Gastroenterology 2020;158:515–526

Development and Validation of a Test to Monitor Endoscopic Activity in Patients With Crohn's Disease Based on Serum Levels of Proteins

Geert D'Haens,[1,*] **Orlaith Kelly**,[2,*] **Robert Battat**,[3,*] Mark S. Silverberg,[4] David Laharie,[5] Edouard Louis,[6] Edoardo Savarino,[7] Giorgia Bodini,[8] Andres Yarur,[9] Brigid S. Boland,[3] Waqqas Afif,[10] Xiao-jun Li,[11] Michael Hale,[11] Jessica Ho,[11] Venkateswarlu Kondragunta,[11] Benjamin Huang,[11] Crystal Kuy,[11] Lauren Okada,[11] Kelly D. Hester,[11] Kurtis R. Bray,[11] Larry Mimms,[11] Anjali Jain,[11] Siddharth Singh,[3] Angelina Collins,[3] Mark A. Valasek,[12] William J. Sandborn,[3] Severine Vermeire,[13,§] and Parambir S. Dulai[3,§]

[1]Department of Gastroenterology, Academic Medical Centre, Amsterdam, The Netherlands; [2]Connolly Hospital Blanchardstown, Dublin, Ireland; [3]Division of Gastroenterology, University of California San Diego, La Jolla, California; [4]Mount Sinai Hospital, Toronto, Canada; [5]CHU Bordeaux, Bordeaux, France; [6]University Hospital Centre Hospitalier Universitaire (CHU) Liege, Liege, Belgium; [7]University of Padua, Padua, Italy; [8]University of Genoa, Genoa, Italy; [9]Medical College of Wisconsin, Milwaukee, Wisconsin; [10]McGill University, Montreal, Canada; [11]Prometheus Laboratories Inc, San Diego, California; [12]Department of Pathology, University of California San Diego, La Jolla, California; and [13]University Hospitals Leuven, Leuven, Belgium

See editorial on page 463.

BACKGROUND & AIMS: Noninvasive tests to measure endoscopic activity in patients with Crohn's disease (CD) have limitations. We aimed to develop a test to identify patients in remission, based on endoscopic analysis, and monitor CD activity based on serum levels of proteins. **METHODS:** We developed a test to measure 13 proteins in blood (ANG1, ANG2, CRP, SAA1, IL7, EMMPRIN, MMP1, MMP2, MMP3, MMP9, TGFA, CEACAM1, and VCAM1), called the endoscopic healing index [EHI], using samples from 278 patients with CD from a multinational training cohort. We validated the test using 2 independent cohorts of patients with CD: 116 biologic-naive patients with early-stage CD (validation cohort 1) and 195 biologic-exposed patients with chronic CD (validation cohort 2). The ability of the test to identify patients with active disease vs patients in remission (defined as a simple endoscopic score for CD of ≤ 2 and ≤ 1 in each segment, or a total CD endoscopic index of severity score <3) was assessed by using area under receiver operating characteristic curve (AUROC) analysis. The diagnostic accuracy of the test was compared with that of measurement of serum C-reactive protein (CRP) and fecal calprotectin. **RESULTS:** The EHI scores range from 0 to 100 units; higher scores indicate more severe CD activity, based on endoscopy findings. The EHI identified patients in remission with an AUROC of 0.962 in validation cohort 1 (95% confidence interval, 0.942–0.982) and an AUROC of 0.693 in validation cohort 2 (95% confidence interval, 0.619–0.767), regardless of CD location or phenotype. A cutoff value of 20 points identified patients in remission with the highest level of sensitivity (97.1% in validation cohort 1 and 83.2% in validation cohort 2), with specificity values of 69.0% and 36.6%, respectively. A cutoff value of 50 points identified patients in remission with the highest level of specificity (100% in validation cohort 1 and 87.8% in validation cohort 2), with sensitivity values of 37.3% and 30.0%, respectively. The EHI identified patients in remission with a significantly higher AUROC value than the test for CRP (0.876, $P < .001$ in validation cohort 1 and 0.624, $P = .109$ in validation cohort 2). In analysis of patients with available FC measurements, the AUROC value for the EHI did not differ significantly from that of measurement of FC (AUROC, 0.950 for EHI vs AUROC, 0.923 for FC; $P = .147$ in validation cohort 1 and AUROC, 0.803 for EHI vs AUROC, 0.854 for FC; $P = .298$ in validation cohort 2). **CONCLUSIONS:** We developed an index called the *EHI* to identify patients with CD in endoscopic remission based on blood levels of 13 proteins. The EHI identified patients with resolution of endoscopic disease activity, with good overall accuracy, although with variation between the 2 cohorts assessed. The EHI AUROC values were comparable to measurement of FC and higher than measurement of serum CRP. The test might be used in practice to assess endoscopic activity in patients with CD.

Keywords: Monitr; IBD; Response to Treatment; Resolution.

***Authors share co-first authorship; §Authors share co-senior authorship.**

Abbreviations used in this paper: **AD, active disease; AIC, Akaike information criterion; AUROC, area under the receiver operating characteristic curve; CD, Crohn's disease; CDEIS, Crohn's Disease Endoscopic Index of Severity; CI, confidence interval; CRP, C-reactive protein; EHI, Endoscopic Healing Index; EMMPRIN, extracellular matrix metalloproteinase inducer; ER, endoscopic remission; ES, effect size; FC, fecal calprotectin; IBD, inflammatory bowel disease; IQR, interquartile range; MH, mucosal healing; MMP, matrix metalloproteinase; NLR, negative likelihood ratio; NPV, negative predictive value; PLR, positive likelihood ratio; PPV, positive predictive value; ROC, receiver operating characteristic; SES-CD, Simple Endoscopic Score for Crohn's Disease.**

Most current article

0016-5085
https://doi.org/10.1053/j.gastro.2019.10.034

Crohn's disease (CD) is a chronic condition characterized by mucosal ulcerations and transmural inflammation anywhere along the gastrointestinal tract. Approximately one third of patients have stricturing or penetrating disease complications at the time of diagnosis, and half of the remaining patients experience an intestinal complication within 20 years of diagnosis.[1] Achieving endoscopic healing, also traditionally referred to as *mucosal healing* (MH), has consistently been associated with reductions in disease-related complications, including corticosteroid use, hospitalization, and surgery.[2–5] For this reason, achieving MH is now considered to be a primary treatment target in CD, and it is recommended that all CD patients initiating immunosuppressive and/or biologic therapy have a follow-up assessment for MH within 6 to 9 months of treatment initiation.[6]

The optimal approach for assessing MH has traditionally been through the use of endoscopy. Alternative methods such as cross-sectional imaging, abdominal ultrasonography, and video capsule imaging are under development, but scoring systems with these techniques have been poorly developed or validated. Therefore, many centers worldwide have adopted endoscopy based treat-to-target monitoring algorithms, in which treatment is optimized, modified, or switched based on serial endoscopic monitoring. Although this strategy has been associated with achieving higher rates of MH,[7,8] it is not without cost, risk, or burden, and endoscopy is ranked as the least acceptable tool for this purpose by CD patients.[9] These limitations are likely to explain why the majority of CD patients treated with biologic therapy have no follow-up endoscopy within the first 24 months after treatment initiation.[10] This suggests that endoscopy based treat-to-target strategies may be difficult to implement in current health care landscapes.

The effect of tight control management on Crohn's disease (CALM) trial recently showed that treatment escalation based on symptoms combined with elevated serum C-reactive protein (CRP) and/or fecal calprotectin (FC) resulted in higher rates of MH compared with symptom-based escalation alone.[11] Despite an overall favorable accuracy,[12,13] the use of FC in clinical practice is somewhat impractical and is currently being done in fewer than 2% of patients with CD in the United States.[10,14] If given an option, patients with CD strongly prefer blood-based biomarkers over fecal biomarkers. However, all prior blood-based biomarkers have had poor accuracy and, therefore, limited utility.[13] A need therefore exists for blood-based biomarkers that accurately quantify mucosal disease activity in CD.

The objective of the current study was to develop and validate a multimarker, serologic, algorithm-based diagnostic test that reliably reflects the severity of endoscopic inflammation in CD. Through a multicenter international collaboration, we derived the Endoscopic Mucosal Healing Index (EHI) (Prometheus Laboratories Inc, San Diego, CA) and subsequently validated it in 2 independent cohorts, showing that it is associated with endoscopic inflammation. We further explored the comparative performance of the EHI against CRP and FC, the responsiveness of the EHI to changes in endoscopic disease state, and the diagnostic performance of the EHI for histologic inflammation on a selected subcohort with limited sample size.

WHAT YOU NEED TO KNOW

BACKGROUND AND CONTEXT

We aimed to develop a test to identify patients in remission from Crohn's disease (CD), based on endoscopic analysis, and monitor CD activity based on serum levels of proteins.

NEW FINDINGS

We developed an index to identify patients with CD in endoscopic remission based on blood levels of 13 proteins. The test (called the EHI) identified patients with resolution of mucosal inflammation, based on endoscopic analysis, with good overall accuracy.

LIMITATIONS

We analyzed data from 2 cohorts; and there was some variation in sensitivity and specificity of detection between the cohorts. We did not validate the test in pediatric patients.

IMPACT

The test might be used in practice to monitor mucosal inflammation, usually evaluated by endoscopy, in patients with CD.

Methods

We followed the PRoBE (prospective-specimen collection, retrospective–blinded-evaluation) study design for evaluating the accuracy of a biomarker used for classification of an outcome (ie, mucosal healing).[15] The results are reported in accordance with the Standards for Reporting of Diagnostic Accuracy Studies guidelines[16] (see Supplementary Appendix).

Patient Selection

Adult patients with CD (≥18 years) were included if they had (1) a confirmed diagnosis of CD based on clinical, endoscopic, and histologic data; (2) documented endoscopic disease activity; and (3) sufficient volume of serum sample available for testing. For the validation cohorts, we required samples to be available within ±45 days of endoscopy. Patients selected for this study were not excluded based on current or prior therapies, prior bowel surgeries, or the presence of an ostomy (ileostomy or colostomy).

Cohorts

The study consisted of 3 independent cohorts of prospectively collected, retrospectively analyzed samples for training and validation. The training cohort included samples obtained from prospectively recruited convenience sampling biobanks between June 2006 and August 2015 at the University of Padua, Italy (July 2011–March 2014); Mount Sinai Hospital (MSH), Toronto, Canada (October 2008–August 2015); University of California San Diego (June 2014–May 2015); and the (STORI) clinical trial, (GETAID), France (June 2006–Jane 2007).[17] Validation cohort 1 included samples collected during the prospective Tailored Treatment With Infliximab for Active Crohn's

Disease (TAILORIX) clinical trial (July 2012–September 2015). These included baseline, week 12, and week 54 samples from 116 biologic-naive patients with CD recruited from 27 centers in Belgium, France, and The Netherlands.[18] Validation cohort 2 included samples collected prospectively from a tertiary referral center in San Diego, California (University of California San Diego, June 2014–January 2018), which were distinct from those included from this institution in the training cohort.

Clinical Data Variables

Data on available variables of interest included patient characteristics (age, sex, ethnicity), disease characteristics (prior surgeries, disease-related complications, Montreal phenotype classification), current and prior treatments (corticosteroids, immunosuppressives, biologics), and clinical disease activity (Patient-Reported Outcomes 2 or Crohn's Disease Activity Index).

Endoscopic Healing Definitions

Endoscopic remission (ER) was defined as a total Simple Endoscopic Score for CD (SES-CD) of $\leq$2 and $\leq$1 in each segment (in the 2 validation cohorts) or a total Crohn's Disease Endoscopic Index of Severity (CDEIS) $<$3 (in the training cohort).[19] Consequently, *active disease* (AD) was defined as CDEIS $\geq$3 or SES-CD $>$2 or SES-CD $=$ 2 if only 1 segment had a score of 2 with a score of 0 in the remaining segments. *Endohistopathologic healing* was defined as achieving both ER and histologic remission (Global Histologic Disease Activity $\leq$ 2).[19–21] Endoscopic scores were derived by using either the SES-CD or the CDEIS.[20,21] Scoring was done locally by site investigators at the time of endoscopy in all data sets except TAILORIX, for which endoscopies were scored by blinded central readers. In the derivation-training cohort, all SES-CD scores were converted to CDEIS scores for consistency during training (Supplementary Figure 1). Histologic disease activity assessments were available in validation cohort 2 and were performed by a pathologist with expertise in gastrointestinal pathology and inflammatory bowel disease (IBD) (MV) who was blinded to endoscopy scores. Four biopsy samples were taken from each intestinal segment (using segments identical to those used to calculate SES-CD score) with matching endoscopic scores. Biopsy samples were taken from the most active endoscopic area, and if no active inflammation was observed, then random biopsy samples were taken.

Sample Storage and Testing

All serum samples and fecal samples were frozen within 24 hours of collection to avoid degradation or loss in biomarkers[22] and were kept frozen at –80°C until testing. Thawed serum and stool samples were tested for all biomarkers in a randomized manner with clinical data blinded to the operator. Further details can be found in the Supplementary Methods.

Endoscopic Healing Index Development

Preliminary serum biomarker candidates were identified from a literature review and assessed by the strength of the corresponding evidence, relevance of their biological functions to CD, and involvement of their signal pathways in CD pathogenesis (Supplementary Table 1). Assays were then developed for the selected biomarker candidates and evaluated for their analytical performance. Biomarker candidates whose assays showed poor analytical reproducibility, low detection rate in serum specimens, and/or lack of correlation to disease severity in preliminary studies (data not shown) were eliminated from further consideration as training progressed to validation. As candidate biomarkers were eliminated from consideration, new panels were developed with progressively fewer target biomarkers such that training panels targeted 47 and 38 biomarkers, and subsequent validation was performed on a 24-analyte panel. Regardless of panel configuration, the EHI algorithm was trained on the same 13 analytes that had robust assays and that showed correlation with clinical disease. These combined training and validation cohorts provided the foundation for validation of a 13-biomarker panel currently reporting the EHI to physicians and patients (Monitr, Prometheus Laboratories Inc, San Diego, CA). An analytical method validation was also performed on the final 13-biomarker panel. The multiple logistic regression method was used to predict endoscopic activity as a function of serum biomarker concentrations proposed as continuous predictors after logarithmic transformation and combined through backward elimination with Akaike information criterion (AIC). Biomarkers were removed one by one by sequentially reducing the AIC value until a minimum AIC was reached, using standard settings in JMP, version 12.0 (SAS Institute, Cary, NC). EHI was obtained by transforming the logistic function in terms of probability of being in active disease. A small fraction of patients (11.5%) contributed more than 1 sample in the training cohort. Samples from the same patients were treated as independent samples, an imperfection limited to the training cohort.

Endpoints

Our primary aim was to assess the sensitivity (proportion of patients above a specific EHI limit among patients with active disease, ie, SES-CD $>$ 2 or SES-CD $=$ 2 if only 1 segment has a score of 2, with a score of 0 in the remaining segments) and specificity (proportion of patients below a specific EHI limit among those in endoscopic remission [SES-CD] of $\leq$ 2 and $\leq$ 1 in each segment) of the EHI at various cutoffs for identifying the presence of endoscopic inflammation. Secondary aims were to explore the diagnostic accuracy of the EHI at various cutoffs for identifying the presence of endohistopathologic inflammation and to compare the diagnostic accuracy of the EHI to CRP and FC. Positive likelihood ratio (PLR), negative likelihood ratio (NLR), positive predictive value (PPV), negative predictive value (NPV), and area under the receiver operating characteristic (ROC) curve (AUROC) were used as secondary measures to assess the performance of the EHI. Finally, we assessed the responsiveness of the EHI compared with endoscopy, CRP, and FC, to assess its utility as a tool for monitoring endoscopic disease activity in patients with CD.

Statistical Analysis

The Delong method was used for computing the 95% confidence interval (CI) of an AUROC and for comparing AUROCs of different biomarkers on paired samples.[23] Exact binomial confidence limits were used for the 95% CIs of sensitivity and specificity. The 95% CIs of PLR and NLR were computed by using formulas provided by Simel et al.[24] A pairwise Wilcoxon rank sum test was used to compare the effect size of different

variables. A P value (2-sided) of .05 or lower was considered significant. All data analysis was carried out using JMP or R, version 3.3.2. A mixed-effect logistic regression modeling was used for validation cohort 1 to assess the performance of EHI, CRP, and FC. The effect sizes[25] of the SES-CD, CDEIS, EHI, FC, and CRP were calculated between baseline and week 12 and between baseline and week 54 to assess the responsiveness of those variables. Further statistical details can be found in the Supplementary Methods.

Sample Size Calculation

No sample size calculation was performed on the training cohort. The EHI had a sensitivity of 90% at the threshold of 20 and a specificity of 95% at the threshold of 50 in the training cohort. We aimed to validate such performance in the validation cohorts with precisions such that the corresponding 1-sided lower 95% confidence limits of sensitivity and specificity were $\geq$80% and $\geq$85%, respectively. Based on exact binomial confidence limits, a minimal of 57 samples from CD patients with AD and a minimum of 40 samples from patients with CD in ER were needed for the validation study.

Ethics

All authors had access to study results and reviewed and approved the final manuscript. Informed consent was obtained from all patients by the study sites according to institutional review board–approved clinical protocols and local regulatory guidance.

Results

Patient Demographics

Serum samples from a total of 589 patients were used (Supplementary Table 2), distributed across the training and 2 distinct validation cohorts without any sample overlap among the training and validation cohorts. The flow chart describing the patients and samples used in the 2 validation cohorts is shown in Supplementary Figure 2. All validation samples used in this study were obtained ±45 days of endoscopy, with approximately 66% (311/470) of samples collected on the same day as the endoscopy: validation cohort 1, 123/275 (44.7%) and validation cohort 2, 188/195 (96.4%). Overall, 90.1% of the training samples were also obtained ±45 days of endoscopy, with 43.9% (147/335) of the samples collected on the same day as the endoscopy. Median elapse from endoscopy to sample collection was 0 days (IQR, 0–14.5) in the training cohort.

The training cohort included 278 patients with a total of 335 endoscopy visits (Supplementary Table 3). The median age was 30.0 years (IQR, 24.9–40.0), and 46.0% were female. The median disease duration was 4 years (IQR, 3.0–12.5). Validation cohort 1 included 116 patients with a median age of 30.2 years (IQR, 22.4–45.2), and 59.5% were female. The cohort included a total of 275 endoscopy visits distributed at baseline (102 visits), week 12 (98 visits), and week 54 (75 visits). FC was available at the same time points (Supplementary Table 4). Validation cohort 2 included 195 patients with 1 endoscopy visit per patient. The median age of the cohort was 38.5 years (IQR, 28.0–52.0), and 49.7% were female. A subcohort of validation cohort 2 (n = 81) patients also had paired fecal calprotectin values obtained from stool samples collected within 45 days of endoscopy (Supplementary Table 5).

The 2 validation cohorts differed significantly in baseline age (30.2 vs 38.5 years; $P < .001$), disease duration (0.7 vs 11.0 years; $P < .001$), disease phenotype (nonstricturing, nonpenetrating: 73% vs 58%; $P = .006$), prior IBD-related bowel surgery (10% vs 46%; $P < .001$) and prior biologic exposure (0% vs 77%, $P < .001$). The median SES-CD (6.0 vs 3.0, $P < .001$) and FC (336 vs 55, $P < .001$) were significantly higher in validation cohort 1 with comparable CRP values (2.5 vs 2.6, $P = 0.460$). All disease locations were represented in the training and both validation cohorts.

Training of the Endoscopic Healing Index

The 47 markers evaluated for the development of EHI are listed in Supplementary Table 1. Initially, 23 of 47 biomarkers were eliminated because of poor assay lot-to-lot reproducibility, poor analytical sensitivity, absence of detectable concentrations in serum, or lack of correlation with endoscopic disease severity. Eleven markers were further eliminated during logistic regression analyses because they did not enhance the performance of the EHI, and the final EHI model included serum concentrations of 13 biomarkers: angiopoietin 1 (ANG1) and 2 (ANG2), carcinoembryonic antigen-related cell adhesion molecule 1 (CEACAM1), CRP, serum amyloid A1 (SAA1), interleukin 7 (IL7), transforming growth factor α (TGFα), vascular cell adhesion molecule 1 (VCAM1), extracellular matrix metalloproteinase inducer (EMMPRIN), and matrix metalloproteinases 1 (MMP1), 2 (MMP2), 3 (MMP3), and 9 (MMP9). EHI was constructed as a scale of 0–100 arbitrary units of EHI activity, a higher score indicating more severe disease activity. Analytical reproducibility of EHI was established using Deming regression (Supplementary Figure 3) with a slope of 1.005 (95% CI, 0.926–1.087) and an intercept of -0.298 (95% CI, –2.790 to 2.144).

The AUROC of the EHI for distinguishing AD from ER in the training cohort was 0.748 (95% CI, 0.696–0.800) (data not shown). Sensitivity and specificity of the EHI were evaluated at increasing cutoffs from 20 to 50 (Supplementary Table 6), covering the clinically relevant region (from high sensitivity to high specificity) as observed in the training cohort. The EHI had a sensitivity and specificity of 90.3% (95% CI, 85.0–94.3) and 95.0% (95% CI, 90.3–97.8) at cutoffs of 20 and 50, respectively, in the training cohort.

Validation Cohort 1

The median global endoscopic SES-CD scores with corresponding median EHI values were 15.0 (IQR, 9.0–22.0) and 55.5 (IQR, 42.0–72.8) at baseline, 4.0 (IQR, 2.0–8.0) and 33.5 (IQR 23.0–41.8) at week 12, 1.0 (IQR, 0.0–3.0) and 25.0 (IQR, 19.0–37.5) at week 52, respectively. The prevalence of ER among patients with paired endoscopy and serum samples in validation cohort 1 was 0% at baseline, 26.5% at

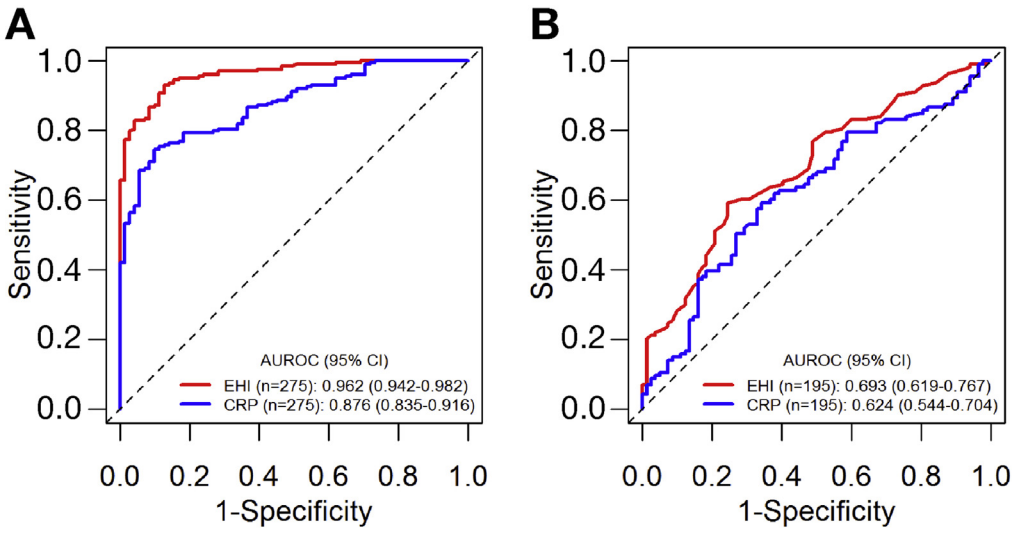


Figure 1. ROC curves of the EHI and CRP for distinguishing active disease vs endoscopic remission in (*A*) validation cohort 1 and (*B*) validation cohort 2. In validation cohort 1, mixed logistic regression models with random intercepts for individual participants were used to combine multiple samples of the same participant.

week 12, and 60% at week 52. The AUROC for distinguishing AD from ER was 0.962 (95% CI, 0.942–0.982) (Figure 1*A*). Sensitivity for ruling out endoscopic inflammation at an EHI cutoff of ≤20 was 97.1% (95% CI, 93.7–98.9) (Table 1). At EHI cutoffs of 40 and 50, the specificity for ruling in AD was 100% (95% CI, 94.9–100.0). At a cutoff of 30, sensitivity and specificity for endoscopic inflammation was 84.8% (95% CI, 79.1–89.4) and 91.5% (95% CI, 82.5–96.8), respectively. The PLRs for detecting AD showed a steady increase from 3.132 (95% CI, 2.212–4.436) to infinity with increasing EHI cutoffs, whereas the NLR was no greater than 0.627 (95% CI, 0.564–0.697). The prevalence of AD at EHI of ≥50 was 100%, and the prevalence of ER at EHI of <20 was 89.1% (Supplementary Figure 2).

Validation Cohort 2

The median global SES-CD score was 3.0 (IQR, 0.0–6.5), and the median EHI value was 32 (IQR, 19.5–46.5). The prevalence of ER was 42.1% (82/195). The AUROC of the EHI for distinguishing AD from ER was 0.693 (95% CI, 0.619–0.767) (Figure 1*C*). Sensitivity was the highest at an EHI cutoff 20 at 83.2% (95% CI, 75.0–89.6) (Table 1). The specificity of the test progressively increased with increasing EHI cutoffs, and at an EHI cutoff of 50, the specificity was observed to be 87.8% (95% CI, 78.7–94.0). The prevalence of AD at EHI of ≥50 was 77.3%, and the prevalence of ER at EHI of <20 was 61.2% (Supplementary Figure 2). PPV and NPV in validation cohorts 1 and 2 at EHI cutoffs 20 and 50 were calculated at assumed AD prevalence ranging from 5% to 75% (Table 2). The presence or absence of prior IBD-related surgeries did not affect the performance of EHI: AUROC in patients with prior surgery: 0.699; 95% CI, 0.588–0.811; AUROC in patients without prior surgery: 0.680; 95% CI, 0.578–0.782; $P = .801$ (Supplementary Figure 4). Paired histology data were available for a subset of validation cohort 2 (N = 79) patients. The AUROC estimate of EHI for distinguishing endohistopathologic healing (defined as endoscopic remission plus histologic remission) from active endoscopic or histologic disease was 0.666 (95% CI, 0.536–0.797) (Supplementary Figure 5).

Table 1. TPs, TNs, Sensitivity, Specificity, PLR, and NLR of the EHI in Distinguishing AD vs ER in validation cohorts 1 and 2

Cohort	EHI threshold	MLG probability[a]	TPs, n	TNs, n	Sensitivity, % (95% CI)	Specificity, % (95% CI)	PLR (95% CI)	NLR (95% CI)
Validation 1:	20	0.550	198	49	97.1 (93.7–98.9)	69.0 (56.9–79.5)	3.13 (2.21–4.44)	0.04 (0.02–0.10)
ER (n = 71)	30	0.746	173	65	84.8 (79.1–89.4)	91.5 (82.5–96.8)	10.04 (4.66–21.63)	0.17 (0.12–0.23)
AD (n = 204)	40	0.876	118	71	57.8 (50.7–64.7)	100.0 (94.9–100.0)	Infinity	0.42 (0.36–0.50)
	50	0.945	76	71	37.3 (30.6–44.3)	100.0 (94.9–100.0)	Infinity	0.63 (0.56–0.70)
Validation 2:	20	—	94	30	83.2 (75.0–89.6)	36.6 (26.2–48.0)	1.31 (1.10–1.58)	0.46 (0.28–0.76)
ER (n = 82)	30	—	74	49	65.5 (56.0–74.2)	59.8 (48.3–70.4)	1.63 (1.21–2.19)	0.58 (0.42–0.79)
AD (n = 113)	40	—	54	65	47.8 (38.3–57.4)	79.3 (68.9–87.4)	2.31 (1.45–3.67)	0.66 (0.54–0.81)
	50	—	34	72	30.1 (21.8–39.4)	87.8 (78.7–94.0)	2.47 (1.29–4.70)	0.80 (0.69–0.92)

MLG, mixed logistic regression; NLR, negative likelihood ratio; PLR, positive likelihood ratio; TN, true negative; TP, true positive.
[a]The population-averaged probability from the MLG model.

Table 2. Sensitivity, Specificity, PPV and NPV of the EHI, CRP, and FC in Distinguishing AD vs ER in the 2 Validation Cohorts Under Different Possible Prevalence of AD

	EHI threshold of 20 units				EHI threshold of 50 units			
	Validation 1		Validation 2		Validation 1		Validation 2	
Sensitivity	0.962		0.917		0.355		0.354	
Specificity	0.641		0.424		1.000		0.909	
Prevalence	PPV	NPV	PPV	NPV	PPV	NPV	PPV	NPV
0.05	0.124	0.997	0.077	0.990	1.000	0.967	0.170	0.964
0.25	0.472	0.981	0.347	0.939	1.000	0.823	0.565	0.808
0.40	0.641	0.962	0.515	0.885	1.000	0.699	0.722	0.679
0.60	0.801	0.918	0.705	0.773	1.000	0.508	0.854	0.484
0.75	0.889	0.849	0.827	0.630	1.000	0.341	0.921	0.319
	CRP Threshold of 3 mg/L				CRP Threshold of 5 mg/L			
	Validation 1		Validation 2		Validation 1		Validation 2	
Sensitivity	0.596		0.625		0.443		0.417	
Specificity	0.938		0.636		0.969		0.727	
Prevalence	PPV	NPV	PPV	NPV	PPV	NPV	PPV	NPV
0.05	0.336	0.978	0.083	0.970	0.429	0.971	0.074	0.960
0.25	0.762	0.874	0.364	0.836	0.826	0.839	0.337	0.789
0.40	0.865	0.777	0.534	0.718	0.905	0.723	0.505	0.652
0.60	0.935	0.608	0.720	0.531	0.955	0.537	0.696	0.454
0.75	0.966	0.436	0.837	0.361	0.977	0.367	0.821	0.294
	FC threshold of 50 μg/g				FC threshold of 250 μg/g			
	Validation 1		Validation 2		Validation 1		Validation 2	
Sensitivity	1.000		0.750		0.683		0.438	
Specificity	0.063		0.788		0.891		1.000	
Prevalence	PPV	NPV	PPV	NPV	PPV	NPV	PPV	NPV
0.05	0.053	1.000	0.157	0.984	0.248	0.982	1.000	0.971
0.25	0.262	1.000	0.541	0.904	0.676	0.894	1.000	0.842
0.40	0.416	1.000	0.702	0.825	0.807	0.808	1.000	0.727
0.60	0.616	1.000	0.841	0.678	0.904	0.652	1.000	0.543
0.75	0.762	1.000	0.914	0.512	0.949	0.484	1.000	0.372

NPV, negative predictive value; PPV, positive predictive value.

Diagnostic Performance of the Endoscopic Healing Index by Disease Location and Phenotype

The AUROC of the EHI for distinguishing AD vs ER was not significantly different across disease locations in both validation cohorts (Supplementary Figure 6*A* and *B*) (pairwise $P \geq .171$ and $P \geq .292$, respectively). EHI performance was also comparable across disease behaviors B1, B2, B3 (Supplementary Figure 7).

Sensitivity and specificity in each location were evaluated at cutoffs that had a high performance in both validation cohort 1 (Supplementary Table 7) and validation cohort 2 (Supplementary Table 8). In validation cohort 1, an EHI cutoff of 20 showed a high sensitivity when the cohort was limited to L1 disease (98.1%; 95% CI, 89.7–100.0), L2 disease (100%, 95% CI, 88.4–100.0%), or L3 disease (95.7%; 95% CI, 90.1–98.6%). The specificity at EHI cutoffs 40 and 50 was 100%, regardless of the disease location. Similarly, in validation cohort 2, sensitivity and specificity at cutoffs 20 and 50, respectively, by disease location was as follows: L1: sensitivity, 84.6%; specificity, 100.0%; L2: sensitivity, 78.9%; specificity, 79.3%; and L3: sensitivity, 85.1%; specificity, 86.2%.

Comparison of the Endoscopic Healing Index to C-reactive protein

The AUROC of EHI to distinguish active endoscopic disease from ER was significantly higher than that of CRP alone in validation cohort 1 (EHI: 0.962; 95% CI, 0.942–0.982; CRP: 0.876; 95% CI, 0.835–0.916; $P < .001$) (Figure 1*A* and *B*). In validation cohort 2, the AUROC of the EHI was numerically better than CRP but did not reach significance (EHI, 0.693; 95% CI, 0.619 0.767; CRP, 0.624; 95% CI, 0.544–0.704; $P = .109$) (Figure 1*C* and *D*). Diagnostic performance of the EHI was also significantly better than the corresponding AUROC of CRP in the training cohort (EHI vs CRP: 0.748 vs. 0.604; $p < .001$; data not shown). The CRP cutoff of 5 mg/L

Table 3. TPs, TNs, Sensitivity, Specificity, PLR, and NLR of the EHI, CRP and FC in Distinguishing AD (n = 183) vs ER (n = 64) in the FC Subcohort of Validation 1

Test	Threshold	MLG Probability[a]	TPs, n	TNs, n	Sensitivity, % (95% CI)	Specificity, % (95% CI)	PLR (95% CI)	NLR (95% CI)
EHI	20	0.542	176	41	96.2 (92.3–98.4)	64.1 (51.1–75.7)	2.68 (1.93–3.72)	0.06 (0.03–0.13)
	30	0.728	153	59	83.6 (77.4–88.7)	92.2 (82.7–97.4)	10.70 (4.60–24.89)	0.18 (0.13–0.25)
	40	0.858	109	64	59.6 (52.1–66.7)	100.0 (94.4–100.0)	INFINITY	0.40 (0.34 0.48)
	50	0.932	65	64	35.5 (28.6–42.9)	100.0 (94.4–100.0)	INFINITY	0.65 (0.58–0.72)
CRP, *mg/L*	3	0.830	109	60	59.6 (52.1–66.7)	93.8 (84.8–98.3)	9.53 (3.66–24.80)	0.43 (0.36–0.52)
	5	0.868	81	62	44.3 (36.9–51.8)	96.9 (89.2–99.6)	14.16 (3.59–55.95)	0.58 (0.50–0.66)
	10	0.908	49	64	26.8 (20.5–33.8)	100.0 (94.4–100.0)	Infinity	0.73 (0.67–0.80)
FC, *μg/g*	50	0.230	183	4	100.0 (98.0–100.0)	6.2 (1.7–15.2)	1.07 (1.00–1.14)	0.00 (0.00)
	150	0.624	144	53	78.7 (72.0–84.4)	82.8 (71.3–91.1)	4.58 (2.66–7.88)	0.26 (0.19–0.35)
	250	0.787	125	57	68.3 (61.0–75.0)	89.1 (78.8–95.5)	6.25 (3.08–12.65)	0.36 (0.28–0.45)

MLG, mixed logistic regression; NLR, negative likelihood ratio; PLR, positive likelihood ratio; TN, true negative; TP, true positive.
[a]The population-averaged probability from the MLG models.

had a sensitivity of 41.7%–44.3% in both validation cohorts 1 and 2 (Tables 3 and 4). At a cutoff of 20, the EHI had a sensitivity of 91.7%–96.2%. PPV and NPV in validation cohorts 1 and 2 at CRP cutoffs of 3 and 5 mg/L were calculated at assumed AD prevalence ranging from 5% to 75% (Table 4) and compared with that of the EHI.

Comparison of the Endoscopic Healing Index to fecal calprotectin

A total of 247 FC assessments were available in validation cohort 1 (Figure 2*A* and *B*), and 81 paired stool samples were available in validation cohort 2 (Figure 2*C* and *D*), for comparison between EHI and FC. The subcohorts with and without FC available from validation cohort 2 were comparable for all baseline characteristics (Supplementary Table 5), although a hidden bias cannot be ruled out. The diagnostic accuracy of the EHI was not significantly different from that of FC in either of the 2 validation cohorts; in validation cohort 1, it was numerically superior to FC (EHI vs FC: AUROC, 0.950 vs 0.923; $P = .147$) but numerically inferior in the validation cohort 2 (EHI vs FC: AUROC, 0.803 vs 0.854; $P = .298$). An FC cutoff of 50 μg/g had 100% sensitivity in validation cohort 1 and 75% sensitivity in validation cohort 2. Corresponding sensitivity for the EHI at a cutoff of 20 was 96% in validation cohort 1 and 92% in validation cohort 2. An FC cutoff of 250 μg/g had 89% specificity in validation cohort 1 and 100% specificity in validation cohort 2. Corresponding specificity for the EHI at a cutoff of 50 was 100% in validation cohort 1 and 91% in validation cohort 2 (Tables 3 and 4). PPV and NPV in validation cohorts 1 and 2 at FC cutoffs 50 and 250 μg/g were calculated at assumed AD prevalence ranging from 5% to 75% (Table 2) and compared with that of the EHI.

Responsiveness of the Endoscopic Healing Index

Endoscopy-paired, longitudinal serum samples were available from 97 patients in validation cohort 1. Effect sizes (ESs) were calculated between baseline and week 12 (n = 70 patients) (Figure 3*A*) and between baseline and

Table 4. True Positives (TPs), True Negatives (TNs), Sensitivity, Specificity, Positive Likelihood Ratio (PLR) and Negative Likelihood Ratio (NLR) of endoscopic Healing Index (EHI), C-Reactive Protein (CRP) and Fecal Calprotectin (FC) in Distinguishing Active Disease (n = 48) vs Endoscopic Remission (n=33) in the FC Sub-Cohort of Validation 2

Test	Threshold	TPs, n	TNs, n	Sensitivity, % (95% CI)	Specificity, % (95% CI)	PLR (95% CI)	NLR (95% CI)
EHI	20	44	14	91.7 (80.0–97.7)	42.4 (25.5–60.8)	1.59 (1.17–2.16)	0.20 (0.07–0.54)
	30	34	24	70.8 (55.9–83.0)	72.7 (54.5–86.7)	2.60 (1.45–4.67)	0.40 (0.25–0.65)
	40	25	28	52.1 (37.2–66.7)	84.8 (68.1–94.9)	3.44 (1.47–8.06)	0.57 (0.41–0.78)
	50	17	30	35.4 (22.2–50.5)	90.9 (75.7–98.1)	3.90 (1.24–12.24)	0.71 (0.56–0.90)
CRP, *mg/L*	3	30	21	62.5 (47.4–76.0)	63.6 (45.1–79.6)	1.72 (1.04–2.84)	0.59 (0.38–0.92)
	5	20	24	41.7 (27.6–56.8)	72.7 (54.5–86.7)	1.53 (0.80–2.93)	0.80 (0.58–1.10)
	10	11	26	22.9 (12.0–37.3)	78.8 (61.1–91.0)	1.08 (0.47–2.50)	0.98 (0.77–1.24)
FC, *μg/g*	50	36	26	75.0 (60.4–86.4)	78.8 (61.1–91.0)	3.54 (1.80–6.97)	0.32 (0.19–0.53)
	150	28	33	58.3 (43.2–72.4)	100.0 (89.4–100.0)	Infinity	0.42 (0.30–0.58)
	250	21	33	43.8 (29.5–58.8)	100.0 (89.4–100.0)	Infinity	0.56 (0.44–0.72)

MLG, mixed logistic regression; NLR, negative likelihood ratio; PLR, positive likelihood ratio; TN, true negative; TP, true positive.

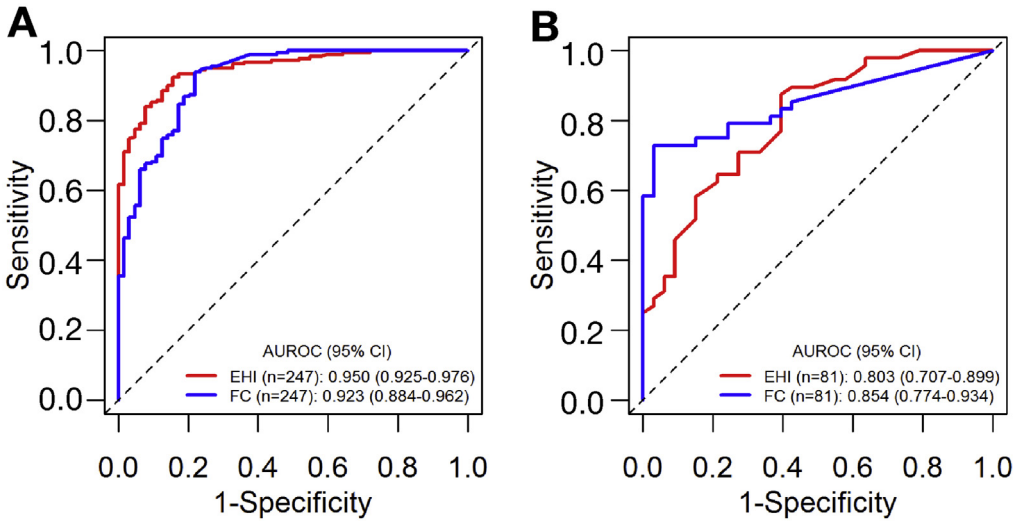


Figure 2. ROC curves of the EHI and FC for distinguishing active disease vs endoscopic remission in (*A*) validation cohort 1 and (*B*) validation cohort 2. In validation cohort 1, mixed logistic regression models with random intercepts for individual participants were used to combine multiple samples of the same participant.

week 54 (n = 59 patients) (Figure 3*B*) for the 2 endoscopic indices (SES-CD and CDEIS) and the 3 biomarkers (EHI, CRP, FC). Between baseline and week 12, the median ES of the EHI (1.10; IQR, 0.52 –1.83) was numerically better than that of FC (0.96; IQR, 0.43–1.96; $P = .423$) and significantly better than that of CRP (0.26; IQR, 0.11–0.51; $P < .001$). Similar results were noted between baseline and week 54, where the median ES of the EHI (1.64; IQR, 0.65–2.29) was numerically better than that of FC (1.16; IQR, 0.51–2.32; $P = .574$) and significantly better than that of CRP (0.21; IQR, 0.09–0.56; $P < .001$). Between both time intervals, the median ES of the EHI was on par with those of the endoscopic scores and mirrored changes in SES-CD and CDEIS (SES-CD between weeks 0 and 12: 1.53; IQR, 0.67–2.23; $P = .077$; SES-CD between weeks 0 and 54: 1.87; IQR, 0.93–2.67; $P = .069$; CDEIS between weeks 0 and 12: 1.29, IQR, 0.8 –2.25; $P = .182$; CDEIS between weeks 0 and 54: 1.50; IQR, 0.81–2.17; $P = .997$).

Discussion

Symptoms are often not representative of disease activity in CD, and endoscopy represents the current gold standard for objective disease assessment. Despite this, the majority of CD patients starting biologic therapy have no follow-up endoscopy within 24 months of treatment initiation.[10] The exact reason for this gap is unknown, but cost, risk, and burden are likely to be drivers of these practice patterns. FC is a stool-based biomarker that is widely available and approved by regulatory agencies to aid in the diagnosis of IBD. Its routine availability lends itself to use as a monitoring tool, particularly considering the recent emerging evidence supporting biomarker-based adjustments in therapy to optimize the achievement of MH.[11] Although its use in Europe is well established, and home FC monitoring kits are available,[26] in the United States, fewer than 2% of patients with established IBD undergo FC testing.[10] A considerable gap remains in monitoring CD patients for mucosal inflammation to guide treatment decisions.

In the current study, we have developed and validated a novel 13-biomarker panel serum-based assay (EHI; Prometheus Laboratories Inc, San Diego, CA) that detects mucosal inflammation in CD. All of the selected 13 markers have documented roles in CD pathophysiology. SAA1, an acute phase reactant like CRP, correlates with lack of mucosal healing and serves as a surrogate marker of disease activity even in those patients in whom CRP is not up-regulated.[27] Compromised barrier function observed in CD may be due to altered barrier permeability caused by MMP family members, recruitment of proinflammatory cytokines, and altered angiogenesis. MMP3 is down-regulated in patients with CD with fibrostenotic phenotype,[28] and MMP9 has been implicated in the pathogenesis of CD[29,30] with anti-MMP9 antibodies considered for targeting active CD.[31] EMMPRIN, an inducer of MMPs, plays a role in wound healing and nutrient transport inflammation and has been suggested as a protein interaction partner of NOD2.[32,33] Transforming growth factor α is a growth factor and epidermal growth factor receptor pathway ligand. Epidermal growth factor receptor pathway ligands have been implicated in immunity, inflammation, and tissue repair.[34] The angiopoietin system, including ANG1 and ANG2, have been proposed as factors to maintain pathologic angiogenesis during the development of IBD.[35,36] IL7 is a critical survival factor for lymphocytes, and its availability determines the size and proliferative state of the resting T-cell pool.[37] Adhesion molecules ICAM-1, VCAM-1, and CEA-CAM family members are known to be up-regulated in IBD. Zundler et al[38] have suggested that the a4b1-VCAM1 axis is involved in mechanisms controlling the homing of T-effector cells to the inflamed gut in CD.

EHI assay was subsequently validated in 2 independent cohorts representing both early disease and biologic-naive patients with CD and those with longer-duration CD with prior bowel surgeries, disease-related complications, and multiple biologic exposures. Across both cohorts, EHI was observed to have an overall favorable diagnostic accuracy for identifying endoscopic inflammation, and in the second

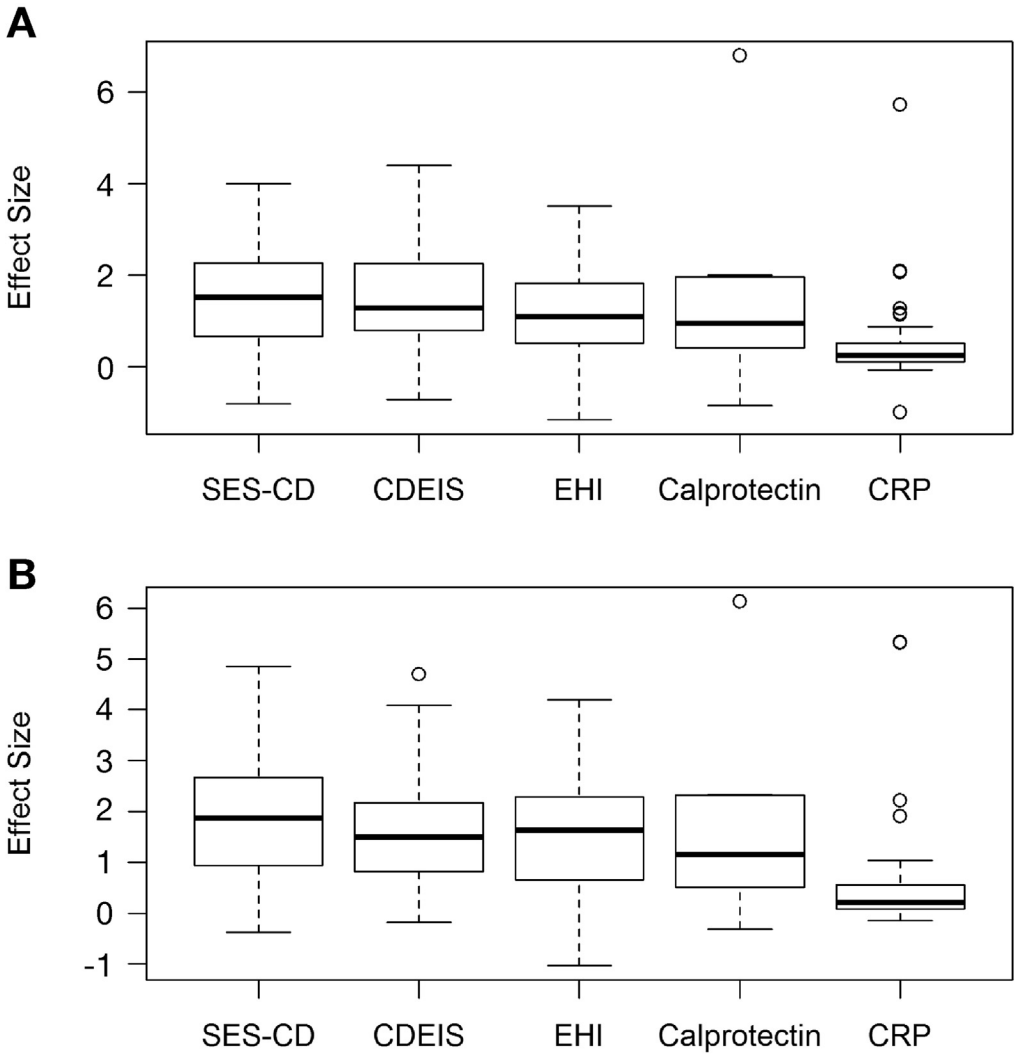


Figure 3. Use of the EHI for monitoring. Boxplots of ES of the SES-CD, CDEIS, EHI, FC, and CRP in monitoring disease changes in validation cohort 1. (*A*) Between baseline and week 12 (n = 70). The median ES of the EHI (1.10; IQR, 0.52–1.83) was on par with those of the SES-CD (1.53; IQR, 0.67–2.23; $P = .077$) and CDEIS (1.29; IQR, 0.80–2.25; $P = .182$), slightly better than that of FC (0.96; IQR, 0.43–1.96; $P = .423$), and significantly better than that of CRP (0.26; IQR, 0.11–0.51; $P < .001$). (*B*) Between baseline and week 54 (n = 59). The median ES of the EHI (1.64; IQR, 0.65–2.29) was on par with that of SES-CD (1.87; IQR, 0.93–2.67, $P = .069$), slightly better than those of CDEIS (1.50; IQR, 0.81–2.17; $P = .997$) and FC (1.16; IQR, 0.51–2.32; $P = .574$), and significantly better than that of CRP (0.21; IQR, 0.09–0.56; $P < .001$).

validation cohort, an early exploratory analysis observed it to have a reasonable diagnostic accuracy for identifying histologic inflammation. Based on these data, a cutoff of 20 was observed to have a high sensitivity for ruling out endoscopic inflammation, and a cutoff of 50 was observed to have a high specificity for ruling in endoscopic inflammation. The sensitivity of CRP was consistently poor across both validation cohorts, indicating that CRP is an unreliable serum marker for ruling out endoscopic inflammation. Most notably, the diagnostic accuracy of EHI was observed to be consistent across disease locations and disease phenotypes, and its performance was comparable to that of FC and superior to CRP.

CD patients strongly prefer blood-based testing over fecal testing,[13] but, to date, there have been no routinely available blood-based tests with a diagnostic performance comparable to that of FC. A recent study examined the diagnostic accuracy of serum calprotectin for differentiating patients with IBD from healthy control individuals, and although serum calprotectin had a favorable diagnostic accuracy for identifying IBD (AUC, 0.87; 95% CI, 0.78–0.97), it was still less accurate than FC (AUC, 0.99; 95% CI, 0.87–1.00; $P = .01$).[39] Our study is, therefore, novel and substantially additive to the current tests available for monitoring patients with CD in routine practice.

When comparing EHI to FC, there are several important observations from our study worth noting. First, the sensitivity and specificity of EHI at cutoffs of 20 and 50, respectively, remained consistent across both validation cohorts (92%–96% and 91%–100%). Although the specificity of FC at a cutoff of 250 μg/g remained stable between validation cohorts (89%–100%), the sensitivity of FC at a cutoff of 50 μg/g was quite different between validation cohort 1 (an early-disease, biologic-naive population; 100%) and validation cohort 2 (routine practice, long-duration, biologic-exposed population; 75%). Second, the diagnostic accuracy of the EHI was consistent across disease locations and phenotypes. Prior literature has shown that the diagnostic performance of FC varies by disease location, and even in the presence of very large ulcers, patients with ileal CD may not have markedly elevated FC levels.[40] Furthermore, the presence of perianal fistulas, even in the absence of colonic inflammation, leads to elevated FC.[41] Third, one of the major limitations of FC is the variability across

platforms, collection techniques, and timing of sample collection, which have downstream implications on the diagnostic performance of FC.[22,26,42–47] The EHI was built to ensure reproducibility and consistency in performance, which was observed throughout the validation process. Although no power analysis was performed for the comparison of the EHI and FC, we note that the sample size for this comparison was somewhat limited, with 247 samples in validation cohort 1 and 81 samples in validation cohort 2.

EHI performance was consistent compared with other serum markers across various endoscopic AD prevalence values for patients with established CD, which ranged from 5% to 75%. At a threshold of 20, the EHI had a high NPV (84.9%–99.7%) in validation cohort 1 across disease prevalence of 5%–75% and an NPV of 63%–99% in validation cohort 2. In contrast, the NPV of CRP decreased with increasing disease prevalence and was as low as 29% at a cutoff of 5 mg/L in validation cohort 2, indicating that CRP is a poor marker to rule out AD. Performance of FC was better than CRP, with a better NPV at a cut-off of 50 μg/g, which was comparable to or lower than that of EHI 20.

Our study has several strengths, including the multicenter, multinational collaboration with varying patient populations and disease characteristics, availability of both endoscopic and histologic disease activity assessments, comparative accuracy assessments against both FC and CRP, and longitudinal comparisons for responsiveness in a prospective clinical trial. Some limitations, however, remain. First, the observed diagnostic accuracy of EHI for identifying histologic inflammation was exploratory and is available in only a subset of patients from validation cohort 2. Given that the test was not trained on a cohort with histologic assessments, the current test is associated with endoscopic MH, and further work is needed to understand how the EHI performs for identifying the evolving definition of MH, which encompasses both endoscopic and histologic activity. Second, we did not have routine cross-sectional imaging assessments, and further analyses are required to understand how the EHI compares against cross-sectional imaging–based assessments of disease activity, particularly for isolated small-bowel CD. Third, although we observed a similar performance of the EHI in patients with and without prior IBD surgeries, we were unable to assess the prognostic value of the EHI for predicting future endoscopic recurrence or disease relapse, particularly in postoperative CD patients, for whom FC has a clearly established role.[48] Fourth, reasons for differences in performance of the EHI in various CD populations need to be further explored. Performance of the EHI in validation cohort 2 was lower than the corresponding performance in validation cohort 1, but the performances of FC and CRP were lower as well, likely reflecting the fact that participants in the cohort were mainly patients with chronic CD with altered cellular signaling pathways. Finally, although the second validation cohort encompassed multiple biologic exposures, because treatment paradigms shift to include additional pathway-targeted therapies, continued validation of the EHI will be needed to ensure generalizability across all populations. In addition, future studies should assess the cost effectiveness of the EHI, relative to both colonoscopy and to other available biomarker tests, such as CRP and FC. Such efforts will need to take into account the accuracy of the various biomarker tests in patient populations with different prevalence of endoscopic disease activity.

In conclusion, we have developed and validated a serum-based assay with a favorable diagnostic accuracy for identifying mucosal inflammation, which is comparable to FC. The serum-based assay was observed to be responsive to changes in endoscopic disease activity, and accuracy was consistent across subgroups. This test could help bridge current gaps in monitoring patients with CD.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at www.gastrojournal.org, and at https://doi.org/10.1053/j.gastro.2019.10.034.

References

1. Peyrin-Biroulet L, Loftus EV Jr, Colombel JF, et al. The natural history of adult Crohn's disease in population-based cohorts. Am J Gastroenterol 2010;105:289–297.
2. Baert F, Moortgat L, Van Assche G, et al. Mucosal healing predicts sustained clinical remission in patients with early-stage Crohn's disease. Gastroenterology 2010;138:463–468.
3. Froslie KF, Jahnsen J, Moum BA, et al. Mucosal healing in inflammatory bowel disease: results from a Norwegian population-based cohort. Gastroenterology 2007;133:412–422.
4. af Björkesten CG, Nieminen U, Sipponen T, et al. Mucosal healing at 3 months predicts long-term endoscopic remission in anti-TNF-treated luminal Crohn's disease. Scand J Gastroenterol 2013;48:543–551.
5. Shah SC, Colombel JF, Sands BE, et al. Systematic review with meta-analysis: mucosal healing is associated with improved long-term outcomes in Crohn's disease. Aliment Pharmacol Ther 2016;43:317–333.
6. Peyrin-Biroulet L, Sandborn W, Sands BE, et al. Selecting Therapeutic Targets in Inflammatory Bowel Disease (STRIDE): determining therapeutic goals for treat-to-target. Am J Gastroenterol 2015;110:1324–1338.
7. Bouguen G, Levesque BG, Feagan BG, et al. Treat to target: a proposed new paradigm for the management of Crohn's disease. Clin Gastroenterol Hepatol 2015;13:1042–1050.
8. Bouguen G, Levesque BG, Pola S, et al. Endoscopic assessment and treating to target increase the likelihood of mucosal healing in patients with Crohn's disease. Clin Gastroenterol Hepatol 2014;12:978–985.
9. Buisson A, Gonzalez F, Poullenot F, et al. Comparative acceptability and perceived clinical utility of monitoring tools: a nationwide survey of patients with inflammatory bowel disease. Inflamm Bowel Dis 2017;23:1425–1433.
10. Limketkai BN, Singh S, Jairath V, et al. US practice patterns and impact of monitoring for mucosal

inflammation after biologic initiation in inflammatory bowel disease. Inflamm Bowel Dis 2019;25:1828–1837.
11. Colombel JF, Panaccione R, Bossuyt P, et al. Effect of tight control management on Crohn's disease (CALM): a multicentre, randomised, controlled phase 3 trial. Lancet 2018;390(10114):2779–2789.
12. Lin JF, Chen JM, Zuo JH, et al. Meta-analysis: fecal calprotectin for assessment of inflammatory bowel disease activity. Inflamm Bowel Dis 2014;20:1407–1415.
13. Mosli MH, Zou G, Garg SK, et al. C-reactive protein, fecal calprotectin, and stool lactoferrin for detection of endoscopic activity in symptomatic inflammatory bowel disease patients: a systematic review and meta-analysis. Am J Gastroenterol 2015;110:802–819.
14. van Deen WK, van Oijen MG, Myers KD, et al. A nationwide 2010–2012 analysis of U.S. health care utilization in inflammatory bowel diseases. Inflamm Bowel Dis 2014;20:1747–1753.
15. Pepe MS, Feng Z, Janes H, et al. Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: standards for study design. J Natl Cancer Inst 2008;100:1432–1438.
16. Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ 2015;351:h5527.
17. Louis E, Mary JY, Vernier-Massouille G, et al. Maintenance of remission among patients with Crohn's disease on antimetabolite therapy after infliximab therapy is stopped. Gastroenterology 2012;142:63–70.
18. D'Haens G, Vermeire S, Lambrecht G, et al. Increasing infliximab dose based on symptoms, biomarkers, and serum drug concentrations does not increase clinical, endoscopic, and corticosteroid-free remission in patients with active luminal Crohn's disease. Gastroenterology 2018;154:1343–1351.
19. Dulai PS, Levesque BG, Feagan BG, et al. Assessment of mucosal healing in inflammatory bowel disease: review. Gastrointest Endosc 2015;82:246–255.
20. Daperno M, D'Haens G, Van Assche G, et al. Development and validation of a new, simplified endoscopic activity score for Crohn's disease: the SES-CD. Gastrointest Endosc 2004;60:505–512.
21. Mary JY, Modigliani R. Development and validation of an endoscopic index of the severity for Crohn's disease: a prospective multicentre study. Groupe d'Etudes Thérapeutiques des Affections Inflammatoires du Tube Digestif (GETAID). Gut 1989;30:983–989.
22. Padoan A, D'Inca R, Scapellato ML, et al. Improving IBD diagnosis and monitoring by understanding preanalytical, analytical and biological fecal calprotectin variability. Clin Chem Lab Med 2018;56:1926–1935.
23. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44:837–845.
24. Simel DL, Samsa GP, Matchar DB. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. J Clin Epidemiol 1991;44:763–770.
25. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Mahwah, NJ: Lawrence Erlbaum, 1988.
26. Heida A, Knol M, Kobold AM, et al. Agreement between home-based measurement of stool calprotectin and ELISA results for monitoring inflammatory bowel disease activity. Clin Gastroenterol Hepatol 2017;15:1742–1749.
27. Yarur AJ, Quintero MA, Jain A, et al. Serum amyloid A as a surrogate marker for mucosal and histologic inflammation in patients with Crohn's disease. Inflamm Bowel Dis 2017;23:158–164.
28. de Bruyn JR, van den Brink GR, Steenkamer J, et al. Fibrostenotic phenotype of myofibroblasts in Crohn's disease is dependent on tissue stiffness and reversed by LOX inhibition. J Crohns Colitis 2018;12:849–859.
29. Baugh MD, Perry MJ, Hollander AP, et al. Matrix metalloproteinase levels are elevated in inflammatory bowel disease. Gastroenterology 1999;117:814–822.
30. Faubion WA Jr, Fletcher JG, O'Byrne S, et al. EMerging BiomARKers in inflammatory bowel disease (EMBARK) study identifies fecal calprotectin, serum MMP9, and serum IL-22 as a novel combination of biomarkers for Crohn's disease activity: role of cross-sectional imaging. Am J Gastroenterol 2013;108:1891–1900.
31. Schreiber S, Siegel CA, Friedenberg KA, et al. A phase 2, randomized, placebo-controlled study evaluating matrix metalloproteinase-9 inhibitor, andecaliximab, in patients with moderately to severely active Crohn's disease. J Crohns Colitis 2018;12:1014–1020.
32. Kumar D, Vetrivel U, Parameswaran S, et al. Structural insights on druggable hotspots in CD147: a bull's eye view. Life Sci 2019;224:76–87.
33. Till A, Rosenstiel P, Brautigam K, et al. A role for membrane-bound CD147 in NOD2-mediated recognition of bacterial cytoinvasion. J Cell Sci 2008;121:487–495.
34. Zaiss DMW, Gause WC, Osborne LC, et al. Emerging functions of amphiregulin in orchestrating immunity, inflammation, and tissue repair. Immunity 2015;42:216–226.
35. Algaba A, Linares PM, Encarnacion Fernandez-Contreras M, et al. The effects of infliximab or adalimumab on vascular endothelial growth factor and angiopoietin 1 angiogenic factor levels in inflammatory bowel disease: serial observations in 37 patients. Inflamm Bowel Dis 2014;20:695–702.
36. Linares PM, Chaparro M, Gisbert JP. Angiopoietins in inflammation and their implication in the development of inflammatory bowel disease. A review. J Crohns Colitis 2014;8:183–190.
37. Martin CE, Spasova DS, Frimpong-Boateng K, et al. Interleukin-7 availability is maintained by a hematopoietic cytokine sink comprising innate lymphoid cells and T cells. Immunity 2017;47:171–182.
38. Zundler S, Fischer A, Schillinger D, et al. The $\alpha 4\beta 1$ homing pathway is essential for ileal homing of Crohn's disease effector T cells in vivo. Inflamm Bowel Dis 2017;23:379–391.
39. Kalla R, Kennedy NA, Ventham NT, et al. Serum calprotectin: a novel diagnostic and prognostic marker in

inflammatory bowel diseases. Am J Gastroenterol 2016; 111:1796–1805.
40. Gecse KB, Brandse JF, van Wilpe S, et al. Impact of disease location on fecal calprotectin levels in Crohn's disease. Scand J Gastroenterol 2015;50:841–847.
41. Stevens T, D'Haens GR, Buskens CJ, et al. P150. Low diagnostic accuracy of faecal calprotectin to predict mucosal inflammation in Crohn's disease patients with active perianal fistulas. Abstract presented at: European Crohn's and Colitis Organization Congress; 2018; Vienna, Austria.
42. Whitehead SJ, French J, Brookes MJ, et al. Between-assay variability of faecal calprotectin enzyme-linked immunosorbent assay kits. Ann Clin Biochem 2013;50:53–61.
43. Oyaert M, Boel A, Jacobs J, et al. Analytical performance and diagnostic accuracy of six different faecal calprotectin assays in inflammatory bowel disease. Clin Chem Lab Med 2017;55:1564–1573.
44. Rodriguez A, Yokomizo L, Christofferson M, et al. Correlation of rapid point-of-care vs send-out fecal calprotectin monitoring in pediatric inflammatory bowel disease. World J Gastrointest Pharmacol Ther 2017;8:127–130.
45. Hejl J, Theede K, Mollgren B, et al. Point of care testing of fecal calprotectin as a substitute for routine laboratory analysis. Pract Lab Med 2018;10:10–14.
46. Whitehead SJ, Ford C, Gama RM, et al. Effect of faecal calprotectin assay variability on the management of inflammatory bowel disease and potential role of faecal S100A12. J Clin Pathol 2017;70:1049–1056.
47. Jang HW, Kim HS, Park SJ, et al. Accuracy of three different fecal calprotectin tests in the diagnosis of inflammatory bowel disease. Intest Res 2016;14:305–313.
48. Wright EK, Kamm MA, De Cruz P, et al. Measurement of fecal calprotectin improves monitoring and detection of recurrence of Crohn's disease after surgery. Gastroenterology 2015;148:938–947.

Received September 25, 2018. Accepted October 26, 2019.

Correspondence
Address correspondence to: Geert D'Haens MD, PhD, Dept of Gastroenterology, Amsterdam University Medical Centers, Meibergdreef 9, 1100 DZ Amsterdam, The Netherlands. e-mail: g.dhaens@amc.uva.nl; fax: +3120 69 17033.

Acknowledgments
Author contributions: All authors approved the final draft submitted. Study concept, design, and supervision: Anjali Jain, Kurtis R. Bray, Geert R. D'Haens, William J. Sandborn, Mark S. Silverberg, Michael Hale, Larry Mimms, and Parambir S. Dulai. Provision of specimens: Geert R. D'Haens, Mark S. Silverberg, David Laharie, Severine Vermeire, William J. Sandborn, Edouard Louis, Edoardo Savarino, Giorgia Bodini, Andres Yarur, and Waqqas Afif. Acquisition of clinical data: Geert R. D'Haens, Robert Battat, Orlaith Kelly, Mark S. Silverberg, David Laharie, Severine Vermeire, William J. Sandborn, Edouard Louis, Edoardo Savarino, Giorgia Bodini, Andres Yarur, Brigid S. Boland, Waqqas Afif, and Parambir S. Dulai. Analytical biomarker development and data acquisition: Michael Hale, Anjali Jain, Jessica Ho, Benjamin Huang, Crystal Kuy, Kelly D. Hester, and Lauren Okada. Statistical analysis: Venkateswarlu Kondragunta, Anjali Jain, Xiao-jun Li, Michael Hale, Lauren Okada, Kurtis R. Bray, and Parambir S. Dulai. Drafting of manuscript: Anjali Jain, Xiao-jun Li, Kurtis R. Bray, Geert R. D'Haens, William J. Sandborn, Kelly D. Hester, and Parambir S. Dulai. Interpretation, critical revisions, and final approval: All authors.

Conflicts of interest
The authors disclose the following: Geert R. D'Haens has received consulting and/or lecture fees from AbbVie, Ablynx, Allergan, Alphabiomics, Amakem, Amgen, AM Pharma, Arena Pharmaceuticals, Biogen, Bristol-Myers Squibb, Boehringer Ingelheim, Celgene/Receptos, Celltrion, Echo Pharmaceuticals, Eli Lilly, Engene, Ferring, Dr Falk Pharma, Galapagos, Genentech/Roche, Gilead, GlaxoSmithKline, Gossamerbio, Pfizer, Immunic, Johnson & Johnson, Kintai Therapeutics, Millennium/Takeda, Medtronics, Mitsubishi Pharma, Merck Sharp Dome, Mundipharma, Nextbiotics, Novo Nordisk, Otsuka, Pfizer/Hospira, Photopill, Prodigest, Prometheus Laboratories/Nestle, Progenity, Protagonist, RedHill, Robarts Clinical Trials, Samsung Bioepis, Sandoz, Seres/Nestle, Setpoint, Shire, Takeda, Teva, Tigenix, Tillotts, Topivert, Versant and Vifor. Orlaith Kelly served on the tofacitinib advisory board Ireland 2018 and received an AbbVie CIHR/CAG Advanced IBD Fellowship Bursary award, Canada 2014 and 2015. Mark S. Silverberg has been a speaker for AbbVie, Janssen, Prometheus, Takeda, Shire, Pfizer/Hospira, Ferring, Novartis, Lilly; served on the advisory boards of AbbVie, Allergan, Janssen, Prometheus, Takeda, Shire, Pfizer/Hospira, Ferring; received research support for AbbVie, Janssen, Prometheus, Takeda, Pfizer/Hospira; and has served as a consultant for AbbVie, Janssen, Prometheus, Takeda, Pfizer/Hospira. David Laharie as served on the board of or received lectures fees from AbbVie, Celgene, Ferring, Janssen, Merck Sharpe & Dohme (MSD), Novartis, Pfizer, Roche, and Takeda. Edouard Louis has received research grants from Takeda and Pfizer; educational grants from AbbVie, MSD, Takeda; and speaker fees from AbbVie, Ferring, MSD, Chiesi, Falk, Takeda, Hospira, Janssen, Pfizer; has served on the advisory board for AbbVie, Ferring, MSD, Mitsubishi Pharma, Takeda, Celltrion, Celgene, Hospira, Janssen and as a consultant for AbbVie. Edoardo Savarino has received consulting/lecture fees from AbbVie, MSD, Takeda, Janssen, Sofar, Malesci. Giorgia Bodini has been an invited speaker for AbbVie, MSD, and Takeda. Andres Yarur has received consulting fees from Takeda Pharmaceuticals and Prometheus Laboratories and has served on the speakers bureau for AbbVie, Takeda Pharmaceuticals, and Prometheus Laboratories. Brigid S. Boland has received consulting fees from AbbVie and Prometheus Laboratories outside of the submitted work. Waqqas Afif has performed consulting for AbbVie, Janssen, Pfizer, Merck, Takeda, Shire, and Allergan and received research support AbbVie, Janssen, Theradiag, Prometheus, and Ferring. Severine Vermeire has received grant support from MSD, AbbVie, Pfizer, Johnson & Johnson, and Takeda; lecture fees from AbbVie, MSD, Ferring Pharmaceuticals, Takeda, Hospira;, and consultancy fees from AbbVie, Takeda, Pfizer, Ferring Pharmaceuticals, Shire Pharmaceuticals Group, Prometheus, MSD, Hospira, Mundipharma, Celgene, Galapagos, and Genentech/Roche. Siddharth Singh has received consulting fees from AbbVie, Pfizer, Takeda, and AMAG Pharmaceuticals. Angelina Collins has received consulting fees, and/or honoraria from AbbVie and Janssen. Mark A. Valasek received research support from Prometheus. William J. Sandborn has received research grants from Atlantic Healthcare Limited, Amgen, Genentech, Gilead Sciences, AbbVie, Janssen, Takeda, Lilly, Celgene/Receptos; consulting fees from AbbVie, Allergan, Amgen, Boehringer Ingelheim, Celgene, Conatus, Cosmo, Escalier Biosciences, Ferring, Genentech, Gilead, Janssen, Lilly, Miraca Life Sciences, Nivalis Therapeutics, Novartis Nutrition Science Partners, Oppilan Pharma, Otsuka, Paul Hastings, Pfizer, Precision IBD, Progenity, Prometheus Laboratories, Ritter Pharmaceuticals, Robarts Clinical Trials (owned by Health Academic Research Trust), Salix, Shire, Seres Therapeutics, Sigmoid Biotechnologies, Takeda, Tigenix, Tillotts Pharma, UCB Pharma, Vivelix; and stock options from Ritter Pharmaceuticals, Oppilan Pharma, Escalier Biosciences, Precision IBD, and Progenity. Parambir S. Dulai received research support, consulting fees, and honoraria from Takeda; received research support from Pfizer, and research support and consulting fees from Janssen. Michael Hale, Jessica Ho, Venkateswarlu Kondragunta, Benjamin Huang, Crystal Kuy, Lauren Okada, Xiao-Jun Li, Kelly D. Hester, Kurtis R. Bray, Larry Mimms, and Anjali Jain were all employees of Prometheus Laboratories Inc.

Funding
The University of California San Diego cohorts were recruited with support from a grant to William Sandborn and Parambir S. Dulai through the Litwin Pioneers Program Crohn's and Colitis Foundation and an unrestricted collaboration grant from Prometheus. The study was partially supported by the NIDDK-funded San Diego Digestive Diseases Research Center (P30 DK120515). Parambir S. Dulai is supported by the AGA Research Scholar Award. The TAILORIX cohort was recruited in a prospective clinical trial by the GETAID, sponsored by Janssen Biologics and Merck Sharp and Dome. Prometheus was responsible for running assays and analyses. Primary clinical data and reference assessments for disease activity for the training and validation cohorts were collected independent of Prometheus by local coordinating sites.

Supplementary Methods

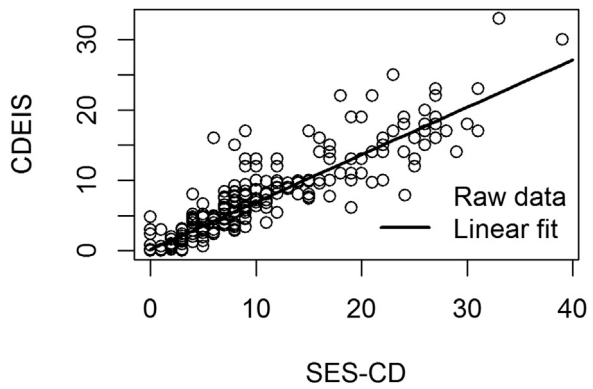
Sample Storage and Testing

Serum CRP was tested by using a turbidity assay (hsCRP, Beckman Coulter, Brea, CA). Other serum biomarkers were measured via multiplexed fluorescent immunoassays. Serum specimens were added to a mixture of color-coded beads, which were precoated with analyte-specific capture antibodies. Biotinylated detection antibodies specific to the target analytes were then added to form an antibody-antigen sandwich. Afterward, phycoerythrin-conjugated streptavidin was added, which bound to the biotinylated detection antibodies. The magnitude of the phycoerythrin-derived signal, which was directly proportional to the amount of target analyte in the sample, was then detected in a flow-based instrument. Values of FC from the STORI and TAILORIX trials were used directly without retesting (Buhlmann Laboratories, Schonenbuch, Switzerland). Values of FC in validation cohort 2 were measured using the QUANTA Lite Calprotectin Extended Range assay (Inova Diagnostics, San Diego, CA).

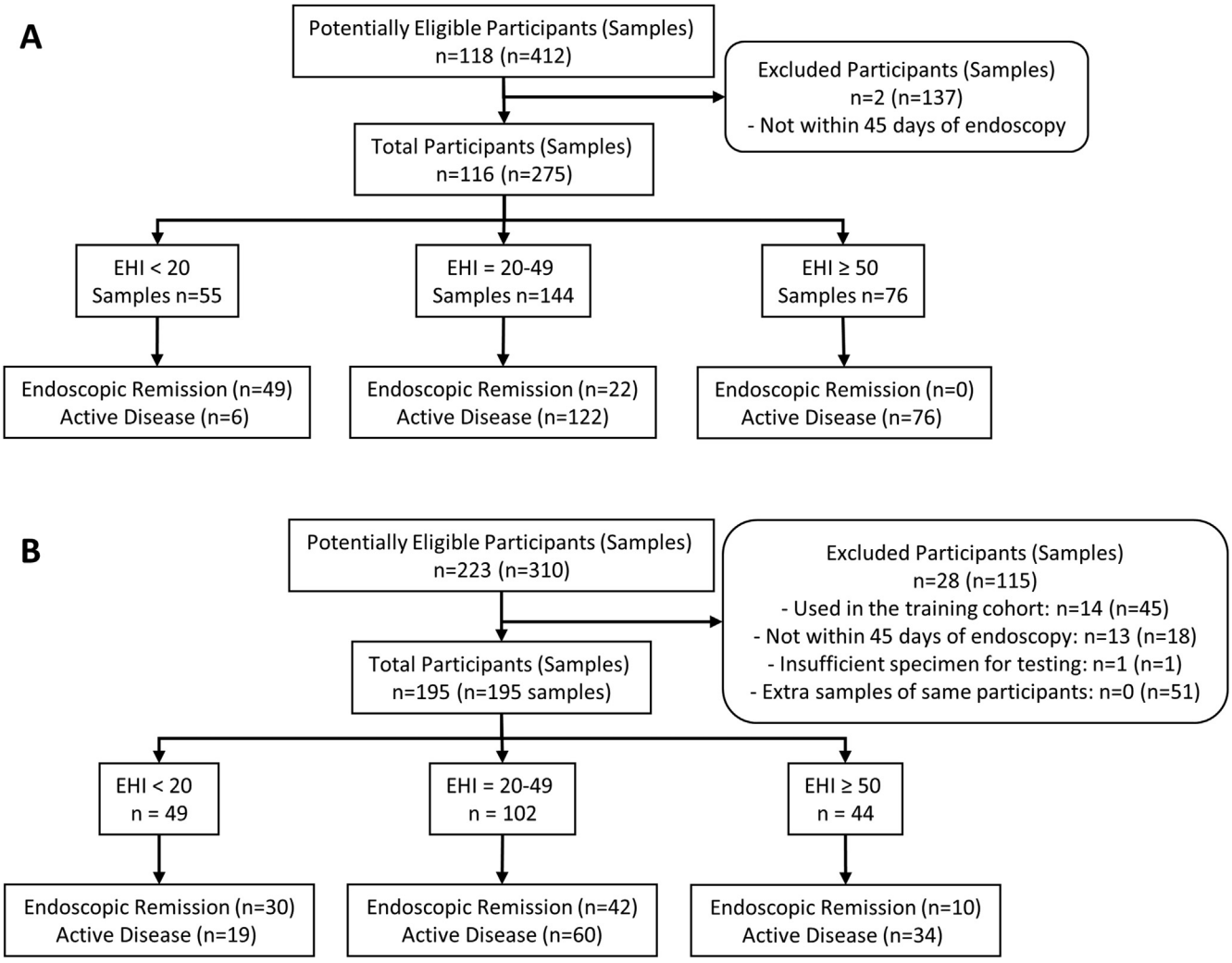
Statistical Analysis

Continuous variables were reported as medians with IQRs, and compared between groups using the Mann-Whitney test. Categorical variables are reported as numbers and percentages and are compared between groups using the Fisher exact test. The Delong method was used for computing the 95% CI of the AUROCs and for comparing AUROCs of different biomarkers on paired samples.[23] Exact binomial confidence limits were used for the 95% CIs of sensitivity and specificity. The 95% CIs of PLR and NLR were computed by using formulae provided by Simel et al.[24] Pairwise Wilcoxon rank sum test was used for comparing the ESs of different variables. A P value (2-sided) of .05 or lower was considered significant. All data analysis was carried out using JMP, version 12.0 (SAS Institute, Cary, NC) or R, version 3.3.2.

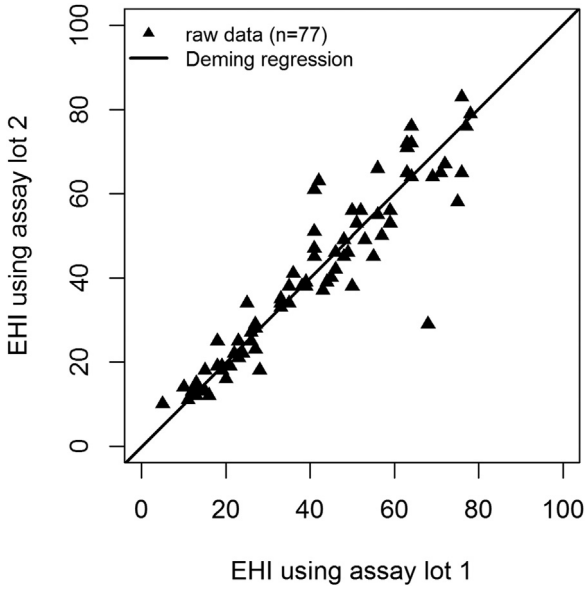
For validation cohort 1, where longitudinal samples were available, a mixed-effect logistic regression modeling was used to assess the performance of the EHI, CRP, and FC (response: disease status of ER or AD; fixed effect: EHI, CRP, or FC; random effect: random intercepts for study participants). The values of CRP and FC were first logarithmic transformed in the modeling. For samples whose FC test result was 0, the corresponding FC values were set to 10 μg/g, which was well below the minimal nonzero FC value of 31 μg/g observed in the cohort. Subsequently, ESs[25] of the SES-CD, CDEIS, EHI, FC, and CRP were calculated between baseline and week 12 and between baseline and week 54 to assess the responsiveness of those variables. Because the underlying data were mostly not normally distributed, the corresponding median and IQR were reported instead of the mean.



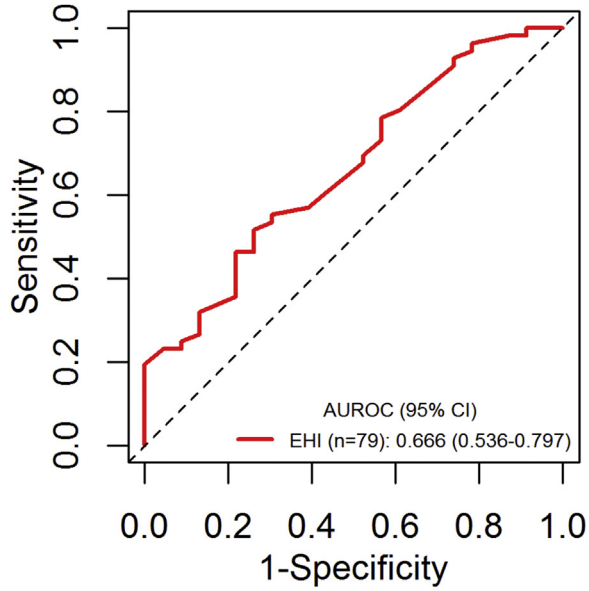
Supplementary Figure 1. SES-CD and CDEIS scores in validation cohort 1. The linear fit (CDEIS $= 0.1569 + 0.6744 \times$ SES-CD) was used to convert SES-CD scores to CDEIS scores for samples that had only SES-CD scores in the training cohort.



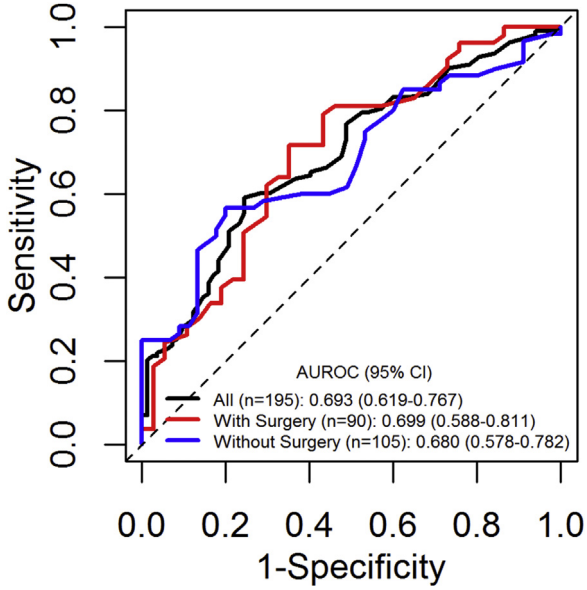
Supplementary Figure 2. Patient and sample flowchart. (*A*) Validation cohort 1. (*B*) Validation cohort 2.



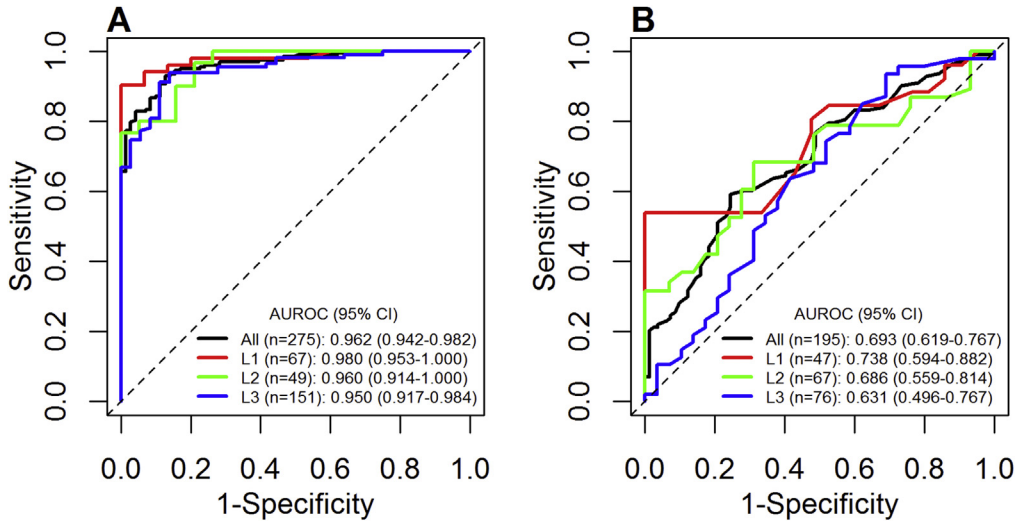
Supplementary Figure 3. Reproducibility of the EHI when the same samples were analyzed using 2 different lots of reagents. Serum samples from patients with clinically diagnosed CD (n = 77) were used to study reproducibility of the EHI. The Deming regression had a slope of 1.005 (95% CI, 0.926–1.087) and an intercept of –0.298 (95% CI, –2.790 to 2.144).



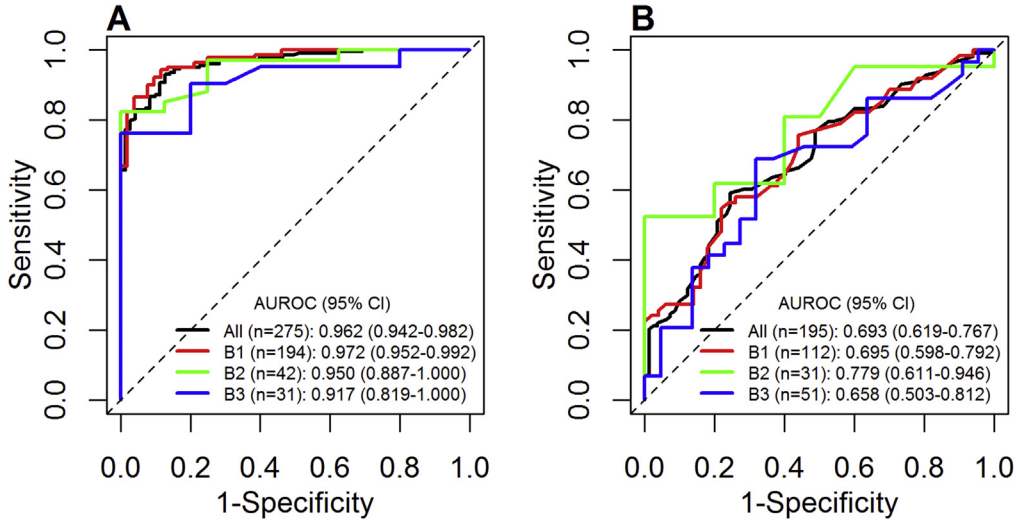
Supplementary Figure 5. Diagnostic accuracy of the EHI in endohistopathologic healing. The ROC curve of the EHI in distinguishing endohistopathologic healing vs non–endohistopathologic healing in validation cohort 2.



Supplementary Figure 4. Diagnostic accuracy of the EHI in patients with surgery. The ROC curves of the EHI in distinguishing AD vs ER in patient subcohorts with (red) or without (blue) a history of IBD-related surgery in validation cohort 2. The AUROC of the EHI in the 2 subcohorts was not significantly different (P = .801).



Supplementary Figure 6. Comparative accuracy of the EHI across disease locations. The ROC curves of the EHI in distinguishing AD vs ER by disease location in (*A*) validation cohort 1 and (*B*) validation cohort 2. A mixed logistic regression model with random intercepts for individual participants was used to combine multiple samples of the same participant in validation cohort 1. The minimum pairwise P values comparing the AUROC of the EHI for patients with different disease locations were (*A*) 0.171 and (*B*) 0.292, respectively, indicating that EHI performance was consistent across disease locations.



Supplementary Figure 7. Comparative accuracy of the EHI across disease behaviors. The ROC curves of the EHI in distinguishing AD vs ER by disease behavior in (*A*) validation cohort 1 and (*B*) validation cohort 2. A mixed logistic regression model with random intercepts for individual participants was used to combine multiple samples of same participant in validation cohort 1. The pairwise P values comparing the AUROC of the EHI on different groups of patients were (*A*) $P \geq 0.290$ and (*B*) $P \geq 0.300$, respectively.

Supplementary Table 1. The 47 Biomarkers Evaluated for the Development of the EHI and Their Corresponding Signaling Pathways, From Which 13 Biomarkers Were Selected for the EHI Assay From Logistic Regression Modeling

Angiogenesis	Cell adhesion	Growth factors	Immune modulation	Inflammation	Matrix remodeling
k = 4	k = 6	k = 10	k = 16	k = 5	k = 6
ANG1[c]	CEACAM1[e]	TGFα[e]	IL7[c]	CRP[e]	EMMPRIN[c]
ANG2[f]	VCAM-1[d]	BTC[a]	GM-CSF[b]	SAA1[e]	MMP-1[e]
VEGFα[a]	ALCAM[a]	EGF[a]	IL1β[b]	ADA[a]	MMP-2[g]
FGF2[b]	α4β7[a]	SCF[a]	IL2[b]	TWEAK[a]	MMP-3[d]
	ICAM-1[a]	AREG[b]	IL5[b]	IFN-γ[b]	MMP-9[c]
	MAdCAM[a]	ANXA13[b]	IL6[b]		Fibronectin[a]
		EREG[b]	IL10[b]		
		HB-EGF[b]	IL12/23p40[b]		
		HGF[b]	IL13[b]		
		TGFβ[b]	IL15[b]		
			IL17a[b]		
			IL17f[b]		
			IL22[b]		
			IL23[b]		
			IL31[b]		
			IL33[b]		

[a]These 11 biomarkers did not enhance the performance of EHI and were not included in the EHI assay.
[b]These 23 biomarkers were eliminated due to poor analytical reproducibility, low detection rate in serum specimens, and/or lack of correlation to disease severity in preliminary studies.
[c]$P < 0.001$.
[d].001 $\leq P <$ .01.
[e].01 $\leq P <$ .05.
[f]$P = .094$.
[g]$P = .121$.

Supplementary Table 2. Participant and Sample Characteristics of the Study Cohorts[a]

Characteristics	Training	Validation 1 (TAILORIX)	Validation 2 (UCSD)	*P* value[b] (training vs validation 1)	P value[b] (training vs validation 2)	*P* value[b] (validation 1 vs 2)
Participants						
n	278	116	195			
Age, *y*	30.0 (24.9–40.0) [2]	30.2 (22.4–45.2)	38.5 (28.0–52.0) [17]	.842	<.001	<.001
Female sex	128 (46.0)	69 (59.5)	92 (49.7) [10]	.020	.449	.123
Race/ethnicity				—	.039	—
African American	5 (1.8)	—	5 (2.6)			
Asian	14 (5.0)	—	3 (1.5)			
White	160 (57.6)	—	125 (64.1)			
Hispanic	5 (1.8)	—	8 (4.1)			
Other	5 (1.8)	—	0 (0.0)			
Unknown	89 (32.0)	—	54 (27.7)			
CD duration, *y*	4.0 (3.0–12.5) [260]	0.7 (0.1–6.9)	11.0 (5.0–19.0) [16]	.002	.018	<.001
Age at diagnosis, *y*	[100]		[12]	<.001	.017	<.001
A1: ≤16	55 (30.9)	6 (5.2)	49 (26.8)			
A2: 17–40	106 (59.6)	86 (74.1)	97 (53.0)			
A3: >40	17 (9.6)	24 (20.7)	37 (20.2)			
CD location	[100]	[5]	[5]	.738	<.001	.002
L1: Ileal	43 (24.2)	27 (24.3)	47 (24.7)			
L2: Colonic	26 (14.6)	20 (18.0)	67 (35.3)			
L3: Ileocolonic	109 (61.2)	64 (57.7)	76 (40.0)			
CD behavior	[186]	[5]	[1]	<.001	<.001	.006
B1: Nonstricturing, nonpenetrating	26 (28.3)	81 (73.0)	112 (57.7)			
B2: Stricturing	57 (62.0)	17 (15.3)	31 (16.0)			
B3: Penetrating	9 (9.8)	13 (11.7)	51 (26.3)			
Perianal disease modifier	29 (17.5)[112]	31 (27.9) [5]	21 (10.8)	.052	.069	<.001
Biologic medication use	—	0 (0)	139 (77.2) [15]	—	—	<.001
History of IBD related surgery	—	12 (10.3)	90 (46.2)	—	—	<.001
Samples						
n	335	275	195			
Endoscopic remission[c]	159 (47.5)	71 (25.8)	82 (42.1)	<.001	.241	<.001
CDEIS	2.8 (0.2–6.0) [202]	4.4 (0.8–9.1)	—	.016	—	—
SES-CD	6.0 (1.0–12.0) [133]	6.0 (2.0–12.0)	3.0 (0.0–6.5)	.321	<.001	<.001
CRP, *mg/L*	2.0 (0.7–6.5)	2.5 (0.5–7.2)	2.6 (0.7–7.1)	.586	.172	.460
Fecal calprotectin, *μg/g*	50.8 (30.1–270.3) [273]	336.0 (100.0–1197.5) [28]	55.0 (0.0–251.1) [114]	<.001	.086	<.001
EHI	32 (20–44)	38 (25–53)	32 (19.5–46.5)	.001	.752	.006

UCSD, University of California San Diego.
[a]Continuous variables are reported as median (IQR), categorical variables are reported as n (%), and numbers of missing data, if any, are listed inside brackets ([n]).
[b]Based on Mann-Whitney test for continuous variables and Fisher exact test for categorical variables.
[c]On the training cohort, SES-CD scores were first converted to CDEIS scores by CDEIS = 0.1569 + 0.6744 × SES-CD (see Supplementary Figure 1). *ER* was then defined as either original or converted CDEIS score of <3. On validation cohorts, *ER* was defined as a total SES-CD of ≤2 and ≤1 in each segment.

Supplementary Table 3. Participant and Sample Characteristics of the Training Cohort[a]

Characteristics	Total	Univ Padua, Italy	MSH, Toronto	STORI	UCSD
Participants					
n	278	18	146	83	31
Age, *y*	30.0 (24.9–40.0) [2]	34.5 (26.5–51.5)	29.0 (23.0–39.2) [2]	31.6 (25.6–39.2)	30.0 (23.5–45.0)
Female sex	128 (46.0)	5 (27.8)	63 (43.2)	45 (54.2)	15 (48.4)
Race/ethnicity					
African American	5 (1.8)	0 (0.0)	5 (3.4)	—	0 (0.0)
Asian	14 (5.0)	0 (0.0)	13 (8.9)	—	1 (3.2)
White	160 (57.6)	18 (100.0)	115 (78.8)	—	27 (87.1)
Hispanic	5 (1.8)	0 (0.0)	3 (2.1)	—	2 (6.5)
Other	5 (1.8)	0 (0.0)	4 (2.7)	—	1 (3.2)
Unknown	89 (28.7)	0 (0.0)	6 (4.1)	—	0 (0.0)
Disease duration, *y*	4.0 (3.0–12.5) [260]	4.0 (3.0–12.5)	—	—	—
Age at diagnosis, *y*	[100]		[17]	[83]	
A1: ≤16	55 (30.9)	0 (0.0)	47 (36.4)	—	8 (25.8)
A2: 17–40	106 (59.6)	13 (72.2)	74 (57.4)	—	19 (61.3)
A3: >40	17 (9.6)	5 (27.8)	8 (6.2)	—	4 (12.9)
CD location	[100]		[17]	[83]	
L1: Ileal	43 (24.2)	9 (50.0)	27 (20.9)	—	7 (22.6)
L2: Colonic	26 (14.6)	3 (16.7)	11 (8.5)	—	12 (38.7)
L3: Ileocolonic	109 (61.2)	6 (33.3)	91 (70.5)	—	12 (38.7)
CD behavior	[186]		[102]	[83]	[1]
B1: Nonstricturing, nonpenetrating	26 (28.3)	7 (38.9)	0 (0.0)	—	19 (63.3)
B2: Stricturing	57 (62.0)	6 (33.3)	44 (100.0)	—	7 (23.3)
B3: Penetrating	9 (9.8)	5 (27.8)	0 (0.0)	—	4 (13.3)
Perianal disease modifier	29 (17.5) [112]	2 (11.1)	25 (17.1)	—	2 (100) [29]
Samples					
n	335	50	157	83	45
ER[b]	159 (47.5)	5 (10.0)	67 (42.7)	63 (75.9)	24 (53.3)
CDEIS	2.8 (0.2–6.0) [202]	6.6 (4.0–17.6)	—	0.7 (0.0–2.8)	—
SES-CD	6.0 (1.0–12.0) [133]	—	6.0 (2.0–12.0)	—	4.0 (0.0–9.0)
CDAI	59.5 (24.2–123.8) [205]	147.0 (120.0–240.0) [3]	—	36.5 (17.2–61.1)	—
CRP, *mg/L*	2.0 (0.7–6.5)	1.3 (0.6–4.4)	3.6 (1.0–6.5)	1.4 (0.6–2.7)	2.0 (0.6–3.9)
Fecal calprotectin, *μg/g*	50.8 (30.1–270.3) [273]	—	—	50.8 (30.1–270.3) [21]	—
EHI	32 (20–44)	42 (33–59)	33 (22–47)	21 (13–29)	39 (30–50)

UCSD, University of California San Diego; Univ, university.

[a]Continuous variables are reported as median (IQR), categorical variables are reported as n (%), and numbers of missing data, if any, are listed inside brackets ([n]).

[b]SES-CD scores were converted to CDEIS scores by: CDEIS = 0.1569 + 0.6744 × SES-CD (see Supplementary Figure 1). Disease activity status was defined by either original or converted CDEIS scores as remission (CDEIS < 3) or active (CDEIS ≥ 3).

Supplementary Table 4. Participant and Sample Characteristics of Validation Cohort 1[a]

Characteristics	Total	Baseline	Week 12	Week 54
Participants[b]				
n	116	102	98	75
Age (years)	30.2 (22.4–45.2)	30.9 (24.1–45.6)	30.9 (22.9–45.7)	30.4 (23.1–44.5)
Female sex	69 (59.5)	61 (59.8)	58 (59.2)	43 (57.3)
CD duration, *y*	0.7 (0.1–6.9)	0.9 (0.1–7.3)	0.5 (0.1–6.1)	0.5 (0.0–5.6)
Age at diagnosis, *y*				
A1: ≤16	6 (5.2)	5 (4.9)	3 (3.1)	3 (4.0)
A2: 17–40	86 (74.1)	77 (75.5)	72 (73.5)	55 (73.3)
A3: >40	24 (20.7)	20 (19.6)	23 (23.5)	17 (22.7)
CD location	[5]	[3]	[4]	[1]
L1: Ileal	27 (24.3)	26 (26.3)	23 (24.5)	18 (24.3)
L2: Colonic	20 (18.0)	16 (16.2)	17 (18.1)	16 (21.6)
L3: Ileocolonic	64 (57.7)	57 (57.6)	54 (57.4)	40 (54.1)
CD behavior	[5]	[3]	[4]	[1]
B1: Nonstricturing, nonpenetrating	81 (73.0)	73 (73.7)	69 (73.4)	52 (70.3)
B2: Stricturing	17 (15.3)	16 (16.2)	13 (13.8)	13 (17.6)
B3: Penetrating	13 (11.7)	10 (10.1)	12 (12.8)	9 (12.2)
Perianal disease modifier	31 (27.9) [5]	26 (26.3) [3]	25 (26.6) [4]	21 (28.4) [1]
History of IBD related surgery	12 (10.3)	12 (11.8)	8 (8.2)	4 (5.3)
Samples				
n	275	102	98	75
Endoscopic remission[c]	71 (25.8)	0 (0.0)	26 (26.5)	45 (60.0)
CDEIS	4.4 (0.8-9.1)	10.0 (7.4-16.0)	3.0 (0.7-5.2)	0.1 (0.0-2.6)
SES-CD	6.0 (2.0–12.0)	15.0 (9.0–22.0)	4.0 (2.0–8.0)	1.0 (0.0–3.0)
CDAI	166.0 (72.5–261.0) [4]	279.5 (233.0–321.8)	112.0 (58.5–181.5) [3]	66.5 (39.8–115.2) [1]
CRP, *mg/L*	2.5 (0.5–7.2)	8.2 (3.5–14.3)	0.9 (0.3–3.5)	0.8 (0.3–2.5)
Fecal calprotectin, *μg/g*	336.0 (100.0–1197.5) [28]	1462.5 (709.4–1800.0) [6]	122.0 (100.0–430.5) [17]	105.2 (100.0–215.8) [5]
EHI	38.0 (25.0–53.0)	55.5 (42.0–72.8)	33.5 (23.0–41.8)	25.0 (19.0–37.5)

[a]Continuous variables are reported as median (IQR), categorical variables are reported as n (%), and numbers of missing data, if any, are listed inside brackets ([n]).
[b]Participants in the 3 time points were subsets of the full cohort that contributed the corresponding samples.
[c]*ER* was defined as a total SES-CD of ≤2 and ≤1 in each segment.

Supplementary Table 5. Participant and Sample Characteristics of Validation Cohort 2 and Subcohorts With or Without Fecal Calprotectin[a]

Characteristics	Full Cohort	Subcohort (with calprotectin)	Subcohort (without calprotectin)	*P* value[b]
n	195	81	114	
Age, *y*	38.5 (28.0–52.0) [17]	37.0 (28.0–47.0) [9]	39.0 (28.0–53.8) [8]	.603
Female sex	92 (49.7) [10]	36 (48.0) [6]	56 (50.9) [4]	.765
Race/ethnicity				<.001
African American	5 (2.6)	1 (1.2)	4 (3.5)	
Asian	3 (1.5)	2 (2.5)	1 (0.9)	
White	125 (64.1)	33 (40.7)	92 (80.7)	
Hispanic	8 (4.1)	3 (3.7)	5 (4.4)	
Unknown	54 (27.7)	42 (51.9)	12 (10.5)	
CD duration, *y*	11.0 (5.0–19.0) [16]	11.0 (5.0–20.0) [8]	11.0 (5.0–18.0) [8]	.684
Age at diagnosis, *y*	[12]	[8]	[4]	.530
A1: ≤16	49 (26.8)	22 (30.1)	27 (24.5)	
A2: 17–40	97 (53.0)	39 (53.4)	58 (52.7)	
A3: >40	37 (20.2)	12 (16.4)	25 (22.7)	
CD location	[5]	[3]	[2]	.566
L1: Ileal	47 (24.7)	22 (28.2)	25 (22.3)	
L2: Colonic	67 (35.3)	28 (35.9)	39 (34.8)	
L3: Ileocolonic	76 (40.0)	28 (35.9)	48 (42.9)	
CD behavior	[1]		[1]	.444
B1: Nonstricturing, nonpenetrating	112 (57.7)	51 (63.0)	61 (54.0)	
B2: Stricturing	31 (16.0)	12 (14.8)	19 (16.8)	
B3: Penetrating	51 (26.3)	18 (22.2)	33 (29.2)	
Perianal disease modifier	21 (10.8)	11 (13.6)	10 (8.8)	.350
Biologic medication use	139 (77.2) [15]	59 (76.6) [4]	80 (77.7) [11]	1.000
History of IBD-related surgery	90 (46.2)	33 (40.7)	57 (50.0)	.244
ER[c]	82 (42.1)	33 (40.7)	49 (43.0)	.771
Endohistopathologic healing[d]	23 (29.1) [116]	11 (28.9) [43]	12 (29.3) [73]	1.000
SES-CD	3.0 (0.0–6.5)	3.0 (0.0–7.0)	3.0 (0.0–6.0)	.609
CDAI PRO2	7.7 (2.9–15.5) [39]	6.9 (2.4–15.4) [10]	8.3 (3.7–15.4) [29]	.457
GHAS	3.0 (1.0–6.0) [116]	3.0 (1.0–6.0) [43]	4.0 (1.0–6.0) [73]	.886
CRP, *mg/L*	2.6 (0.7–7.1)	3.2 (0.7–7.5)	2.4 (0.8–6.2)	.476
Fecal calprotectin, *μg/g*	55.0 (0.0–251.1) [114]	55.0 (0.0–251.1)	—	—
EHI	32 (19.5–46.5)	32 (21–47)	32.0 (19.0–46.0)	0.576

GHAS, Global Histologic Disease Activity; PRO2, Patient-Reported Outcomes 2.
[a]Continuous variables are reported as median (IQR), categorical variables are reported as n (%), and numbers of missing data, if any, are listed inside brackets ([n]).
[b]Based on Mann-Whitney test for continuous variables and Fisher exact test for categorical variables.
[c]*ER* was defined as a total SES-CD of ≤2 and ≤1 in each segment.
[d]*Endohistopathologic healing* was defined as achieving both ER and histologic remission (GHAS ≤ 2).

Supplementary Table 6. TPs, TNs, Sensitivity, Specificity, PLR, and NLR of the EHI in Distinguishing AD (n = 176) vs ER (n = 159) vs in the training cohort

EHI Threshold	TPs, n	TNs, n	Sensitivity, % (95% CI)	Specificity, % (95% CI)	PLR (95% CI)	NLR (95% CI)
20	159	62	90.3 (85.0–94.3)	39.0 (31.4–47.0)	1.48 (1.30–1.69)	0.25 (0.15–0.41)
30	129	100	73.3 (66.1–79.7)	62.9 (54.9–70.4)	1.98 (1.58–2.46)	0.43 (0.32–0.56)
40	86	131	48.9 (41.3–56.5)	82.4 (75.6–88.0)	2.78 (1.92–4.01)	0.62 (0.53–0.73)
50	54	151	30.7 (24.0–38.1)	95.0 (90.3–97.8)	6.10 (3.00–12.41)	0.73 (0.66–0.81)

NLR, negative likelihood ratio; PLR, positive likelihood ratio; TN, true negative; TP, true positive.

Supplementary Table 7. TPs, TNs, Sensitivity, Specificity, PLR, and NLR of the EHI in Distinguishing AD vs ER by Disease Location in Validation Cohort 1

CD Location	EHI Threshold	MLG Probability[a]	TPs, n	TNs, n	Sensitivity,% (95% CI)	Specificity, % (95% CI)	PLR (95% CI)	NLR (95% CI)
All	20	0.550	198	49	97.1 (93.7–98.9)	69.0 (56.9–79.5)	3.13 (2.21–4.44)	0.04 (0.02–0.10)
ER (n = 71)	30	0.746	173	65	84.8 (79.1–89.4)	91.5 (82.5–96.8)	10.04 (4.66–21.63)	0.17 (0.12–0.23)
AD (n = 204)	40	0.876	118	71	57.8 (50.7–64.7)	100.0 (94.9–100.0)	Infinity	0.42 (0.36–0.50)
	50	0.945	76	71	37.3 (30.6–44.3)	100.0 (94.9–100.0)	Infinity	0.63 (0.56–0.70)
L1	20	0.550	51	12	98.1 (89.7–100.0)	80.0 (51.9–95.7)	4.90 (1.78–13.50)	0.02 (0.00–0.17)
ER (n = 15)	30	0.746	48	14	92.3 (81.5–97.9)	93.3 (68.1–99.8)	13.85 (2.08–92.12)	0.08 (0.03–0.21)
AD (n = 52)	40	0.876	29	15	55.8 (41.3–69.5)	100.0 (78.2–100.0)	Infinity	0.44 (0.33–0.60)
	50	0.945	17	15	32.7 (20.3–47.1)	100.0 (78.2–100.0)	Infinity	0.67 (0.56–0.81)
L2	20	0.550	30	12	100.0 (88.4–100.0)	63.2 (38.4–83.7)	2.71 (1.51–4.89)	0.00 (0.00–)
ER (n = 19)	30	0.746	23	18	76.7 (57.7–90.1)	94.7 (74.0–99.9)	14.57 (2.14–99.15)	0.25 (0.13–0.48)
AD (n = 30)	40	0.876	17	19	56.7 (37.4–74.5)	100.0 (82.4–100.0)	Infinity	0.43 (0.29–0.65)
	50	0.945	13	19	43.3 (25.5–62.6)	100.0 (82.4–100.0)	Infinity	0.57 (0.41–0.78)
L3	20	0.550	110	25	95.7 (90.1–98.6)	69.4 (51.9–83.7)	3.13 (1.91–5.13)	0.06 (0.03–0.15)
ER (n = 36)	30	0.746	96	32	83.5 (75.4–89.7)	88.9 (73.9–96.9)	7.51 (2.97–18.99)	0.19 (0.12–0.29)
AD (n = 115)	40	0.876	67	36	58.3 (48.7–67.4)	100.0 (90.3–100.0)	Infinity	0.42 (0.34–0.52)
	50	0.945	43	36	37.4 (28.5–46.9)	100.0 (90.3–100.0)	Infinity	0.63 (0.54–0.72)

MLG, mixed logistic regression; NLR, negative likelihood ratio; PLR, positive likelihood ratio; TN, true negative; TP, true positive.
[a]The population-averaged probability from the MLG model.

Supplementary Table 8. TPs, TNs, Sensitivity, Specificity, PLR, and NLR of the EHI in Distinguishing AD vs ER by Disease Location in Validation Cohort 2

CD Location	EHI Threshold	TPs, n	TNs, n	Sensitivity, % (95% CI)	Specificity, % (95% CI)	PLR (95% CI)	NLR (95% CI)
All	20	94	30	83.2 (75.0–89.6)	36.6 (26.2–48.0)	1.31 (1.09–1.58)	0.46 (0.28–0.76)
ER (n = 82)	30	74	49	65.5 (56.0–74.2)	59.8 (48.3–70.4)	1.63 (1.21–2.19)	0.58 (0.42–0.79)
AD (n = 113)	40	54	65	47.8 (38.3–57.4)	79.3 (68.9–87.4)	2.31 (1.45–3.67)	0.66 (0.54–0.81)
	50	34	72	30.1 (21.8–39.4)	87.8 (78.7–94.0)	2.47 (1.29–4.70)	0.80 (0.69–0.92)
L1	20	22	8	84.6 (65.1–95.6)	38.1 (18.1–61.6)	1.37 (0.94–1.99)	0.40 (0.14–1.16)
ER (n = 21)	30	14	17	53.8 (33.4–73.4)	81.0 (58.1–94.6)	2.83 (1.09–7.32)	0.57 (0.36–0.91)
AD (n = 26)	40	10	21	38.5 (20.2–59.4)	100.0 (83.9–100.0)	Infinity	0.62 (0.45–0.83)
	50	8	21	30.8 (14.3–51.8)	100.0 (83.9–100.0)	Infinity	0.69 (0.54–0.90)
L2	20	30	9	78.9 (62.7–90.4)	31.0 (15.3–50.8)	1.15 (0.85–1.54)	0.68 (0.30–1.54)
ER (n = 29)	30	26	15	68.4 (51.3–82.5)	51.7 (32.5–70.6)	1.42 (0.92–2.19)	0.61 (0.34–1.10)
AD (n = 38)	40	23	21	60.5 (43.4–76.0)	72.4 (52.8–87.3)	2.19 (1.15–4.17)	0.55 (0.35–0.86)
	50	16	23	42.1 (26.3–59.2)	79.3 (60.3–92.0)	2.04 (0.91–4.55)	0.73 (0.53–1.01)
L3	20	40	11	85.1 (71.7–93.8)	37.9 (20.7–57.7)	1.37 (1.01–1.87)	0.39 (0.17–0.90)
ER (n = 29)	30	32	15	68.1 (52.9–80.9)	51.7 (32.5–70.6)	1.41 (0.92–2.16)	0.62 (0.36–1.07)
AD (n = 47)	40	19	20	40.4 (26.4–55.7)	69.0 (49.2–84.7)	1.30 (0.68–2.48)	0.87 (0.62–1.21)
	50	8	25	17.0 (7.6–30.8)	86.2 (68.3–96.1)	1.23 (0.41–3.74)	0.96 (0.79–1.17)

MLG, mixed logistic regression; NLR, negative likelihood ratio; PLR, positive likelihood ratio; TN, true negative; TP, true positive.