

# Assessing Clinical Disease Recurrence Using Laboratory Data in Surgically Resected Patients From the TOPPIC Trial

Akbar K. Waljee, MD, MS,<sup>\*†‡¶||</sup> Shirley Cohen-Mekelburg, MD, MS,<sup>\*†¶||</sup> Yumu Liu, MS,<sup>§</sup> Boang Liu, MS,<sup>§</sup> Ji Zhu, PhD,<sup>§</sup> and Peter D.R. Higgins, MD, PhD, MSc<sup>†</sup>

**Background:** Machine learning methodologies play an important role in predicting progression of disease or responses to medical therapy. We previously derived and validated a machine learning algorithm to predict response to thiopurines in an inflammatory bowel disease population. We aimed to apply a modified algorithm to predict postsurgical treatment response using clinical trial data.

**Methods:** TOPPIC was a multicenter randomized double-blinded placebo-controlled trial of 240 patients, evaluating the effectiveness of 6-mercaptopurine in preventing or delaying postsurgical Crohn disease recurrence. We adapted a well-established machine learning algorithm to predict clinical recurrence postresection using age and multiple laboratory-specific covariates, and compared this to the thiopurine metabolite, 6-thioguanine.

**Results:** The random forest machine learning algorithm demonstrates a mean under the receiver operator curve (AuROC) of 0.62 [95% confidence interval (CI) 0.47, 0.78]. Similar results were evident when adding thiopurine metabolite (6-thioguanine) results. Alanine aminotransferase/mean corpuscular volume (ALT/MCV) and potassium  $\times$  alkaline phosphatase (POT  $\times$  ALK) predicted endoscopic and biologic recurrence, respectively, with AuROCs of 0.714 (95% CI 0.601, 0.827) and 0.730 (95% CI 0.618, 0.841).

**Conclusions:** A machine learning algorithm with laboratory data from within the first 3 months postsurgically does not discriminate clinical recurrence well. Alternative noninvasive measures should be considered and further evaluated.

## Lay Summary

We applied a machine learning algorithm to predict postsurgical treatment response using clinical trial data. However, unlike medical treatment response, a machine learning algorithm did not discriminate clinical recurrence well. Other noninvasive methods of monitoring postsurgical recurrence are necessary.

**Key Words:** machine learning, inflammatory bowel disease, surgery

## INTRODUCTION

Thiopurines are common worldwide and continue to play an important role in the treatment of patients with moderate-to-severe ulcerative colitis and Crohn disease.<sup>1,2</sup> Thiopurines have a narrow therapeutic window, with shunting leading to the formation of toxic metabolites, and adverse effects often leading to drug withdrawal or subtherapeutic drug levels and incomplete treatment response. The importance of drug optimization has become clearer in the last decade with

therapeutic drug monitoring becoming more commonplace in clinical practice.<sup>3</sup>

Predictive modeling using innovative machine learning methodologies has demonstrated an important role in medicine in predicting progression of disease or responses to medical therapy.<sup>4-9</sup> We previously derived a machine learning algorithm to predict clinical response [under the receiver operator curve (AuROC) 0.856; 95% confidence interval (CI) 0.793–0.919] and objective remission (AuROC 0.79; 95% CI 0.78, 0.81) to

Received for publications July 16, 2020; Editorial Decision August 27, 2020.

\*VA Center for Clinical Management Research, VA Ann Arbor Health Care System, Ann Arbor, Michigan, USA; †Department of Internal Medicine, Division of Gastroenterology and Hepatology, Michigan Medicine, Ann Arbor, Michigan, USA; ‡Michigan Integrated Center for Health Analytics and Medical Prediction (MiCHAMP), Ann Arbor, Michigan, USA; §Department of Statistics, University of Michigan, Ann Arbor, Michigan, USA

¶Co-first authors.

Funding: No funding was received for this work.

Conflict of Interest: The Regents of the University of Michigan, along with authors Peter Higgins, Akbar Waljee and Ji Zhu, have a patent on the application of machine learning algorithms using the complete blood count and differential and the comprehensive chemistry panel to the prediction of clinical response, adherence

and shunting among patients on thiopurines. The patent was granted 2/28/2012. The remaining authors disclose no conflicts of interest.

Address correspondence to: Akbar K. Waljee, MD, MSc, Division of Gastroenterology, Department of Internal Medicine, Michigan Medicine, Ann Arbor Veterans Affairs Medical Center, 2215 Fuller Road, 111D, Ann Arbor, MI 48105 ([awaljee@med.umich.edu](mailto:awaljee@med.umich.edu)).

Published by Oxford University Press on behalf of Crohn's & Colitis Foundation 2020. This work is written by (a) US Government employee(s) and is in the public domain in the US.

This Open Access article contains public sector information licensed under the Open Government Licence v2.0 (<http://www.nationalarchives.gov.uk/doc/open-government-licence/version/2/>).

doi: 10.1093/crocol/otaa088  
Published online 23 October 2020

thiopurines and internally validated our model in a single-center inflammatory bowel disease population.<sup>8,10</sup> The top 10 variables for each of these models are reported in [Supplementary Table 1](#). We externally validated our model utilizing clinical trial data from SONIC, a landmark clinical trial of azathioprine, infliximab, or combination infliximab and azathioprine, which demonstrated an AuROC of 0.764 (95% CI 0.602, 0.926) for objective remission on azathioprine monotherapy.<sup>11</sup> While work to date has demonstrated the effectiveness of using a machine learning algorithm to predict treatment response, this model has not been applied to postsurgical management.

The TOPPIC trial evaluated the efficacy of 6-mercaptopurine in preventing or delaying postsurgical Crohn disease recurrence using a cohort of patients in the United Kingdom.<sup>12</sup> 6-Mercaptopurine protected against clinical recurrence among smokers with an adjusted hazard ratio of 0.13 (95% CI 0.04–0.46). We aimed to adapt an established machine learning algorithm to predict postsurgical clinical, endoscopic, and biologic disease recurrence using TOPPIC clinical trial data.

## METHODS

Data from TOPPIC were made available through the University of Edinburgh through a data-sharing agreement. TOPPIC was a multicenter randomized double-blinded placebo-controlled trial of 240 patients, evaluating the effectiveness of 6-mercaptopurine in preventing or delaying postsurgical Crohn disease recurrence. The initial trial randomized 128 patients to azathioprine and 120 patients to placebo. The primary outcome was clinical recurrence as defined by a Crohn's disease activity index (CDAI) of >150 or an increase in CDAI by 100 points requiring medical or surgical treatment. Among smokers, 10% in the 6-mercaptopurine treated arm and 45% in the placebo arm developed a clinical recurrence. Secondary endpoints included endoscopic recurrence, defined by a Rutgeerts score of  $\geq 2$  or greater, with 43% in the 6-mercaptopurine treated arm and 49% in the placebo arm achieving this endpoint.<sup>12</sup> The study also evaluated Crohn's Disease Endoscopic Index of Severity (CDEIS) score.<sup>12</sup>

We focused on the treated patients with a valid unique visit 3 (week 6) postrandomization record. We tested predictions of 5 different main outcomes that represent markers of clinical or objective disease recurrence, based on the original TOPPIC study<sup>12</sup>: Rutgeerts score, CDAI, C-reactive protein (CRP), CDEIS, and fecal calprotectin. A CDAI >150 was defined as a clinical recurrence. A Rutgeerts score of  $\geq 2$  or greater, a CDEIS of greater than 3, or a fecal calprotectin >100 was consistent with endoscopic recurrence. Biologic recurrence was defined by a CRP >5 mg/L. We adapted our institutional derived and externally validated algorithm to TOPPIC data making use of available variables. The predictive model was constructed to incorporate age and multiple laboratory-specific covariates, including white blood cell count (WBC), hemoglobin (HGB), hematocrit (HCT), mean corpuscular volume (MCV), sodium

(SOD), potassium (POT), blood urea nitrogen (BUN), creatinine (CR), albumin (ALB), alanine aminotransferase (ALT), alkaline phosphatase (ALK), total bilirubin (TB), basophil count, eosinophil count, monocyte count, neutrophil count, and red cell count. We also included combinations of highly important variables (eg, products) based on prior clinical experience and published data.<sup>13</sup> In addition, 6-thioguanine (6-TGN) at visit 4 (week 13) was tested as a predictor.

## Statistical Analysis

We compared 3 different statistical approaches: (1) a random forest model built from a previously defined cohort, (2) a random forest model built using TOPPIC data, and (3) a lasso penalized logistic regression model built using TOPPIC data. Model parameters were set with the number of trees was set to 500 and the number of candidate variables at each split was  $\sqrt{p}$  where  $p$  is the number of variables. For each of the 5 outcomes (Rutgeerts, CDAI, CRP, CDEIS, and fecal calprotectin), we first trained a random forest model on a previously published dataset, limited to only the variables available in the TOPPIC dataset, and then used this model to give each subject in the TOPPIC cohort a predicted probability of achieving each outcome.<sup>8</sup> An AuROC was calculated to evaluate discrimination for the defined outcomes. Next, we removed the observations with missing outcomes and directly used the TOPPIC dataset to train random forest models for the defined outcomes, constructed the Receiver operating curve (ROC) curves, and calculated the out-of-bag AuROC to compare model performance. Finally, we randomly split the data into a training set incorporating 70% of the data and a testing set incorporating the remaining 30% of the data, stratified by response. We then trained a lasso penalized logistic regression model on the training set for each outcome variable separately, with tuning parameters selected by cross-validation within the training set, and computed the AuROC under the testing set. This procedure was repeated 100 times and the mean AuROC over the 100 replications reported. At each replication, we compared the AuROC curves of our machine learning algorithm to a single predictor model with 6-TGN using Delong test for ROC curves. We reported the median  $P$  value. This study was reviewed by the University of Michigan IRB and was considered exempt from IRB review. All analyses were performed using R statistical software.

## RESULTS

The final TOPPIC dataset contained 117 subjects, after 5 were removed due to missing lab values. This cohort consisted of 41.1% males, with a median age of 38.9 years old, and a median weight of 70.3 kg ([Table 1](#)). The AuROC was used to evaluate algorithm performance. Our machine learning algorithm based on lasso penalized logistic regression predicted endoscopic remission poorly by Rutgeerts score (AuROC 0.58; 95% CI 0.50–0.72), CDEIS (AUC 0.60; 95% CI 0.51–0.68), and fecal calprotectin (AuROC of 0.60; 95% CI 0.52–0.80). It predicted clinical remission

**TABLE 1. Descriptive Characteristics (N = 112)**

Demographic	Value
Male gender (n, %)	46 (41.1%)
Age (years) (median, IQR)	38.9 (28.5–50.2)
Smoking (n, %)	86 (76.8%)
Weight (kg) (median, IQR)	70.3 (59.4–82.4)
Previous surgery (n, %)	31 (27.7%)
Duration of CD (years) (median, IQR)	6.4 (0.6–9.5)
	27 (24.1%) missing

CD, Crohn's disease; IQR, interquartile range.

poorly by CDAI with an AuROC of 0.61 (95% CI 0.51–0.71) and biologic remission poorly by CRP with an AuROC of 0.62 (95% CI 0.52–0.76). The random forest models built from the original cohort and the TOPPIC cohort had similar or poorer performances. These results are presented in [Table 2](#).

Separately, 6-TGN at visit 4 predicted clinical remission by CDAI poorly with an AuROC of 0.55 (95% CI 0.41–0.70). It also poorly predicted endoscopic remission by Rutgeerts score (AuROC 0.55; 95% CI 0.43–0.68), CDEIS (AuROC of 0.60; 95% CI 0.47–0.73), fecal calprotectin (AuROC of 0.52; 95% CI 0.39–0.66), and biological remission by CRP with an AuROC of 0.52 (95% CI 0.36–0.67) ([Table 2](#)). The *P* values ([Table 2](#)) suggest that the ROC curves using 6-TGN as predictor did not significantly differ from the lasso penalized logistic regression models. On the other hand, the PLT × ALT product predicted clinical recurrence by CDAI with an AuROC of 0.714 (95% CI 0.602, 0.825), and the ALT/MCV ratio predicted endoscopic recurrence by CDEIS with an AuROC: 0.714 (95% CI 0.601, 0.827). Finally, the POT × ALK product predicted biologic recurrence with an AuROC: 0.730 (95% CI 0.618, 0.841). We evaluated the performances using out-of-bag AuROC for the random forest models and testing set AuROC under random splitting for the lasso penalized regression model while tuning using cross-validation within the training set. These techniques give fair evaluation of the performances of the models and reduce overfitting. Variable importance tables for each of the random forest models built for the 5 outcomes are reported in [Figure 1](#).

### DISCUSSION

While we previously demonstrated that a machine learning algorithm can predict response to thiopurines, this does not hold true for predicting postsurgical outcomes when thiopurines are utilized using a prophylactic strategy. In fact, laboratory values in the first 3 months of a surgical resection are not predictive of postsurgical clinical or endoscopic recurrence among patients with Crohn disease.

Further, the AuROC does not suggest good discrimination of the 5 main outcomes. The patients in this study are largely in surgically induced remission, so that the effects of

**TABLE 2. AuROC by Outcome**

Response Measure	Success Definition	Missing N's	Number Success	Original Model AuROC	TOPPIC Random Forest	*TOPPIC Lasso Penalized Logistic Regression	Mean <i>P</i> for Comparing Lasso Penalized Logistic Regression and Visit 4 6-TGN	
							6-TGN at Visit 4 AuROC	Visit 4 6-TGN
Clinical remission	≤150	11	101	0.55 (0.39, 0.71)	0.50 (0.35, 0.64)	0.61 (0.51, 0.71)	0.55 (0.41, 0.70)	0.512 (0.151, 0.907)
Endoscopic remission	≤1	26	86	0.57 (0.44, 0.70)	0.51 (0.39, 0.63)	0.58 (0.50, 0.72)*	0.55 (0.43, 0.68)	0.370 (0.037, 0.697)
CDEIS	<3	27	85	0.56 (0.43, 0.70)	0.58 (0.45, 0.71)	0.60 (0.51, 0.68)	0.60 (0.47, 0.73)	0.535 (0.101, 0.950)
Fecal calprotectin	≤100	34	78	0.52 (0.38, 0.67)	0.55 (0.41, 0.70)	0.60 (0.52, 0.80)	0.52 (0.39, 0.66)	0.456 (0.030, 0.956)
Biologic remission	≤5	21	91	0.58 (0.45, 0.72)	0.62 (0.47, 0.78)	0.62 (0.52, 0.76)	0.52 (0.36, 0.67)	0.427 (0.721, 0.889)

\*CIs based on 100 random-splits and corresponding quantiles.

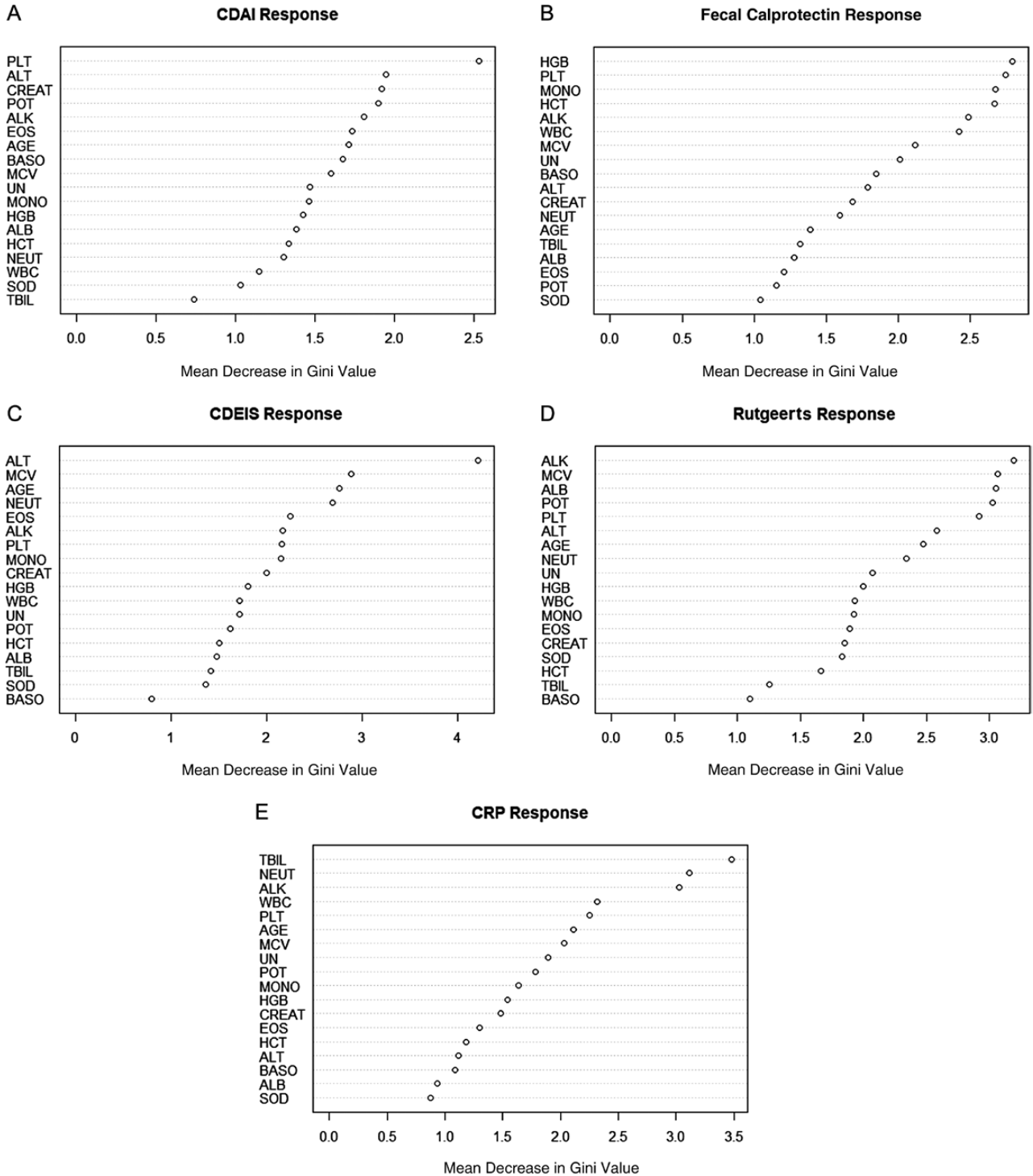


FIGURE 1. Variable importance table for each of the random forest models built for the 5 outcomes. A, CDAI response. B, Calpro response. C, CDEIS response. D, Rutgeerts response. E, CRP response.

thiopurines on clinical outcomes, and the associations between biologic changes in laboratory values induced by thiopurines and clinical outcomes are weak. Most of these patients will remain in surgically induced clinical remission for some time,

so that the predictive value of biologic changes induced by thiopurines is poor.

Though TOPPIC was a prospectively executed study, our retrospective analysis of this clinical data has limitations.

We are also limited by the study population, which might lack generalizability outside of the clinical trial environment. This study has a relatively small sample size which may limit machine learning methods. However, rather than newly training a model, we utilized an established algorithm making this more robust. There is also an inherent risk of overfitting, though we used cross-validation techniques to overcome this limitation. For the rescue therapy outcome, the logistic regression has convergence issues likely due to the lack of “success” response cases after we split the data into training and testing. The logistic regressions with lasso penalty may have convergence problems when selecting the tuning parameter through cross-validation. The mean AuROC is computed over the converged splits.

Other noninvasive, low-cost methods of monitoring postsurgical recurrence are necessary. It would be interesting to study the impact of fecal calprotectin on postsurgical recurrence, as this was promising in the Postoperative Crohn's Endoscopic Recurrence (POCER)<sup>14</sup> data, but this dataset does not allow this analysis. Future studies should focus on longer time points postsurgically, including 6- to 12-month follow-up to evaluate the effectiveness of a machine learning algorithm (likely including fecal calprotectin) in predicting disease recurrence as compared to current standards.

### SUPPLEMENTARY MATERIAL

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

### DATA AVAILABILITY

Data from TOPPIC were made available through the University of Edinburgh through a data-sharing agreement.<sup>12</sup>

The data can be requested via <https://datashare.is.ed.ac.uk/handle/10283/2196>. No new data were created or analyzed.

### REFERENCES

1. Adam L, Phulukdaree A, Soma P. Effective long-term solution to therapeutic remission in Inflammatory Bowel Disease: role of Azathioprine. *Biomed Pharmacother*. 2018;100:8–14.
2. Fraser AG, Orchard TR, Jewell DP. The efficacy of azathioprine for the treatment of inflammatory bowel disease: a 30 year review. *Gut*. 2002;50:485–489.
3. Bastida Paz G, Nos Mateu P, Aguas Peris M, et al. [Optimization of immunomodulatory treatment with azathioprine or 6-mercaptopurine in inflammatory bowel disease]. *Gastroenterol Hepatol*. 2007;30:511–516.
4. Konerman MA, Beste LA, Van T, et al. Machine learning models to predict disease progression among veterans with hepatitis C virus. *PLoS One*. 2019;14:e0208141.
5. Waljee AK, Lipson R, Wiitala WL, et al. Predicting hospitalization and outpatient corticosteroid use in inflammatory bowel disease patients using machine learning. *Inflamm Bowel Dis*. 2017;24:45–53.
6. Waljee AK, Liu B, Sauder K, et al. Predicting corticosteroid-free biologic remission with Vedolizumab in Crohn's disease. *Inflamm Bowel Dis*. 2018;24:1185–1192.
7. Waljee AK, Liu B, Sauder K, et al. Predicting corticosteroid-free endoscopic remission with vedolizumab in ulcerative colitis. *Aliment Pharmacol Ther*. 2018;47:763–772.
8. Waljee AK, Sauder K, Patel A, et al. Machine learning algorithms for objective remission and clinical outcomes with thiopurines. *J Crohns Colitis*. 2017;11:801–810.
9. Waljee AK, Wiitala WL, Govani S, et al. Corticosteroid use and complications in a US inflammatory bowel disease cohort. *PLoS One*. 2016;11:e0158017.
10. Waljee AK, Joyce JC, Wang S, et al. Algorithms outperform metabolite tests in predicting response of patients with inflammatory bowel disease to thiopurines. *Clin Gastroenterol Hepatol*. 2010;8:143–150.
11. Waljee AK, Sauder K, Zhang Y, et al. External validation of a thiopurine monitoring algorithm on the SONIC clinical trial dataset. *Clin Gastroenterol Hepatol*. 2018;16:449–451.
12. Mowat C, Arnott I, Cahill A, et al.; TOPPIC Study Group. Mercaptopurine versus placebo to prevent recurrence of Crohn's disease after surgical resection (TOPPIC): a multicentre, double-blind, randomised controlled trial. *Lancet Gastroenterol Hepatol*. 2016;1:273–282.
13. Stidham RW, Guentner AS, Ruma JL, et al. Intestinal dilation and platelet:albumin ratio are predictors of surgery in stricturing small bowel Crohn's disease. *Clin Gastroenterol Hepatol*. 2016;14:1112–1119.e2.
14. De Cruz P, Kamm MA, Hamilton AL, et al. Efficacy of thiopurines and adalimumab in preventing Crohn's disease recurrence in high-risk patients—a POCER study analysis. *Aliment Pharmacol Ther*. 2015;42:867–879.