AGA CLINICAL PRACTICE GUIDELINE ON THE ROLE OF BIOMARKERS FOR THE MANAGEMENT OF CROHN’S DISEASE

Ashwin N. Ananthakrishnan¹*, Jeremy Adler²,³*, Karen A. Chachu⁴*, Nghia H. Nguyen⁵, Shazia M. Siddique⁶,⁷, Jennifer M. Weiss⁸, Shahnaz Sultan⁹, Fernando S. Velayos¹⁰, Benjamin L. Cohen¹¹, Siddharth Singh¹²,¹³, on behalf of the AGA Clinical Guidelines Committee

¹Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts; ²Division of Pediatric Gastroenterology, C.S. Mott Children's Hospital, Michigan Medicine, University of Michigan, Ann Arbor, Michigan; ³Susan B. Meister Child Health Evaluation and Research Center, University of Michigan, Ann Arbor, Michigan; ⁴Division of Gastroenterology, Department of Medicine, Duke University, Durham, North Carolina; ⁵Division of Gastroenterology, Kaiser Permanente Medical Group, Riverside, California; ⁶Division of Gastroenterology, Department of Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania; ⁷Center for Evidence-Based Practice, University of Pennsylvania Health System, Philadelphia, Pennsylvania; ⁸Division of Gastroenterology and Hepatology, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin; ⁹Division of Gastroenterology, Hepatology, and Nutrition, University of Minnesota, Minneapolis, Minnesota and Veterans Affairs Healthcare System, Minneapolis, Minnesota; ¹⁰Division of Gastroenterology, Kaiser Permanente Medical Group, San Francisco, California; ¹¹Division of Gastroenterology, Hepatology, and Nutrition, Digestive Disease Institute, Cleveland Clinic, Cleveland, Ohio; ¹²Division of Gastroenterology and Hepatology and ¹³Division of Biomedical Informatics, Department of Medicine, University of California San Diego, La Jolla, California;

*Denotes co-first authors

Address for Correspondence:
Chair, Clinical Guidelines Committee,
American Gastroenterological Association
National Office, 4930 Del Ray Avenue
Bethesda, Maryland 20814
E-mail: clinicalpractice@gastro.org
Telephone: (301) 941-2618

Word count:
Abstract – 339
Manuscript – 11,195
References – 48
Tables and Figures – 7+0
eTables and eFigures (online supplement) – 2+8

Funding: These guidelines were fully funded by the AGA Institute. Dr. Ananthakrishnan is supported by NIDDK R21DK127227 and R01 DK127171, in addition to grants from the Leona M and Harry B Helmsley Charitable Trust and the Chleck Family Foundation. Dr. Siddique is supported by NIDDK K08DK120902. Dr. Singh is supported by NIDDK K23DK117058 and R03DK129631.

Conflicts of Interest:
- Ashwin N Ananthakrishnan – None
- Jeremy Adler – Research grants from Janssen Research & Development, LLC.
- Karen A. Chachu – None
- Nghia H. Nguyen – None
- Shazia M. Siddique – None
- Jennifer Weiss – None
- Shahnaz Sultan – None
- Fernando S. Velayos – None
- Benjamin L. Cohen receives the following financial support: Consulting fees from Abbvie, Celgene-Bristol Myers Squibb, Lilly, Pfizer, Sublimity Therapeutics, Takeda, TARGET RWE; CME Companies: Cornerstones, Vindico; Speaking: Abbvie; Educational Grant: Pfizer
• Siddharth Singh’s Institution has received research grants from Pfizer and AbbVie, and he has received personal fees from Pfizer (for ad hoc grant review).

A full list of conflicts active at the time of guideline development can be accessed at AGA’s National Office in Bethesda, MD.

Acknowledgement: We would like to thank Ms. Caitlin Bakker, University of Minnesota Libraries for designing and conducting the literature search.
ABSTRACT

Background & Aims: Biomarkers are frequently used for evaluation and monitoring of patients with Crohn’s disease (CD). This American Gastroenterological Association (AGA) guideline is intended to support practitioners in decisions about the use of biomarkers for the management of CD.

Methods: A multi-disciplinary panel of content experts and guideline methodologists used the Grading of Recommendations Assessment, Development and Evaluation (GRADE) framework to formulate patient-centered clinical questions and review evidence on the performance of fecal calprotectin, serum C-reactive protein (CRP), and endoscopic healing index (EHI) in patients with established CD who were asymptomatic, had symptoms of varying severity or were in surgically induced remission. Biomarker performance was assessed against the gold standard of endoscopic activity defined as SES-CD ≥ 3. The panel used the GRADE Evidence-to-Decision framework to develop recommendations for use of biomarkers in various settings. Implementation considerations were formulated for each recommendation to inform clinical practice.

Results: The guideline panel made 11 conditional recommendations. In patients with CD in symptomatic remission, the panel suggests use of a biomarker- and symptom-based monitoring strategy over symptoms alone. In patients in symptomatic remission, a fecal calprotectin < 150µg/g and normal CRP rules out active inflammation, avoiding endoscopic evaluation for assessment of disease activity. However, elevated biomarkers in this setting merit confirmation with endoscopy prior to treatment adjustment. In patients with CD with mild symptoms, neither normal nor elevated biomarkers alone are sufficiently accurate to determine endoscopic activity. In patients with CD with moderate-to-severe symptoms, elevated fecal calprotectin or serum CRP suggests endoscopic activity, precluding routine endoscopic assessment for disease activity. In patients with CD in surgically induced remission, in low-risk patients on pharmacologic prophylaxis, a normal fecal calprotectin reliably rules out endoscopic recurrence. In other post-operative settings, the panel suggests endoscopic assessment for establishing post-operative recurrence.

Conclusions: In patients with CD, fecal calprotectin and serum CRP can inform disease management in both asymptomatic and symptomatic disease. Discordance between
symptom-assessment and biomarker value may merit endoscopic evaluation for confirmation of status of disease activity.

**Key words:** Inflammatory bowel disease; monitoring; endoscopic remission; treat to target; evidence synthesis
INTRODUCTION

Inflammatory Bowel Diseases, comprising Crohn’s disease (CD) and ulcerative colitis (UC) are rising in incidence and prevalence worldwide\(^1\),\(^2\). They often have an onset in young adulthood and are characterized by a protracted relapsing-remitting course with progressive permanent bowel damage\(^3\). The therapeutic armamentarium for CD has expanded over the past decade with multiple new mechanisms of action. Despite such progress, nearly half the individuals with CD require at least one surgery over the course of their illness\(^4\),\(^5\). To durably modify the natural history of CD, an important concept that has emerged is the need for early institution of effective treatment and assessment of attainment of therapeutic target to improve long-term outcomes and prevent disease related disability\(^6\)-\(^9\). Cross-sectionally, symptoms in CD have correlated poorly with endoscopic disease activity making symptom-based disease activity assessment suboptimal\(^10\)-\(^12\). Thus, objective assessment of inflammation has traditionally relied on endoscopic assessment of disease activity and demonstration of mucosal healing. In a population-based cohort from Norway, patients with CD who achieved endoscopic healing at 1 year had superior long-term outcomes up to five years after diagnosis with a reduced need for corticosteroids and fewer CD-related hospitalizations\(^13\). However, reliance solely on endoscopy for repeated assessment of disease activity is limited by cost and resource utilization, invasiveness, and reduced patient acceptability. In the CALM trial comparing a symptom-based therapeutic strategy against a biomarker-based strategy, the use of frequent biomarker measurement to guide therapy escalation was associated with improved patient outcomes over two years\(^14\). The performance of serum and fecal biomarkers of disease activity as well as robust determination of thresholds that can function as surrogates of endoscopic activity assessment have not been comprehensively examined, leading to significant variability in clinical practice in optimal use of these biomarkers.

Objective

The objective of this guideline is to inform the role of commonly used serum and fecal biomarkers as surrogates for endoscopic disease activity for both cross-sectional assessment as well as longitudinal monitoring of patients with an established diagnosis.
of CD. The scope of this guideline was restricted to biomarkers that are widely available commercially within the United States. This guideline also separately examined the predictive value of biomarkers for assessment of post-operative recurrence in CD. The panel did not examine role of biomarkers in the diagnostic pathway for patients with suspected CD. The role of biomarkers in UC have also been examined in a recent guideline\textsuperscript{15}.

**Target Audience**

The target audience for these guidelines are gastroenterology health care professionals, primary care and emergency/urgent care providers, patients, and policy makers. Recommendations are provided for common clinical scenarios in typical patients with CD. However, individual patients may have unique circumstances which must be accounted for in implementing these guidelines. Each recommendation in this guideline is accompanied by key implementation considerations and qualifying remarks that should be considered an integral part of the recommendation statement and should not be omitted. Discussions about benefits and harms are important in shared decision-making particularly for conditional recommendations where patient values and specific tradeoffs are important to consider.

**METHODS**

**Overview**

This document represents the official recommendations of the American Gastroenterological Association (AGA) and was developed using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) framework for diagnostic tests and strategies and adheres to best practices in guideline development, as outlined by the National Academy of Medicine (formerly Institute of Medicine)\textsuperscript{16}. The development of this guideline was fully funded by the AGA Institute.

**Guideline Panel Composition and Conflict of Interest**

Members of the guideline panel and evidence synthesis panel were selected on the basis of their clinical and methodological expertise after a thorough vetting process. Panel
members disclosed all potential conflicts of interest. Conflicts were managed according to AGA policies, the National Academy of Medicine, and Guidelines International Network standards. Guideline chair (Chachu) and co-chair/senior methodologist (Singh) had no direct conflicts of interest. No panel members were excluded due to a disqualifying conflict. The evidence synthesis panel consisted of 2 content experts with expertise in CD (Ananthakrishnan, Adler), a senior guideline methodologist with expertise in evidence synthesis and GRADE (Singh) and two junior guideline methodologists (Nguyen, Siddique). The guideline panel was multidisciplinary, consisting of a general gastroenterologist (Weiss), gastroenterologists with expertise in CD (Chachu, Cohen, Velayos), and guideline methodologists (Singh, Sultan, Nguyen, Siddique). The input of a patient representative on the role of biomarkers in the management of IBD was also considered in framing recommendations. A full list of conflicts can be accessed at AGA’s National Office in Bethesda, MD.

**Scope**

The guideline panel defined biomarkers as molecules that are quantifiable in tissue, blood, stool, or urine that represent an underlying biological disease process\(^{17}\). Biomarkers have been studied in CD in various clinical context including assessing likelihood of a diagnosis of CD in patients with suggestive symptoms, predicting clinical course of CD including need for surgery, development of stricturing or penetrating disease, and quantifying disease activity, etc. The panel focused on biomarkers that are widely used for assessing disease activity and making treatment decisions, measurable in easily accessible tissue or body fluid compartments, and commercially available in the United States. We examined the performance of individual biomarkers in unselected cohorts of patients with CD as well as for initial assessment of post-operative recurrence of CD following surgically-induced remission. The panel examined the cross-sectional performance of each biomarker against endoscopic assessment of disease activity as the gold standard. Biomarkers with demonstrated utility only in research studies but not available for widespread commercial use were outside the scope of this guideline. We also did not examine biomarkers that have been developed solely for prediction of likelihood of disease progression such as development of stricturing or penetrating disease. However, the panel examined performance of biomarkers that may predict future
disease activity in the context of longitudinal monitoring. Based on these criteria, we focused on serum C-reactive protein (CRP), fecal calprotectin, and the endoscopic healing index (EHI, Monitr®).

**Formulation of Clinical Questions**

Through an iterative process, the guideline and evidence synthesis panels developed focused clinical questions deemed relevant for clinical practice to be addressed in the guideline (Table 1). These related to diagnostic accuracy and utility of commonly used serum or stool biomarkers that are commercially available. Each guideline statement was derived from a focused clinical question that comprised of a well-defined statement using the PICO format (patients, intervention, comparator and outcome). These statements were used to formulate the study inclusion and exclusion criteria for review, guided the literature search and informed the final guideline recommendation. The AGA Governing Board approved the final set of questions in October 2021.

**Search Strategy**

An experienced medical librarian conducted a comprehensive search of the following databases (Ovid Medline In-Process & Other Non-Indexed Citations, Ovid MEDLINE, EMBASE, and Wiley Cochrane Library) from inception to November 21, 2021, using a combination of controlled vocabulary terms supplemented with keywords (eTable 1); an updated search of Ovid MEDLINE was performed on September 1, 2022. The search was limited to English language and humans. The bibliography of prior guidelines and the included references were searched to identify relevant studies that may have been missed. In addition, content experts helped identify any additional or ongoing studies.

**Study Selection, Data Abstraction and Statistical Analysis**

We included randomized controlled trials (RCTs) or observational studies of diagnostic accuracy meeting the following inclusion criteria: (a) performed in patients with CD, (b) provided adequate description of biomarker (CRP, fecal calprotectin, EHI) with cut-off corresponding to detection of endoscopically active CD (generally corresponding to SES-CD score > 3 [mild to severe inflammation]), (c) with ileocolonoscopy as gold standard (or MR enterography, video capsule endoscopy or balloon-assisted enteroscopy for
patients with small bowel CD not adequately examined on colonoscopy), and (d) provided sufficient data to allow estimation of diagnostic accuracy of biomarker for detection of endoscopic activity. For wider applicability and generalizability, we preferentially chose cut-offs most commonly used in clinical practice. These cut-offs were: C-reactive protein (5±5mg/L or 0.5±0.5mg/dl), fecal calprotectin (250±50µg/g, 150±50µg/g and 50±50µg/g) and EHI (> 50 or < 20).

We abstracted data on patient population and phenotype, biomarker, reference standard outcome and test performance for each eligible study. Paired values of sensitivity and specificity were pooled using a bivariate regression random-effects model proposed by Reitsma et al using STATA 14.0 software (College Station, TX)\textsuperscript{18}. Statistical assessment of heterogeneity was performed using the inconsistency index ($I^2$), which estimates what proportion of total variation across studies was due to heterogeneity rather than chance\textsuperscript{19}.

**Outcomes of Interest and Illustrative Clinical Scenarios**

For PICO$s$ focusing on biomarker cut-offs to either detect or rule out mild to severe endoscopic activity, the preferred outcome was direct consequences on patient-important outcomes (i.e., implications of true positive [TP], false positive [FP], true negative [TN], false negative [FN] results for patients – see below). As none of the studies assessed these outcomes directly, we used TP, FP, TN and FN rates as surrogate outcomes and inferred downstream consequences on patient-important outcomes. We opted to focus on detection of mild to severe endoscopic activity (SES-CD > 3), rather than only detection of moderate to severe endoscopic activity since most studies reported the performance of biomarkers at this cut-off and treatment adjustments in CD are recommended in response to presence of any ulcers, rather than focusing only on patients with SES-CD >6\textsuperscript{20-22}.

For questions focusing on *ruling out endoscopically active CD*, our outcome was minimizing rates of FN (i.e., patients incorrectly being labeled as being in remission, when they actually have active endoscopic inflammation) to a level <5% in general with reasonable rates of TP, FP and TN (**Figure 1**). For questions focusing on *detecting endoscopic activity*, our outcome was minimizing rates of FP (i.e., patients incorrectly
labeled as having active endoscopic inflammation, when their disease is actually in remission) (eFigure 2). The threshold of 5% FN and FP rate is similar to that used in the UC guideline and was consistent with patient preference for choosing stool-based biomarkers over endoscopic assessment for monitoring inflammation.

Overall TP, FP, TN and FN rates are dependent on pre-test probability. We derived illustrative prevalence of any endoscopic activity (SES-CD > 3) based upon combination of abdominal pain (AP) and stool frequency score (SFS), two of the most commonly used patient-reported outcomes, which were used to calculate PRO2 disease activity scores. Prevalence of any endoscopic activity (SES-CD ≥ 3), and of endoscopic remission (SES CD < 3), for different combinations of cut-offs of PRO2 (based on AP and SFS), at varying time points after treatment initiation/adjustment were derived from existing literature based on individual participant data from phase 2 and 3 clinical trial programs of biologic agents in patients with moderate to severely active CD (unpublished data), as well as referral center observational cohorts with prospective assessment of clinical disease scores and endoscopic activity.

For our analysis, we used four illustrative scenarios, two pertaining to patients with CD in symptomatic remission and two for those with active symptoms:

- **Low pre-test probability** of having endoscopically active inflammation comprised of asymptomatic patients with CD (PRO2 < 8, ≤ 3 very loose/watery stools per day AND absent or mild abdominal pain; or PRO3 < 13). This population was further subdivided into two subgroups based on whether the patient had recently (within 3 years) undergone endoscopic assessment of disease activity. Patients in whom endoscopic activity status is unknown (i.e., assessment of endoscopic activity was > 3 years ago), we estimated a higher prevalence of endoscopically active inflammation in asymptomatic patients of 45% based on data from clinical trials and observational cohorts. In contrast, in patients with recent confirmation of endoscopic remission without subsequent change in clinical status and therapy were estimated to have a probability of active endoscopic inflammation of 20%, given low likelihood of endoscopic progression in the absence of any change in clinical status and therapy.
- **Intermediate pre-test probability** of having endoscopically active inflammation included patients with mild symptoms of CD (PRO2 8-13, with 3-5 loose/watery stools per day AND mild abdominal pain, or PRO3 score 13-21). The estimated prevalence of mild to severe endoscopic inflammation in these patients was ~65%.
- **High pre-test probability** of having endoscopically active inflammation. These include patients with moderate to severe symptoms of active CD (PRO2 score > 13, with > 5 loose/watery stools per day AND/OR moderate-severe abdominal pain, or PRO3 score > 21). The estimated prevalence of mild to severe endoscopic inflammation in these patients was ~80%.

For the detection of post-operative endoscopic recurrence, we examined the performance of each biomarker in detecting significant endoscopic recurrence defined as Rutgeerts' endoscopic score > i2^25, 26. The panel assumed that initial assessment of endoscopic activity in asymptomatic patients with CD following surgically induced remission would be 6-12 months after the resection. We used three illustrative clinical scenarios to determine likelihood of endoscopic recurrence at this assessment based on individual patient factors influencing risk of post-operative recurrence at time of surgery, and use of post-operative pharmacologic prophylaxis. The risk factors include early age of CD diagnosis, smoking, long-segment disease, prior bowel resection and penetrating disease behavior. Risk of post-operative recurrence is further modified by use of post-operative pharmacologic prophylaxis. Typically, use of post-operative prophylaxis with immunosuppressive therapies lowers the risk of post-operative recurrence by 50-70%.

- **Low risk of post-operative endoscopic recurrence** comprised of patients without any risk factors associated with a greater likelihood of post-operative recurrence who were on post-operative prophylactic therapy. The estimated likelihood of endoscopic recurrence in this population was ~10%.
- **Intermediate risk of post-operative endoscopic recurrence** comprised of patients who had one or more risk factors for post-operative recurrence but were on post-operative prophylactic therapy associated with reducing risk of recurrence. The estimated likelihood of endoscopic recurrence in this population was ~30%.
- **High risk of post-operative endoscopic recurrence** comprised of patients who had one or more risk factors associated with high likelihood of recurrence who were
not on post-operative prophylactic therapy. The estimated likelihood of endoscopic recurrence in this population was ~60%.

**Consequences of diagnostic test results on patient-important outcomes**
The panel considered downstream consequences in important patient outcomes corresponding to each possible outcome of a diagnostic test (TP, FP, TN, FN) (Table 2). Healthcare providers should be aware of test performance at an individual patient level in each of these scenarios and balance the FN and FP rate with the downstream change in treatment plan that would result from each of these scenarios.

**Certainty of the Evidence**
We rated the certainty of evidence using the GRADE approach for diagnostic tests and strategies. In this approach, all evidence from RCTs (comparing different diagnostic tests or cut-offs of same test) and observational diagnostic accuracy studies start at high-quality, but can be rated down for any of the following factors:

- Risk of bias in included studies (inferred based on QUADAS-2 instrument).
- Indirectness (deemed present if there are important differences between the populations studied and those for whom the recommendation is intended). In this updated GRADE approach for diagnostic accuracy studies, TP, FP, TN, and FN derived from sensitivity/specificity are not considered surrogate outcomes.
- Inconsistency (deemed present if there were considerable differences between studies in the accuracy estimates that were not explained, or if cut-offs for biomarkers corresponding to endoscopic activity were not pre-specified but primarily obtained post-hoc corresponding to area under the receiver operating characteristic curve).
- Imprecision (deemed present if there were wide confidence intervals for TP, FN, TN, FP rates).
- Publication bias, if strongly suspected.

Evidence profiles were developed for each intervention using the GRADEpro Guideline Development Tool (https://gradepro.org).

**Translating Evidence to Recommendations**
The guideline panel and evidence synthesis panel met face to face on May 8, 2023 to discuss the evidence and formulate the guideline recommendations. Based on the evidence-to-decision framework, the panel considered the certainty of evidence, balance of benefit and harms, patient values and preferences, and (when applicable) feasibility, acceptability, equity, and resource use. For all recommendations, the panel reached consensus. The certainty of evidence and the strength of recommendation are provided for each clinical question. As per GRADE methodology, recommendations are labeled as “strong” or “conditional”. The words “we recommend” indicate strong recommendations and “we suggest” indicate conditional recommendations provide the suggested interpretation of strong and weak recommendations for patients, clinicians, and healthcare policymakers.

**Review Process**

This guideline was submitted for public comment, and peer review. All comments were reviewed and addressed by the full panel. Final recommendations were approved by the AGA Governing Board.

**DISCUSSION OF RECOMMENDATIONS**

A summary of all the recommendations is provided in Table 3 and discussed below. Key implementation considerations when considering using biomarkers in CD are discussed below and in Table 4.

**Key considerations for implementing these recommendations in clinical practice**

The recommendations in the guideline provide a framework for use of serum or fecal biomarkers in the management of patients with CD to inform treatment. It is important to recognize the limitations of the available data as well as incorporate both patient and provider thresholds for FP and FN when confronted with a specific clinical setting to implement each guideline recommendation.

1. **Considerations of test performance and specificity of biomarkers:** Neither serum CRP nor fecal calprotectin are specific for CD activity.
CRP may be elevated in systemic inflammatory processes and does not always represent luminal CD activity. Fecal calprotectin is more specific for gut inflammation but may be elevated in the setting of concomitant gastrointestinal infections. In patients with CD who present with elevated biomarkers and disease-related symptoms, stool testing for C. difficile and other enteric pathogens is important to help rule out other sources of GI infections.

2. Role of endoscopic evaluation for other indications: In certain situations, endoscopic assessment may be required for reasons other than assessment of disease activity. Thus, decision for replacement of endoscopic evaluation by biomarker measurement should consider other information provided by endoscopy.

   Endoscopic evaluation is warranted for determining the extent and severity of inflammation, dysplasia detection and surveillance, evaluation and endoscopic treatment of stricturing disease, and ruling out cytomegalovirus colitis; biomarkers are not helpful in these situations.

3. Association between treatment target and biomarker performance: The panel debated comparing biomarker performance against any endoscopically active CD (SES-CD ≥ 3) or moderate-to-severe endoscopic activity (SES-CD ≥ 6). We eventually elected to benchmark biomarker performance against SES-CD ≥ 3 given that this was the threshold used to report biomarker performance in most studies and there was paucity of data evaluating biomarker performance for detecting moderate to severe endoscopic inflammation (SES-CD ≥ 6). In addition, current CD treatment guidelines recommend a target of endoscopic remission defined as an SES-CD score < 320-22. It is likely that biomarker performance would be superior against a higher threshold of SES-CD > 6. In contrast, the accuracy of biomarkers to detect complete endoscopic healing defined as a SES-CD of 0 may be less optimal. Yzet et al. compared the outcomes of patients with CD who achieved a Crohn’s disease endoscopic index of severity (CDEIS) score of 0 to those with a CDEIS score 1-428. On longitudinal follow-up, patients with a CDEIS of 0 had lower rates of treatment failure (25%) compared to those with a score of 1-3 (48%, p=0.047). None of the patients with a CDEIS of 0
underwent surgical resection compared to 11% of patients with a CDEIS score 1-4. Furthermore, transmural healing defined on cross-sectional imaging or intestinal ultrasound may also have prognostic significance in patients with CD. Diagnostic performance of a combination of symptoms and biomarkers to detect these more rigorous endpoints was not assessed in this guideline but are likely to have inferior performance given differences in pre-test probability. In this guideline, we focused on accuracy of biomarkers to detect active endoscopic inflammation (mild to severe activity) defined as the absence of endoscopic remission to be consistent with clinical practice and RCT endpoints.

- **Test performance of all biomarkers in this guideline reflect their ability to rule out active endoscopic inflammation (SES-CD > 3).** Biomarkers may be suboptimal for detecting more rigorous treatment targets such as SES-CD of 0, transmural healing, or histologic remission. On the other hand, biomarker performance may be better than reported in the guidelines to detect moderate-to-severe endoscopic activity (SES-CD > 6).

4. **Influence of disease location on performance of fecal biomarkers, and correlation with symptoms:** Elevation of fecal calprotectin may be influenced by the extent and location of inflamed surface. The panel identified limited data comparing the performance of fecal calprotectin by disease location. A systematic review by Simon *et al.* identified 16 eligible studies that examined the sensitivity and specificity of fecal calprotectin by disease location. The sensitivity of fecal calprotectin for small bowel inflammation ranged from 43% - 100% while that for large bowel disease ranged from 67 – 100%. Three studies noted fecal calprotectin correlated with endoscopic severity in the large bowel alone while two other studies demonstrated similar correlation in both small bowel and large bowel disease locations. From 11 studies that compared performance across disease locations, four studies demonstrated superior performance in large bowel CD while seven other studies found no difference in performance between small bowel and large bowel locations. Similarly, studies have suggested that the correlation between symptoms and endoscopic activity may be influenced by disease location and prior
bowel surgery, with a stronger correlation between symptoms and endoscopic activity observed in patients with CD with colon-dominant disease vs. patients with predominantly small bowel involvement, and in patients without prior intestinal resection. We were unable to critically analyze the diagnostic performance of fecal calprotectin in various clinical scenarios by disease location, and hence, opted to report diagnostic performance for small bowel and colonic CD together. The guideline panel posited that beyond disease location, the extent and severity of involved segments may have a significant impact on test performance, independent of disease location, similar to observations in patients with UC where the performance of fecal calprotectin may be inferior in patients with limited proctitis vs. extensive colitis. Unfortunately, studies did not report the performance of fecal calprotectin by disease extent, separate from disease location.

- **Gastrointestinal symptoms may correlate less accurately with endoscopic activity in patients with small bowel CD or those with prior intestinal resection, compared with patients with predominant or extensive colonic involvement.**

- **Fecal biomarkers may be modestly less accurate in detecting endoscopic inflammation in small bowel CD or upper gastrointestinal disease than patients with predominant or extensive colonic involvement. In order to interpret results of fecal biomarkers in patients with predominantly small bowel involvement, close anchoring of symptoms and biomarkers with endoscopic findings (i.e. measuring biomarker and endoscopic activity simultaneously) in patients with active disease, and in remission, is preferred.**

5. **Interpreting biomarker performance for low-risk vs. high-risk treatment adjustments:** The acceptable threshold for performance of biomarkers may differ based on the absolute and/or perceived cost and risk of the proposed interventions in response to biomarker thresholds. For example, in patients with CD with symptoms, a higher rate of FP (i.e., patients incorrectly labeled as having active endoscopic inflammation based on biomarkers, when their disease is actually in remission) may be acceptable for lower risk treatment adjustments such as a brief
course of steroids in individuals at low-risk for adverse effects. On the other hand, it is reasonable to accept lower FP rates for interventions that may be associated with significant cost (dose escalation of biologic therapy) or risk (change in therapy).

- Application of all biomarkers in clinical practice should be guided by downstream implications, including risk of consequent treatment decisions (low-risk treatment adjustment vs. high-risk treatment adjustment). Test performance thresholds (acceptable false positive and false negative rates) may vary for patient-provider teams depending on what treatment adjustment is being considered.

6. Inter- and intra-assay test variability: Biomarker levels may vary between laboratories. Thus, use of the same assay type and laboratory are preferred for accurate comparison of biomarker trajectory.

- Fecal calprotectin assays may not be interchangeable and the same assay should be used for a given patient to compare results over time. Since there can be substantial within-stool, and within-day variation of fecal calprotectin measurements from a single patient, confidence in any single measurement may be limited. Hence, if there is uncertainty of results (such as borderline or unexpected results), repeat fecal calprotectin testing or endoscopic evaluation for confirmation may be required.

7. Inter-individual heterogeneity in biomarker responsiveness: Biomarkers including CRP and fecal calprotectin demonstrate heterogeneity between individuals. Not all patients will demonstrate an elevation in these biomarkers in the setting of endoscopically active disease. Consequently, it is important to anchor the performance of a biomarker against endoscopic assessment for a given patient both in active disease and in remission. Biomarker accuracy is likely superior and of greater clinical value in a patient where the biomarker was shown to be elevated in the setting of endoscopically active disease and normalizes with
resolution of inflammation. In patients where this correlation has not been demonstrated before, interpretation of biomarker result may merit more caution.

➢ There are inter-individual differences in biomarker elevation in patients with intestinal inflammation, and in a subset of patients, biomarkers may correlate poorly with endoscopic activity. The overall performance and confidence in the use of biomarkers for treatment decisions in a particular patient may be higher when these biomarkers have been longitudinally observed to correlate with the patient’s endoscopic disease activity (both during active disease and remission).

GUIDELINE RECOMMENDATIONS

PATIENTS WITH CROHN’S DISEASE IN SYMPTOMATIC REMISSION

Question 1. In patients with CD in symptomatic remission, is interval biomarker-based monitoring superior to symptom-based monitoring to improve long-term outcomes?

Recommendation:

#1 In patients with CD in symptomatic remission, the AGA suggests a monitoring strategy that combines biomarkers and symptoms, rather than relying on symptoms alone (Conditional recommendation, low certainty in evidence)

Comment: Patients who place a higher value in avoiding burden of biomarker testing, over a potentially higher risk of flare and disease progression caused by missing subclinical inflammation, may reasonably choose interval symptom-based monitoring

Implementation considerations:

• Interval biomarker monitoring may be performed every 6-12 months in patients in symptomatic remission.

• Biomarker-based monitoring may be particularly useful in patients where biomarkers have historically correlated with endoscopic disease activity.

Summary of the evidence
The panel compared a biomarker-based monitoring strategy with routine and systematic checking of biomarkers against monitoring of symptoms alone to guide treatment changes in patients with established CD in symptomatic remission. **eFigure 3** lays out the schematic for the proposed comparison. We did not identify any RCTs that directly compared the two strategies and could inform our recommendations. While the CALM trial compared a symptom-based treatment adjustment strategy to biomarker-based treatment adjustment, all participants had active disease at study entry\(^30\). Similarly, the STARDUST trial compared symptom-based vs. symptom and biomarker-based treatment escalation with early endoscopic assessment to guide increase in dose of ustekinumab but all patients were symptomatically active at study entry\(^31\). Thus, both studies did not directly inform this study question. We subsequently examined cohort studies in patients with CD in symptomatic remission comparing rates of disease relapse over long-term follow-up were compared between those with elevated and normal biomarkers. A meaningfully higher risk of relapse in those with elevated biomarkers in symptomatic remission would support a biomarker-based monitoring strategy in CD. We identified 12 cohort studies comprising 982 patients in symptomatic remission with unknown endoscopic activity at enrollment. All these studies examined fecal calprotectin as the biomarker. One-third of patients (38%) had elevated fecal calprotectin defined variably as > 200-300\(\mu\)g/g (**eFigure 4**). On median follow-up of 1 year, patients with elevated fecal calprotectin were 4.8 times more likely to have disease relapse compared with patients with normal fecal calprotectin (95% CI, 2.81 – 8.17), with a high degree of heterogeneity (\(I^2=82\%\)). With an observed median annual risk of relapse of 11% in patients with CD in symptomatic remission and normal fecal calprotectin, the risk of relapse over 12 months in those with elevated biomarkers and symptomatic remission was 52.7% (95% CI 30.9% - 89.9%) (**Table 5**).

**Benefits and Harms (Downsides)**

*Symptom-based monitoring strategy:* The benefit of a symptom-based monitoring strategy is that it relies on symptoms usually assessed as part of routine ongoing clinical care. However, the potential harms of this strategy would be a potentially higher risk of relapse due to potentially missing ongoing endoscopically active and clinical meaningful inflammation in asymptomatic individuals.
Biomarker-based monitoring strategy: The benefit of a biomarker-based strategy may be more accurate prognostication of disease outcomes over 1 year by identifying individuals who may have ongoing endoscopically active disease despite the absence of symptoms, potentially allowing for early treatment adjustment prior to symptomatic relapse. The potential harms of a biomarker-based strategy are the cost and inconvenience, particularly for stool-based biomarkers. Elevated biomarkers in asymptomatic individuals may also lead to anxiety and increased costs due to need for downstream testing to determine FP rates (see question#2, recommendation#4).

Certainty of Evidence
When examining cohort studies comparing long-term outcomes in patients with CD in symptomatic remission with elevated vs. normal biomarkers, there was low confidence in effect estimates supporting the use of a biomarker-based monitoring strategy over a symptom-based monitoring strategy. Evidence was rated down for risk of bias in included studies, and inconsistency in effect estimates with variability in cut-offs of fecal calprotectin. There was limited data on prognostic value of other biomarkers like serum CRP in patients with asymptomatic CD.

Rationale
Using the GRADE Evidence-to-Decision framework, incorporating the potential benefits and downsides of both strategies, the guideline panel conditionally recommended a biomarker-based monitoring strategy over symptom-based monitoring alone. Some patients who prefer to avoid the burden of biomarker-based monitoring in terms of cost and inconvenience may reasonably decide to adopt a symptom-based monitoring strategy alone. The panel determined that an interval of 6-12 months for monitoring biomarkers would be reasonable to reflect routine clinic follow-ups for most patients with CD. As biomarkers may not perform equally well in all patients, the recommendation for biomarker-based monitoring is best suited to those who have previously demonstrated a good correlation between their endoscopic inflammation and biomarker elevation. The panel could not identify literature supporting efficacy of downstream treatment adjustments in response to biomarker elevation alone in asymptomatic individuals, particularly those in sustained remission. The panel acknowledged that it may be
reasonable, in a subset of patients, to follow-up an elevated biomarker measure with serial monitoring rather than treatment escalation or immediate endoscopic assessment. The literature was inadequate to examine the relative prognostic value of different cut-offs for elevated biomarkers, and the panel acknowledged that the prognostic value of a markedly elevated biomarker may differ from mild elevation.

Question 2. In patients with CD in symptomatic remission, at what (A) fecal calprotectin, and (B) serum C-reactive protein cut-off can we accurately rule out active inflammation, obviating routine endoscopic assessment?

Recommendations:

#2 In patients with CD in symptomatic remission with recent confirmation of endoscopic remission (without any change in clinical status, on stable therapy) the AGA suggests using fecal calprotectin $<150 \mu g/g$ and/or CRP $<5mg/L$ (or below cut-off for normal range for the laboratory) to rule out active inflammation, and avoid routine endoscopic assessment of disease activity (Conditional recommendation, low to moderate certainty in evidence)

#3 In patients with CD in symptomatic remission without recent confirmation of endoscopic remission, the AGA suggests endoscopic evaluation to rule out active inflammation, rather than relying solely on fecal calprotectin or CRP (Conditional recommendation, low to moderate certainty in evidence)

Implementation considerations:

- The panel considered recent confirmation of endoscopic or radiologic remission to ideally have been within 3 years.
- Radiologic assessment of disease activity may be a reasonable alternative to endoscopic assessment for patients with predominantly small bowel involvement.
#4 In patients with CD in symptomatic remission, with elevated biomarkers of inflammation (fecal calprotectin >150µg/g, CRP >5mg/L), the AGA suggests endoscopic assessment of disease activity rather than empiric treatment adjustment. (Conditional recommendation, low certainty in evidence)

Implementation considerations:

- In patients with CD in sustained symptomatic remission but elevated biomarkers, repeat measurement of biomarkers (in 3-6 months) may be a reasonable alternative to endoscopic (or radiologic) assessment, especially if the latter has been performed recently.
- Lack of normalization of biomarkers (or persistently elevated biomarkers) in patients whose symptoms recently resolved following initial treatment of symptomatically active CD, likely suggests active inflammation, and may warrant treatment adjustment, without need for endoscopic (or radiologic) evaluation.

Summary of the evidence

Diagnostic performance of fecal calprotectin: The evidence synthesis team examined three diagnostic cut-offs for fecal calprotectin – 50µg/g (± 50µg/g), 150µg/g (± 50µg/g) and 250µg/g (± 50µg/g) through a systematic review and meta-analysis of published studies (Table 6A). To minimize bias due to selective reporting of optimized cut-offs (as is common in diagnostic accuracy studies), we included only studies which reported diagnostic accuracy of pre-selected fecal calprotectin cut-offs or reported the performance across two or more pre-determined cut-offs. The gold-standard for comparison was either endoscopically active CD (SES-CD score ≥3) or endoscopic remission (SES-CD < 3). The sensitivity and specificity of fecal calprotectin cut-off of 50±50µg/g was 88% (95% CI, 79-94) and 67% (95% CI, 51-80), respectively, based on 16 cohorts. The corresponding sensitivity and specificity of 150±50µg/g cut-off (11 cohorts) was 81% (95% CI, 74-87) and 72% (95% CI, 61-81), respectively, and of 250±50µg/g cut-off (14 cohorts) was 76% (95% CI, 70-82) and 74% (95% CI, 67-80), respectively (eFigure 5).
Test performance in patients with CD with known recent endoscopic remission within preceding 3 years (with 20% prevalence of active inflammation in asymptomatic patients): We estimated that among patients with CD in endoscopic remission on stable therapy, and without any change in clinical status, less than 20% of patients will have progression to endoscopic inflammation over 2-3 years. In applying fecal calprotectin performance cut-offs to this scenario, approximately 2.4%, 3.8%, and 4.8% (FN rate) with fecal calprotectin <50µg/g, <150µg/g and <250µg/g, respectively, may be misclassified as having endoscopic remission when they actually have endoscopically active inflammation (Table 6A). In contrast, elevated fecal calprotectin >50µg/g, >150µg/g and >250µg/g in asymptomatic patients had significantly elevated FP rates of 26.4%, 22.4%, and 20.8% whereby patients may be misclassified as having endoscopically active disease despite being in endoscopic remission.

Test performance in patients with CD with unknown endoscopic remission status (with 45% prevalence of endoscopically active inflammation): In this scenario, approximately 5.4%, 8.5% and 10.8% patients (FN rate) with fecal calprotectin <50µg/g, <150µg/g and <250µg/g, respectively, may be misclassified as having endoscopic remission while they have endoscopically active disease (Table 6A). In contrast, elevated fecal calprotectin >50µg/g, >150µg/g and >250µg/g also had high false positive rates of 18.1%, 15.4% and 14.3%, i.e., a significant proportion of patients with endoscopic remission may be falsely classified as having endoscopic activity.

Diagnostic performance of serum CRP: We identified 20 studies reporting on the performance of CRP in this setting, with most studies using a cut-off > 5mg/L. The sensitivity of elevated CRP to detect endoscopically active disease was 67% (95% CI, 54 – 77) with a specificity of 73% (95% CI 65 – 80). (eFigure 6).

Known recent endoscopic remission within preceding 3 years (20% prevalence of endoscopic activity): In applying this cut-off (elevated CRP, generally >5mg/L), approximately 6.6% of patients would have a FN result and be mislabeled as being in endoscopic remission while having endoscopically active disease (Table 6B). Elevated CRP (>5mg/L) in this setting had a FP rate of 21.6% suggesting that nearly one-quarter
of patients may be mislabeled as having endoscopic active disease while being in remission.

**Unknown endoscopic remission or endoscopic remission confirmation >3 years ago** (45% prevalence of endoscopically active inflammation): In this scenario, 14.8% of patients would have a FN result and be mislabeled as being in remission while having endoscopically active disease. A similar proportion (14.8%) would have a FP result and be wrongly characterized as having endoscopically active disease while in remission.

**Certainty of Evidence**

There was no direct evidence comparing how different biomarker cut-offs and accompanying treatment decisions impact downstream patient-important outcomes. However, we did not rate down for indirectness since the presence of endoscopic activity is a close surrogate for unfavorable patient outcomes, and an indication for treatment adjustment.

**Fecal calprotectin:** There was moderate certainty of evidence supporting the use of fecal calprotectin cut-offs of <50 µg/g (evidence rated down for inconsistency due to selective inclusion of studies reporting specific cut-offs and high heterogeneity for summary sensitivity/specificity), and low certainty of evidence supporting the use of fecal calprotectin cut-offs of <150 µg/g and <250 µg/g (evidence rated down for inconsistency, and imprecision since 95% CI exceeded the maximal tolerable FN rate of 5%) to rule out endoscopic inflammation in patients with known endoscopic remission. In patients with unknown endoscopic remission status, the corresponding certainty of evidence is low for calprotectin <50 µg/g, and very low for cut-offs of <150 µg/g and <250 µg/g (evidence rated down for heterogeneity and very serious imprecision since both the point estimate and 95% CI is higher than FN threshold of 5%). In patients with CD in symptomatic remission, either with known endoscopic remission or unknown endoscopic remission status, there was very low certainty of evidence supporting the use of any proposed cut-off of elevated fecal calprotectin to rule in endoscopic inflammation, due to unacceptably high rates of FP (very serious imprecision) and inconsistency.
**Serum CRP**: There was low certainty of evidence supporting the use of CRP < 5mg/L to rule out endoscopic inflammation in patients with CD in symptomatic remission and with known endoscopic remission. Evidence was rated down for inconsistency due to selective reporting of cut-offs in studies optimized for best performance and high heterogeneity for summary sensitivity/specificity, and for serious imprecision since 95% CI exceeded the maximal tolerable FN rate of 5%. In patients in whom endoscopic remission status is unknown, evidence supporting the use of CRP <5mg/L to rule out endoscopic inflammation was very low due to unacceptably high rates of FN (very serious imprecision) and inconsistency. In both endoscopic remission scenarios, there was very low certainty of evidence supporting the use of elevated CRP to rule in endoscopic inflammation, due to unacceptably high rates of FP (very serious imprecision) and inconsistency.

**Rationale**

For the appropriate use of biomarkers in the assessment of patients with CD, patients and healthcare providers should incorporate both test performance as well as downstream consequence of FP and FN rates. The panel acknowledged that there may be instances where patients and providers may be willing to accept higher (> 5%) FN rates depending on the downstream consequences. For ease of implementation in clinical practice and for consistency with UC guidelines\(^{15}\), the guideline panel preferred choosing a single fecal calprotectin cut-off (<150µg/g) that is applicable to multiple scenarios. However, there may be specific clinical situations where a higher or lower cut-off may have an acceptable performance such as cut-off of fecal calprotectin <250µg/g in asymptomatic patients with known endoscopic remission. Conceivably, in patients with small bowel dominant CD where correlation between symptoms and endoscopic activity is less strong and where performance of fecal calprotectin may be modestly lower, lower fecal calprotectin thresholds such as <150µg/g or <50µg/g may yield lower FN rates.

There was limited data regarding the predictive value of serially measured biomarkers. In asymptomatic patients with elevated biomarkers, the FP rate for an
elevated fecal calprotectin or CRP was sufficiently high that the panel recommended endoscopic assessment prior to treatment adjustment to minimize likelihood of overtreatment. However, in this clinical scenario, the panel also recognized that it may be reasonable, in some patients, to consider serial monitoring of biomarkers and determine trajectory of elevation as a factor in informing downstream actions. There was also insufficient data to inform examination of combinations of biomarkers. For example, does the presence of an elevated fecal calprotectin and an elevated CRP increase the likelihood of endoscopically active disease beyond elevation of either biomarker alone. In general, the sensitivity of fecal calprotectin was greater when compared to CRP. However, CRP may be more readily measured and integrated into routine clinical practice.

**PATIENTS WITH SYMPTOMATICALLY ACTIVE CROHN’S DISEASE**

| Question 3. In patients with symptomatically active CD, is an evaluation strategy that combines biomarkers and symptoms superior to symptom-based evaluation for making treatment adjustments? |
| Recommendation: |

#5 *In patients with symptomatically active CD, the AGA suggests a biomarker-based assessment and treatment adjustment strategy, rather than relying on symptoms alone* *(Conditional recommendation, moderate certainty in evidence)*

**Comment:** Patients who place a higher value on avoiding burden of biomarker testing, over a potentially higher risk of over- or under-treatment if relying only on symptoms, may consider choosing interval symptom-based treatment adjustment while being treated for active symptoms

**Implementation considerations:**

- Interval biomarker assessment and treatment adjustment may be performed every 2-4 months in patients being treated for active symptoms.
- After resolution of symptoms (and normalization of biomarkers), endoscopic (and/or radiologic) evaluation should be performed to rule out active inflammation, typically 6-12 months after treatment initiation or adjustment. The patient may then transition to guidance for patients in symptomatic remission.
Summary of the Evidence

A biomarker-based evaluation strategy involves checking non-invasive biomarkers of inflammation in patients with symptomatically active CD to inform ongoing management; in contrast, symptom-based evaluation would involve treatment decisions being driven solely based on symptoms. One RCT, the CALM study, directly compared a biomarker-based evaluation strategy vs. symptom-based evaluation for patients with symptomatically active CD\textsuperscript{30} (Table 7). In this multicenter open-label RCT, Colombel, et al. recruited adults with moderate-to-severely active non-stricturing, non-penetrating CD who were naïve to immune suppressive therapy other than prednisone, and had endoscopic and biochemical evidence of inflammation. All patients were treated with prednisone, and were randomized after 9 weeks to “tight control,” in which treatment escalation (initiation and subsequent escalation of adalimumab) was based on fecal calprotectin >250\( \mu \)g/g and/or CRP >5mg/L and/or symptoms suggestive of CD, versus “clinical management,” in which treatment escalation was based on symptoms alone, assessed every 12 weeks. Of the 244 included patients, more patients in the tight control than clinical management groups (37\% vs. 23\%) achieved deep remission (defined as clinical remission [CDAI <150], endoscopic remission [CDEIS <4 and no deep ulcers], absence of draining fistula and discontinuation of corticosteroids for ≥8 weeks) by 48 weeks.

Benefits and Harms (Downsides)

Symptom-based evaluation strategy: Potential benefit of a symptom-based monitoring strategy is the convenience of relying only on patient-reported outcomes, cost, and faster decision-making. However, potential harms related to relying only on symptoms for treatment decisions are higher rates of inappropriate treatment adjustments/overtreatment and treatment-related complications; approximately 20-35\% of patients with gastrointestinal symptoms suggestive of CD may be in endoscopic remission.

Biomarker-based evaluation strategy: Potential benefits of a biomarker-based evaluation strategy is more accurate detection of inflammation than symptoms alone, to facilitate optimal treatment decisions and treatment escalation in patients with persistently elevated
biomarkers, while simultaneously avoiding overtreatment. Potential harms of a biomarker-based evaluation strategy are the costs and inconvenience of sample collection, particularly stool-based tests, and potential delays in treatment which happen due to the extra step of test completion and awaiting results.

**Certainty of evidence**

From the CALM RCT, there was moderate certainty evidence supporting the use of biomarker-based evaluation strategy in patients with symptomatically active CD with evidence rated down for imprecision due to low event rate.

**Rationale**

Using the GRADE Evidence-to-Decision framework, the guideline panel conditionally recommended in favor of a strategy that combines biomarkers and symptoms compared with symptom-based evaluation alone in patients with symptomatically active CD. The panel recognized that adding an extra step of biomarker testing in patients with symptomatically active CD may potentially delay treatment adjustments for patients, particularly those with limited access to healthcare resources. The panel recognized the value of shared decision-making in these patients; some patients, particularly those with severe symptoms, who place high value in avoiding burden of biomarker testing, may reasonably choose symptom-based evaluation for treatment decisions, acknowledging potentially higher risk of inappropriate overtreatment with symptom-based evaluation alone. This may be particularly true if treatment decisions are considered low risk by the treating provider-patient team.

In the CALM RCT, interval biomarker assessment was performed every 3 months in patients with symptomatically active CD. Optimal management strategy in case of discrepancy between symptoms and biomarkers is unclear. In patients with typical symptoms suggestive of CD, normal biomarkers may not exclude lack of active inflammation, and endoscopic assessment may be a preferred approach. A subset of patients who were symptomatically active, and now have resolving symptoms on therapy but have persistently elevated biomarkers, likely have ongoing inflammation. Treatment adjustments in response to elevated biomarkers is acceptable in this scenario. In this treat-to-target strategy in whom symptoms and biomarkers normalize with iterative
treatment adjustments in response to biomarker-based monitoring, endoscopic confirmation of remission is warranted to facilitate ongoing biomarker-based monitoring in asymptomatic patients as recommended above.

Question 4. In patients with symptomatically active CD, at what (A) fecal calprotectin and (B) serum C-reactive protein cut-off can we accurately diagnose active inflammation, obviating routine endoscopic assessment?

Recommendations:

#6 In patients with CD with mild symptoms and elevated biomarkers of inflammation (fecal calprotectin >150µg/g, CRP >5mg/L), the AGA suggests endoscopic assessment of disease activity rather than empiric treatment adjustment. (Conditional recommendation, very low certainty in evidence)

Implementation consideration:

- Lack of normalization of biomarkers (or persistently elevated biomarkers) in patients whose symptoms partially improve following initial treatment of active CD, likely suggests active inflammation, and may warrant treatment adjustment, without need for endoscopic (or radiologic) evaluation.

#7 In patients with CD with mild symptoms and normal biomarkers of inflammation (fecal calprotectin <150µg/g, CRP <5mg/L), the AGA suggests endoscopic assessment of disease activity rather than empiric treatment adjustment. (Conditional recommendation, very low certainty in evidence)

#8 In patients with CD with moderate to severe symptoms, the AGA suggests using fecal calprotectin >150µg/g or CRP>5mg/L, to rule in active inflammation and inform treatment adjustment and avoid routine endoscopic assessment of disease activity. (Conditional recommendation, low to moderate certainty in evidence)

#9 In patients with CD with moderate to severe symptoms with normal biomarkers of inflammation (fecal calprotectin <150µg/g, CRP <5mg/L), the AGA suggests endoscopic assessment of disease activity rather than empiric treatment adjustment. (Conditional recommendation, low certainty in evidence)
Summary of the evidence

**Diagnostic performance of fecal calprotectin:** Summary sensitivity and specificity of fecal calprotectin for detecting endoscopic inflammation is reported above in question 2. 

*High pre-test probability scenario* (patients with moderate-to-severe symptoms suggestive of CD flare [PRO2 >13, or PRO3 >21], with 80% prevalence of inflammation): In applying these fecal calprotectin cut-offs in high pre-test probability scenarios, approximately 6.6%, 5.6% and 5.2% patients (FP rate) with fecal calprotectin >50µg/g, >150µg/g and >250µg/g, respectively, may be misclassified as having endoscopic activity when actually these patients are in endoscopic remission (Table 8A). In contrast, fecal calprotectin <50µg/g, <150µg/g and <250µg/g in this high pre-test probability scenario, had significantly high rates of being FN (9.6%, 15.2% and 19.2%, respectively), i.e., a significant proportion of symptomatic patients with fecal calprotectin below thresholds who have endoscopic activity, may be incorrectly classified as being in endoscopic remission.

*Intermediate pre-test probability scenario* (patients with mild symptoms of CD [PRO2 score 8-13 or PRO3 score 13-21], with 65% prevalence of endoscopic inflammation): In an intermediate pre-test probability scenario, approximately 11.5%, and 9.8% and 9.1% patients (FP rate) with fecal calprotectin >50µg/g, >150µg/g and >250µg/g, respectively, may be misclassified as having endoscopic activity when they actually are in endoscopic remission (Table 8A). Additionally, fecal calprotectin <50µg/g, <150µg/g and <250µg/g in this intermediate pre-test probability scenario, had high rates of being FN (7.8%, 12.3% and 15.6%, respectively), i.e., a significant proportion of mildly symptomatic patients who have endoscopic activity may be incorrectly classified as being in endoscopic remission.

**Diagnostic performance of serum CRP:** Summary sensitivity and specificity of serum CRP for detecting active endoscopic inflammation is reported above in question 2. 

*High pre-test probability scenario* (patients with moderate-to-severe symptoms suggestive of CD flare [PRO2 >13, or PRO3 >21], with 80% prevalence of inflammation): In applying this cut-off (elevated CRP, generally >5mg/L) to a high pre-test probability scenario, only approximately 5.4% patients (FP rate) with elevated CRP may be
misclassified as having endoscopic activity while in endoscopic remission (Table 8B). In contrast, normal CRP (<5mg/L) had significantly high rates of being FN (26.4%). i.e., a high proportion of symptomatic patients with normal CRP who have endoscopic activity, may be incorrectly classified as being in endoscopic remission.

Intermediate pre-test probability scenario (patients with mild symptoms of CD [PRO2 score 8-13 or PRO3 score 13-21] with 65% prevalence of endoscopic inflammation): In an intermediate pre-test probability scenario, FP rate of elevated CRP was 9.4%, and FN rate of normal CRP was 21.4% (Table 8B). i.e., a high proportion of symptomatic patients who have endoscopic activity may be incorrectly classified as being in endoscopic remission.

Certainty of Evidence
Even though there was no direct data comparing how different biomarker cut-offs and accompanying treatment decisions impact downstream patient-important outcomes, we did not rate down for indirectness since the presence of endoscopic activity is a close surrogate for unfavorable patient outcomes and an indication for treatment adjustment.

Fecal calprotectin: There was low certainty of evidence supporting the use of any proposed fecal calprotectin cut-off to rule in endoscopic inflammation in a high pre-test probability setting (evidence rated down for inconsistency due to selective inclusion of studies reporting specific cut-offs and high heterogeneity for summary sensitivity/specificity, and imprecision since 95% CI exceeded the maximal tolerable FP rate of 5%). In contrast, in the intermediate probability scenario of patients with mild symptoms, there was very low certainty of evidence supporting the use of any proposed fecal calprotectin cut-off to rule in endoscopic inflammation, due to unacceptably high rates of FP (inconsistency, and very serious imprecision since both the point estimate and 95% CI is higher than FP threshold of 5%). Similarly, in both the high and intermediate probability scenarios, there was very low certainty of evidence supporting the use of fecal calprotectin <150µg/g or <250 µg/g as cut-offs to rule out endoscopic inflammation, due to unacceptably high rates of FN (very serious imprecision) and inconsistency. However, the fecal calprotectin cutoff of <50µg/g performed slightly better, although it still was rated
low certainty of evidence (inconsistency, and imprecision since 95% CI exceeded the maximal tolerable FN rate of 5%).

**Serum CRP:** There was low certainty of evidence supporting the use of elevated CRP to rule in endoscopic inflammation in the high pre-test probability setting. Evidence was rated down for inconsistency due to selective inclusion of studies reporting specific cut-offs and high heterogeneity for summary sensitivity/specificity, and imprecision since 95% CI exceeded the maximal tolerable FP rate of 5%. In an intermediate pre-test probability scenario, the certainty of evidence was very low for ruling out endoscopic inflammation (inconsistency, and very serious imprecision since both the point estimate and 95% CI is higher than FP threshold of 5%). In contrast, in both the intermediate and high probability scenario, there was very low certainty of evidence supporting the use of normal serum CRP to rule out endoscopic inflammation (inconsistency, very serious imprecision).

**Rationale**

The guideline panel determined *a priori* the maximal tolerable FP thresholds at 5% for patients with symptomatically active CD. However, the panel deemed there may be circumstances where patients and providers may be willing to accept higher rates of FP, depending on risk of downstream consequences including the nature of treatment adjustment. Thus, the panel promotes shared decision-making with conditional recommendations. For example, while endoscopic evaluation would be warranted for most patients with mild symptoms, decision-based on elevated biomarkers may be acceptable if there are logistical delays in obtaining endoscopic evaluation and if patients and providers are willing to accept negative consequences associated with low-risk treatment adjustments.

As noted earlier, for ease of implementation in clinical practice, the guideline panel felt choosing a single fecal calprotectin cut-off (>150µg/g) that is broadly applicable across a wide range of clinical scenarios is preferable, rather than reporting multiple different cut-offs for different scenarios. Higher fecal calprotectin cut-offs may have modestly lower rates of FP with modest improvement in confidence of decision-making.
Question 5. In patients with CD in surgically-induced remission, at what (A) fecal calprotectin, and (B) serum C-reactive protein cut-off can we accurately rule out post-operative endoscopic recurrence, obviating routine endoscopic assessment?

Recommendations:

#10 In asymptomatic patients with CD following surgically-induced remission within the past 12 months and who are either on post-operative pharmacologic prophylaxis, OR at low risk of post-operative recurrence, the AGA suggests using fecal calprotectin <50 µg/g, to avoid routine endoscopic assessment of disease activity (Conditional recommendation, moderate certainty in evidence)

Comment: Patients, particularly those with multiple prior surgeries, and/or with failure of multiple advanced therapies prior to surgery, who value more accurate assessment of endoscopic recurrence, over the inconvenience and costs of colonoscopy, may reasonably choose endoscopic assessment of disease activity within 12m after surgery.

#11 In asymptomatic patients with CD following surgically-induced remission within the past 12 months, who are at high baseline risk of recurrence AND are not receiving post-operative pharmacologic prophylaxis, the AGA suggests endoscopic evaluation, rather than relying solely on biomarkers, for assessing endoscopic recurrence. (Conditional recommendation, low to moderate certainty in evidence)

Implementation consideration:

- Risk stratification schemes to classify patients’ risk of endoscopic recurrence after surgically-induced remission are not well-defined. Risk factors typically associated with low risk of recurrence 6-12m after surgically-induced remission include older age at surgery (>50 years old), non-smoking, long-standing disease (>10 years), and first surgery for a short segment of fibrostenotic disease (<10-20cm). Risk factors typically associated with high risk of recurrence 6-12m after surgically-induced remission include two
or more prior surgeries, penetrating or perianal disease, smoking, young age at surgery with long-segment of small bowel resection however, these risk factors may not be additive.

- In patients at low baseline risk of recurrence, who are also receiving post-operative pharmacologic prophylaxis, fecal calprotectin \(<150\mu g/g\) may also rule out endoscopic recurrence.
- Normal CRP in patients with asymptomatic CD in surgically-induced remission is not able to rule out endoscopic recurrence accurately.
- There is limited data on ongoing biomarker monitoring alone in patients with CD in surgically-induced remission. Colonoscopic evaluation may be warranted beyond 12m after surgery in patients where biomarker-based monitoring is being pursued.

**Summary of the Evidence**

**Diagnostic performance of fecal calprotectin:** We conducted a systematic review to identify cross-sectional and cohort studies in patients with established CD in surgically-induced remission which reported the diagnostic accuracy of fecal calprotectin for detecting endoscopic recurrence with active inflammation reported as Rutgeerts’ score \(i2\) or higher in most studies. To minimize bias due to selective reporting of optimized cut-offs, we included only studies which reported diagnostic accuracy of pre-selected fecal calprotectin cut-offs or reported the performance across two or more pre-determined cut-offs. Overall, 22 cohorts met these criteria. Using this approach, the sensitivity and specificity of fecal calprotectin cut-off of \(50 \pm 50 \mu g/g\) was 86% (95% CI, 75-93) and 50% (95% CI, 34-66), respectively, based on 11 cohorts; corresponding sensitivity and specificity of \(150 \pm 50 \mu g/g\) cut-off (6 cohorts) was 64% (95% CI, 51-75) and 71% (95% CI, 63-77), respectively, and of \(250 \pm 50 \mu g/g\) cut-off (6 cohorts) was 52% (95% CI, 40-64) and 79% (95% CI, 66-88), respectively (eFigure 7).

**Low Pre-test Probability scenario** (low baseline risk of endoscopic recurrence, on pharmacologic prophylaxis after surgery, estimated 10% prevalence of endoscopic recurrence) (eFigure 8): In applying fecal calprotectin diagnostic cut-offs to this low pre-
test probability scenario, approximately 1.4%, 3.6% and 4.8% patients (FN rate) with fecal calprotectin <50µg/g, <150µg/g and <250µg/g, respectively, may be misclassified as being in endoscopic remission (Rutgeerts’ score i0 or i1) when they actually have endoscopic recurrence (Rutgeerts’ score i2 or higher) (Table 9A). In contrast, elevated fecal calprotectin >50µg/g, >150µg/g and >250µg/g in this low pre-test probability scenario, had significantly high rates of being FP (45.0%, 26.1% and 18.9%, respectively), i.e., a significant proportion of patients who are in endoscopic remission may be incorrectly classified as having endoscopic recurrence.

Intermediate Pre-test Probability scenario (asymptomatic patients with high baseline risk of endoscopic recurrence who are receiving post-operative prophylaxis, or patients with low baseline risk of endoscopic recurrence, who are not receiving pharmacologic prophylaxis, estimated 30% prevalence of endoscopic recurrence): In this intermediate pre-test probability scenario, approximately 4.2%, 10.8% and 14.4% patients (FN rate) with fecal calprotectin <50µg/g, <150µg/g and <250µg/g, respectively, may be misclassified as being in endoscopic remission when they actually have endoscopic recurrence (Table 9A). In contrast, elevated fecal calprotectin >50µg/g, >150µg/g and >250µg/g in this intermediate pre-test probability scenario, had significantly high rates of being FP (35.0%, 20.3% and 14.7%, respectively), i.e., a significant proportion of patients who are in endoscopic remission may be incorrectly classified as having endoscopic recurrence.

High Pre-test Probability scenario (asymptomatic patients with high baseline risk of endoscopic recurrence and are not receiving post-operative pharmacologic prophylaxis, with estimated 60% prevalence of endoscopic recurrence): In this high pre-test probability scenario, approximately 8.4%, 21.6% and 28.8% patients (FN rate) with fecal calprotectin <50µg/g, <150µg/g and <250µg/g, respectively, may be misclassified as being in endoscopic remission when they actually have endoscopic recurrence (Table 9A). In contrast, elevated fecal calprotectin >50µg/g, >150µg/g and >250µg/g in this high pre-test probability scenario, had significantly high rates of being FP (20.0%, 11.6% and 8.4%, respectively),
respectively), i.e., a significant proportion of patients who are in endoscopic remission may be incorrectly classified as having endoscopic recurrence.

**Diagnostic performance of serum CRP:** We only relied on studies that simultaneously reported both fecal calprotectin and CRP data. We identified 4 studies reporting the diagnostic accuracy of serum CRP for detecting post-operative endoscopic recurrence. Most studies reported endoscopic recurrence as i2 or higher. Summary sensitivity and specificity of elevated CRP for detecting endoscopic recurrence was 30% (95% CI, 21-40) and 90% (95% CI, 84-94).

**Low Pre-test Probability scenario** (low baseline risk patients who are receiving pharmacologic prophylaxis, estimated 10% prevalence of endoscopic recurrence): In applying this cut-off (elevated CRP, generally >5mg/L) to a low pre-test probability scenario, approximately 7.0% patients (FN rate) with normal CRP (<5mg/L) may be misclassified as having endoscopic remission when they actually have endoscopic recurrence (**Table 9B**). In contrast, elevated CRP (>5mg/L), in this low pre-test probability scenario, had moderate rates of being FP (9.0%), i.e., 9.0% patients who have endoscopic remission may be incorrectly classified as having endoscopic recurrence.

**Intermediate Pre-test Probability scenario** (asymptomatic patients with high baseline risk of recurrence, receiving post-operative prophylaxis or low baseline risk patients not receiving prophylaxis, estimated 30% prevalence of endoscopic recurrence): In an intermediate pre-test probability scenario, approximately 21.0% patients (FN rate) with normal CRP (<5mg/L) may be misclassified as having endoscopic remission when they actually have endoscopic recurrence (i2 or higher) (**Table 9B**). In contrast, elevated CRP (>5mg/L), in this intermediate pre-test probability scenario, had moderate rates of being false positive (7.0%).

**High Pre-test Probability scenario** (asymptomatic patients with high risk of recurrence not receiving post-operative prophylaxis, with estimated 60% prevalence of endoscopic recurrence): In a high pre-test probability scenario, approximately 42.0% patients (FN
rate) with normal CRP (<5mg/L) may be misclassified as having endoscopic remission when they actually have endoscopic recurrence (Table 9B). In contrast, elevated CRP (>5mg/L), in this high pre-test probability scenario, had low rates of being false positive (4.0%).

Certainty of the Evidence

Fecal calprotectin: There was moderate certainty of evidence supporting the use of fecal calprotectin <50µg/g and <150µg/g to rule out post-operative recurrence in a low pre-test probability scenario (evidence rated down for inconsistency due to selective inclusion of studies reporting specific cut-offs and high heterogeneity for summary sensitivity/specificity) and low certainty of evidence supporting the use of fecal calprotectin <250µg/g to rule out post-operative recurrence in a low pre-test probability scenario (evidence rated down for inconsistency and imprecision since 95% CI of the FN crosses the established threshold of 5%). In contrast, in the intermediate and high pre-test probability scenarios, low certainty of evidence supported the use of fecal calprotectin <50µg/g (inconsistency and imprecision), and very low certainty of evidence supported the use of fecal calprotectin <150µg/g and <250µg/g to rule out post-operative recurrence (inconsistency, and very serious imprecision since both the point estimate and 95% CI is higher than FN threshold of 5%).

In all probability scenarios, there was very low certainty of evidence supporting the use of fecal calprotectin >50µg/g, >150µg/g or >250 µg/ to rule in endoscopic recurrence in asymptomatic patients, due to unacceptably high rates of FP (very serious imprecision) and inconsistency. However, the fecal calprotectin cutoff of >250µg/g performed slightly better, although it still was rated low certainty of evidence (inconsistency, and imprecision since 95% CI exceeded the maximal tolerable FP rate of 5%).

Serum CRP: In all probability scenarios, there was very low certainty of evidence supporting the use of normal CRP <5mg/L to rule out endoscopic recurrence in asymptomatic patients, due to unacceptably high rates of FN (very serious imprecision) and inconsistency. In contrast, low certainty supported the use of elevated CRP >5mg/L
to rule in endoscopic recurrence in asymptomatic patients in a high- and intermediate pre-
test probability scenario (inconsistency and imprecision since 95% CI exceeded the
maximal tolerable FP rate of 5%); there was very low certainty evidence supporting its
use in a low pre-test probability scenario.

Rationale
Using the GRADE Evidence-to-Decision framework, the guideline panel conditionally
recommended in favor of a strategy that uses fecal calprotectin-based monitoring, using
a cut-off of \(<50\mu g/g\), over routine endoscopic evaluation to rule out post-operative
recurrence in asymptomatic patients with CD following surgically-induced remission
within the past 12 months and who are either receiving post-operative pharmacologic
prophylaxis or at low baseline risk of post-operative recurrence (regardless of post-
operative prophylaxis). In asymptomatic patients with CD following surgically-induced
remission within the past 12 months, who are at high baseline risk of recurrence and are
not receiving post-operative pharmacologic prophylaxis, the panel conditionally
recommended endoscopic evaluation, rather than relying solely on biomarkers, for
assessing endoscopic recurrence.

The guideline panel recognizes the challenge in clinical practice of using a different
fecal calprotectin cut-off (<50\mu g/g) in the evaluation of asymptomatic patients with CD
following surgically-induced remission within the past 12 months. It is worth noting that
fecal calprotectin performed well at <50\mu g/g to rule out endoscopic recurrence and active
inflammation in patients at low and intermediate baseline risk of recurrence. In patients
at low baseline risk of recurrence, who are also receiving post-operative pharmacologic
prophylaxis, fecal calprotectin <150\mu g/g may also rule out endoscopic recurrence.
However, in patients at high baseline risk of recurrence, normal fecal calprotectin does
not rule out endoscopic recurrence and an endoscopic evaluation is recommended in
these patients over biomarker-based assessment.

In contrast to the utility of a normal fecal calprotectin value, the test performance
of elevated fecal calprotectin was not sufficient to recommend relying only on this to make
treatment decisions in asymptomatic patients with CD following surgically-induced
remission within the past 12 months, regardless of pre-test probability scenarios. In this
setting, the guideline panel recommends endoscopic evaluation to confirm presence and
severity of endoscopic recurrence prior to treatment adjustments. The data for CRP did not support its use (either normal or elevated value) to determine endoscopic recurrence accurately in patients with asymptomatic CD in surgically-induced remission.

There is limited data on ongoing biomarker monitoring alone in patients with CD in surgically-induced remission. The panel noted that given the limited data, endoscopic evaluation may be warranted beyond 12m after surgery in patients where biomarker-based monitoring is being pursued. For patients with surgically-induced remission with symptoms, the panel recommends management strategy outlined in recommendation statements #6-9 for patients with mild or moderate-severe symptoms.
ENDOSCOPIC HEALING INDEX (Monitr®)

Question 6. In patients with CD in symptomatic remission, at what EHI cut-off can we accurately rule out active inflammation, obviating routine endoscopic assessment?

Question 7. In patients with symptomatically active CD, at what EHI cut-off can we accurately diagnose active inflammation, obviating routine endoscopic assessment?

Question 8. In patients with CD in surgically-induced remission, at what EHI cut-off can we accurately rule out post-operative endoscopic recurrence, obviating routine endoscopic assessment?

Recommendation: #12 In patients with CD, the AGA suggests neither in favor of nor against the use of endoscopic healing index (EHI, Monitr®) for monitoring inflammation and treatment decisions (No recommendation, Knowledge gap)

Summary of the Evidence

Diagnostic performance of EHI in luminal CD: The endoscopic healing index (EHI, Monitr™) measures 13 proteins in blood (ANG1, ANG2, CRP, SAA1, IL7, EMMPRIN, MMP1, MMP2, MMP3, MMP9, TGFA, CEACAM1, and VCAM1) and was developed as a diagnostic test to reflect the severity of endoscopic inflammation in CD. A single derivation-validation study was identified which evaluated the test performance of EHI in two validation cohorts. The first validation cohort, Tailored Treatment With Infliximab for Active Crohn's Disease (TAILORIX), consisted of 116 patients with prospectively collected data, 26% of whom were in endoscopic remission and 10% had a history of prior IBD surgery. The second validation cohort included samples prospectively collected from a tertiary referral center (University of California San Diego), 46% of whom had a history of prior IBD surgery.

Two cut-offs were determined for evaluation of luminal CD. A cut-off of EHI <20 was optimized to rule out active inflammation, defined as SES-CD < 3. In the TAILORIX
cohort, the sensitivity and specificity of EHI <20 were 96% and 64% respectively. In the UCSD cohort, the sensitivity and specificity of EHI <20 were 92% and 42% respectively. A cut-off of EHI >50 was optimized to rule in active inflammation, defined as SES-CD ≥3. In the TAILORIX cohort, the sensitivity and specificity of EHI >50 were 36% and 100% respectively. In the UCSD cohort, the sensitivity and specificity of EHI >50 were 35% and 91% respectively.

With this optimized test performance, EHI <20 had very low rates of FN in asymptomatic patients with luminal CD, regardless of whether patients were in known (FN 1.2%), or unknown endoscopic remission status (FN 2.7%). In these clinical scenarios, EHI >20 had very high rates of FP (>25%), implying that a high proportion of asymptomatic patients in endoscopic remission will be incorrectly classified as having endoscopic activity when EHI >20 (Table 10A). In patients with symptomatically active CD, with mild symptoms, or with moderate to severe symptoms, EHI >50 had very low rates of FP (<2%). In these symptomatic patients, EHI <50 had very high rates of FN (>40%), implying that a high proportion of symptomatic patients with endoscopically active disease will be incorrectly classified as being in remission when EHI<50 is used as a cut-off (Table 10B). While the test cut-offs were optimized, a large number of results of the test were in the 20-50 range, making them indeterminate.

Diagnostic performance of EHI in post-operative CD: A single study was identified evaluating the use of EHI in post-operative CD. This study was a secondary analysis of the Post Operative Crohn’s Endoscopic Recurrence (POCER) trial. At 6 months, the sensitivity and specificity of EHI<20 for endoscopic recurrence (i2 or higher) was 82% and 50% respectively (Table 10C). At this optimized cut-off, EHI <20 was able to accurately rule out endoscopic recurrence (FN rate <6%) in patients at low- and intermediate pre-test probability scenarios, including asymptomatic patients with CD following surgically-induced remission within the past 12 months and who are either receiving post-operative pharmacologic prophylaxis, or at low baseline risk of post-operative recurrence (regardless of post-operative prophylaxis).

Certainty of the Evidence
The overall body of evidence supporting the use of EHI in different clinical scenarios was rated as very low quality due to overall paucity of studies. Only one derivation and two accompanying validation studies examined the performance of EHI in patients with luminal CD; there was only one study evaluating the performance of EHI in patients with post-operative CD. Consequently, the body of evidence was rated down for very serious imprecision, and possibly reporting/publication bias.

Rationale
Using the GRADE Evidence-to-Decision framework, the guideline panel decided to make a recommendation neither in favor of nor against the use of the EHI test in CD. In arriving at this recommendation, the guideline panel weighed the performance of the test in the reported validation cohorts against the paucity of independent data despite the test being commercially available since 2020. Based on this, together with limited access and feasibility and risk of exacerbating inequity for this proprietary test, the guideline panel opted not to make a recommendation in favor of, or against its use, identifying this as a knowledge gap. The availability of more generalizable data demonstrating high accuracy from independent data sets in sufficiently heterogenous populations, as well as evidence of its feasibility and cost-effectiveness relative to other widely available tests would merit reconsideration of the recommendation.

### BIOMARKER- vs. ENDOSCOPY-BASED MONITORING STRATEGY IN CD

**Question 9. In patients with established CD, is interval biomarker-based monitoring superior to endoscopy-based monitoring to improve long-term outcomes?**

**Recommendation:**

**#13 In patients with CD, the AGA makes no recommendation in favor of, or against, a biomarker-based monitoring strategy over an endoscopy-based monitoring strategy to improve long-term outcomes. (No recommendation, Knowledge gap)**

**Summary of the Evidence**

A biomarker-based monitoring strategy involves routine assessment of symptoms and non-invasive biomarkers of inflammation in patients with CD in symptomatic remission to
inform ongoing management. In this situation, normalization of biomarkers is an adequate treatment target – asymptomatic patients with normal biomarkers would continue current management without endoscopy, whereas those with elevated biomarkers would undergo endoscopy. In contrast, an endoscopy-based monitoring strategy involves routine endoscopic assessment to confirm endoscopic remission of CD periodically. eFigure 9 lays out the schematic for proposed comparison. We did not identify any RCTs that compared a biomarker-based monitoring strategy vs. an endoscopy-based monitoring strategy. Normalization of CRP and reduction of fecal calprotectin are recognized as short-term treatment targets in managing CD in expert consensus statements, assessed early in treatment course. Early achievement of these biomarker outcomes is associated with favorable longer term outcomes including risk of relapse, as well as likelihood of achieving endoscopic improvement. Potential benefits of a biomarker-based monitoring strategy are convenience and low resource utilization due to avoidance of routine and recurrent endoscopic assessment. Potential harms of a biomarker-based monitoring strategy are insufficient assessment and suboptimal performance for achieving deeper remission endpoints such as complete endoscopic or transmural remission which may be associated with more favorable long-term outcomes. Hence, the guideline panel felt there is insufficient evidence to inform between the choice of a biomarker-based monitoring strategy vs. an endoscopy-based monitoring strategy in patients with CD in symptomatic remission. This was identified as a knowledge gap which warrants further study.

Limitations of Current Evidence and Future Directions
The evidence panel identified numerous knowledge gaps in the literature where there is insufficient data to inform recommendations.

1. **Biomarker-based treat-to-target strategy in CD:** Treatment strategy trials such as CALM have demonstrated that incorporating biomarker assessment as part of the treat-to-target strategy is beneficial especially in patients with active disease.\(^\text{14}\) However, such treatment strategy trials have relied on a rigid set of pre-specified criteria that would result in treatment escalation. There is a need for examination of various biomarker cut-off threshold to guide therapy escalation, examination of the role of combination of biomarkers, role of biomarker-based treat-to-target
strategy in asymptomatic patients, as well as potential harm of not dose escalating
in the setting of mild biomarker abnormality to robustly inform biomarker-based
treatment strategies. There have not been any studies comparing a biomarker-
based strategy to an endoscopy-based strategy for assessment and monitoring of
endoscopic remission. This was identified as a knowledge gap by the panel.

2. **Magnitude of elevation of biomarkers**: The guideline panel focused on
examination of performance of biomarkers at commonly reported cut-offs that are
widely used in clinical practice. Consequently, management recommendations
could only be made based on whether the value was above the cut-off for that
biomarker but did not factor in the degree of abnormality. A single measurement
demonstrating marked elevation of a biomarker may, for a given patient, carry a
different prognostic implication, than a more modest elevation. For example, in
individuals with mild symptoms, fecal calprotectin > 2500\(\mu\)g/g may carry different
implications for management than fecal calprotectin of 251\(\mu\)g/g.\(^{37}\) There was
insufficient data to guide nuanced decision making in this context. There are
several novel biomarkers, including biomarker panels, of disease activity and
prognosis that have been studied in research settings, but require more robust
clinical validation before widespread adoption. The paucity of data on this was also
identified as a knowledge gap by the panel, requiring further research.

3. **Choice of treatment target**: Consistent with existing clinical guidelines defining
endoscopic remission in CD as SES-CD < 3\(^{20}\), this guideline examined the
diagnostic accuracy of biomarkers in determining either presence or absence of
inflammation at this threshold. Studies have demonstrated that a more rigorous
treatment target of SES-CD score of 0 may be associated with better outcomes\(^{28}\).
It is likely that biomarker performance will not be as robust against the more
rigorous treatment target. There was limited data on performance of biomarkers
against other treatment goals such as histologic remission or transmural healing
on radiologic assessment. This was identified as a knowledge gap. Conversely,
some patients and physicians may elect to optimize therapy only for moderate-to-
severe disease activity (SES-CD > 6); the performance of the biomarker may be
superior against that endpoint compared to that reported in this guideline.
4. **Influence of disease location and extent:** There is significant heterogeneity in extent and location of involved segment in CD; this may directly influence biomarker sensitivity and specificity as well as its accuracy. The correlation between symptoms and endoscopic activity may be weaker for small bowel CD which would lead to a lower prevalence of endoscopically active disease for given symptoms. This would reduce accuracy of the biomarker in individuals with symptomatic CD. There are no widely accepted validated scoring systems for endoscopic assessment of mucosal inflammation in CD involving the proximal small bowel in isolation (i.e. beyond the reach of colonoscope). Thus, the panel determined this to be a knowledge gap in the performance of biomarkers. Please see more detailed discussion of the impact of disease location on biomarker performance under key implementation considerations above.

5. **Biomarker performance in diverse populations:** The panel recognized the lack of robust data in specific clinical situations including mild CD, CD involving the J pouch or in patients with an ostomy, and in geographically and ethnically diverse patient populations, where there exist only few studies examining the role of biomarkers to date.

6. **Comparison against other disease activity assessment modalities:** Cross-sectional imaging is increasingly being utilized to define transmural healing in CD with growing use of CT, MR, and intestinal ultrasound-based assessments. Intestinal ultrasound, in particularly, is attractive as a point-of-care test without radiation exposure and limited preparation. There were few studies comparing biomarker performance against these imaging-based assessments.

**What do other guidelines say?**

There has been limited discussion on the role of non-invasive biomarkers in the management of CD in clinical guidelines. The American College of Gastroenterology society guidelines published in 2018 on the management of CD suggested fecal calprotectin and serum C-reactive protein may have adjunctive role in assessing inflammation in patients with CD but did not provide specific cut-offs or recommendations for use\(^3\). ECCO-ESGAR (European Crohn’s and Colitis Organization and the European Society of Gastrointestinal and Abdominal Radiology) guidelines on the diagnostic
assessment of IBD recognized that asymptomatic patients with elevated biomarkers of inflammation, mainly fecal calprotectin and CRP may suggest imminent flare and recommended endoscopic or radiologic evaluation.\textsuperscript{39} In patients with clinical response to medical therapy, the guidelines recommend evaluating for mucosal healing either via endoscopy or fecal calprotectin. None of these guidelines discussed performance of specific cut-offs and downstream implications involved in decision-making which are critical to using these biomarkers in clinical practice. Similar to the current guideline, the AGA guideline for the use of biomarkers in UC similarly suggests that a normal biomarker in asymptomatic patients or an elevated biomarker in those with moderate-to-severe symptom can reliably rule out or rule in the presence of endoscopically active disease respectively, thereby avoiding endoscopy solely for assessment of disease activity\textsuperscript{15}. However, there are some key distinctions between the two guidelines. First, in CD, symptoms correlate less well with endoscopic activity. Thus, biomarker performance was acceptable only in asymptomatic individuals who had recently confirmed endoscopic remission; in those without recent endoscopic assessment, test performance was suboptimal, and the guideline suggests endoscopic assessment as the preferred strategy for assessing disease activity. Second, the weaker correlation between symptoms and endoscopic activity in CD also reduced the utility of biomarker measurement to infer disease activity in those with mild symptoms.

**Plans for updating this guideline**

Guidelines are living products. To remain useful, they need to be updated regularly as new information accumulates. This document will be updated when major new research is published. The need for update will be determined no later than 2026 and, if appropriate, we will update the guidelines to incorporate updated recommendations as new evidence, without duplicating or creating a new comprehensive guideline.
REFERENCES

17. Sands BE. Biomarkers of Inflammation in Inflammatory Bowel Disease. Gastroenterology 2015;149:1275-+


