Evaluating Discriminative Accuracy of Biomarkers in Relation to Binary Study Outcomes: First Validate, Then Celebrate?

Arno R. Bourgonje

1 Department of Gastroenterology and Hepatology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands

Correspondence:
Arno R. Bourgonje
Department of Gastroenterology and Hepatology, University Medical Center Groningen
Hanzeplein 1, 9713 GZ, Groningen, the Netherlands
Tel: +31 50 361 26 20
E-mail: a.r.bourgonje@umcg.nl

© The Author(s) 2022. Published by Oxford University Press on behalf of European Crohn’s and Colitis Organisation.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
Letter to the Editor

With interest I read the study by Swaminathan et al., in which the performance of faecal myeloperoxidase [fMPO] as biomarker of endoscopic disease activity was assessed in patients with inflammatory bowel disease [IBD]. Additionally, performance was assessed in comparison with C-reactive protein [CRP] and faecal calprotectin [fCal], being established biomarkers of disease activity. The study concluded that fMPO is an accurate biomarker of endoscopic activity in IBD and predicted a more complicated IBD course during follow-up.

Patients with IBD suffer from fluctuating disease activity that is difficult to detect and monitor. To avoid repeated endoscopies, other means to monitor disease activity are widely used, e.g. clinical indices, CRP, and fCal. However, these often demonstrate inconsistent associations and lack specificity, emphasizing the need for additional biomarkers capable of reflecting intestinal inflammation. To assess the performance of fMPO, fCal and CRP in predicting endoscopic disease activity, the authors leveraged receiver operating characteristics [ROC] statistics and ROC curves, the latter visualizing performance of a biomarker/model across different classification thresholds.

Although ROC curves are commonly used to assess biomarker performance, this assessment can be optimistically biased if based on a single evaluation of data. Therefore, testing the ability of a biomarker/model to generalize to new cases, i.e. a sample (“test sample”) independent of the sample used to predict the dependent variable (“training sample”), is recommended. Using all cases from the original analysis, instead of this train/test-approach, may result in overly optimistic estimates of biomarker performance. To overcome this issue, external validation of biomarker performance in independent cohorts is pivotal to determine generalizability. In addition, internal validation or cross-validation methods (e.g. leave-pair-out cross-validation [LPO], k-fold cross-validation or bootstrap resampling) can be used to generate more realistic estimates, yielding more variance to build broader confidence intervals. These methods split the original “training” dataset into multiple subsets, which are then iteratively used as “training” and “test” subsets, either randomly or not, and averaged to roughly indicate how biomarker performance is affected by changes in “training” data.
Aside from cross-validation, there are also accepted methods to effectively compare correlated AUROC estimates, e.g. DeLong’s test, and to quantify the maximized sensitivity/specificity balance for “optimal cut-off points”, e.g. Youden’s J-statistic. Although the presented results do hint to these methods, the authors do not clearly state what criteria were used to compare fMPO to fCal and CRP or to assess optimal cut-off points. Finally, standard errors or confidence intervals were lacking, which could facilitate readers to better judge biomarker performances.

The efforts of Swaminathan et al. in identifying novel biomarkers for disease activity are important and can only be applauded. However, validation - whether internally or externally performed, or both - remains critical to precisely determine the clinical utility of biomarkers.
Funding

No funding was received for this work.

Conflicts of Interest

ARB has no conflicts of interest to declare.
References


