Corresponding Author: Parambir S. Dulai, MD, University of California at San Diego, Division of Gastroenterology, 9500 Gilman Dr., La Jolla, CA 92093, pdulai@ucsd.edu.

Author Contributions:

1. Study concept and design (PSD, KL, CC)

2. Acquisition of data (PSD, BSB, SS, KC, JLKP, GK, MPP, ES, JH, PC, JM, DW, RH, ACW, LGK, FP, NN, KS, AS, MB, DL, DH)

3. Statistical analysis (PSD, VJ, GYZ, CC)

4. Drafting of manuscript (PSD, CC)

5. Critical revision of the manuscript for important intellectual content (PSD, JTC, JRN, VJ, GYZ, BGF, BS, CAS, EVL, SK, BES, JFC, WJS, KL)

6. Study supervision (PSD, KL, CC)

Development and Validation of a Scoring System to Predict Outcomes of Vedolizumab Treatment in Patients with Crohn's Disease

Parambir S. Dulai[1], **Brigid S. Boland**[1], **Siddharth Singh**[1], **Khadija Chaudrey**[2], **Jenna L. Koliani-Pace**[3], **Gursimran Kochhar**[4], **Malav P. Parikh**[4], **Eugenia Shmidt**[5], **Justin Hartke**[6], **Prianka Chilukuri**[6], **Joseph Meserve**[1], **Diana Whitehead**[3], **Robert Hirten**[7], **Adam C. Winters**[5], **Leah G. Katta**[5], **Farhad Peerani**[5,8], **Neeraj Narula**[5,9], **Keith Sultan**[7], **Arun Swaminath**[10], **Matthew Bohm**[6], **Dana Lukin**[11], **David Hudesman**[12], **John T. Chang**[1], **Jesus Rivera-Nieves**[1], **Vipul Jairath**[13], **GY Zou**[13], **Brian G. Feagan**[13], **Bo Shen**[4], **Corey A. Siegel**[3], **Edward V. Loftus Jr**[2], **Sunanda Kane**[2], **Bruce E. Sands**[5], **Jean-Frederic Colombel**[5], **William J. Sandborn**[1], **Karen Lasch**[14], and **Charlie Cao**[14]

[1]University of California - San Diego, La Jolla, CA, USA

[2]Mayo Clinic, Rochester, MN, USA

[3]Dartmouth-Hitchcock Medical Center, Lebanon, NH, USA

[4]Cleveland Clinic Foundation, Cleveland, OH, USA

[5]Icahn School of Medicine at Mount Sinai, New York, NY, USA

[6]Indiana University, Indianapolis, IN, USA

[7]North Shore University Hospital, Manhasset, NY, USA

[8]University of Alberta, Edmonton, AB, Canada

[9]McMaster University Medical Centre, Hamilton, ON, Canada

[10]Lenox Hill Hospital, New York, NY, USA

[11]Montefiore Medical Center, New York, NY, USA

[12]New York University (NYU), New York, NY, USA

[13]University of Western Ontario, London, ON, Canada

[14]Takeda Pharmaceuticals USA Inc., Deerfield, IL, USA

Abstract

Background & Aims: As more treatment options for inflammatory bowel diseases become available, it is important to identify patients most likely to respond to different therapies. We created and validated a scoring system to identify patients with Crohn's disease (CD) who respond to vedolizumab.

Methods: We collected data from GEMINI 2 phase 3 trial of patients with active CD treated with vedolizumab for 26 weeks (n=814) and performed logistic regression analysis to identify factors associated with clinical, steroid-free, and durable remission (derivation set). We used these data to

develop a clinical decision support tool, which we validated using data from 366 participants in a separate clinical practice observational cohort of patients with active CD treated with vedolizumab for 26 weeks (the VICTORY cohort). We evaluated the ability of this tool to identify patients in clinical remission or corticosteroid-free remission, or those with mucosal healing (MH), clinical remission with MH, or corticosteroid-free remission with MH after vedolizumab therapy using receiver operating characteristic area under the curve (AUC) analyses. The primary outcome was to develop and validate a list of factors associated with achieving remission by vedolizumab in patients with active CD.

Results: In the derivation analysis, we identified absence of previous treatment with a tumor necrosis factor antagonist (+3 points), absence of prior bowel surgery (+2 points), absence of prior fistulizing disease (+2 points), baseline level of albumin (+0.4 points per g/L), and baseline concentration of C-reactive protein (reduction of 0.5 points for values between 3.0–10.0 mg/L and 3.0 points for values > 10.0 mg/L) as factors associated with remission. In the validation set, our model identified patients in clinical remission with an AUC of 0.67, patients in corticosteroid-free remission with an AUC of 0.66, patients with MH with an AUC of 0.72, patients in clinical remission with MH with an AUC of 0.73, and patients in corticosteroid-free clinical remission with MH with an AUC of 0.75. A cut-off value of 13 points identified patients in clinical remission after vedolizumab therapy with 92% sensitivity, patients in corticosteroid-free remission with 94% sensitivity, patients with MH with 98% sensitivity, patients in deep remission with 100% sensitivity, and patients with corticosteroid-free clinical remission with MH with 100% sensitivity.

Conclusions: We developed and validated a scoring system to identify patients with CD most likely to respond to 26 weeks of vedolizumab therapy. Further studies are needed to optimize its accuracy in select populations and determine its cost effectiveness.

Graphical Abstract
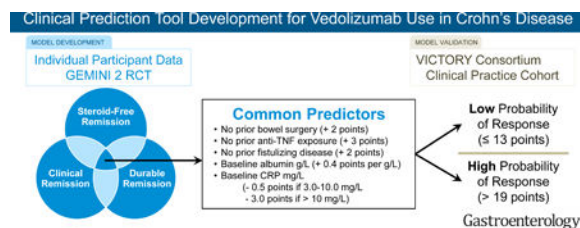
INTRODUCTION

Vedolizumab (VDZ), a humanized anti-$\alpha_4\beta_7$ integrin monoclonal antibody that selectively targets lymphocyte trafficking to the gut, is currently indicated for the treatment of adult patients with moderately to severely active Crohn's disease (CD) who have failed corticosteroids, immunomodulators, or tumor necrosis factor-alpha (TNFα) antagonist therapy. In the GEMINI 2 phase III trial, approximately one-third of patients were in clinical remission (CREM) or corticosteroid-free clinical remission (CSF-REM) following treatment

with VDZ.[1] Although these results are clinically important, the GEMINI 2 trial did not assess for mucosal healing (MH), and the strict inclusion criteria used may limit the generalizability of the results to routine clinical practice.[2]

Real-world data from multiple jurisdictions are now available for VDZ therapy in CD and outcomes are fairly consistent across clinical practice. In the US-based VICTORY consortium that evaluated 212 CD patients, CREM, CSF-REM, and deep remission (CREM + MH) were seen in 35%, 34%, and 26% of patients, respectively, by 12 months.[3] In real world cohorts and in the GEMINI 2 trial, prior exposure to TNFα-antagonists negatively affected treatment outcomes.[3–5] However, the magnitude of this effect and the influence of other clinical factors on treatment outcomes varied across these studies, making it difficult for clinicians to interpret their relevance to practice.

Clinical prediction models utilize baseline characteristics to provide an estimate of the value of a therapy on treatment outcomes for an individual patient. Furthermore, the transformation of these models into decision support tools facilitates their application as a component of 'precision medicine'.[6, 7] With the evolving landscape of biologic therapy in CD and increasing treatment choice, a validated prognostic tool for treatment outcomes with VDZ would be of considerable value.[8] We aimed to address this gap by deriving and validating a multivariable clinical prediction model within the GEMINI 2 clinical trial dataset. To improve the ease with which this prediction model can be used at the 'bedside', we transformed it into a prognostic clinical decision support tool (CDST) and validated this tool in a cohort of CD patients treated with VDZ in routine clinical practice.

METHODS

We developed and validated a multivariable model to predict CREM with VDZ treatment for patients with active CD.[1, 9, 10] We further assessed model prediction for MH and deep remission in VDZ-treated CD patients with endoscopically active disease at baseline.[3] Finally, we transformed this prediction model into a CDST for use in routine practice. This study is reported according to the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement.[9]

Data Sources and Participants

We used two data sources to derive and validate the prediction model and CDST.[1, 3] First, the GEMINI 2 trial was used to derive the CREM prediction model, and to derive the CDST. Second, data from the VICTORY consortium were used to externally validate the CREM prediction model, to assess the performance of the prediction model for predicting MH and deep remission, and to validate the CDST.

A subset of participants in GEMINI 2 (n=814) was used in the post hoc analysis. To mimic a treat-straight-through cohort design, patients were included if they received VDZ induction therapy and did not receive placebo during maintenance therapy, irrespective of Week 6 response status (Supplementary Material).

CD patients from the VICTORY consortium were included in the validation cohort (n=336) if they had: started VDZ therapy for clinically and/or endoscopically active CD, completed VDZ induction therapy, had a clinical or endoscopic assessment of disease activity after VDZ induction therapy, and had a minimum follow-up period of 26 weeks after VDZ initiation (Supplementary Material).

Outcomes and Definitions

The primary objective was to develop and validate a prediction model for CREM with VDZ in patients with active CD. Three definitions of CREM were used for this purpose: (1) CREM after 26 weeks of VDZ therapy; (2) CSF-REM after 26 weeks of VDZ for subjects receiving corticosteroid at baseline; and (3) sustained CREM during 52 weeks of VDZ, defined as 10 of 12 study visits from Week 10 to Week 52, including the final visit (Week 52). Clinical disease activity and remission were assessed using the Crohn's disease activity index (CDAI; CREM defined as CDAI 150) in GEMINI 2, and the physician global assessment (CREM defined as absence of all CD-related symptoms) in the VICTORY consortium.[11]

A secondary analysis was performed to assess the performance of the model for predicting MH and deep remission after 26 weeks of VDZ in endoscopically active CD patients. MH could only be assessed in the VICTORY consortium, and was defined as the absence of ulcers or erosions on ileocolonoscopy, or absence of findings of inflammation on cross-sectional imaging in patients who could not be adequately assessed with ileocolonoscopy. Deep remission was defined as achieving both CREM and MH, and CSF-deep remission was defined as achieving both CSF-REM and MH (Supplementary Material).[12]

Statistical Analysis

Model Derivation—Individual multi-variable logistic regression prediction models were built in the GEMINI 2 derivation cohort for the outcomes of: CREM after 26 weeks of VDZ; CSF-REM after 26 weeks of VDZ; and sustained CREM during 52 weeks of VDZ therapy. [13–18] From these three individual CREM prediction models, a set of variables were chosen for inclusion in a single final single prediction model based on: (1) being predictive in at least two of the three individual models; (2) review of the literature;[19, 20] and/or (3) expert opinion of the author group. These models were transformed into a single final prediction model, as opposed to being individually validated externally, because it was felt that having 3 separate models to predict separate outcomes with the same intervention (ie, VDZ) would lead to uncertainty in interpretation in clinical practice, potentially diverging predictions, and ultimately poor uptake of the CDST.

To build this single final prediction model, we weighted the coefficient of regression of each variable from each individual model by an estimation of the inverse variance for each coefficient estimate.[21, 22] This approach does not take into account the clinical importance of each outcome from each individual model, or individual provider opinions that may influence treatment decisions. Therefore, a sensitivity analysis was performed using survey results from VICTORY consortium investigators on how each outcome should be weighted for the final single prediction model equation (Supplementary Material).

Albumin and CRP were both retained in the final model and have been previously shown to be the strongest predictors of exposure-efficacy relationships for VDZ in CD.[23] However, these 2 variables were observed to explain less than 30% of the variability in VDZ exposure. Therefore, a second sensitivity analysis was performed to more accurately account for variations in VDZ exposure by replacing baseline albumin and CRP with calculated VDZ clearance profiles within the GEMINI 2 derivation cohort. Calculated VDZ clearance for individual participants from the GEMINI 2 trial were obtained from previously published population pharmacokinetic-pharmacodynamics modeling and exposure-efficacy relationship studies.[23, 24] A comparison of accuracy for predicting CREM, CSF-REM, and sustained CREM were then performed within the GEMINI 2 derivation cohort to determine if albumin and CRP should be replaced by the population pharmacokinetic-pharmacodynamic model equation.

Model Validation—The final single prediction model was then validated externally in the VICTORY consortium cohort. Discriminative ability was assessed by receiver operating characteristic (ROC) curve analysis and is presented as AUC. The values are between 0.5 and 1, with 0.5 denoting that the model does not discriminate and 1 denoting that it perfectly discriminates. Calibration was tested by the Hosmer–Lemeshow goodness-of-fit test after splitting the sample into quintiles. This test assesses whether or not the observed event rates match expected event rates in subgroups of the model population, with P-values <0.05 indicating evidence of poor fit. The overall performance of the models was evaluated with the Nagelkerke R^2 and the Brier score. Nagelkerke R^2 is a measure between 0 and 1, with 0 denoting that the model does not explain any variation and 1 denoting that it perfectly explains the observed variation. The Brier score is a measure between 0 and 1 of prediction with the mean squared difference between the predicted probability and the actual outcome. A lower Brier score indicates better performance, and the Brier score for a model can range from 0 for a perfect model to 0.25 for a non-informative model.[25]

Clinical Decision Support Tool—The final single prediction model was then transformed into a CDST, and prognostic scores were calculated for each patient.[26] Cut-points for a low, intermediate, and high probability of response were determined using the GEMINI 2 cohort. Stratifying GEMINI 2 participants based on response status derived these cut-points. The top 25% were considered high probability, bottom 25% considered low probability, and all others intermediate. These cut-points that differentiated the low, intermediate, and high probability groups were then applied to the GEMINI 2 intention-to-treat population to understand how the probability of response with VDZ compared with placebo-treated participants. Finally, the CDST cut-points were applied to the VICTORY consortium, and the sensitivity, specificity, positive likelihood ratio (PLR), and negative likelihood ratio (NLR) of the scoring tool were calculated to identify patients with a low or high probability of responding to VDZ. (Supplementary Material)

Role of Sponsor

The protocol was designed by the study investigators in collaboration with scientists employed by Takeda Pharmaceuticals. All statistical analyses on the GEMINI 2 dataset were

performed by Takeda Pharmaceuticals employees. All statistical analyses on the VICTORY consortium dataset were performed by VICTORY consortium study investigators.

RESULTS

Patient Demographics

Of the 477 CD patients within the VICTORY consortium who were started on VDZ for clinically and/or endoscopically active disease, 104 patients were excluded for lacking baseline C-reactive protein (CRP) data within 4 weeks of initiation of VDZ therapy, and another 36 were excluded for having a follow-up of less than 26 weeks. Characteristics were similar between the entire VICTORY CD consortium cohort and the VICTORY CD consortium cohort selected for validation (Table 1). Baseline age, gender, and BMI were similar between the GEMINI and VICTORY cohorts. Patients in the GEMINI clinical trial had shorter disease duration (9 vs 14 years), and fewer patients had prior bowel surgeries (44 vs 59%) or prior exposure to a TNFα-antagonist (66 vs 92%) (Table 1).

Model Derivation

The univariable models from the GEMINI 2 derivation cohort were used to derive separate models for CREM at Week 26, CSF-REM at Week 26, and sustained CREM (Supplementary Material). The absence of prior exposure to TNFα-antagonist therapy, absence of prior bowel surgery, absence of prior fistulizing disease, baseline albumin, and baseline CRP were included in the final single prediction model (Table 2). Weighted 7-day liquid or very soft stool score was identified as an independent predictor in all 3 individual models, but this was not included in the final single prediction model as the confidence interval closely approached or included 1.00 for estimates, and this variable was felt to be too subjective and therefore at risk for misclassification in routine practice.

Model Validation

The variables were fitted and re-run on the GEMINI 2 cohort to generate a final single weighted model equation (Supplementary Material):

$$Y = -3.0722 + [0.3483 \text{ if no prior TNFα-antagonist exposure}] + [0.2305 \text{ if no prior bowel surgery}] + [0.1979 \text{ if no prior fistulizing disease}] + [0.0436 \times \text{baseline albumin in g/L}] - [0.0098 \times \text{baseline CRP concentration in mg/L}]$$

A sensitivity analysis was performed by weighting the model based on the VICTORY consortium investigator surveys (Supplementary Material):

$$Y = -3.2679 + [0.3048 \text{ if no prior TNFα-antagonist exposure}] + [0.2299 \text{ if no prior bowel surgery}] + [0.2452 \text{ if no prior fistulizing disease}] + [0.0428 \times \text{baseline albumin in g/L}] - [0.0108 \times \text{baseline CRP concentration in mg/L}]$$

These equations were both externally validated on the VICTORY consortium cohort, and both demonstrated fair to good overall performance with no considerable differences between the 2 equations (Table 3, Supplementary Material).

A sensitivity analysis was then performed by replacing albumin and CRP with calculated VDZ clearance profiles for GEMINI 2 participants based on population modeling and exposure-efficacy relationship assessments.[23, 24] Performance of this model within the GEMINI 2 derivation cohort was comparable to the original model that included albumin and CRP (Supplementary Material).

The final single model equation used was therefore based on the 5 selected baseline variables weighted by an estimate of the inverse variance for each coefficient (Table 3).

Clinical Decision Support Tool

The final single model equation was transformed into a CDST and points were assigned to each variable based on multiplication of β coefficient by 10 and rounding to the nearest value. Non-linearity in distribution was observed for CRP and therefore it was transformed into a categorical value for the CDST, and an assessment of accuracy revealed that the modified CDST with CRP as a categorical value had higher sensitivity than the model with CRP as a continuous value (Supplementary Material and Table 4). The final CDST therefore had CRP as a categorical value (Figure 1).

Among the intention-to-treat population of GEMINI 2, the difference in CREM rates between VDZ and placebo at Week 6 was smaller in participants who would have been classified as low probability (13 points) for responding to VDZ (4.7% vs 2.8%; difference 1.9%), compared with participants who would have classified as high probability (>19 points) for responding to VDZ (23.9% vs 14.3%; difference 9.6%). Furthermore, among re-randomized Week 6 responders, the difference in CREM rates at Week 52 between VDZ every 8 weeks and placebo was smaller between the low probability group (25.8% vs 3.6%; difference 22.2%) and the high probability group (61.7 vs 33.3%; difference 28.4%).

Patients in the VICTORY validation cohort were stratified into low (13 points), intermediate (>13 to 19 points), and high (>19 points) probability for responding to VDZ, with good discriminative performance for the CDST (Figure 1). A cut-off of 13 points had high sensitivity for identifying patients who were likely to respond to VDZ and a cut-off of 19 points had a modest-high specificity for identifying patients who were likely to respond to VDZ (Table 4). Treatment outcomes among the VICTORY consortium patients in the low probability group were similar irrespective of number of prior TNFα-antagonists used, prior TNFα-antagonist failure versus intolerance, or concomitant immunosuppressive use (*P*>0.05 for all comparisons).

Among the intermediate (>13 to 19 points) to high probability (>19 points) groups, rates of MH were significantly higher in patients with no prior fistulizing disease (83% vs. 68%, p=0.05), and trended towards significance for those with no prior surgery (55% vs. 39%, p=0.08). When stratified by CRP categories, patients with an intermediate to high probability score had lower rates of CREM (38.1% vs. 29.4% vs 20%, p=0.01), CSF-REM (32.4% vs 20.8% vs. 15.2%, p=0.08), and corticosteroid free deep remission (18% vs 11.8% vs. 8.7%, p=0.32).

DISCUSSION

Prior prediction models in CD have largely focused on estimating an individual patient's probability of developing complications to help decide whether or not highly effective therapeutic approaches, such as early combined immunosuppression with a biologic, would be preferred strategies.[27–30] However, prediction models for treatment outcomes with biologic therapy in CD are limited. Two clinical models have been built for predicting primary non-response to TNFα-antagonists; however, performance of these models was limited by inadequate precision due to low event rates and sample sizes, and lack of external validation.[31, 32] Therefore, high-quality clinical prediction models to guide therapeutic decisions in CD are lacking. We derived and externally validated a multivariable prediction model that accurately predicts CREM, CSF-REM, MH, and deep remission with VDZ therapy in patients with CD. We further transformed this prediction model into an easy to use prognostic CDST, which demonstrated a good predictive value for identifying patients with either a low or high probability of responding to VDZ. Prospective use of the CDST is likely to improve the overall cost-effectiveness of biologic therapy in CD, by identifying patients who will most likely respond to VDZ and allow therapies to be selected accordingly. Furthermore, discussion of the results of the CDST with patients offers providers an opportunity to more readily engage in shared decision-making and personalization of treatment decisions.

Based on our model, the optimal positioning of VDZ within current treatment paradigms is prior to TNFα-antagonist exposure (supporting its use as a first-line biological agent), or early in the disease course prior to the development of disease-related complications (eg, fistulizing disease, bowel surgery). A risk score of 13 points had a very high sensitivity for identifying those less likely to respond to therapy, and a score of > 19 points had a modest specificity for identifying those more likely to respond to therapy. Among patients with an intermediate (>13 to 19 points) or high (>19 points) probability of response to VDZ, having an elevated baseline CRP (> 3 mg/L) was associated with lack of response to VDZ and was the predominant factor for misclassification in these groups. These observations could be due to several reasons including: an inherent resistance to responding to any biologic therapy in these refractory patients,[33] alterations in immune phenotype with prior TNFα-antagonist exposure,[34–36] or variations in pharmacokinetics and VDZ drug exposure. [23, 24] To address the latter of these, we performed a sensitivity analysis by including individual participant VDZ clearance profiles based on population modeling, and observed no differences in model performance or accuracy for predicting outcomes. This would suggest that these variables are not simply surrogates for VDZ clearance and exposure but rather true prognostic markers of therapeutic response.

The strengths of this study include its size, scope, and external validation in an independent dataset. The derivation and validation of the data from multiple sites and jurisdictions, along with the simplicity of the CDST, enhances its generalizability and performance. In addition, the model is able to simultaneously predict outcomes of importance to patients (CREM and CSF-REM at 26 weeks) as well as outcomes associated with improved disease course (MH and deep remission) after initiation of VDZ, which are of importance to patients and physicians. Several limitations should be acknowledged. Firstly, these were post hoc

analyses of the GEMINI 2 clinical trial dataset, which was not powered for subgroup analyses and did not have endoscopic confirmation of efficacy outcomes. Caution should be taken when interpreting the differences in outcomes across subgroups and the results of the intention-to-treat population results stratified by probability of response. Secondly, although the prediction model had a favorable diagnostic accuracy, it explains less than 10% of the variability in outcomes. Prior work has demonstrated that the combination of serologic or genetic markers may help to improve the diagnostic accuracy of prediction models in CD. [29, 32] Therefore, future studies are still needed to determine if biomarkers can be identified to help improve the accuracy of our clinical prediction model, as well as its ability to explain variability in outcomes with VDZ. Thirdly, the model had a poor fit when validated in the VICTORY consortium. The poor fit of the model may be due in part to the variability in outcome definitions, known risk of disease activity classification in CD with symptom based indices, or the combining of several models into a single model for ease of clinical interpretation and integration. Although the model exhibited poor calibration for matching predictive and observed probabilities throughout the whole range of potential prognostic estimates, the CDST demonstrated a high sensitivity for identifying those at risk for treatment failure. This highlights that predictive models and CDSTs are simply one piece of a multi-step process that includes physician assessments and patient preferences, to choose the most appropriate therapies and optimize treatment effectiveness on an individualized basis in routine practice. Future studies are needed to better calibrate the model and assess the impact of this prediction model and CDST on clinical practice, health care resource utilization, and treatment outcomes, when implemented at the population level.

In summary, we have derived and externally validated a clinical prediction model for treatment outcomes with VDZ, with good overall performance and accuracy. We have further transformed this prediction model into an easy to use CDST to help guide patients and providers on the appropriate positioning of VDZ within current treatment algorithms. Based on our model the ideal positioning of VDZ may be early in the disease course prior to exposure to TNF-antagonists and prior to the development of CD related complications. Further studies are needed to determine how our prediction model can be used to optimize treatment outcomes at the population level.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

Abbreviations:

AUC	area under the curve

CDST	clinical decision support tool
CI	confidence interval
CREM	clinical remission
CRP	C-reactive protein
CSF-REM	corticosteroid-free remission
MH	mucosal healing
NLR	negative likelihood ratio
PLR	positive likelihood ratio
ROC	receiver operating characteristic
TNFα	tumor necrosis factor-alpha
VDZ	vedolizumab

References

1. Sandborn WJ, Feagan BG, Rutgeerts P, Hanauer S, Colombel JF, Sands BE, Lukas M, Fedorak RN, Lee S, Bressler B, Fox I, Rosario M, Sankoh S, Xu J, Stephens K, Milch C, Parikh A, Group GS. Vedolizumab as induction and maintenance therapy for Crohn's disease. N Engl J Med 2013;369:711–21. [PubMed: 23964933]

2. Ha C, Ullman TA, Siegel CA, Kornbluth A. Patients enrolled in randomized controlled trials do not represent the inflammatory bowel disease patient population. Clin Gastroenterol Hepatol 2012;10:1002–1007. [PubMed: 22343692]

3. Dulai PS, Singh S, Jiang X, Peerani F, Narula N, Chaudrey K, Whitehead D, Hudesman D, Lukin D, Swaminath A, Shmidt E, Wang S, Boland BS, Chang JT, Kane S, Siegel CA, Loftus EV, Sandborn WJ, Sands BE, Colombel JF. The Real-World Effectiveness and Safety of Vedolizumab for Moderate-Severe Crohn's Disease: Results From the US VICTORY Consortium. Am J Gastroenterol 2016;111:1147–55. [PubMed: 27296941]

4. Stallmach A, Langbein C, Atreya R, Bruns T, Dignass A, Ende K, Hampe J, Hartmann F, Neurath MF, Maul J, Preiss JC, Schmelz R, Siegmund B, Schulze H, Teich N, von Arnim U, Baumgart DC, Schmidt C. Vedolizumab provides clinical benefit over 1 year in patients with active inflammatory bowel disease - a prospective multicenter observational study. Aliment Pharmacol Ther 2016;44:1199–1212. [PubMed: 27714831]

5. Sands BE, Feagan BG, Rutgeerts P, Colombel JF, Sandborn WJ, Sy R, D'Haens G, Ben-Horin S, Xu J, Rosario M, Fox I, Parikh A, Milch C, Hanauer S. Effects of vedolizumab induction therapy for patients with Crohn's disease in whom tumor necrosis factor antagonist treatment failed. Gastroenterology 2014;147:618–627.e3. [PubMed: 24859203]

6. Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. Ann Intern Med 2006;144:201–9. [PubMed: 16461965]

7. Toll DB, Janssen KJ, Vergouwe Y, Moons KG. Validation, updating and impact of clinical prediction rules: a review. J Clin Epidemiol 2008;61:1085–94. [PubMed: 19208371]

8. Dulai PS, Singh S, Casteele NV, Boland BS, Sandborn WJ. How Will Evolving Future Therapies and Strategies Change How We Position the Use of Biologics in Moderate to Severely Active Inflammatory Bowel Disease. Inflamm Bowel Dis 2016;22:998–1009. [PubMed: 26835982]

9. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med 2015;162:55–63. [PubMed: 25560714]

10. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. Bmj 2009;338:b605. [PubMed: 19477892]

11. Best WR, Becktel JM, Singleton JW, Kern F, Jr. Development of a Crohn's disease activity index. National Cooperative Crohn's Disease Study. Gastroenterology 1976;70:439–44. [PubMed: 1248701]

12. Peyrin-Biroulet L, Sandborn W, Sands BE, Reinisch W, Bemelman W, Bryant RV, D'Haens G, Dotan I, Dubinsky M, Feagan B, Fiorino G, Gearry R, Krishnareddy S, Lakatos PL, Loftus EV, Jr., Marteau P, Munkholm P, Murdoch TB, Ordas I, Panaccione R, Riddell RH, Ruel J, Rubin DT, Samaan M, Siegel CA, Silverberg MS, Stoker J, Schreiber S, Travis S, Van Assche G, Danese S, Panes J, Bouguen G, O'Donnell S, Pariente B, Winer S, Hanauer S, Colombel JF Selecting Therapeutic Targets in Inflammatory Bowel Disease (STRIDE): Determining Therapeutic Goals for Treat-to-Target. Am J Gastroenterol 2015;110:1324–38. [PubMed: 26303131]

13. Bleeker SE, Moll HA, Steyerberg EW, Donders AR, Derksen-Lubsen G, Grobbee DE, Moons KG. External validation is necessary in prediction research: a clinical example. J Clin Epidemiol 2003;56:826–32. [PubMed: 14505766]

14. Bursac Z, Gauss CH, Williams DK, Hosmer DW. Purposeful selection of variables in logistic regression. Source Code Biol Med 2008;3:17. [PubMed: 19087314]

15. Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. Stat Med 2016;35:214–26. [PubMed: 26553135]

16. Ogundimu EO, Altman DG, Collins GS. Adequate sample size for developing prediction models is not simply related to events per variable. J Clin Epidemiol 2016;76:175–82. [PubMed: 26964707]

17. Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. Am J Epidemiol 2007;165:710–8. [PubMed: 17182981]

18. Wynants L, Bouwmeester W, Moons KG, Moerbeek M, Timmerman D, Van Huffel S, Van Calster B, Vergouwe Y. A simulation study of sample size demonstrated the importance of the number of events per variable to develop prediction models in clustered data. J Clin Epidemiol 2015;68:1406–14. [PubMed: 25817942]

19. Mocko P, Kawalec P, Smela-Lipinska B, Pilc A. Effectiveness and safety of vedolizumab for treatment of Crohn's disease: a systematic review and meta-analysis. Arch Med Sci 2016;12:1088–1096. [PubMed: 27695501]

20. Chandar AK, Singh S, Murad MH, Peyrin-Biroulet L, Loftus EV, Jr. Efficacy and Safety of Natalizumab and Vedolizumab for the Management of Crohn's Disease: A Systematic Review and Meta-analysis. Inflamm Bowel Dis 2015;21:1695–708. [PubMed: 25856769]

21. Pennells L, Kaptoge S, White IR, Thompson SG, Wood AM. Assessing risk prediction models using individual participant data from multiple studies. Am J Epidemiol 2014;179:621–32. [PubMed: 24366051]

22. Marin-Martinez F, Sanchez-Meca J. Weighting by inverse variance or by sample size in random-effects meta-analysis. Educ Psychol Meas 2010;70:56–73.

23. Rosario M, French JL, Dirks NL, Sankoh S, Parikh A, Yang H, Danese S, Colombel JF, Smyth M, Sandborn WJ, Feagan BG, Reinisch W, Sands BE, Sans M, Fox I. Exposure-efficacy relationships for vedolizumab induction therapy in patients with ulcerative colitis or Crohn's disease. J Crohns Colitis 2017;11:921–29. [PubMed: 28333288]

24. Rosario M, Dirks NL, Gastonguay MR, Fasanmade AA, Wyant T, Parikh A, Sandborn WJ, Feagan BG, Reinisch W, Fox I. Population pharmacokinetics-pharmacodynamics of vedolizumab in patients with ulcerative colitis and Crohn's disease. Aliment Pharmacol Ther 2015;42:188–202. [PubMed: 25996351]

25. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, Pencina MJ, Kattan MW. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology 2010;21:128–38. [PubMed: 20010215]

26. Hu MY, Katchar K, Kyne L, Maroo S, Tummala S, Dreisbach V, Xu H, Leffler DA, Kelly CP. Prospective derivation and validation of a clinical prediction rule for recurrent Clostridium difficile infection. Gastroenterology 2009;136:1206–14. [PubMed: 19162027]

27. Nos P, Hinojosa J, Mora J, Garrigues V, Ponce J. Validation of a simplified clinical index to predict evolving patterns in Crohn's disease. Eur J Gastroenterol Hepatol 2002;14:847–51. [PubMed: 12172404]

28. Lichtenstein GR, Targan SR, Dubinsky MC, Rotter JI, Barken DM, Princen F, Carroll S, Brown M, Stachelski J, Chuang E, Landers CJ, Stempak JM, Singh S, Silverberg MS. Combination of genetic and quantitative serological immune markers are associated with complicated Crohn's disease behavior. Inflamm Bowel Dis 2011;17:2488–96. [PubMed: 21391291]

29. Siegel CA, Horton H, Siegel LS, Thompson KD, Mackenzie T, Stewart SK, Rice PW, Stempak JM, Dezfoli S, Haritunians T, Levy A, Baek M, Milgrom R, Dulai PS, Targan SR, Silverberg MS, Dubinsky MC, McGovern DP. A validated web-based tool to display individualised Crohn's disease predicted outcomes based on clinical, serologic and genetic variables. Aliment Pharmacol Ther 2016;43:262–71. [PubMed: 26567467]

30. Dias CC, Rodrigues PP, Coelho R, Santos PM, Fernandes S, Lago P, Caetano C, Rodrigues A, Portela F, Oliveira A, Ministro P, Cancela E, Vieira AI, Barosa R, Cotter J, Carvalho P, Cremers I, Trabulo D, Caldeira P, Antunes A, Rosa I, Moleiro J, Peixe P, Herculano R, Goncalves R, Goncalves B, Sousa HT, Contente L, Morna H, Lopes S, Magro F. Development and Validation of Risk Matrices for Crohn's Disease Outcomes in Patients Who Underwent Early Therapeutic Interventions. J Crohns Colitis 2016.

31. Billiet T, Papamichael K, de Bruyn M, Verstockt B, Cleynen I, Princen F, Singh S, Ferrante M, Van Assche G, Vermeire S. A Matrix-based Model Predicts Primary Response to Infliximab in Crohn's Disease. J Crohns Colitis 2015;9:1120–6. [PubMed: 26351386]

32. Barber GE, Yajnik V, Khalili H, Giallourakis C, Garber J, Xavier R, Ananthakrishnan AN. Genetic Markers Predict Primary Non-Response and Durable Response To Anti-TNF Biologic Therapies in Crohn's Disease. Am J Gastroenterol 2016;111:1816–1822. [PubMed: 27596696]

33. Gisbert JP, Marin AC, McNicholl AG, Chaparro M. Systematic review with meta-analysis: the efficacy of a second anti-TNF in patients with inflammatory bowel disease whose previous anti-TNF treatment has failed. Aliment Pharmacol Ther 2015;41:613–23. [PubMed: 25652884]

34. Biancheri P, Di Sabatino A, Rovedatti L, Giuffrida P, Calarota SA, Vetrano S, Vidali F, Pasini A, Danese S, Corazza GR, MacDonald TT. Effect of tumor necrosis factor-alpha blockade on mucosal addressin cell-adhesion molecule-1 in Crohn's disease. Inflamm Bowel Dis 2013;19:259–64. [PubMed: 23328772]

35. Danese S, Panes J. Development of drugs to target interactions between leukocytes and endothelial cells and treatment algorithms for inflammatory bowel diseases. Gastroenterology 2014;147:981–9. [PubMed: 25220794]

36. Chowers Y, Sturm A, Sans M, Papadakis K, Gazouli M, Harbord M, Jahnel J, Mantzaris GJ, Meier J, Mottet C, Peyrin-Biroulet L, Allez M. Report of the ECCO workshop on anti-TNF therapy failures in inflammatory bowel diseases: biological roles and effects of TNF and TNF antagonists. J Crohns Colitis 2010;4:367–76. [PubMed: 21122531]

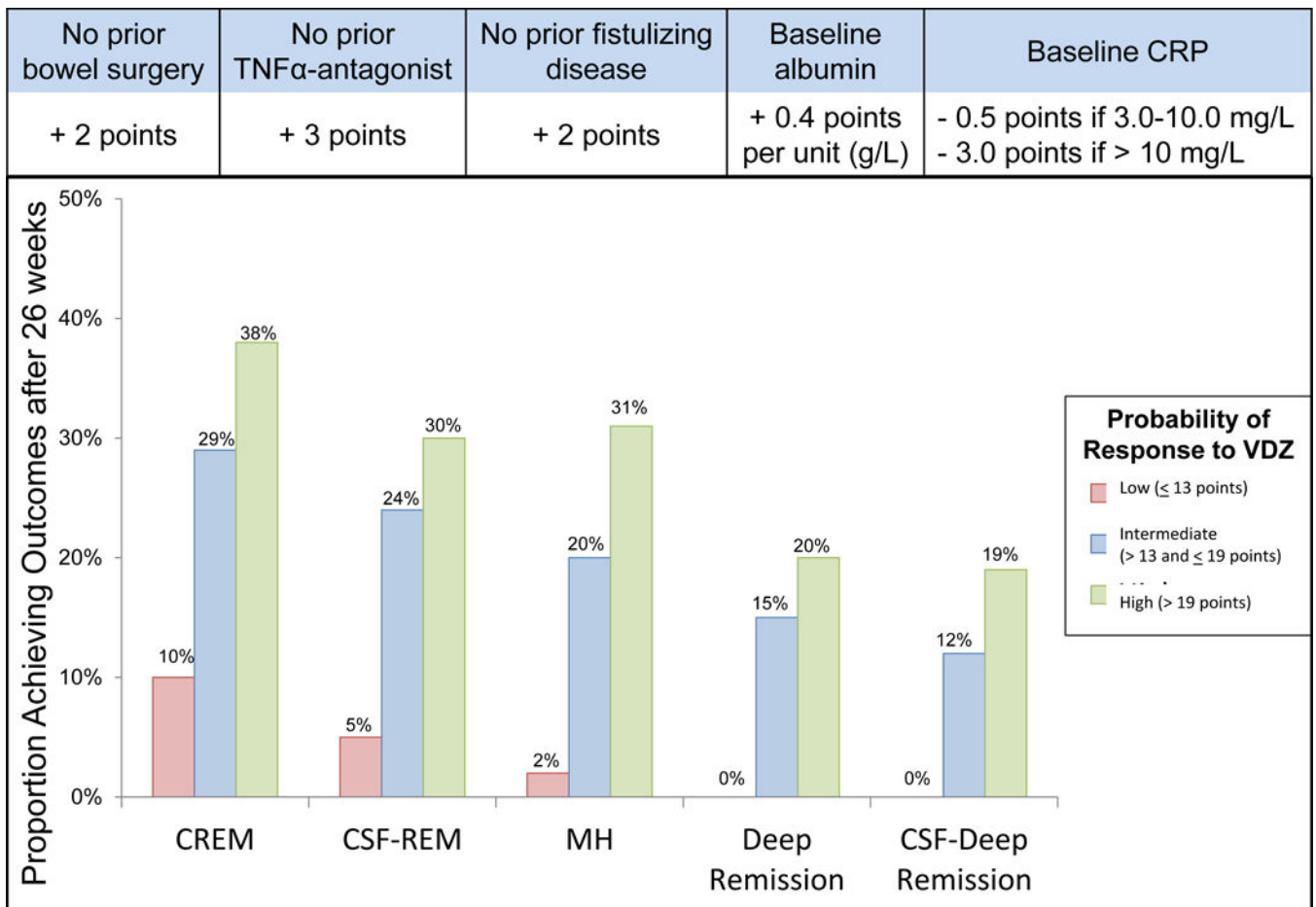| No prior bowel surgery	No prior TNFα-antagonist	No prior fistulizing disease	Baseline albumin	Baseline CRP
+ 2 points	+ 3 points	+ 2 points	+ 0.4 points per unit (g/L)	- 0.5 points if 3.0-10.0 mg/L - 3.0 points if > 10 mg/L

Figure 1.
Prognostic Clinical Decision Support Tool with stratified treatment outcomes in VICTORY consortium. CD, Crohn's disease; CRP, C-reactive protein; CREM, clinical remission; CSF-REM, corticosteroid-free remission; ds, disease; MH, mucosal healing; CSF-Deep Remission, corticosteroid-free deep remission; TNF, tumor necrosis factor; VDZ, vedolizumab.

Table 1.

Comparison of Demographics Between GEMINI 2 and VICTORY Consortium Cohorts

	GEMINI 2	VICTORY consortium	
	Overall cohort (n=814)	Overall cohort (n=477)	Validation cohort (n=336)
Female sex, n (%)	435 (53)	279 (58)	192 (57)
Mean age, years (SD)	35.5 (11.9)	39.3 (15.4)	38.1 (15.1)
Mean BMI, kg/m^2 (SD)	24.0 (6.0)	25.4 (6.8)	25.5 (6.9)
Mean disease duration, years (SD)	9.1 (7.5)	14.4 (11)	14.1 (10.6)
Concomitant CS only, n (%)	280 (34)	133 (28)	96 (28)
Concomitant IMMs only, n (%)	133 (16)	90 (19)	66 (20)
Concomitant CS and IMMs, n (%)	137 (17)	105 (22)	75 (22)
Prior TNFα antagonist exposure, n (%)	535 (66)	441 (92)	309 (92)
Prior TNFα antagonist failure, n (%)	497 (61)	374 (78)	263 (78)
Median CRP, mg/L (IQR)	10.6 (4.5–31.6)	–	5.0 (1.0–19.4)
Mean albumin, g/L (SD)	34.9 (5.7)	–	38.2 (5.7)
Disease localization, n (%) [*] Ileum only Colon only Ileocolonic	141 (17) 230 (28) 443 (54)	78 (16) 108 (23) 291 (61)	48 (14) 76 (23) 212 (63)
Prior surgery for CD, n (%)	355 (44)	287 (60)	199 (59)
Prior fistulizing disease, n (%)	297 (36)	183 (38)	136 (40)

[*] As per patient history at the time of enrolment into GEMINI 2 or the initiation of vedolizumab therapy in the VICTORY consortium.

BMI, body mass index; CD, Crohn's disease; CRP, C-reactive protein; CS, corticosteroid; IMM, immunomodulator; SD, standard deviation; TNFα, tumor necrosis factor-alpha.

Table 2.

Multivariable Analyses for Remission in the GEMINI 2 Derivation Cohort

Independent Predictor*	OR†	Confidence Interval
Clinical Remission		
No prior CD related hospitalization*	1.43	1.08,1.89
No baseline EIM	1.43	1.09, 1.88
No previous bowel surgery	1.49	1.12,1.98
No previous TNFα antagonist	2.18	1.65, 2.88
Baseline CRP	0.99	0.98, 1.00
Ethnicity (Other vs. Non-hispanic/Latino)	0.30	0.12, 0.75
Weighted 7-day liquid or very soft stool	0.99	0.99, 1.00
Baseline Albumin	1.04	1.01,1.08
Corticosteroid-Free Remission		
No history of fistulizing disease	1.85	1.13, 3.03
No previous TNFα antagonist	1.80	1.16, 2.80
Baseline CRP	0.98	0.97, 1.00
Weighted 7-day liquid or very soft stool	0.99	0.99, 1.00
Sustained Clinical Remission		
No history of fistulizing disease	1.82	1.22, 2.71
No previous bowel surgery	1.71	1.18, 2.49
No previous TNFα antagonist	1.42	1.01, 2.00
Baseline Albumin	1.07	1.04,1.10
Weighted 7-day liquid or very soft stool	0.995	0.99, 1.00

*CD-related hospitalization within the previous 12 months.

CRP, C-reactive protein; OR, odds ratio; TNFα, tumor necrosis factor-alpha.

Table 3.

Performance of Final Single Prediction Model in VICTORY Consortium Validation Cohort

	Clinical remission 26 weeks	Corticosteroid-free remission 26 weeks[*]	Mucosal healing 26 weeks	Deep Remission 26 weeks	Corticosteroid-free Deep Remission 26 weeks[*]
Nagelkerke R^2	0.06	0.06	0.07	0.06	0.07
Brier score	0.24	0.18	0.16	0.11	0.09
ROC-AUC (95% CI)	0.67 (0.63, 0.73)	0.66 (0.57, 0.76)	0.72 (0.64, 0.80)	0.73 (0.64, 0.81)	0.75 (0.64, 0.86)

Hosmer–Lemeshow goodness-of-fit test: The model-exhibited poor fit (P<0.05) for all outcomes.

[*] Performed for subset of patients who were on corticosteroids at baseline.

Final single prediction model based on inverse variance weighting and includes: no prior bowel surgery, no prior TNFα-antagonist exposure, no history of fistulizing disease, baseline albumin, and baseline CRP score. Nagelkerke R-squared is a measure between 0 and 1, with 0 denoting that model does not explain any variation and 1 denoting that it perfectly explains the observed variation. The Brier score is a measure between 0 and 1 of prediction with the mean squared difference between the predicted probability and the actual outcome. The Brier score for a model can range from 0 for a perfect model to 0.25 for a non-informative model. ROC-AUC curve values are between 0.5 and 1, with 0.5 denoting that the model does not discriminate and 1 denoting that it perfectly discriminates. In the Hosmer–Lemeshow goodness-of-fit test, observed event rates are tested against expected event rates by decile of fitted values for prediction; P-values <0.05 indicate evidence of poor fit.

AUC, area under the curve; CD, Crohn's disease; CRP, C-reactive protein; ROC, receiver operating characteristic; CI: confidence interval.

Table 4.

Diagnostic Accuracy of Prognostic Clinical Decision Support Tool for Identifying Patients Likely to Respond to Vedolizumab using CRP as categorical value

	Sensitivity (95% CI)	Specificity (95% CI)	PLR (95% CI)	NLR (95% CI)
13 points				
Clinical remission after 26 weeks	92% (85–97%)	25% (20–31%)	1.24 (1.12–1.36)	0.30 (0.14–0.64)
Corticosteroid-free remission after 26 weeks [*]	94% (81–99%)	30% (22–38%)	1.34 (1.17–1.54)	0.19 (0.05–0.76)
Mucosal healing after 26 weeks	98% (88–100%)	30% (23–37%)	1.39 (1.25–1.54)	0.08 (0.01–0.55)
Deep Remission after 26 weeks	100% (88–100%)	28% (22–35%)	1.39 (1.28–1.51)	0.00 (–)
Corticosteroid-Free Deep Remission after 26 weeks [*]	100% (75–100%)	31% (23–41%)	1.45 (1.28–1.65)	0.00 (–)
19 points				
Clinical remission after 26 weeks	33% (24–44%)	80% (74–85%)	1.65 (1.12–2.42)	0.84 (0.72–0.98)
Corticosteroid-free remission after 26 weeks [*]	37% (22–55%)	77% (69–84%)	1.62 (0.95–2.75)	0.82 (0.62–1.07)
Mucosal healing after 26 weeks	40% (25–56%)	80% (74–86%)	2.01 (1.26–3.20)	0.75 (0.59–0.97)
Deep Remission after 26 weeks	38% (21–58%)	79% (73–84%)	1.78 (1.05–3.04)	0.79 (0.59–1.06)
Corticosteroid-Free Deep Remission after 26 weeks [*]	46% (19–75%)	78% (69–85%)	2.07 (1.05–4.09)	0.69 (0.42–1.16)

[*] Performed for subset of patients who were on corticosteroids at baseline. Stratified analysis by baseline steroid dose performed with no significant changes in performance.

PLR, positive likelihood ratio; NLR, negative likelihood ratio.