ORIGINAL ARTICLES—ALIMENTARY TRACT

Algorithms Outperform Metabolite Tests in Predicting Response of Patients With Inflammatory Bowel Disease to Thiopurines

AKBAR K. WALJEE,* JOEL C. JOYCE,‡ SIJIAN WANG,§ ADITI SAXENA,* MARGARET HART,* JI ZHU,‖ AND PETER D. R. HIGGINS*

*Division of Gastroenterology, Department of Internal Medicine, ‡General Clinical Research Center, and ‖Department of Statistics, University of Michigan, Ann Arbor, Michigan; and §Department of Statistics, University of Wisconsin, Madison, Wisconsin

BACKGROUND & AIMS: Levels of the thiopurine metabolites 6-thioguanine nucleotide (6-TGN) and 6-methylmercaptopurine commonly are monitored during thiopurine therapy for inflammatory bowel disease despite this test's high cost and poor prediction of clinical response (sensitivity, 62%; specificity, 72%). We investigated whether patterns in common laboratory parameters might be used to identify appropriate immunologic responses to thiopurine and whether they are more accurate than measurements of thiopurine metabolites in identifying patients who respond to therapy. **METHODS:** We identified 774 patients with inflammatory bowel disease on thiopurine therapy using metabolite and standard laboratory tests over a 24-hour time period. Machine learning algorithms were developed using laboratory values and age in a random training set of 70% of the cases; these algorithms were tested in the remaining 30% of the cases. **RESULTS:** A random forest algorithm was developed based on laboratory and age data; it differentiated clinical responders from nonresponders in the test set with an area under the receiver operating characteristic (AUROC) curve of 0.856. In contrast, 6-TGN levels differentiated clinical responders from nonresponders with an AUROC of 0.594 ($P < .001$). Algorithms developed to identify thiopurine nonadherence (AUROC, 0.813) and thiopurine shunters (AUROC, 0.797) were accurate. **CONCLUSIONS: Algorithms that use age and laboratory values can differentiate clinical response, nonadherence, and shunting of thiopurine metabolism among patients who take thiopurines. This approach was less costly and more accurate than 6-TGN metabolite measurements in predicting clinical response. If validated, this approach would provide a low-cost, rapid alternative to metabolite measurements for monitoring thiopurine use.**

View this article's video abstract at www.cghjournal.org.

Mesalamine compounds fail to maintain remission in approximately 30% of ulcerative colitis (UC)[1] patients and approximately 91% of Crohn's disease (CD) patients.[2] Thiopurines are widely used immunomodulators that have proven benefits in inflammatory bowel disease (IBD) patients who have failed mesalamine compounds.[3] These immunomodulators can effectively induce both clinical and endoscopic remission and are steroid-sparing.[4,5] Unfortunately, thiopurines have a narrow therapeutic index, and their pharmacokinetics vary widely between individuals. Traditionally, the balance between efficacy and risk with thiopurines has been managed by monitoring complete blood counts (CBC) and chemistry panels. This approach is inexpensive (~$30) and produces rapid results from local laboratories. However, this approach does not have an established algorithm and is dependent on expert assessment, which makes it difficult to reproduce.

More recently, a reproducible approach to thiopurine risk management and evaluation of clinical efficacy has been introduced: the monitoring of 6-thioguanine nucleotide (6-TGN) and 6-methylmercaptopurine (6-MMP) metabolites. 6-TGN levels greater than 235 pmol/10^8 red blood cells are associated with clinical response and 6-MMP levels greater than 5700 pmol/10^8 red blood cells are associated with an increased risk of hepatotoxicity.[6–8] However, a recent meta-analysis of studies of these metabolites shows that their sensitivity for clinical response is only 62% and their specificity is only 72%.[8] In addition, monitoring these metabolites is costly ($268),[9] and slow (typical time to return results, 5 days). In part because traditional monitoring does not have a reproducible algorithm, the performance of monitoring with CBC and chemistries has never been tested directly against the performance of metabolite monitoring.

Metabolite levels correlate poorly with dosing of thiopurines, reflecting significant variation in pharmacokinetics between individuals. Although the metabolite levels control for this pharmacokinetic variation, they predict clinical response relatively poorly. We hypothesized that the poor performance of metabolite levels in predicting clinical response was due to pharmacodynamic variation between individuals, who have different responses of their immune systems and body chemistries to the same 6-TGN metabolite levels. If this is the case, measurement of crude markers of the biologic response to thiopurines, the blood counts and chemistries, might be more predictive of clinical response than metabolite measurements. We further hypothesized that experienced users of thiopurines in

Abbreviations used in this paper: AuROC, area under the receiver operating characteristic curve; CBC, complete blood count; CD, Crohn's disease; IBD, inflammatory bowel disease; MCV, mean corpuscular volume; 6-MMP, 6-methylmercaptopurine; 6-TGN, 6-thioguanine nucleotide; UC, ulcerative colitis; WBC, white blood cell.

1542-3565/10/$36.00
doi:10.1016/j.cgh.2009.09.031

Table 1. Dubinsky Criteria for Defining Clinical Responders and Nonresponders for CD and UC

CD		UC	
Clinical nonresponders	Clinical responders	Clinical nonresponders	Clinical responders
1. mHBI ≥4 on or off steroids, or, 2. mHBI <4, but still requiring steroids for maintenance of remission (steroid-dependent), or, 3. The presence of draining fistulas for fistulizing patients	1. CD patients must have been in remission as defined by a mHBI <4, off steroids, and no open fistulae for at least 3 weeks[20]	1. mUCDAI ≥4 on or off steroids, or, 2. mUCDAI <4, but still requiring steroids for maintenance of remission (steroid-dependent)	1. UC patients must have been in remission as defined by a mUCDAI <4 off steroids

mHBI, modified Harvey Bradshaw index[19]; mUCDAI, modified ulcerative colitis disease activity index.
Criteria from Dubinsky et al.[13]

IBD have an internal algorithm that they use to evaluate common laboratory tests in thiopurine monitoring and that this algorithm can be identified through machine learning.

Recently, Mayer[10] called for a direct comparison of inexpensive, rapid monitoring with CBC and chemistries with the more expensive, slower metabolite monitoring. Previous investigators have shown that low white blood cell (WBC) levels and increasing red blood cell size (mean corpuscular volume [MCV]) are associated with clinical response.[11,12] Therefore, a defined algorithm using routine laboratory values might differentiate clinical response rapidly and inexpensively. To this end, our study objectives were as follows: (1) to determine whether components of the complete blood count, chemistry panel, and patient age can perform the following: (a) differentiate clinical response, (b) identify thiopurine nonadherence, and (c) identify patients (shunters) who produce a low (<20) 6-TGN/6-MMP ratio when they metabolize thiopurines; and (2) to compare the test characteristics of algorithms based on common laboratory tests with the test characteristics of 6-TGN and with logistic regression models using WBC and MCV.

Methods

Patient Population

The study sample included all patients who had thiopurine metabolite analysis, CBC, and a comprehensive chemistry panel drawn within a 24-hour period at the University of Michigan between May 1, 2004, and August 31, 2006. This study was approved by the University of Michigan Medical Institutional Review Board with a waiver of explicit consent from the subjects. The patient sample included 774 cases, in a total of 346 individuals. For the analysis of the outcome of clinical response to thiopurines, 5 exclusion criteria were applied: exclusion of patients who did not have IBD, exclusion of patients who had not started on thiopurines at the time when metabolites were measured, exclusion of patients on biologic anti–-tumor necrosis factor- therapy, exclusion of patients without documentation of their clinical status at the time of laboratory measurement, and exclusion of patients who had an infection that confounded assessment of clinical response.

Data Extraction

Data were obtained though a query of the University of Michigan Clinical Data Repository. Additional patient data fields necessary to determine clinical response were extracted from clinical records separately by 2 of the authors (A.K.W. and J.C.J.) who were blinded to the metabolite monitoring values, and these fields were entered separately into a Microsoft Access (Microsoft Corp, Redmond, WA) database. The 2 data sets were compared with Epi Info 3.3.2 (CDC, Atlanta, GA) to identify discrepancies. Discrepancies were resolved by consensus between the 2 reviewers, with the assistance of the senior author (P.D.R.H.).

Dependent Outcome Variable Definitions

Three dependent outcome variables were measured as follows: clinical response, nonadherence, and preferential shunting to 6-MMP rather than metabolism to 6-TGN. We determined clinical response through medical record review of the 395 cases eligible for the clinical response outcome. Cases with CD or UC/indeterminate colitis then were classified as clinical responders (216), nonresponders (179), or unknown responders (1) using the criteria established by Dubinsky et al.[13] Table 1 details the criteria used to classify patients for this outcome. Cases were classified separately on these criteria by 2 authors (A.K.W. and J.C.J.). If there were any discrepancies in data interpretation or classification (occurred in <3% of cases), these were resolved by review of the clinical notes and consultation with the senior author (P.D.R.H.).

Nonadherence was defined as a 6-TGN level of ≤25 because no adherent patients on 25 mg of azathioprine daily (the lowest recorded dose in our data set) had a 6-TGN level of less than 30. Shunting of thiopurine metabolism away from 6-TGN was defined a priori as a 6-TGN/6-MMP ratio of less than 0.05 in patients who had 6-TGN levels of greater than 25.

Independent Predictor Variables

Independent predictor variables were derived from the CBC (with automated differential) values, chemistries, patient age, 6-TGN, and 6-MMP serum values. Patient demographic information, CBC, chemistries, and 6-TGN and 6-MMP serum values were obtained from the University of Michigan Clinical Data Repository. The age was calculated as the number of years between the date of birth and the date of the laboratory draw at which thiopurine metabolites were obtained.

Statistical Analysis

Machine learning approaches were used to develop algorithms for the 3 dependent variables: clinical response, nonadherence, and shunting. The machine learning methods used included boosted trees,[14] RuleFit,[15] and Random Forest[16] meth-

ods, which produced similar predictive results. Only the results of the random forest methods are presented for the sake of brevity. The machine learning approaches were used to develop algorithms on a model development data set, which was a random sample of 70% of the data. The algorithms developed then were validated by testing on the remaining 30% of the data, the model validation data set. All areas under the receiver operating characteristic curve (AuROCs) and 95% confidence intervals (CIs) presented are calculated based on the model validation data set. The AuROCs were calculated using the Bioconductor package ROC. When the AuROC is p, the variance was estimated with the formula $s^2 = p*(1 - p)/n$, where n is the number of observations in the training set. The 95% CI for the AuROC was calculated with $p \pm 1.96(s)$. In comparing different AuROCs, we determined whether the 95% CIs overlapped, which is a 2-sided test. A graph of the relative importance of the independent variables is presented for each of the 3 algorithms.

R version 2.8.1, with the software package "randomForest", was used for machine learning algorithm development. The effect of varying the number of trees in the random forest approach was tested, and the results are stable with at least 500 trees, which was the number of trees used for each of the 3 algorithms. The number of input variables randomly selected at each split also was tuned, using a 10-fold cross-validation based on the training data. The selected numbers of variables were 3, 5, and 5, respectively, for clinical response, adherence, and shunting. All predictor variables were continuous in nature, and the importance of each variable computed from the random forests is conditional on the existence of other variables. The final algorithms, consisting of 500 trees each, are not presented here for the sake of brevity.

The AuROC produced by the 6-TGN metabolite as a predictor of clinical response was compared with the AuROC of the machine learning algorithm based on the laboratory tests. To evaluate the accuracy of the machine learning model in comparison with the measurement of 6-TGN in predicting clinical response, we used the results of the 6-TGN criterion (6-TGN > 235 = likely clinical responder[7]) for the initial classification of cases, and the machine learning model was used for reclassification. The gain in diagnostic accuracy was assessed with the net improvement in reclassification statistic.[17] To test whether the 6-TGN metabolite would add to the value of the machine learning algorithm, a new random forest algorithm was created with the laboratory tests, age, and 6-TGN, to determine the marginal benefit (in terms of AuROC) of adding the 6-TGN result.

Because many practitioners use reductions in WBC and/or increases in MCV as inexpensive guides to thiopurine therapy, we also evaluated how valuable these 2 variables were in predicting the 3 outcomes of clinical response, adherence, and shunting with logistic regression modeling. The models were evaluated using AuROC graphs, and their calibration was checked with Hosmer–Lemeshow tables. Because the MCV and WBC were reasonably predictive for the outcome of clinical response, we developed a simple prediction rule that practitioners can use to calculate the probability of clinical response from MCV and WBC. We did not test a machine learning approach for these models with few predictors because machine learning has little advantage over simpler logistic models when there are only a few predictors.

Because steroids can alter some of the common laboratory parameters (increase glucose, decrease eosinophils), and steroids were part of the outcome definition of Dubinsky et al,[13] there was a danger of circular reasoning in our model development (steroid use could alter laboratory values [ie, white blood cell count or glucose], which would differentiate steroid use/failure of thiopurines). We therefore assessed whether the algorithms also were effective to differentiate clinical responders from nonresponders in the subset of patients not on steroids. All machine learning methods were performed with algorithms in the statistical language R (version 2.8.1), with the package randomForest,[18] by S.W. and J.Z. All data preparation, logistic regression, and graphing was performed using Stata 10.0 (StataCorp, College Station, TX) by A.K.W. and P.D.R.H., and 2-sided *P* values less than .05 were considered statistically significant.

Table 2. Study Population Demographics (n = 346 individuals)

Sex, M:F	175:171
Percent Caucasian	88.1% [28 race not reported]
Median age, *y* (IQR) [range]	23 (16–40) [5–86]
Disease type	
CD	218
UC	93
Indeterminate colitis	12
Not IBD	23
Disease location	
CD	
Small bowel	26
Large bowel	24
Both	168
UC	
Proctitis	2
Left-sided	15
Pancolitis	76
Median disease duration, *y* (IQR)	4.12 (2.10–7.96)
Thiopurine type	
Azathioprine (N = 234)	
Median dose, *mg/d* (IQR)	100.00 (80.00–150.00)
Median duration of current dose, *mo*	5.42
6-mercaptopurine (N = 90)	
Median dose, *mg/d* (IQR)	75.00 (50.00–100.00)
Median duration of current dose, *mo* (IQR)	10.28 (2.61–21.86)
Median duration of current dose for all individuals on thiopurines, *mo* (N = 324)	6.51
Not taking thiopurines on date of blood draw (N = 22 individuals)	
Ever steroid use?	
Yes	338
No	8
Current steroid use?	
Yes	101
No	245
Median concomitant steroid dose, *mg/d* (IQR) in those currently using steroids	20.00 (10.00–30.00)
Concomitant mesalamine use	118 of 346

IQR, interquartile range.

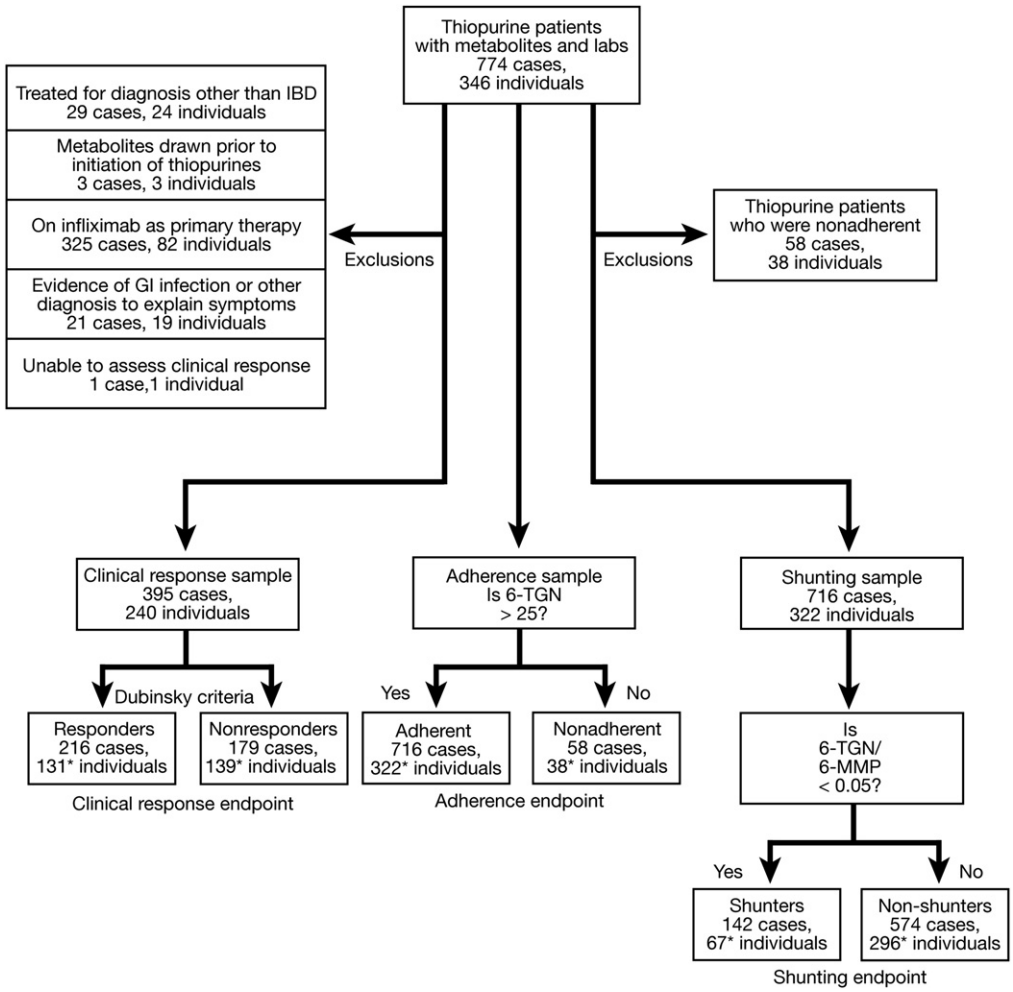

Figure 1. Sample selection for the 3 study end points. The flow diagram illustrates the exclusions and definitions of the clinical response, nonadherence, and shunting end points. *Note that individuals can appear more than once as distinct cases, so that 30 individuals appear as both clinical responders and clinical nonresponders at different points in time, as distinct cases in the data set. Similarly, 14 individuals appear as adherent or nonadherent at different points in time in this data set, and 41 individuals meet the criteria for shunting at one point in time, but not at other points in time in the data set.

Results

Table 2 describes the characteristics of our total study sample by patient, which included 774 cases, in a total of 346 individuals. Of note, no patients were found to be on concomitant allopurinol. The application of the exclusion criteria for clinical response, as illustrated in the flowchart in Figure 1, narrowed the study population for the clinical response end point to 395 cases with a total of 240 individuals in the study population. Of note, 30 individuals appear as both clinical responders and clinical nonresponders at different points in time, as distinct cases in the data set. No patients were excluded for the outcome of nonadherence, and thus this study sample included all 774 cases in a total of 346 individuals. An additional exclusion criterion was applied for the shunting outcome, which was to exclude patients who were nonadherent. These exclusions narrowed the study sample for the shunting end point to 716 cases in a total of 322 individuals.

The median duration of disease and median dose of thiopurine were similar between the clinical responder and nonresponder groups. A detailed breakdown of the characteristics of the 119 clinical responders and 121 clinical nonresponders is presented in Table 3. These 2 groups were largely similar, although the thiopurine nonresponders were more likely to be on steroids, tended to be older (median age, 28 vs 18 y), and had been on thiopurines and their current thiopurine dose for a shorter time than clinical responders. There was no evidence of significant hepatotoxicity associated with the medications in our sample.

Predicting Clinical Response

A random forest algorithm using laboratory values and patient age differentiated clinical response from nonresponse in the model validation data set with an AuROC of 0.856 (95% CI, 0.793–0.919) (Figure 2*A*). In comparison, 6-TGN levels differentiated clinical response from nonresponse with an AuROC of 0.594 (95% CI, 0.546–0.642). The difference between the 2 areas was highly significant, with a *P* value of less than .001. The most important independent variables in differentiating clinical responders from nonresponders were neutrophil count, alkaline phosphatase, red cell distribution width, age, and WBC count (Figure 2*B*). Adding the 6-TGN independent variable as an input to the model did not significantly improve the AuROC, which changed from 0.856 to 0.862 (95% CI, 0.800–0.924).

Because steroid use was part of the definition of nonresponse, we considered the possibility that steroid use could alter laboratory values (ie, increase glucose, lower eosinophil count, and increase neutrophil count) in reproducible ways that would allow the algorithm to identify patients as clinical response failures largely by identifying those on steroids. We tested whether the algorithm still was effective in identifying

Table 3. Study Population Characteristics by Responder Status

	Responders	Nonresponders
No. of individuals	119	121
Sex, M:F	67:52	56:65
Percent Caucasian	84.8%	78.5% (8 race not reported)
Median age, *y* (range)	18 (5–76)	28 (5–86)
IBD, CD:UC	76:43	76:45
Median disease duration, *y* (IQR)	3.66 (2.50–6.17) 6 missing	3.80 (1.78–9.00) 13 missing
Disease location		
UC		
Left-sided	9	14
Pancolitis	30	27
Unknown	4	4
CD		
Perianal	4	1
Large bowel	12	9
Small bowel	15	18
Both	35	38
Unclear	10	10
Median duration of thiopurines, *y* (IQR)	2.04 (0.75–3.31) 26 missing	0.99 (0.33–2.85) 33 missing
Median thiopurine dose (IQR)	100.00 (64.58–131.25) 4 missing	100.00 (75.00–150.00) 5 missing
Median duration of current thiopurine dose, *y* (IQR)	0.75 (0.33–2.15)	0.33 (0.10–0.92)
Percent on steady-state thiopurine dose	97.1	82.4
Prior steroids		
No	10	5
Yes	92	109
Unknown	17	7
Concomitant mesalamine	50 of 119	55 of 121
Concomitant steroid dose, *mg/d*		
None	118	42
<10	0	21
≥10	0	58
Unknown	1	0

NOTE. Each individual only appears once in this table, based on his or her clinical response status at his or her first presentation in the data set.
IQR, interquartile range

clinical response failures who were not on steroids, and when excluding steroid users, found that the AuROC was reduced only slightly, from 0.856 to 0.807 (95% CI, 0.721–0.893). This value is still significantly greater than the AuROC of 6-TGN, which was 0.555 (95% CI, 0.497–0.613) when tested on the subjects not on steroids.

Reclassification is a more concrete measure of improved clinical discrimination the AuROC. We evaluated the proportion of the patients misclassified by the 6-TGN model who would be reclassified correctly with the random forest model for the clinical response end point. The net improvement in reclassification was 0.362 ($P < .001$). Table 4 illustrates the reclassification of results from 6-TGN classification by the random forest algorithm.

In addition, we evaluated simple logistic regression models to explain nonadherence using the predictors MCV and WBC because these are used commonly in clinical practice to estimate the response to thiopurines. We found that the ratio of MCV to

Table 4. Reclassification Table for Clinical Response to Thiopurines

	True clinical responders: reclassification by RFA			True clinical nonresponders: reclassification by RFA		
Initial test classification based on 6-TGN ≥235	RFA test responders	RFA test nonresponders	Totals classified by 6-TGN	RF test responders	RF test nonresponders	Totals classified by 6-TGN
6-TGN test responders	74 (0.79) TP →	*20 (0.21) FN*	94 TP	18 (0.33) FP →	**37 (0.67) TN**	55 FP
6-TGN test nonresponders	**81 (0.66) TP**	← 41 (0.34) FN	122 FN	*24 (0.19) FP*	← 100 (0.81) TN	124 TN
Totals classified by RFA	155 (0.72) TP	61 (0.28) FN	216 CR	42 (0.23) FP	137 (0.77) TN	179 CNR

NOTE. The net improvement in reclassification was 0.36 ($P < .0001$) by the RFA in contrast to the 6-TGN result. Arrows indicate direction of reclassification.
Bold, correctly reclassified; italic, incorrectly reclassified.
CR, true clinical responders; CNR, true clinical nonresponders; RFA, Random Forest Algorithm; TP, true positive; FN, false negative; FP, false positive; TN, true negative.

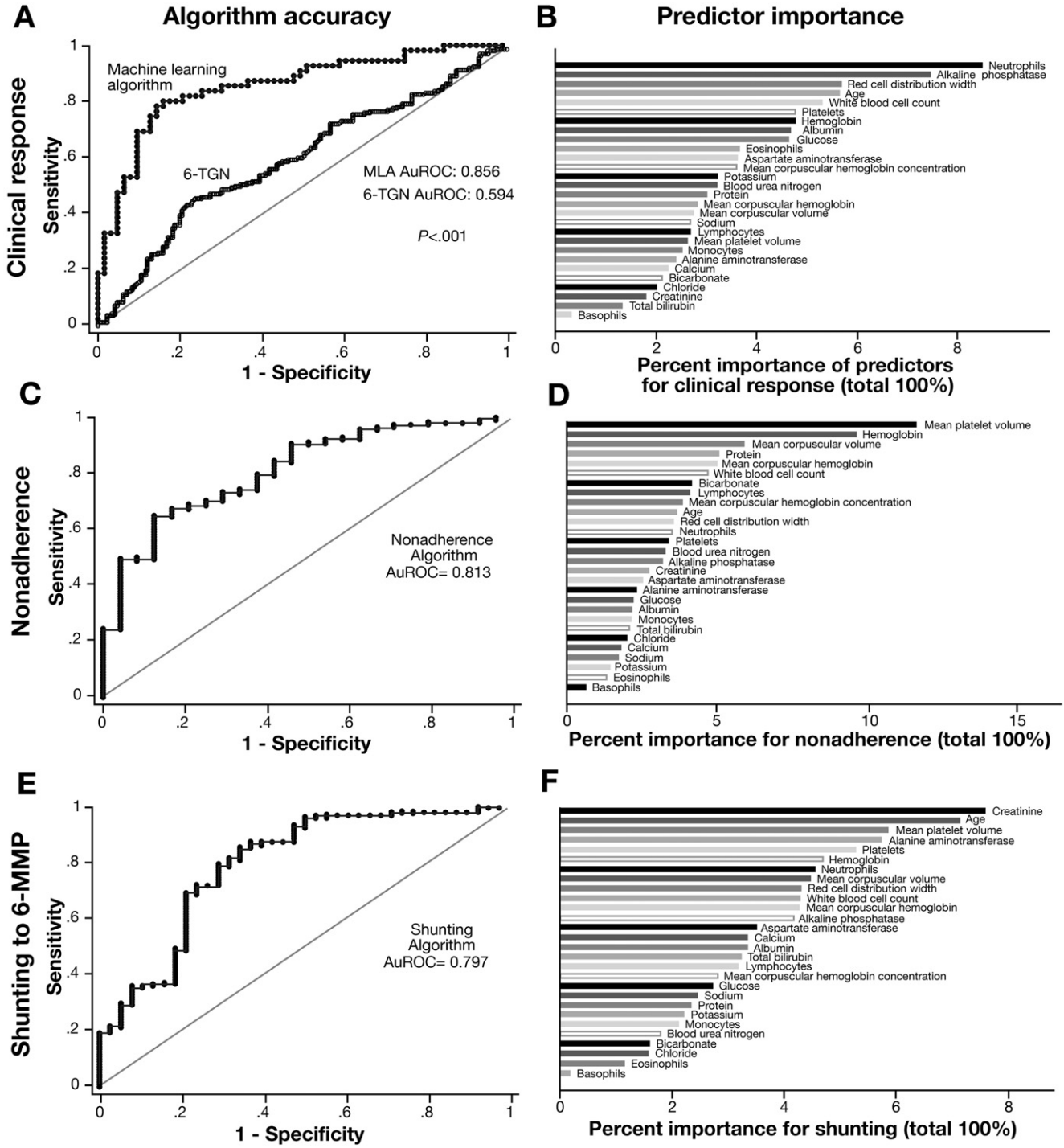

Figure 2. Algorithm accuracy and predictor importance of random forest machine learning models based on common laboratory tests. (*A*) AuROC for clinical response with the random forest clinical response algorithm and with 6-TGN alone. (*B*) Predictor importance for the random forest clinical response algorithm. (*C*) AuROC for nonadherence with the random forest nonadherence algorithm. (*D*) Predictor importance for the random forest nonadherence algorithm. (*E*) AuROC for shunting to 6-MMP with the random forest shunting algorithm. (*F*) Predictor importance for the random forest shunting algorithm.

WBC was a simple predictor of clinical response. The MCV/WBC ratio had an AuROC of 0.715 (95% CI, 0.597–0.833). This model was developed and tested on the entire data set, biasing toward good fit and a high AuROC. This predictor is significantly inferior to the random forest model, and is not significantly better than 6-TGN alone. However, for clinicians who would like to use these 2 laboratory values to identify patients likely to have good clinical response, the MCV/WBC ratio can be used with a cut-off point of 12 or greater predicting clinical response. Subjects with ratios of 12 or greater had a probability

Table 5. Clinical Response Classification by 6-TGN, MCV/WBC Ratio, and Random Forest MLA

	6-TGN		MCV/WBC ratio		MLA scores	
Test classification	Cutoff	% Clinical response	Cutoff	% Clinical response	Cutoff	% Clinical response
Responder	≥235	63	≥12	68	≥−0.16	75
Nonresponder	<235	50	<12	34	<−0.16	19

MLA, machine learning algorithm.

of clinical response of 0.67, and those less than 12 had a probability of clinical response of 0.35. These results are compared with the cut-off point of 6-TGN of 235 or greater and the random forest algorithm in Table 5.

Predicting Nonadherence

A random forest algorithm using laboratory values and patient age differentiated thiopurine adherence from nonadherence with an AuROC of 0.813 (95% CI, 0.763–0.863) (Figure 2*C*). The proportional importance of each input variable in the random forest model is shown in Figure 2*D*. The most important independent variables in differentiating adherent from nonadherent patients were as follows: mean platelet volume, hemoglobin, MCV, protein, and mean corpuscular hemoglobin. A 2-predictor logistic regression model to explain nonadherence was constructed using the predictors WBC and MCV. This model had an AUC of 0.704 with a 95% CI from 0.637 to 0.771. Although this result is biased in favor of the logistic model, it is less effective than the random forest model.

Predicting Shunting to 6-MMP

A random forest algorithm using laboratory values and patient age differentiated thiopurine shunters from nonshunters with an AuROC of 0.797 (95% CI, 0.743–0.850) (Figure 2*E*). The proportional importance of each input variable in the random forest model is show in Figure 2*F*. The most important independent variables in differentiating shunting from nonshunting patients were as follows: creatinine, age, mean platelet volume, alanine aminotransferase, and platelet counts. A 2-predictor logistic regression model to predict shunting also was constructed, using the predictors WBC and MCV. This model had an AuROC of 0.613 with a 95% CI from 0.561 to 0.665. This model is not nearly as good as the random forest model, and its lower confidence bound approaches an AuROC of 0.5 (approximates a coin flip in predictive ability).

Discussion

Historically, thiopurine monitoring with blood counts and chemistries in IBD has been based on a gestalt assessment of common laboratory values by experienced practitioners. Metabolite monitoring has offered an appealingly reproducible alternative, although at a substantial increase in cost. The value of metabolite monitoring has been questioned, but no reproducible algorithm for monitoring thiopurines with standard laboratories has been available for comparison with metabolite testing. In this study, we have developed and tested a machine learning algorithm for differentiating thiopurine clinical responders from nonresponders and found that it performs significantly better than metabolite monitoring in predicting clinical response. Additional machine learning algorithms were able to identify nonadherent and shunting patients with good accuracy. There is no additional cost associated with implementation of the algorithm. The only expense incurred is for standard monitoring with blood counts and chemistries, typically performed every 3 months while patients are on stable doses of the medication.

One limitation of this study was the patient sample, obtained from a single tertiary care center, because the results may not be generalizable to all patients with IBD. It is possible that patients in whom metabolites perform poorly were more likely to be referred to our tertiary center, and this may have led to the reduced predictive value of 6-TGN in this sample (AuROC, 0.594), compared with the 62% sensitivity and 72% specificity of metabolites for clinical response in the meta-analysis by Osterman et al.[8] A second limitation of this study was that it was cross-sectional, rather than longitudinal. This prevented us from assessing whether the results of the clinical response algorithm could be used prospectively to direct dose changes of thiopurines or changes in therapy.

The strengths of this study included the use of the same clinical response definitions used to define clinical response by Dubinsky et al;[7] the use of common, inexpensive, and readily available laboratory tests as the inputs for the algorithms; and the use of the open source software R for implementation. We have been able to implement the calculation of the probabilities of clinical response, nonadherence, and shunting in our electronic medical records system. Because R is freely available, this should allow ready implementation in many clinical settings.

The results of this study suggest that pharmacodynamic variation is an important part of the variability in the clinical response to thiopurines, that clinical response to thiopurines only weakly is predicted by metabolite levels or reductions in the WBC count and increases in the mean corpuscular volume. We find it striking that mild increases in the eosinophil count are an important predictor of clinical response, suggesting that effective doses of thiopurines may be polarizing the patient's immune system toward Th2 immune responses. The sometimes surprising results of modeling can generate new hypotheses about the mechanism of action of old medications.

By showing that there are patterns in common laboratory values that can differentiate accurately between thiopurine responders and nonresponders, this study raises the possibility that these patterns could be used to guide thiopurine dosing (or to help decide when thiopurines simply will not be effective) in the future. This would require an appropriately powered prospective clinical trial, and we currently are pursuing funding for this project.

We hypothesize, based on the data presented here, that the physiologic parameters measured in common laboratory tests

and incorporated in these algorithms are surrogate markers for changes in the immune system induced by effective thiopurine therapy. It seems likely that patients with inadequate responses in these parameters are underdosed, and that some patients, even at high doses, will not achieve these physiologic changes, because of differences in pharmacodynamics. Other patients may achieve these physiologic changes without a clinical response. These patients likely would have either symptoms caused by noninflammatory causes, or have an autoimmune physiology that is not amenable to thiopurine therapy, which would prompt a change to a different therapy (methotrexate, biologics, and so forth). We speculate that the eventual clinical utility of this machine learning approach, depending on a positive outcome of a prospective clinical trial, will be in guiding thiopurine dosing and timely changes to a different class of therapy.

Supplementary data

To access the supplementary material accompanying this article, visit the online version of *Clinical Gastroenterology and Hepatology* at www.cghjournal.org, and at doi:10.1016/j.cgh.2009.09.031.

References

1. Cohen RD, Woseth DM, Thisted RA, et al. A meta-analysis and overview of the literature on treatment options for left-sided ulcerative colitis and ulcerative proctitis. Am J Gastroenterol 2000;95:1263–1276.
2. Hanauer SB. The case for using 5-aminosalicyclates in Crohn's disease: pro. Inflamm Bowel Dis 2005;11:609–612.
3. Domenech E. Inflammatory bowel disease: current therapeutic options. Digestion 2006;73(Suppl 1):67–76.
4. Lopez-Sanroman A, Bermejo F, Carrera E, et al. Efficacy and safety of thiopurinic immunomodulators (azathioprine and mercaptopurine) in steroid-dependent ulcerative colitis. Aliment Pharmacol Ther 2004;20:161–166.
5. Ardizzone S, Maconi G, Russo A, et al. Randomised controlled trial of azathioprine and 5-aminosalicylic acid for treatment of steroid dependent ulcerative colitis. Gut 2006;55:47–53.
6. Dervieux T, Meyer G, Barham R, et al. Liquid chromatography-tandem mass spectrometry analysis of erythrocyte thiopurine nucleotides and effect of thiopurine methyltransferase gene variants on these metabolites in patients receiving azathioprine/6-mercaptopurine therapy. Clin Chem 2005;51:2074–2084.
7. Dubinsky MC, Lamothe S, Yang HY, et al. Pharmacogenomics and metabolite measurement for 6-mercaptopurine therapy in inflammatory bowel disease. Gastroenterology 2000;118:705–713.
8. Osterman MT, Kundu R, Lichtenstein GR, et al. Association of 6-thioguanine nucleotide levels and inflammatory bowel disease activity: a meta-analysis. Gastroenterology 2006;130:1047–1053.
9. Mayo Medical Laboratories. Prometheus thiopurine metabolites. 2008. Available at: http://www.mayomedicallaboratories.com/test-catalog/print.php?unit_code=91564. Accessed: November 1, 2008.
10. Mayer L. When is too much enough? Gastroenterology 2006;130:1352–1354.
11. Thomas CW Jr, Lowry PW, Franklin CL, et al. Erythrocyte mean corpuscular volume as a surrogate marker for 6-thioguanine nucleotide concentration monitoring in patients with inflammatory bowel disease treated with azathioprine or 6-mercaptopurine. Inflamm Bowel Dis 2003;9:237–245.
12. Achkar JP, Hanauer SB. Medical therapy to reduce postoperative Crohn's disease recurrence. Am J Gastroenterol 2000;95:1139–1146.
13. Dubinsky MC, Yang H, Hassard PV, et al. 6-MP metabolite profiles provide a biochemical explanation for 6-MP resistance in patients with inflammatory bowel disease. Gastroenterology 2002;122:904–915.
14. Freund Y, Schapire RE. Experiments with a new boosting algorithm. In: Saitta L, ed. Machine Learning: Proceedings of the Thirteenth International Conference. San Francisco, CA: Morgan Kaufmann Publishers, Inc, 1996:148–156.
15. Friedman J, Popescu B. Predictive learning via rule ensembles. Palo Alto, CA: Stanford University Department of Statistics Technical Report, 2005.
16. Breiman L. Random forest. Mach Learn 2001;45:5–32.
17. Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, et al. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Stat Med 2008;27:157–172, 207–212.
18. Ihaka R, Gentleman R. R: a language for data analysis and graphics. J Comput Graph Stat 1996;5:299–314.
19. Bonner GF, Fakhri A, Vennamaneni SR. A long-term cohort study of nonsteroidal anti-inflammatory drug use and disease activity in outpatients with inflammatory bowel disease. Inflamm Bowel Dis 2004;10:751–757.
20. Present DH, Rutgeerts P, Targan S, et al. Infliximab for the treatment of fistulas in patients with Crohn's disease. N Engl J Med 1999;340:1398–1405.

Reprint requests

Address requests for reprints to: Peter D. R. Higgins, MD, PhD, MSc, Division of Gastroenterology, University of Michigan, 6520 MSRB I, Box 0682, 1150 West Medical Center Drive, Ann Arbor, Michigan 48109. e-mail: phiggins@umich.edu; fax: (734) 763-2535.

Conflicts of interest

The authors disclose the following: The Regents of the University of Michigan, along with authors Peter Higgins, Akbar Waljee, Joel Joyce, Sijian Wang, and Ji Zhu, have applied for a patent on the application of machine learning to patterns in the complete blood count and differential and the comprehensive chemistry panel to the prediction of clinical response to thiopurines. As of December 20, 2009, no patent has yet been granted. The remaining authors disclose no conflicts.