

*Drafting of the manuscript:* Waljee, Cohen-Mekelburg, Y. Liu.

*Critical revision of the manuscript for important intellectual content:* Waljee, Wallace, Cohen-Mekelburg, B. Liu, Sauder, Stidham, Zhu, Higgins.

*Statistical analysis:* Waljee, Y. Liu, B. Liu, Stidham, Zhu, Higgins.

*Obtained funding:* Higgins.

*Administrative, technical, or material support:* Cohen-Mekelburg, Sauder, Stidham, Higgins.

*Supervision:* Waljee, Higgins.

**Conflict of Interest Disclosures:** Dr Higgins reported serving as a consultant for Takeda Pharmaceuticals International Inc; serving as a speaker for the GI Health Foundation and PRIME Medical Education; serving as a consultant and an advisory board member for AbbVie, Allergan, Amgen, Eli Lilly, Janssen, Lycera, Takeda, and UCB; and receiving research funding from AbbVie, Amgen, Arena, Ascentage Pharma, BCBS of Michigan, Buhmann, Eli Lilly, Genentech, Janssen, Lycera, Medimmune, Nestle, Pfizer, RedX Pharma, Seres, Shire, Takeda, and UCB. No other disclosures were reported.

**Funding/Support:** This study was supported by career development grant CDA 11-217 and merit review award IIR 16-024 from the US Department of Veterans Affairs Health Services Research and Development Service (Dr Waljee), by grant R01 GMO97117 from the National Institutes of Health (Drs Higgins and Waljee), and by grant T32AR007080-38 from the University of Michigan (Dr Wallace).

**Role of the Funder/Sponsor:** The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

**Disclaimer:** The content is solely the responsibility of the authors and does not necessarily represent the official views of the University of Michigan, the Department of Veterans Affairs, or the National Institutes of Health.

## REFERENCES

1. Shivashankar R, Tremaine WJ, Harmsen WS, Loftus EV Jr. Incidence and prevalence of Crohn's disease and ulcerative colitis in Olmsted County, Minnesota from 1970 through 2010. *Clin Gastroenterol Hepatol*. 2017;15(6):857-863. doi:10.1016/j.cgh.2016.10.039
2. Katsanos KH, Papamichael K, Feuerstein JD, Christodoulou DK, Cheifetz AS. Biological therapies in inflammatory bowel disease: beyond anti-TNF therapies [published online March 12, 2018]. *Clin Immunol*.
3. Yanai H, Lichtenstein L, Assa A, et al. Levels of drug and antidrug antibodies are associated with outcome of interventions after loss of response to infliximab or adalimumab. *Clin Gastroenterol Hepatol*. 2015;13(3):522-530.e2. doi:10.1016/j.cgh.2014.07.029
4. Papamichael K, Gils A, Rutgeerts P, et al. Role for therapeutic drug monitoring during induction therapy with TNF antagonists in IBD: evolution in the definition and management of primary nonresponse. *Inflamm Bowel Dis*. 2015;21(1):182-197. doi:10.1097/MIB.000000000000202
5. Yanai H, Hanauer SB. Assessing response and loss of response to biological therapies in IBD. *Am J Gastroenterol*. 2011;106(4):685-698. doi:10.1038/ajg.2011.103
6. Chan HC-H, Ng SC. Emerging biologics in inflammatory bowel disease. *J Gastroenterol*. 2017;52(2):141-150. doi:10.1007/s00535-016-1283-0
7. Feagan BG, Sandborn WJ, Gasink C, et al; UNITI-IM-UNITI Study Group. Ustekinumab as induction and maintenance therapy for Crohn's disease. *N Engl J Med*. 2016;375(20):1946-1960. doi:10.1056/NEJMoa1602773
8. Waljee AK, Liu B, Sauder K, et al. Predicting corticosteroid-free biologic remission with vedolizumab in Crohn's disease. *Inflamm Bowel Dis*. 2018;24(6):1185-1192. doi:10.1093/ibd/izy031
9. Waljee AK, Liu B, Sauder K, et al. Predicting corticosteroid-free endoscopic remission with vedolizumab in ulcerative colitis. *Aliment Pharmacol Ther*. 2018;47(6):763-772. doi:10.1111/apt.14510
10. Adedokun OJ, Xu Z, Gasink C, et al. Pharmacokinetics and exposure response relationships of ustekinumab in patients with Crohn's disease. *Gastroenterology*. 2018;154(6):1660-1671. doi:10.1053/j.gastro.2018.01.043
11. The YODA Project. <https://yoda.yale.edu/>. Accessed November 6, 2018.
12. Reinisch W, Wang Y, Oddens BJ, Link R. C-reactive protein, an indicator for maintained response or remission to infliximab in patients with Crohn's disease: a post-hoc analysis from ACCENT I. *Aliment Pharmacol Ther*. 2012;35(5):568-576. doi:10.1111/j.1365-2036.2011.04987.x
13. Mosli MH, Zou G, Garg SK, et al. C-reactive protein, fecal calprotectin, and stool lactoferrin for detection of endoscopic activity in symptomatic inflammatory bowel disease patients: a systematic review and meta-analysis. *Am J Gastroenterol*. 2015;110(6):802-819. doi:10.1038/ajg.2015.120
14. Breiman L. Random forests. *Mach Learn*. 2001;45(1):5-32. doi:10.1023/A:1010933404324
15. Liaw A, Wiener M. Classification and regression by random Forest. *R News*. 2002;2(3):18-22. [https://www.r-project.org/doc/Rnews/Rnews\\_2002-3.pdf](https://www.r-project.org/doc/Rnews/Rnews_2002-3.pdf). Accessed June 1, 2018
16. Vermeire S, Van Assche G, Rutgeerts P. C-reactive protein as a marker for inflammatory bowel disease. *Inflamm Bowel Dis*. 2004;10(5):661-665. doi:10.1097/OO054725-200409000-00026
17. Kochhar G, Lashner B. Utility of biomarkers in the management of inflammatory bowel disease. *Curr Treat Options Gastroenterol*. 2017;15(1):105-115. doi:10.1007/s11938-017-0129-z
18. Metro Medical. Stelara 90MG/1ML PFS each. <https://www.metromedicalorder.com/stelara-90mg-1ml-pfs-each.html>. Accessed November 6, 2018.

## SUPPLEMENT.

**eTable.** Quantitative Laboratory Tests Included in Week-8 Model

**eFigure 1.** Consort Diagram

**eFigure 2.** Partial Dependence Plots for Baseline Model Predictors

**eFigure 3.** Partial Dependence Plots for Week-8 Model Predictors