

DR AKBAR K WALJEE (Orcid ID : 0000-0003-1964-8790)

DR RYAN WILLIAM STIDHAM (Orcid ID : 0000-0001-9638-2186)

Article type : Original Scientific Paper

**Title:** Predicting Corticosteroid Free Endoscopic Remission with Vedolizumab in Ulcerative Colitis

**Short Title:** Predicting Vedolizumab Outcomes in UC

**Authors:**

Akbar K. Waljee, M.D., MS (1,2)

Boang Liu, M.S. (3)

Kay Sauder (2)

Ji Zhu, Ph.D. (3)

Shail M. Govani, M.D., MS (2)

Ryan W. Stidham, M.D., MS (2)

Peter D.R. Higgins, M.D., Ph.D., M.Sc. (2)

- 1) VA Center for Clinical Management Research, VA Ann Arbor Health Care System, Ann Arbor, MI, USA.
- 2) Department of Internal Medicine, Division of Gastroenterology and Hepatology, University of Michigan Health System, Ann Arbor, MI, USA.
- 3) Department of Statistics, University of Michigan, Ann Arbor, MI, USA.

**Corresponding Author Contact Information:**

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/apt.14510](https://doi.org/10.1111/apt.14510)

This article is protected by copyright. All rights reserved

Akbar K. Waljee

2215 Fuller Road

Gastroenterology 111D

Ann Arbor, MI 48105, USA

Phone: 734-845-5865

Fax: 734-845-3237

Email: [awaljee@med.umich.edu](mailto:awaljee@med.umich.edu)

References: 15

Tables: 2

Figures: 3

Appendices: 0

Abstract Word Count: 247

Manuscript Word Count: 3369

**Abbreviations:**

VDZ – Vedolizumab

CSDR – Clinical Study Data Request website

UC – Ulcerative colitis

IRB – Institutional Review Board

FCP – Faecal Calprotectin

ROC – Receiver Operating Characteristic

**ABSTRACT**

**Background**

Vedolizumab is an effective therapy for ulcerative colitis (UC), but costly and slow to work. New clinical responses occur after 30 weeks of therapy.

**Aims**

We aimed to enable physicians, patients, and insurers to predict whether a patient with UC will respond to vedolizumab at an early time point after starting therapy.

## **Methods**

The Clinical Study Data Request website provided the phase 3 clinical trial data for vedolizumab. Random forest models were trained on 70% and tested on 30% of the data to predict corticosteroid-free endoscopic remission at week 52. Models were constructed using baseline data, or data through week 6 of vedolizumab therapy from 491 subjects.

## **Results**

The AuROC for prediction of corticosteroid-free endoscopic remission at week 52 using baseline data was only 0.62 (95% CI: 0.53 – 0.72), but was 0.73 (95% CI: 0.65 – 0.82) when using data through week 6. 47% of subjects were predicted to be remitters, and 59% of these subjects achieved corticosteroid-free endoscopic remission, in contrast to 21% of the predicted non-remitters. A week 6 prediction using FCP <234µg/g was nearly as accurate.

## **Conclusions**

A machine learning algorithm using laboratory data through week 6 of vedolizumab therapy was able to accurately identify which UC patients would achieve corticosteroid-free endoscopic remission on vedolizumab at week 52. Application of this algorithm could have significant implications for clinical decisions on whom to continue on this costly medication when the benefits of the vedolizumab are not clinically apparent in the first 6 weeks of therapy.

## **BACKGROUND**

Ulcerative colitis (UC) affects over 700,000 people in the United States.<sup>1</sup> For many patients with moderate to severe disease, the gut-selective alpha-4-beta-7 integrin therapy, vedolizumab (VDZ), has proven effective, but is expensive and relatively slow to produce remission<sup>2 3 4</sup>

Given the high cost of VDZ, insurers are often reluctant to pay for this therapy, which produces remission in roughly one-third of patients with UC in clinical trials. Patients treated with VDZ often do not respond immediately, and additional new clinical responses continue to accumulate even after 30 weeks of therapy.<sup>5</sup> This encourages physicians and patients to continue VDZ therapy at great expense in the hope of a late remission, though the likelihood of this outcome is low. Physicians, patients, and insurers would like to be able to predict whether a given patient with UC will respond to VDZ at baseline, or at some early time point after starting therapy, rather than waiting up to 30 weeks (7 doses,

or \$43,782.76 in Average Wholesale Price drug costs)<sup>6</sup> to determine whether a satisfactory clinical response will occur.

With a pipeline of new therapies in IBD<sup>7</sup>, the growing number of treatment options for each patient raises the question of which therapy is most likely to work for each patient, and how long should we try a new, slow-acting therapy before switching drugs, increasing dose intensity, or adding a combination therapy. There is an increasing need to target therapies to the patients most likely to respond, given the high cost and growing number of therapeutic options in inflammatory bowel disease.

Leveraging clinical trial data may help elucidate which UC patients are most likely to benefit from each therapy, and make clinical decisions about starting or continuing therapies. The Clinical Study Data Request (CSDR) website<sup>8</sup> was implemented to provide researchers the opportunity to conduct further analysis with anonymized data from previously completed clinical studies. Through the CSDR, we obtained access to the phase 3 patient-level clinical trial data for the induction and maintenance of UC using VDZ. These data were used to predict whether baseline data, or data through week 6, could be predictive of week 52 corticosteroid-free endoscopic remission in UC patients treated with VDZ.

## **METHODS**

### **Overview**

We obtained confirmation from our Institutional Review Board (IRB) that IRB approval was not necessary to evaluate previously collected and de-identified clinical trial data from the CSDR for the phase 3 clinical trial data for the induction and maintenance of UC using VDZ (HUM00118527). Predictors and outcomes from the clinical trial dataset were used to develop and test predictive models of the outcome.

### **Cohort and demographics**

Our initial cohort consisted of 895 subjects. Subjects were excluded from model development if they were on placebo (N=275), leaving 620 subjects. Additional subjects were excluded if they had missing predictor variables (N=125) or were missing the outcome (N=4). A final dataset of 491 de-identified subjects was used for modeling. Table 1 shows the demographics of the final cohort, as compared to the original cohort in the two arms that received VDZ.

### **Predictor variables**

Baseline model predictor variables included patient age, gender, race, height, weight, VDZ Interval (VDZ Interval: dosing every 4 or 8 weeks), Immunomodulator use at the start of the trial (ImmAtStart), Steroid use at the start of the trial (SteroidAtStart), previous exposure to anti-TNF therapy (PriorTNF) and all available quantitative laboratory tests at baseline.

A model at week 6 included the baseline variables mentioned above with one exception: quantitative laboratory test results were included from week 6 (or nearest earlier date if week 6 results were not available), rather than from baseline. The VDZ drug level at week 6, as well as calculated longitudinal variables, were also included as predictors.

Longitudinal variables that were calculated included the slope of Faecal Calprotectin (FCP)  $((\text{week 6 FCP} - \text{FCP prior to initiation of medication})/6)$ , and the slope of the VDZ drug level  $((\text{VDZ drug level at week 6} - \text{VDZ drug level at week 2})/4)$ . The slope, acceleration, mean and maximum of each of the other laboratory predictors were tested, but provided no added improvement to the AuROC of the week 6 model, and were removed.

Additional clinical predictors, including disease extent, baseline sigmoidoscopic severity, and disease duration, were tested, but added no improvement to the AuROC, and were removed. A model including values from week 0 through week 14 was also tested, but was only slightly better than the week 6 model, so this model is not reported here. A list of all predictors used for each model, as well as variables that were tested and removed due to no improvement of AuROC, can be found in Supplemental Table 1.

### **Definition of outcomes**

The primary outcome was corticosteroid-free endoscopic remission at week 52, defined by no use of corticosteroid medications (including prednisone and budesonide) at week 52, and a Mayo Sigmoidoscopy Score of 0 or 1 at week 52. Subjects without a visit at week 52, having a sigmoidoscopy score greater than 1, or using steroids (Prednisone EQ Dose or Budesonide EQ Dose > 0 at week 52) were defined as failures. Patients who had a visit but who did not have a sigmoidoscopy score at week 52 were defined as a missing outcome, and were removed from the cohort (N=4).

### **Statistical analysis and model development**

Random forest (RF) machine learning<sup>9</sup> was used to construct these algorithms. This method of prediction uses decision trees<sup>9,10</sup> to classify a new observation. Each observation is run through each of

the trees in the forest and each tree gives a classification (which can also be thought of as a vote). The forest combines the votes from all of the trees to compute a predicted score of the outcome. By choosing a value of the predicted score as a cutoff, one can obtain the desired balance of sensitivity and specificity for the outcome. We developed two random forest models using the predictor variables, one using only baseline data and one using data through week 6. To validate the predictive models, each dataset was split into a 70% training dataset for model development, and a 30% testing dataset. This split method was preferred over an out-of-bag validation in order to show a true training and validation cohort, and to generate misclassification tables.

A baseline random forest model of 1000 trees with baseline variables was fit on the training set, and was used to produce the predictions on the test set. An additional week 6 random forest model of 1000 trees using cross sectional data at the final time point and longitudinal predictor variables was fit, the predictions tested on the test set, and the results compared across models.

#### *Training and testing cohorts*

Training and testing datasets were derived by splitting the data randomly into 70% and 30% subsets. This was done 50 times, and the random forest model was fit on the training dataset and tested on the testing dataset each time. The AuROC (area under the receiver operating characteristic curve) values from the 50 replication test sets were averaged in order to obtain a mean AuROC. Subsequently, the one split of the 50 that produced an AuROC closest to the average AuROC for the week 6 model was selected as a representative split (training and testing cohorts) for both the week 6 and baseline models, which was used for ROC (receiver operating characteristic) plots, representative AuROCs, cutoff point selection, and misclassification tables. We then built baseline and week 6 random forest models on the entire dataset in order to have the most accurate models for future use, and we calculated variable importance and produced partial dependence plots based on these models.

#### *Model performance*

The AuROC was used to evaluate the performance of each model. All statistical analyses were performed using the statistical language R (version 3.3), using the packages *randomForest*<sup>9</sup>, and *pROC*.

#### *Cutoff point selection*

The optimal cutoff was defined as the point on the ROC plot that is closest to the perfect point where both sensitivity and specificity are 1. In other words, we minimize the following criterion:  $(1 - \text{sensitivity})^2 + (1 - \text{specificity})^2$ .

#### *Variable importance*

We evaluated the importance of predictors based on a random forest model built on the entire dataset. 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 of the nodes containing that variable. Predictor variables that provided the greatest combined discrimination have higher importance.

#### *Partial dependence plots*

Partial dependence plots were constructed to demonstrate how individual predictors can affect the probability of success, i.e. corticosteroid-free endoscopic remission at week 52. For each predictor, we used the random forest model on the entire dataset to predict the probability of success for each 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.

#### *Simplified models*

Given the complexity of these models, simpler models were considered more attractive for routine clinical use, if accurate. Informed by the Variable Importance Plots, we hypothesized that two simplified models for week 6 might be helpful, one with the week 6 faecal calprotectin alone, and the second using the FCP/VDZ level ratio at week 6. We also hypothesized that the baseline FCP might be helpful as a single predictor. We evaluated these single variable predictor models on the entire (100%) dataset to determine whether they could predict the week 52 corticosteroid free endoscopic remission outcome.

#### *Reproducible Research Code Repository*

The code used to produce this analysis in R is available in a public Github repository at: <https://github.com/higgi13425/vedoUC>. Note that access to the trial data can only be obtained through the CSDR website at <https://clinicalstudydatarequest.com/>.

## RESULTS

### **Predicting week 52 steroid free endoscopic remission at baseline**

The average AuROC for the baseline model over 50 replications is 0.63. The AuROC for the baseline model under the selected training and testing split is 0.62 (95% CI: 0.53 – 0.72), as shown in Figure 1A. The variable importance plot for the baseline model is shown in Figure 1B. The 5 strongest baseline predictors of corticosteroid-free endoscopic remission at week 52 were: faecal calprotectin, albumin, neutrophils, white blood cell count, and absolute lymphocyte count. Notably weak baseline predictors of this outcome were the interval between VDZ doses, prior use of anti-TNF therapy, use of immunomodulators or corticosteroids at baseline, race, and sex. The best cutoff, number of predicted successes and failures, and the proportion of subjects with true success in the testing set (N=148) within these two predicted classes are displayed in Figure 3A and Table 2. True positives and true negatives are also displayed in Table 2, which provides the sensitivity of 0.63 and specificity of 0.62.

### **Effect of Predictor Variables at Baseline**

The effects of the predictors on the outcome of corticosteroid free endoscopic remission at week 52 can be illustrated by partial dependence plots. Many of these are not linear. These are displayed in multiple panels in Supplemental figure 1. Faecal calprotectin levels below 811.5  $\mu\text{g/g}$  at baseline predict success, and the rate of success increases steeply with lower FCP levels. Lower uric acid levels, perhaps a marker of ongoing bowel damage and cell death, also predict higher rates of success with VDZ in UC.

### **Predicting week 52 steroid free endoscopic remission at week 6**

The average AuROC for the week 6 model over 50 replications is 0.73. The AuROC for the week 6 model under the selected representative training and testing split is 0.73 (95% CI: 0.65 – 0.82), as shown in Figure 2A and Table 2. The p-value for the AuROC of the week 6 model vs. an AuROC of 0.5 is  $1.23 \times 10^{-6}$ . The variable importance graph for the week 6 model is shown in Figure 2B. The 5 strongest predictors of corticosteroid-free endoscopic remission at week 52 were faecal calprotectin at week 6, the slope of VDZ level, the slope of FCP, albumin at week 6, and VDZ level at week 6. Notably weak predictors at week 6 for this outcome were the interval between VDZ doses, prior use of anti-TNF therapy, use of immunomodulators or corticosteroids at baseline, race, and sex. The best cutoff, number



of predicted successes and failures, and the proportion of subjects of true success within these two predicted classes are displayed in Table 2 and Figure 3B. This model classifies subjects into two groups: those likely to succeed (47%), and those likely to fail (53%). Those classified as likely to succeed achieve corticosteroid-free endoscopic remission in 58.6%, while those classified as likely to fail only achieve this outcome in 20.5%. The true positives and true negatives are also displayed in Table 2, which provides the sensitivity of 0.72 and specificity of 0.68.

### **Effect of Predictor Variables at Week 6**

The effects of the Week 6 predictors on the outcome of corticosteroid free endoscopic remission at week 52 can be illustrated in partial dependence plots. These are displayed in multiple panels in Supplemental Figure 2. Faecal calprotectin levels below 233 at week 6 predict success, and the rate of success increases steeply with lower FCP levels. A significant fall in FCP between week 0 and week 6 (steeper negative slope) is a good prognostic marker. Lower albumin levels, perhaps a marker of chronicity and severity of ulceration, and colonic leak of VDZ, predict lower rates of success. As expected, higher VDZ levels predict success, but a falling VDZ level between week 2 and week 6 is a negative prognostic marker.

### **Simpler Pragmatic Models**

Based on the variable importance plots, we tested simpler models using a single predictor on the entire dataset, to see if simplified models could have comparable predictive performance. A simple model composed of very few variables would be easier to acquire and use in clinical practice. For comparison to the baseline model, we tested a simpler model using faecal calprotectin measured before the first dose of VDZ (FCP baseline). A baseline FCP of  $<811.5 \mu\text{g/g}$  was a weak predictor of success. This model was less accurate than the full baseline model, with an AuROC of 0.58 (95% CI: 0.52 – 0.63).

For comparison to the week 6 model, we tested two simpler models. A week 6 FCP/VDZ ratio  $<12.35$  predicted success reasonably well, and this single predictor had an AuROC of 0.71 (95% CI: 0.67 – 0.76). A single predictor of week 6 FCP  $<233.67 \mu\text{g/g}$  predicted success, and this predictor had a similar AuROC of 0.71 (95% CI: 0.66 – 0.76). Other single predictors were not as strong (Slope of VDZ AuROC: 0.65, Slope of FCP AuROC: 0.57, Week 6 Albumin AuROC: 0.65, Week 6 VDZ AuROC: 0.63). The best cutoff, number of predicted successes and failures, and the proportion of patients of true success within each predicted class for these simplified models are displayed in Figures 3C and D, and Table 2. The true positives and true negatives are also displayed in Table 2, which provides the sensitivity and specificity.

## DISCUSSION

Leveraging access to the GEMINI 1 clinical trial data<sup>2</sup> through the CSDR, we were able to apply machine learning tools to develop and validate predictive models of corticosteroid-free endoscopic remission in response to VDZ in UC. While the baseline model is relatively inaccurate, with a sensitivity of 63% and a specificity of 62%, it is helpful for clinicians to know that patients with a very elevated baseline faecal calprotectin (>811.5) are at higher risk of VDZ failure.

The week 6 model is numerically more accurate, with an AuROC of 0.73 (95% CI: 0.65 – 0.82). This model incorporates the change in faecal calprotectin over time, VDZ levels, and the slope of the VDZ concentration along with laboratory values at week 6. This model can classify patients into two distinct groups, who achieve corticosteroid-free endoscopic remission at week 52 in very different proportions of 59% and 21%. Simpler models are not quite as accurate, but it can be helpful to clinicians to know that values of either a Week 6 FCP/VDZ ratio of < 12.35, or a week 6 FCP < 233.67 µg/g, predict success. Takeda has provided various laboratories with VDZ level results and serum samples for validation of external assays, making the Takeda assay used in the clinical trials the *de facto* gold standard.

These models can be implemented as a cloud-based service for laboratories with HL7-compatible laboratory information systems<sup>11</sup>, which can securely submit lab values and accept returned calculated algorithmic results without exposing protected health information. This study is limited to the data available from GEMINI 1, and is only as generalizable as these data which resulted in the FDA approval of VDZ for use in the treatment of UC.

Given the expense of biologic therapies, being able to identify patients with a low rate of success after a short trial of therapy is valuable, as these patients can then move on to a different therapy, or a different mechanism of action in a timely fashion. This would reduce the time period of their active symptoms, likely reduce their exposure to steroids, and reduce the expense of a biologic therapy that is likely to be futile in a given patient.

Being able to identify UC patients who are at higher risk of failure with VDZ at baseline or at week 6 also provides an opportunity to improve the outcomes for these patients. Patients who are predicted as likely to fail VDZ might benefit from addition of a “booster” therapy after a model prediction of high risk of VDZ failure at week 6, possibly including an anti-TNF therapy<sup>12 13</sup>, a JAK inhibitor<sup>14</sup>, or an anti-IL23 therapy<sup>15</sup>. These severe UC patients, once they achieve low faecal

calprotectin levels and adequate VDZ levels, may well be able to continue on VDZ maintenance monotherapy successfully. Future randomized clinical trials in this subset of patients are needed.

The limitations of this study include: 1) 224 of the 491 patients were already corticosteroid-free at baseline. However, to enter into the trial, subjects had to have active disease, defined by a Mayo score of at least 6 (with an endoscopy score of 2-3). Achieving steroid-free remission in these subjects is an important clinical endpoint, whether they started on steroids or not; 2) patients included in all clinical trials are not necessarily representative of the general UC patient population; and 3) the simple predictors (Week 6 FCP/VDZ ratio, and Week 6 FCP) presented here were developed *post hoc*, and should be evaluated in an external dataset to determine their external validity.

Strengths of this study include: 1) the inclusion of patients from multiple sites and across multiple countries; 2) the data used in this study were the same data used for FDA approval of VDZ in UC; and 3) our internal validation used a model developed on a randomly selected 70% of subjects, and validated this model by testing it on the remaining 30% of subjects.; 3) It should also be noted that the simpler pragmatic model of a week 6 FCP/VDZ ratio  $<12.35$  provides a convenient measure for point-of-care use as its accuracy is close to the full model. Because this simple ratio is quite accurate, independent of previous anti-TNF use and body weight, this suggests that any patient can be treated effectively with Vedolizumab if sufficient drug for their inflammatory load is provided. These results suggest that titration of VDZ dosing during induction to produce a low FCP/VDZ level ratio could significantly increase the efficacy of VDZ induction in ulcerative colitis.

## CONCLUSION

A machine learning algorithm was able to distinguish IBD patients who are highly likely to achieve week 52 corticosteroid-free endoscopic remission with VDZ from those who are likely to fail VDZ using data collected by week 6 of therapy. A baseline model was numerically less accurate in predicting this outcome with VDZ. The ability to make early and accurate predictions of outcomes could help reduce costs by targeting this expensive therapy to the UC patients most likely to benefit, or could target additional interventions to patients who are likely to fail VDZ. While these algorithms are imperfect, they are arguably better than the current practice.

## Acknowledgements:

We would like to thank Clinical Study Data Request for making this data publically available.

Akbar K. Waljee, the submission's guarantor, takes responsibility for the integrity of the work as a whole, from inception to published article.

Individual author contributions are listed as follows:

Akbar K. Waljee: Concept and design, Data interpretation, Writing, Figures, Critical revision of the manuscript, Final approval

Boang Liu: Data Analysis, Data interpretation, Critical revision of the manuscript, Final approval

Kay Sauder: Data collection, Figures, Critical revision of the manuscript, Final approval

Ji Zhu: Data Analysis, Data interpretation, Critical revision of the manuscript, Final approval

Shail M. Govani: Critical revision of the manuscript, Final approval

Ryan W. Stidham: Critical revision of the manuscript, Final approval

Peter D.R. Higgins: Concept and design, Data interpretation, Figures, Critical revision of the manuscript, Final approval

All authors approved the final version of the manuscript.

#### **Statement of Interests:**

Peter D.R. Higgins has served as a consultant for Takeda Pharmaceuticals International, Inc. All other authors have no personal interests to declare.

Akbar K. Waljee is supported by a career development grant award (CDA 11-217) from the United States Department of Veterans Affairs Health Services Research and Development Service. Peter D.R. Higgin's and Akbar K. Waljee's research is supported by NIH R01 GM097117. The content is solely the responsibility of the authors and does not necessarily represent the official views of the University of Michigan, the Veterans Affairs, or the National Institutes of Health.

#### **References:**

1. Loftus EV. Clinical epidemiology of inflammatory bowel disease: Incidence, prevalence, and environmental influences. *Gastroenterology*. 2004;126(6):1504-1517.

This article is protected by copyright. All rights reserved

2. Feagan BG, Rutgeerts P, Sands BE, et al. Vedolizumab as induction and maintenance therapy for ulcerative colitis. *N Engl J Med*. 2013;369(8):699-710. doi:10.1056/NEJMoa1215734.
3. Stallmach A, Langbein C, Atreya R, et al. Vedolizumab provides clinical benefit over 1 year in patients with active inflammatory bowel disease – a prospective multicenter observational study. *Aliment Pharmacol Ther*. 2016;44(11-12):1199-1212. doi:10.1111/apt.13813.
4. Amiot A, Serrero M, Peyrin-Biroulet L, et al. One-year effectiveness and safety of vedolizumab therapy for inflammatory bowel disease: a prospective multicentre cohort study. *Aliment Pharmacol Ther*. 2017;46(3):310-321. doi:10.1111/apt.14167.
5. Dulai PS, Singh S, Jiang X, et al. The real-world effectiveness and safety of vedolizumab for moderate-severe Crohn's disease: Results from the US VICTORY consortium. *The American journal of gastroenterology*. 2016;111(8):1147-1155. doi:10.1038/ajg.2016.236.
6. ENTYVIO (VEDOLIZUMAB) FOR INJ 300MG SDV. [metromedicalorder.com](https://www.metromedicalorder.com). <https://www.metromedicalorder.com/entyvio-vedolizumab-for-inj-300mg-sdv.html>. Accessed March 20, 2017.
7. Cohen BL, Sachar DB. Update on anti-tumor necrosis factor agents and other new drugs for inflammatory bowel disease. *BMJ*. 2017;357:j2505. doi:10.1136/bmj.j2505.
8. CSDR: Clinical Study Data Request. <https://clinicalstudydatarequest.com/>.
9. Breiman L. Random forests. 2001;45(1):5-32.
10. Liaw A, Wiener M. Classification and Regression by randomForest. *R news*. 2002;2(3):18-22.
11. Romero J, Lopez P, Vazquez Noguera JL, Cappo C, P Pinto-Roa D, Villalba C. Integrated, Reliable and Cloud-Based Personal Health Record : A Scoping Review. *HIJ*. 2016;5(2/3):01-20. doi:10.5121/hij.2016.5301.
12. Stidham RW, Lee TCH, Higgins PDR, et al. Systematic review with network meta-analysis: the efficacy of anti-tumour necrosis factor-alpha agents for the treatment of ulcerative colitis. *Aliment Pharm Ther*. 2014;39(7):660-671. doi:10.1111/apt.12644.
13. Stidham RW, Lee TCH, Higgins PDR, et al. Systematic review with network meta-analysis: the efficacy of anti-TNF agents for the treatment of Crohn's disease. *Aliment Pharmacol Ther*. 2014;39(12):1349-1362. doi:10.1111/apt.12749.
14. Olivera P, Danese S, Peyrin-Biroulet L. JAK inhibition in inflammatory bowel disease. *Expert Review of Clinical Immunology*. 2017;13(7):693-703. doi:10.1080/1744666X.2017.1291342.

15. Furfaro F, Gilardi D, Allocca M, et al. IL-23 Blockade for Crohn s disease: next generation of anti-cytokine therapy. *Expert Review of Clinical Immunology*. 2017;13(5):457-467.  
doi:10.1080/1744666X.2017.1279055.

Author Manuscript

---

## Table Titles and Figure Legends

### Table 1: Subject Demographics

Table 1 compares the subject demographics of the original clinical trial cohort (which combines both VDZ arms) and the final modeling cohort (where any subject with missing data or missing outcomes was removed) to demonstrate that the modeling cohort is similar to the original clinical trial cohort.

### Table 2: Model Performance

This table presents the performance details for each model. The Area under the Receiver Operating Characteristic curve (AuROC) provides the discriminative power of each model. The best cut off for each model was based on the ROC plot to optimize the balance between the sensitivity and specificity. The number of cases predicted to be a success or failure are listed for each model, along with the respective true success rate, sensitivity, and specificity.

### Figure 1: Baseline Model ROC Plot and Variable Importance Plot

- 1A. Receiver Operating Characteristic (ROC) plot for the baseline model
- 1B. Variable Importance plot for the baseline model shows the relative importance of each predictor variable

### Figure 2: Week 6 Longitudinal Model ROC Plot and Variable Importance Plot

- 2A. Receiver Operating Characteristic (ROC) plot for the week 6 longitudinal model
- 2B. Variable Importance plot for the longitudinal model shows the relative importance of each predictor variable

### Figure 3: Endpoint Success

Each flow diagram shows the percentages of endpoint success (corticosteroid-free and endoscopically healed subjects) at week 52 within the predicted success and predicted failure groups.

- 3A. Success Rate with Baseline Model
- 3B. Success Rate with Week 6 Model
- 3C. Success Rate with Simplified Week 6 Model – Week 6 FCP/VDZ
- 3D. Success Rate with Simplified Week 6 Model – Week 6 FCP

Table 1: Subject Demographics

| Variable                               | Original Cohort<br>(N=620) | Final Modeling Cohort<br>(N=491) |
|--|----------------------------|----------------------------------|
| Mean Age in Years                      | 40.1±13.1                  | 40.2±13.4                        |
| Percentage Male Sex                    | 58.7%                      | 57.4%                            |
| Percentage White Race                  | 83.5%                      | 83.1%                            |
| Mean Body Weight (kg)                  | 73.4±18.3                  | 73.0±18.5                        |
| Percentage Current Smoker              | 5.8%                       | 5.9%                             |
| Mean Duration of Disease in Years      | 6.7±6.0                    | 6.5±5.9                          |
| Mean Mayo Clinic Score at Baseline     | 8.6±1.8                    | 8.5±1.7                          |
| Median Faecal Calprotectin at Baseline | 844 (346-1727)             | 872 (371-1727)                   |
| Site of Disease (%)                    |                            |                                  |
| Rectum and sigmoid colon only          | 13.7%                      | 14.1%                            |



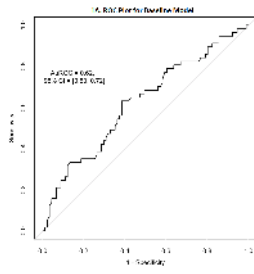
|  |                             |                             |
|--|-----------------------------|-----------------------------|
| Left colon   | 36.6%                       | 37.7%                       |
| Proximal to splenic flexure  | 11.9%                       | 11.4%                       |
| Pancolitis   | 37.7%                       | 36.9%                       |
| Median prednisone equivalent dose in people who were on prednisone (mg)  | 20.0 (10.0-30.0)<br>(N=316) | 20.0 (10.0-30.0)<br>(N=261) |
| Median budesonide equivalent dose in people who were on budesonide (mg)* | 5.0 (5.0-9.0)<br>(N=11)     | 5.0 (5.0-9.0)<br>(N=6)      |
| Percentage with prior anti-TNF therapy                                   | 50.2%                       | 49.1%                       |
| Baseline concomitant medications (%)                                     |                             |                             |
| Glucocorticoids only   | 36.5%                       | 37.5%                       |
| Immunosuppressants only  | 18.4%                       | 17.3%                       |
| Glucocorticoids and<br>Immunosuppressants                                | 16.0%                       | 16.9%                       |
| No Glucocorticoids or<br>Immunosuppressants                              | 29.2%                       | 28.3%                       |

\*Budesonide prescriptions of 5mg were reported in this dataset.

Table 2: Model Performance

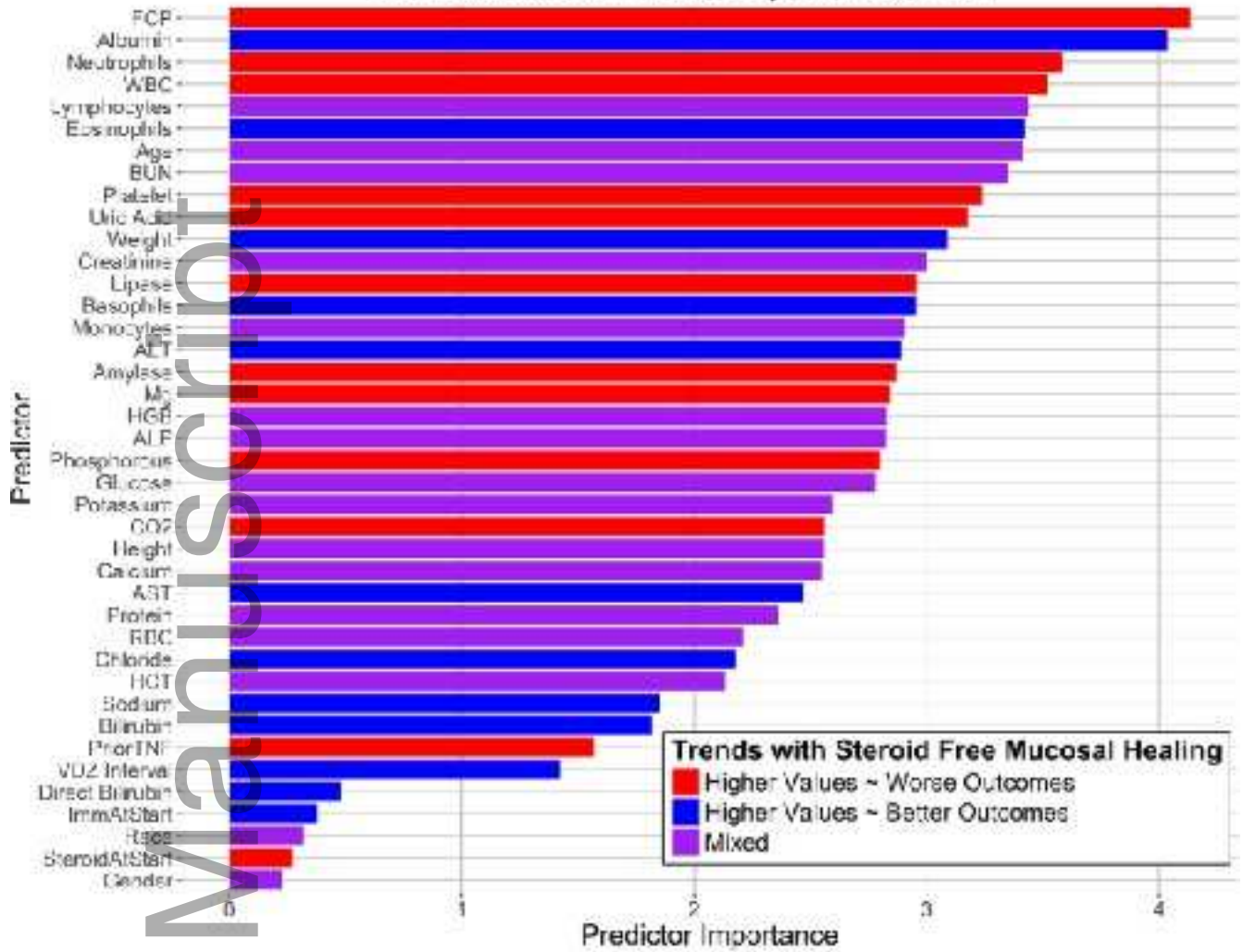
| Model   | Validation sample size | AuROC (95% CI)       | Best cutoff | Prediction category         | Predicted cases | True success rate | Sensitivity | Specificity |
|---|------------------------|----------------------|-------------|-----------------------------|-----------------|-------------------|-------------|-------------|
| Baseline model                                  | 148                    | 0.62<br>(0.53, 0.72) | 0.35        | Predicted success (>cutoff) | 71              | 50.7%             | 0.63        | 0.62        |
|   |                        |                      |             | Predicted failure (<cutoff) | 77              | 27.3%             |             |             |
| Faecal Calprotectin before 1 <sup>st</sup> dose | 491                    | 0.58<br>(0.52, 0.63) | 811.50      | Predicted success (<cutoff) | 236             | 41.9%             | 0.57        | 0.57        |
|   |                        |                      |             | Predicted failure (>cutoff) | 255             | 29.0%             |             |             |
| Week 6 model                                    | 148                    | 0.73<br>(0.65, 0.82) | 0.32        | Predicted success (>cutoff) | 70              | 58.6%             | 0.72        | 0.68        |
|   |                        |                      |             | Predicted failure (<cutoff) | 78              | 20.5%             |             |             |
| Week 6 FCP/VDZ level ratio                      | 491                    | 0.71<br>(0.67, 0.76) | 12.35       | Predicted success (<cutoff) | 225             | 52.0%             | 0.68        | 0.66        |
|   |                        |                      |             | Predicted failure (>cutoff) | 266             | 21.1%             |             |             |
| Week 6 Faecal calprotectin                      | 491                    | 0.71<br>(0.66, 0.76) | 233.67      | Predicted success (<cutoff) | 186             | 54.3%             | 0.58        | 0.73        |
|   |                        |                      |             |                             |                 |                   |             |             |

# Author Manuscript



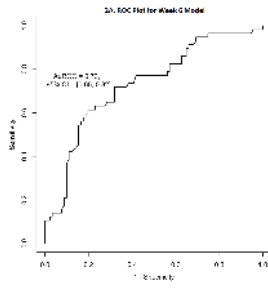
apt\_14510\_f1a.tif

### 1B. Baseline Variable Importance Plot



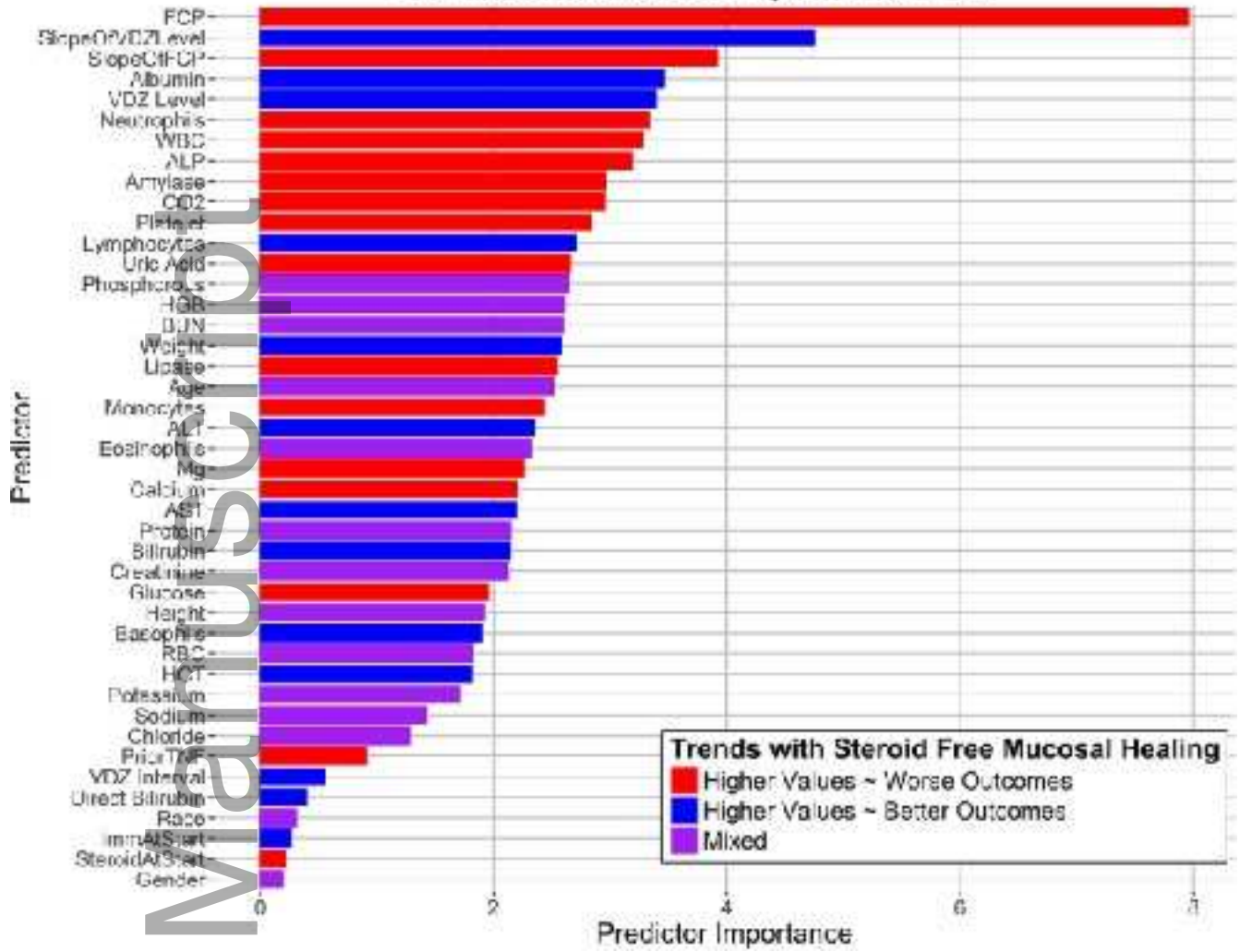
apt\_14510\_f1b.tif

# Author Manuscript



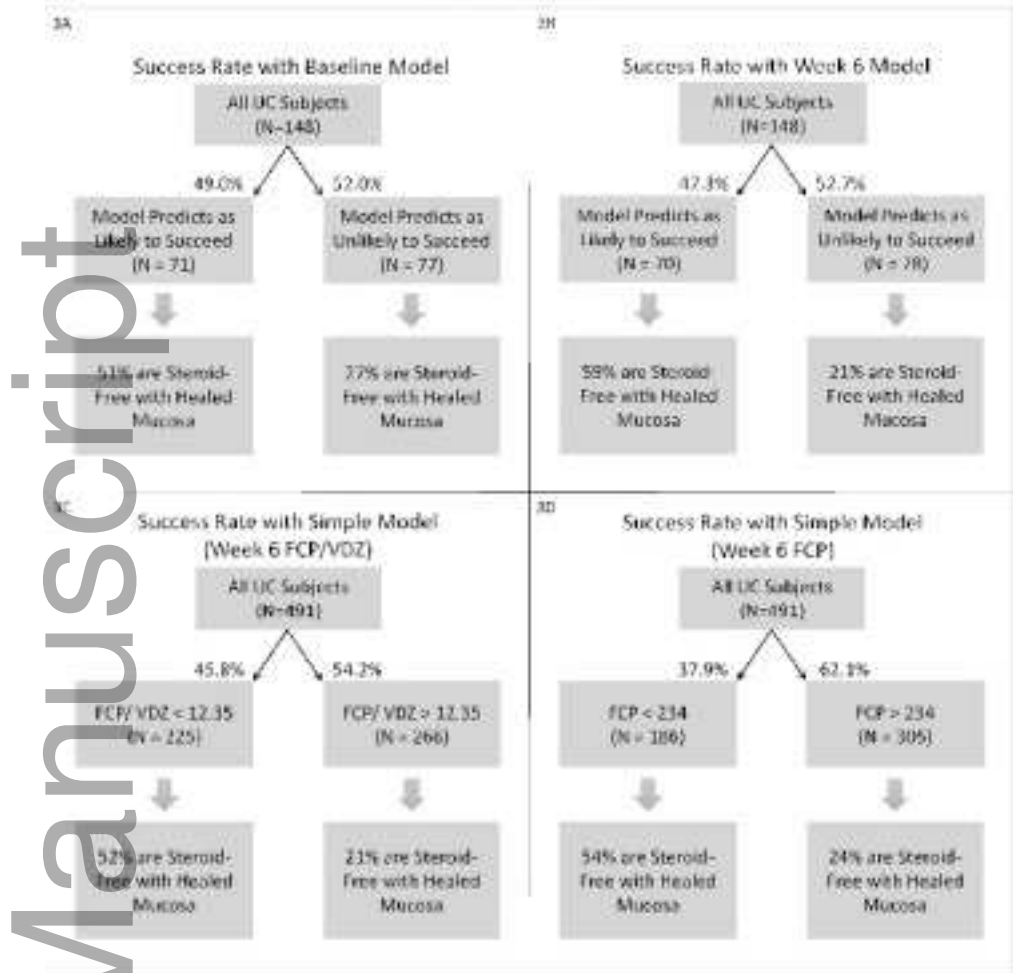
apt\_14510\_f2a.tif

## 2B. Week 6 Variable Importance Plot



apt\_14510\_f2b.tif

Endpoint Success



apt\_14510\_f3.tiff