EXPERT REVIEW OF GASTROENTEROLOGY & HEPATOLOGY
2025, VOL. 19, NO. 4, 359–363
https://doi.org/10.1080/17474124.2025.2482980

Taylor & Francis
Taylor & Francis Group

SPECIAL REPORT

Misinterpretation and misapplication of biomarkers in inflammatory bowel disease: how do we avoid this?

Robert B. Varley[a] and James C. Lee[a,b,c]

[a]Department of Gastroenterology, Royal Free Hospital, London, UK; [b]Genetic Mechanisms of Disease Lab, The Francis Crick Institute, London, UK; [c]Institute for Liver and Digestive Health, Division of Medicine, University College London, London, UK

ABSTRACT

Introduction: The management of inflammatory bowel disease (IBD) has evolved substantially over the past decade, with the emergence of new advanced therapies presenting unprecedented challenges in clinical decision-making. While these therapies provide patients with more opportunities to get better, biomarkers to guide their use remain elusive.

Areas covered: This article highlights the challenges associated with biomarker discovery, interpretation, and application in IBD – based on literature review, first-hand experience of biomarker discovery, and personal opinion. We highlight problems including the misinterpretation of predictive capabilities, lack of independent validation, and reverse causation in retrospective studies, and explain why associations with clinical parameters or seropositivity to microbial antigens often fail to meet the rigorous performance metrics required for clinical utility. The relative need for different biomarkers is also discussed – particularly in light of recent evidence from the PROFILE trial, which emphasizes the considerably greater risk posed by uncontrolled disease than by the potential side-effects of medications.

Expert opinion: Despite multiple challenges, the potential of biomarkers for precision medicine in IBD remains promising, particularly in combination with other clinical and biochemical parameters. Further research into combinatorial biomarker approaches is needed, but must be combined with learning how to communicate results that are inherently uncertain.

ARTICLE HISTORY
Received 9 January 2025
Accepted 18 March 2025

KEYWORDS
IBD; biomarkers; prediction; prognosis; drug response; side effects

1. Introduction

Less than 10 years ago, the management of IBD was relatively straightforward. Only one class of advanced therapy was available and these drugs were simply too expensive to use in anyone who had a chance of responding to cheaper, less effective treatments. Now, however, things are much more complicated – a situation that has parallels in rheumatoid arthritis (RA), systemic lupus erythematosus (SLE) and multiple sclerosis (MS), to name just a few. Having many more advanced therapies means many more decisions have to be made, and all with little advance knowledge of which drug would be best for individual patients. Indeed, recent improvements in IBD care have been driven more by closer monitoring than by individualized care [1]. While research has therefore increasingly focused on finding more sensitive measures of subclinical disease activity – to provide better 'targets' for a 'Treat-to-Target' approach – such biomarkers, including promising candidates like serum leucine-rich α2-glycoprotein [2], are simply measures of current disease activity. For this reason, efforts to develop prognostic and predictive biomarkers continue, both to predict the future disease course and the likely outcome of potential therapy (including efficacy and side effects). In this article, we discuss the errors that are commonly made in the application and interpretation of biomarkers in IBD. For a more granular overview of available biomarkers and their strengths and limitations, please see our previous publications [3,4].

2. What is prediction?

The commonest error with predictive biomarkers is perhaps also the most surprising: thinking you have one when you don't. Predicting anything implies that the outcome has not already happened. Thus, previously requiring bowel surgery and/or advanced therapies or having developed disease complications are not 'predictors' of aggressive IBD [5–7] – just as having joint deformities or renal failure are not predictors of aggressive RA or SLE. Rather, these are clear indicators that aggressive disease is already present. Association is also commonly confused with prediction, with many statistically robust associations being cited as potential predictors when it is clear that their predictive performance would actually be poor [3]. Indeed, very few reported predictors have sufficient sensitivity, specificity, positive and negative predictive values to be useful predictors. A good example of this are clinical parameters, including younger age at diagnosis, early need for steroids and perianal disease in Crohn's Disease (CD) or extensive disease in ulcerative colitis (UC), which are still commonly used in clinical practice – perhaps because of a lack of alternatives – but which have long been known to be poor predictors [8–16].

CONTACT James C. Lee james.lee@crick.ac.uk Genetic Mechanisms of Disease Lab, The Francis Crick Institute, 1 Midland Rd, London NW1 1AT, UK

Article highlights

- Prediction is often confused with association, but requires distinct performance characteristics.
- Prediction is inherently uncertain – recognizing and effectively communicating will be critical.
- Biomarker discovery must involve independent validation.
- Caution is needed when trying to identify biomarkers retrospectively in case of reverse causation.
- Predicting drug response and side-effects should be a greater priority than predicting prognosis in light of the results of PROFILE.

3. Interpreting predictions

A fundamental aspect of prediction is that the projected outcome is not certain. This uncertainty – and risk in general – is often difficult for patients (and some clinicians) to understand. For biomarkers to be successfully adopted in clinical practice, the ability to understand and explain their results – and more importantly, their implications – will be essential. Even good biomarkers will appear to get it wrong in some cases (Figure 1). In these situations, the risk of misclassification should be weighed against the risk of not using the biomarker. For example, a companion diagnostic was recently reported to identify patients with a higher chance of responding to Tulisokibart, an anti-TL1A monoclonal antibody [17]. However, 76% of patients were negative for the biomarker, and the difference between the remission rate in a biomarker positive group (32%) and the unstratified cohort (without considering the biomarker; 26%) was small. As such, use of such biomarkers – to select patients for anti-TL1A therapy – would risk denying some patients the chance of receiving a therapy that could have been effective.

4. Challenges in identifying reliable biomarkers

Experience in oncology has highlighted the many challenges associated with identifying reliable biomarkers [18]. Indeed, while the potential of biomarkers to deliver precision medicine is undisputed – promising prediction of diagnosis, prognosis, treatment response and overall outcomes – the reality is that most proposed biomarkers have failed to fulfill that potential. There are many reasons for this, although some common themes have emerged from previous experience.

4.1. Lack of independent validation

A leading cause of biomarker failure is the lack of adequate independent replication. This is especially important for putative biomarkers identified in '–omic' studies, since the number of parameters measured in these studies (p) vastly exceeds the number of samples (n), resulting in a high likelihood of overfitting (the phenomenon by which spurious associations are detected by chance due to the sheer number of data-points considered) [19]. The situation where $p >> n$ is a recognized challenge in classifier development and can only be confidently overcome by confirming predictive performance in independent cohorts [20]. This is not a trivial task, since recruiting new patient cohorts is both time-consuming and costly, and inherently risks that a biomarker may not validate. Indeed, even when associations do replicate, the effect size observed in replication studies is usually smaller than in the index studies in which they were first reported ('winner's curse'). Nonetheless, since the performance, and thus value, of unvalidated biomarkers is essentially unknown, independent validation must be a pre-requisite for all future biomarkers.

4.2. Reverse causation

To rapidly obtain large study cohorts with extended follow-up data, many investigators have opted for a retrospective approach to biomarker discovery. While this is undoubtedly easier than recruiting prospective cohorts and waiting for

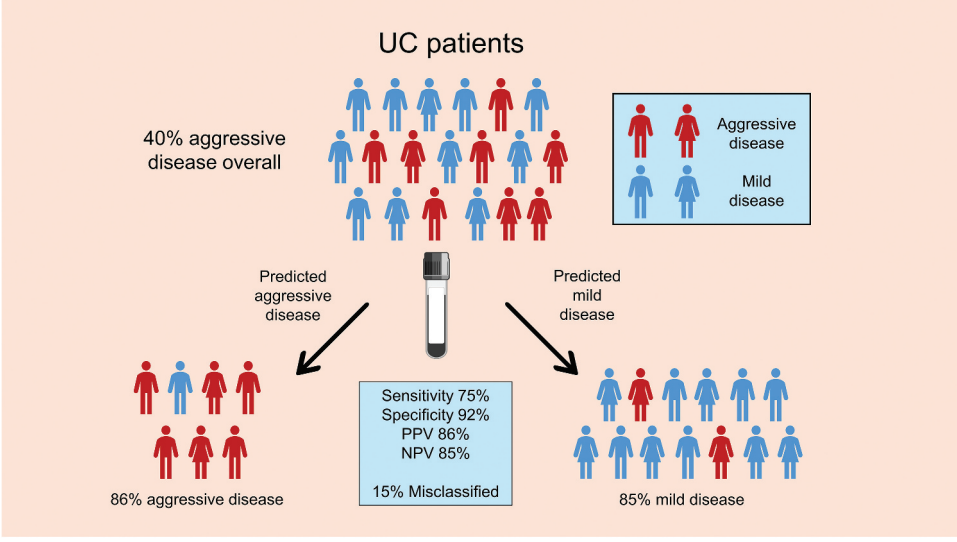

Figure 1. Understanding biomarker performance.

A hypothetical prognostic biomarker can identify a cohort of UC patients who have either an 86% chance or a 15% chance of developing aggressive disease. The biomarker's performance characteristics are therefore excellent, but 15% of patients are 'misclassified.' Blood vial image: NIAID Visual & Medical Arts. (10/7/2024). BDBlood Vial. NIAID NIH BIOART Source. bioart.niaid.nih.gov/bioart/52

follow-up time to elapse, there are several risks inherent with this strategy. A major concern is reverse causation, in which an apparent predictor of an outcome is in fact the result of that outcome. For example, seropositivity to various microbial antigens has been reported to predict more complicated CD and higher need for surgery [20–23]. However, most of these studies were based on serologic testing of samples from CD patients after many years of disease (and correlation with prior disease course). Subsequent work has highlighted that seropositivity increases over time, and while there are undoubtedly some CD patients who are seropositive at diagnosis, our experience is that this is a minority – consistent with the subsequent poor predictive performance reported in attempted replication studies [24,25]. Large prospective studies of newly diagnosed CD patients will therefore be needed to determine the true utility of anti-microbial antibodies as prognostic biomarkers. In the meantime, other serological markers – including anti-αvβ6 antibodies which can precede the diagnosis of ulcerative colitis by up to 10 years – may prove more useful by facilitating prevention studies [26].

5. Using biomarkers appropriately

When interpreting biomarkers, it is important to recognize the contribution of other factors that may influence either the result itself or at least the interpretation. An obvious example is the influence of disease location on fecal calprotectin results. It is now well-recognized that fecal calprotectin levels are generally lower in patients with isolated small bowel disease compared to those with colonic involvement, leading to a reduction in sensitivity [27–29]. It is also important to consider the pretest probability prior to conducting a test, as this may alter the interpretation of the result. For example, in UC patients in symptomatic remission, it has been estimated that the overall rate of moderate or severe inflammation is approximately 15% [30]. In this setting, a fecal calprotectin result of < 250mcg/g carries a 92% negative predictive value. On the other hand, a symptomatic UC patient with rectal bleeding and an elevated stool frequency has an 85% chance of moderate or severe inflammation. Using the same calprotectin cutoff in this setting results in a value < 250mcg/g only having a 26% negative predictive value [30].

6. What are the key performance parameters?

The perfect biomarker is unlikely to ever exist. Nevertheless, there are some aspects of biomarker performance that are more important than others. Exactly which aspects are most important will vary depending on the clinical context, but this should be carefully considered when developing, optimizing and validating biomarkers. This typically requires an assessment of risk – for example, which is more dangerous for a patient: the risk of inadequately treated disease, or the potential risks of the proposed therapy? If uncontrolled disease carries a greater risk, then negative prediction is more important (i.e. reliably identifying some patients who do not need the treatment, even if this means that a few patients could be over-treated) [31]. Alternatively, if the treatment carries a greater risk than the disease, then the reverse will be true and it will be more important to identify patients who would definitely benefit, even if this does not pick up every patient in this group. In IBD, it is increasingly recognized that risks of uncontrolled inflammation outweigh those of most therapies, and so negative prediction is likely to be more important in most scenario [31,32].

7. What biomarkers do we actually need?

In recent years, we and others have highlighted the need for both prognostic biomarkers (to predict disease course) and predictive biomarkers (to identify the most suitable treatment for an individual) [4]. However, the results of the PROFILE trial have provided a new perspective on the relative importance of these aims [32]. In PROFILE – a biomarker-stratified trial designed to test a prognostic biomarker – newly diagnosed CD patients were randomized to accelerated step-up or top-down therapy (infliximab and azathioprine) after biomarker testing. While the biomarker did not show prognostic utility, there was an overwhelming benefit of top-down therapy, with higher endoscopic remission rates than have ever been reported in any Phase 3 trial. Moreover, approximately 1 in 20 patients treated with accelerated step-up required surgery for disease complications within the first year, while the only surgery required in the top-down arm was for a gallstone ileus. Notably there was also no excess of serious infections in newly diagnosed patients treated with top-down therapy – consistent with other reports that infection risk in IBD patients may be driven as much by physiological deconditioning due to longstanding uncontrolled disease than by treatments per se [33,34]. While PROFILE only enrolled patients who had both symptoms and some degree of detectable inflammation (and so cannot be generalized to patients diagnosed incidentally) the message is clear: top-down therapy should be considered the standard of care for most newly diagnosed CD patients, especially with the advent of biosimilar drugs, which mean this is no longer unaffordable in many healthcare settings. Importantly, while some clinicians will be more comfortable with adverse outcomes arising from the underlying disease, rather than from the drugs they have prescribed, it is undoubtedly true that the greatest harm to patients comes from IBD itself [35,36].

8. Expert opinion

Despite considerable ongoing research, the exact role that predictive and prognostic biomarkers will play in the future of IBD management remains unclear. Some investigators have even questioned whether predictive or prognostic biomarkers will ever be realistically implemented into clinical practice – especially following the observed lack of biomarker utility in PROFILE. While this result clearly highlights the challenges associated with developing biomarkers – and emphasizes the importance of formally testing biomarkers in stratified trials (of which PROFILE remains the only one to date in IBD) – it is worth remembering that this situation is not dissimilar to the experience in oncology. Indeed, despite oncology being frequently cited as an exemplar of biomarker success, many more cancer biomarkers have failed than have succeeded. Similar

results have also been seen in other immune-mediated diseases [37]. The result of PROFILE should therefore not deter future biomarker discovery efforts, and provides a useful framework for how putative biomarkers can and should be tested. The other important lesson from PROFILE was that there was clear benefit in using top-down therapy from diagnosis in CD patients, even without a biomarker. This benefit included both disease control and avoidance of complications, and was not associated with a clear downside with respect to side effects or infection risk. This result therefore mandates a reevaluation of the relative need for different types of biomarker and – ironically – suggests that a prognostic biomarker may not be so important after all. Selecting which therapy to use first, however, remains a major unmet need, but we suspect this will not be solved with a single biomarker. Indeed, even determining whether a specific drug would be suitable for an individual patient will require both prediction of efficacy and the risk of side-effects. It is also conceivable that some biomarkers will provide more information when combined with clinical features than they provide alone. These possibilities can, and should, be formally tested to provide patients with the best chance of receiving the right treatment at the right time. Aside from challenges associated with the development of biomarkers, clearly communicating their results to patients will bring additional challenges. Like any forecast, biomarker-based predictions will not be definitive and being able to communicate the associated uncertainty will be important for maintaining doctor-patient relationships, and managing expectations.

So where will the biomarker field be in 10 years? Based on the time it takes to develop biomarkers, validate their performance, and conduct stratified trials to show whether their use would improve care, it seems possible that we may still be having the same discussions regarding the unfulfilled potential of IBD biomarkers. However, there are several promising studies that are currently underway, and – providing we can learn the lessons that have been painfully learnt in oncology – it is equally possible that we may finally be in a position to match patients to the treatment(s) that are most suitable for them.

Funding

JC Lee is funded by the Francis Crick Institute, which receives its core funding from Cancer Research UK [CC2219], the UK Medical Research Council [CC2219], and the Wellcome Trust [CC2219]. JC Lee is a Lister Institute Prize Fellow.

Declaration of interest

JC Lee reports consultancy and/or speaker fees from AbbVie, AgPlus Diagnostics, PredictImmune, Pfizer, Johnson & Johnson, and C4X Discovery outside the submitted work; and grants from GSK outside the submitted work. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

Reviewer disclosures

Peer reviewers on this manuscript have no relevant financial or other relationships to disclose.

References

Papers of special note have been highlighted as either of interest (•) or of considerable interest (••) to readers.

1. Colombel J-F, D'haens G, Lee W-J, et al. Outcomes and strategies to support a Treat-to-target approach in inflammatory bowel disease: a systematic review. J Crohn's Colitis. 2020;14(2):254–266. doi: 10.1093/ecco-jcc/jjz131
2. Shimoyama T, Yamamoto T, Yoshiyama S, et al. Leucine-Rich Alpha-2 glycoprotein is a reliable serum biomarker for evaluating clinical and endoscopic disease activity in inflammatory bowel disease. Inflamm Bowel Dis. 2023;29(9):1399–1408. doi: 10.1093/ibd/izac230
3. Verstockt B, Parkes M, Lee JC. How do we predict a Patient's disease course and whether they will respond to specific treatments? Gastroenterology. 2022;162(5):1383–1395. doi: 10.1053/j.gastro.2021.12.245
   - **A more detailed overview of biomarkers in IBD with their strengths and limitations.**
4. Noor NM, Verstockt B, Parkes M, et al. Personalised medicine in Crohn's disease. The Lancet Gastroenterol Hepatol. 2020;5(1):80–92. doi: 10.1016/S2468-1253(19)30340-1
5. Dulai PS, Boland BS, Singh S, et al. Development and validation of a scoring system to predict outcomes of vedolizumab treatment in patients with Crohn's disease. Gastroenterology. 2018;155(3):687–695.e10. doi: 10.1053/j.gastro.2018.05.039
6. Molnár T, Lakatos PL, Farkas K, et al. Predictors of relapse in patients with Crohn's disease in remission after 1 year of biological therapy. Alimentary Pharmacol Ther. 2013;37(2):225–233. doi: 10.1111/apt.12160
7. Billiet T, Papamichael K, de Bruyn M, et al. A matrix-based Model predicts primary response to infliximab in Crohn's disease. J Crohn's Colitis. 2015;9(12):1120–1126. doi: 10.1093/ecco-jcc/jjv156
8. Wolters FL, Russel, M. G., Sijbrandij, J. et al. Phenotype at diagnosis predicts recurrence rates in Crohn's disease. Gut. 2006;55(8):1124–1130. doi: 10.1136/gut.2005.084061
9. Beaugerie L, Seksik P, Nion-Larmurier I, et al. Predictors of Crohn's disease. Gastroenterology. 2006;130(3):650–656. doi: 10.1053/j.gastro.2005.12.019
10. Loly C, Belaiche J, Louis E. Predictors of severe Crohn's disease. Scand J Gastroenterol. 2008;43(8):948–954. doi: 10.1080/00365520801957149
11. Torres J, Caprioli F, Katsanos KH, et al. Predicting outcomes to optimize disease management in inflammatory bowel diseases. J Crohn's Colitis. 2016;10(12):1385–1394. doi: 10.1093/ecco-jcc/jjw116
12. Cosnes J, Bourrier A, Nion-Larmurier I, et al. Factors affecting outcomes in Crohn's disease over 15 years. Gut. 2012;61(8):1140–1145. doi: 10.1136/gutjnl-2011-301971
13. Solberg IC, Lygren I, Jahnsen J, et al. Clinical course during the first 10 years of ulcerative colitis: results from a population-based inception cohort (IBSEN study). Scand J Gastroenterol. 2009;44(4):431–440. doi: 10.1080/00365520802600961
14. Ananthakrishnan AN, Issa M, Beaulieu DB, et al. History of medical hospitalization predicts future need for colectomy in patients with ulcerative colitis. Inflamm Bowel Dis. 2009;15(2):176–181. doi: 10.1002/ibd.20639
15. Targownik LE, Singh H, Nugent Z, et al. The epidemiology of colectomy in ulcerative colitis: results from a population-based cohort. 2012;107:1228–1235. doi: 10.1038/ajg.2012.127
16. Hart AL, Lomer M, Verjee A, et al. What are the top 10 research questions in the treatment of inflammatory bowel disease? A priority setting partnership with the James Lind Alliance. J Crohn's Colitis. 2017;11(2):204–211. doi: 10.1093/ecco-jcc/jjw144
17. Sands BE, Feagan BG, Peyrin-Biroulet L, et al. Phase 2 trial of anti-TL1A monoclonal antibody tulisokibart for ulcerative colitis. N Engl J Med. 2024;391(12):1119–1129. doi: 10.1056/NEJMoa2314076
18. Kern SE. Why your new cancer biomarker may never work: recurrent patterns and remarkable diversity in biomarker failures. Cancer Res. 2012;72(23):6097–6101. doi: 10.1158/0008-5472.CAN-12-3232
   - **A good overview of challenges in biomarker development and common pitfalls.**

19. Subramanian J, Simon R. Overfitting in prediction models – is it a problem only in high dimensions? Contemp Clin Trials. 2013;36(2):636–641. doi: 10.1016/j.cct.2013.06.011
• **A good explanation of the issue of overfitting and why it commonly happens with -omic data from small numbers of patients.**
20. Ransohoff DF. Rules of evidence for cancer molecular-marker discovery and validation. Nat Rev Cancer. 2004;4(4):309–314. doi: 10.1038/nrc1322
21. Ferrante M, Henckaerts L, Joossens M, et al. New serological markers in inflammatory bowel disease are associated with complicated disease behaviour. Gut. 2007;56(10):1394–1403. doi: 10.1136/gut.2006.108043
22. Amre DK, Lu S-E, Costea F, et al. Utility of serological markers in predicting the early occurrence of complications and surgery in pediatric Crohn's disease patients. Am J Gastroenterol. 2006;101(3):645–652. doi: 10.1111/j.1572-0241.2006.00468.x
23. Seow CH, Stempak JM, Xu W, et al. Novel anti-glycan antibodies related to inflammatory bowel disease diagnosis and phenotype. Am J Gastroenterol. 2009;104(6):1426–1434. doi: 10.1038/ajg.2009.79
24. Israeli E, Grotto, I., Gilburd, B. et al. Anti-Saccharomyces cerevisiae and antineutrophil cytoplasmic antibodies as predictors of inflammatory bowel disease. Gut. 2005;54(9):1232–1236. doi: 10.1136/gut.2004.060228
25. Bodecker-Zingmark L, Widbom L, Hultdin J, et al. Anti-Saccharomyces cerevisiae antibodies are only modestly more common in subjects later developing Crohn's disease. Dig Dis Sci. 2023;68(2):608–615. doi: 10.1007/s10620-022-07630-5
26. Kuwada T, Shiokawa M, Kodama Y, et al. Identification of an anti–integrin αvβ6 autoantibody in patients with ulcerative colitis. Gastroenterology. 2021;160(7):2383–2394.e21. doi: 10.1053/j.gastro.2021.02.019
27. Kawashima K, Ishihara S, Yuki T, et al. Fecal calprotectin level correlated with both endoscopic severity and disease extent in ulcerative colitis. BMC Gastroenterol. 2016;16(1):47. doi: 10.1186/s12876-016-0462-z
28. Sonoyama H, Kawashima K, Ishihara S, et al. Capabilities of fecal calprotectin and blood biomarkers as surrogate endoscopic markers according to ulcerative colitis disease type. J Clin Biochem Nutr. 2019;64(3):265–270. doi: 10.3164/jcbn.18-92
29. Sakuraba A, Nemoto N, Hibi N, et al. Extent of disease affects the usefulness of fecal biomarkers in ulcerative colitis. BMC Gastroenterol. 2021;21(1):197. doi: 10.1186/s12876-021-01788-4
30. Singh S, Ananthakrishnan AN, Nguyen NH, et al. AGA clinical practice guideline on the role of Biomarkers for the management of ulcerative colitis. 2023;164:344–372. doi: 10.1053/j.gastro.2022.12.007
31. Gibson G. Going to the negative: genomics for optimized medical prescription. 2019;20:1–2. doi: 10.1038/s41576-018-0061-7
•• **An excellent summary of the relative importance of different biomarker performance characteristics in different clinical settings.**
32. Noor NM, Lee JC, Bond S, et al. A biomarker-stratified comparison of top-down versus accelerated step-up treatment strategies for patients with newly diagnosed Crohn's disease (PROFILE): a multicentre, open-label randomised controlled trial. The Lancet Gastroenterol Hepatol. 2024;9(5):415–427. doi: 10.1016/S2468-1253(24)00034-7
•• **The first biomarker-stratified trial in IBD, which highlighted the value of top-down therapy from diagnosis.**
33. Khan N, Vallarino C, Lissoos T, et al. Risk of infection and types of infection among elderly patients with inflammatory bowel disease: a retrospective database analysis. Inflamm Bowel Dis. 2020;26:462–468. doi: 10.1093/ibd/izz065
34. Kochar B, Cai W, Cagan A, et al. Pretreatment frailty is independently associated with increased risk of infections after immunosuppression in patients with inflammatory bowel diseases. Gastroenterology. 2020;158(8):2104–2111.e2. doi: 10.1053/j.gastro.2020.02.032
35. Yokoyama Y, Ohta Y, Ogasawara S, et al. The long-term effect of biologics in patients with ulcerative colitis emerging from a large Japanese cohort. Sci Rep. 2022;12(1):21060. doi: 10.1038/s41598-022-25218-x
36. Siegel CA, Hur C, Korzenik JR, et al. Risks and benefits of infliximab for the treatment of Crohn's disease. Clin Gastroenterol Hepatol. 2006;4(8):1017–1024; quiz 976. doi: 10.1016/j.cgh.2006.05.020
37. Turina MC, Yeremenko N, van Gaalen F, et al. Serum inflammatory biomarkers fail to identify early axial spondyloarthritis: results from the SpondyloArthritis caught early (SPACE) cohort. RMD Open. 2017;3(1):e000319. doi: 10.1136/rmdopen-2016-000319