

# Patient-reported outcome measures poorly correlate with objective inflammatory bowel disease activity measures: a systematic review

Xavier Calvet<sup>\*,1,2,3,4</sup>, Maria Giovanna Ferrario<sup>4,5</sup>, Vanessa Marfil<sup>4,6</sup>, Santos Armenteros<sup>†,5,6</sup>, Manuel Barreiro-de Acosta<sup>6,7</sup>

<sup>1</sup>Gastroenterology Unit, Parc Taulí University Hospital, Parc Taulí Research and Innovation Institute (I3PT-CERCA), Sabadell, Spain

<sup>2</sup>Department of Medicine, Autonomous University of Barcelona, Bellaterra, Spain

<sup>3</sup>CIBERehd, Carlos III Health Institute, Madrid, Spain

<sup>4</sup>Medical Statistics Consulting SL, Valencia, Spain

<sup>5</sup>Galapagos Biopharma Spain SL, Madrid, Spain

<sup>6</sup>Inflammatory Bowel Disease Unit, University Clinical Hospital of Santiago de Compostela, Santiago de Compostela, Spain

\*Corresponding author: Xavier Calvet, Gastroenterology Unit, Parc Taulí University Hospital and Research and Innovation Institute (I3PT-CERCA), 1 08208, Sabadell, Spain (xcalvet@tauli.cat).

Conference presentation: 19th Congress of the European Crohn's and Colitis Organisation (ECCO), February 21-24, 2024, Stockholm, Sweden (poster presentation)

<sup>†</sup>Former association; employee of Alfasigma S.p.A. at the time of publication.

## Abstract

**Background and Aims:** We investigated the correlations between patient-reported outcome measures (PROMs) and other measures of inflammatory bowel disease (IBD) activity.

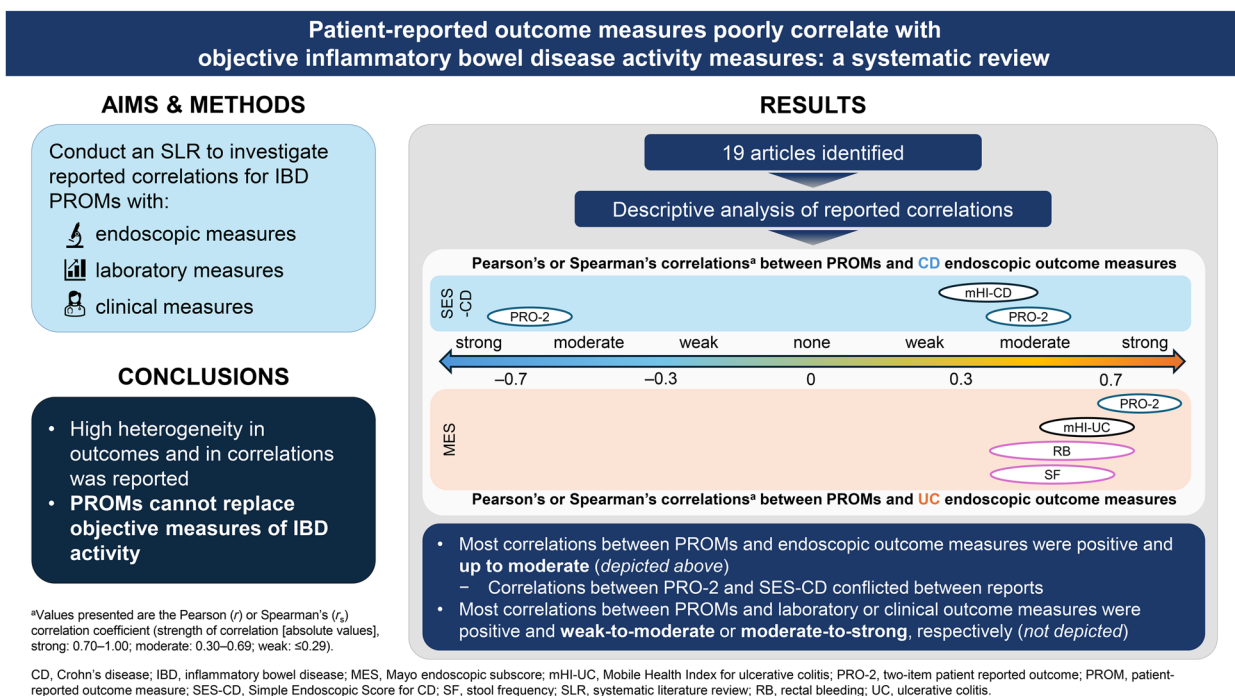
**Methods:** A systematic literature review was performed up to June 2022. Searches were conducted in PubMed, Scopus, and Web of Science. A descriptive analysis was performed. The search protocol was registered in PROSPERO (CRD42022383899).

**Results:** Nineteen studies assessed correlations between PROMs and clinical, endoscopic, and laboratory measures of disease activity in IBD. In Crohn's disease (CD), weak positive correlations were reported for PROMs (eg, the 2 item patient-reported outcome [PRO-2], mobile Health Index [mHI] for CD) and endoscopic scores, more often the Simple Endoscopic Score for CD (SES-CD). In ulcerative colitis (UC), PROMs like PRO-2, the Monitor IBD at Home rectal bleeding item, and the mHI showed weak-to-moderate correlations with the Mayo endoscopic subscore (MES). PROMs also demonstrated limited concordance with laboratory measures such as fecal calprotectin (FCP) and C-reactive protein (CRP) in both CD and UC. The substantial heterogeneity in study designs precluded a structured analysis.

**Conclusions:** Although current PROMs offer valuable complementary insights into IBD control from the patient's perspective, they cannot replace objective measures of IBD activity. Future research should focus on refining PROMs and generating composite indices to improve their accuracy and usefulness.

**Keywords:** PROM; Crohn's disease; ulcerative colitis.

## Graphical abstract



## 1. Introduction

Inflammatory bowel disease (IBD) is characterized by flares of bowel inflammation leading to heterogeneous symptoms that follow a chronic and recurrent course.<sup>1</sup> The term IBD includes 2 main diseases: Crohn's Disease (CD) and ulcerative colitis (UC).<sup>2,3</sup>

In IBD management, patient-centered care is increasingly important. Healthcare professionals are no longer requested to direct patient care; by contrast, they should recognize patients' autonomy by providing accurate information allowing patients to make their own therapeutic decisions tailored to individual needs and preferences.<sup>4,5</sup> This approach improves clinical results<sup>6–8</sup> and increases patient satisfaction and treatment adherence.<sup>9–11</sup>

Patient-centered care includes measuring IBD activity from the patient's point of view using patient-reported outcomes (PROs). PROs are defined as any outcome directly reported by the patient without the interpretation of a clinician. They capture information about the patient's health, quality of life, or functional status associated with healthcare or treatment.<sup>12</sup> Specific, validated and standardized, self-completed instruments to assess PROs are termed patient-reported outcome measures (PROMs). PROMs are intended to support clinical decisions, to complement other measurements of disease activity, to compare treatment outcomes, and to evaluate medical policies and practices.<sup>13</sup> PROMs are increasingly used in IBD to help to monitor disease, in both clinical trials and clinical practice.<sup>14,15</sup> They are also used to identify which outcomes matter most to each patient, which is crucial in providing patient-centered care.<sup>16</sup> Nonetheless, there is some controversy regarding the use of PROMs in the management of IBD and their clinical validity. The main concerns are subjectivity and

variability of self-reported data and the potentially limited sensitivity of PROMs in the detection of subtle changes in disease activity.<sup>17</sup>

In fact, imaging techniques, especially endoscopy, are accepted as the most reliable and objective tools to evaluate IBD activity because they allow direct evaluation of mucosal inflammation. In IBD, endoscopy holds a key position in diagnosing the severity and extent of mucosal lesions and/or assessing mucosal healing and, therefore, guiding therapeutic decisions. Simple Endoscopic Score for Crohn's Disease (SES-CD)<sup>18</sup> and the Crohn's Disease Endoscopic Index of Severity (CDEIS)<sup>19</sup> are common endoscopic indices used in CD, while Mayo endoscopic subscore (MES) is the index most frequently used in UC.<sup>20</sup> Endoscopy, however, is an invasive and expensive technique. Therefore, patient discomfort and risks related to endoscopy prevent its use for regular routine follow-up. For this reason, clinicians rely on a series of indirect measurements of disease activity.<sup>21</sup> These measurements include inflammatory biomarkers (eg, C-reactive protein [CRP] and fecal calprotectin [FCP]) and activity indices assessed by the clinician (eg, the Harvey-Bradshaw Index [HBI],<sup>22</sup> the CD Activity Index [CDAI],<sup>23</sup> and the Simple Clinical Colitis Activity Index [SCCAI]<sup>24</sup>). In conclusion, while endoscopic indices allow direct evaluation of mucosal inflammation and are accepted as the gold standard for evaluating IBD activity, their limitations require noninvasive tools to monitor CD disease activity. Between these tools, PROMs are gaining popularity, although their usefulness to predict endoscopic activity is controversial.

The aim of this study was to assess the reliability of IBD-specific PROMs in measuring disease activity by analyzing the correlation and/or association of these PROMs with validated clinical, laboratory, or endoscopic IBD activity measures.

## 2. Materials and methods

### 2.1 Literature search strategy

A systematic literature review was carried out to identify relevant studies focusing on the development, validation, or utilization of PROMs designed to monitor clinical manifestations in adult patients with IBD. The search was performed in PubMed, Scopus, and the databases included in the Web of Science, up to June 30, 2022. Also, multiple searches were performed in Google Scholar to confirm the searches were exhaustive. The search strategy utilized a Boolean combination of terms as detailed in [Supplementary Table 1](#). Specific search strategies tailored to each database are provided in [Supplementary Table 2](#).

### 2.2 Study selection

Following the initial search, articles were screened for eligibility according to predefined criteria. Studies were excluded if (1) the PROM analyzed was not validated in the IBD population, (2) if the PROM was not available in English or Spanish, (3) if the PROM was not used to measure study outcomes (ie, used solely for baseline patient classification), (4) if they were not original research articles, (5) if the PROM measured aspects unrelated to disease activity (eg, quality of life, or social or work performance), (6) if the PROM was used for pediatric populations, (7) if the PROM was designed for a specific condition that is not applicable for most patients with IBD (eg, PROM designed for ileorectal pouch evaluation or fecal incontinence), or (8) if the PROM was a generic instrument not specifically targeting IBD patients.

The screening, selection, and inclusion processes were carried out by 3 investigators, with a fourth investigator overseeing the selection process. Any disagreements or uncertainties regarding inclusion were discussed and resolved through consensus, or by consulting the fourth reviewer when necessary.

### 2.3 Data extraction

Data extraction was performed by the same 3 investigators, using a standardized extraction table in Excel format. The extracted data included details on study characteristics, participant demographics, PROM characteristics, and outcomes related to PROM validation and utilization.

### 2.4 Descriptive analysis of correlation and predictive performance

The reported correlations between PROMs and other IBD activity measures were identified from each selected article and reported. Correlations were classified based on the absolute Pearson or Spearman's correlation coefficients as follows: strong (0.70-1.00), moderate (0.30-0.69), and weak (below 0.30).

Also, predictive performance indicators were extracted and reported. Measures included: sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). In some articles, the area under the receiver operating characteristic curve (AUROC) was used for comparing each PROM with other measures of disease activity, such as the proportion of patients in endoscopic remission. An AUROC of 0.50 indicated no discrimination, 0.60-0.70 was considered acceptable, 0.80-0.90 was considered good, and 0.90 or

greater was excellent. Additionally, inter-observer agreement was reported in some studies using Cohen's kappa. The interpretation of kappa values was as follows: less than 0 (no agreement beyond chance), 0-0.20 (slight agreement), 0.21-0.40 (fair agreement), 0.41-0.60 (moderate agreement), 0.61-0.80 (substantial agreement), and 0.81-1.00 (almost perfect agreement).

### 2.5 Quality assessment

The critical appraisal skills programme (CASP) tool was employed to assess the quality of evidence presented in each included study.<sup>25</sup> This tool facilitated the evaluation of study design, methodology, and potential biases.

### 2.6 Protocol registration and reporting guidelines

The search protocol was registered in PROSPERO (CRD42022383899).<sup>26</sup> The systematic literature review adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines to ensure transparency and completeness in reporting the review process.<sup>27</sup>

## 3. Results

### 3.1 Selection process and study population description

A summary of the systematic search strategy and article selection is shown in [Figure 1](#). Initially, 2099 articles were found by database and manual searches, while an additional article, not previously identified, was manually added when performing free searches in Google Scholar. After elimination of duplicates, 1385 unique articles were identified. Subsequent screening of titles, abstracts and full texts led to the selection of 76 articles that reported the use of PROMs for measuring IBD activity. Reasons for article exclusion are reported in [Supplementary Table 3](#). Of the 76 articles selected, 22 articles were independently identified by 3 reviewers in parallel. Three of these 22 articles were subsequently excluded following assessment by the fourth reviewer, because they evaluated the long-term prognostic value of PROM without a direct comparison with endoscopic, clinical, or biological activity. The remaining 19 articles fulfilled inclusion criteria because they analyzed correlations between PROMs and other disease activity measures, and/or the accuracy of PROMs in predicting remission or healing measures and were thus included in the final selection.<sup>28-46</sup>

### 3.2 Characteristics of included articles

[Supplementary Table 4](#) summarizes the design features of the 19 included articles and relevant validation studies, including those tools that were not validated within the same study.<sup>28-46</sup> Among these, 4 studies included a CD population only, 5 included a UC population only, while the remaining 10 studies included both CD and UC patients. Fifteen studies had a real-world design (10 prospective cohort studies; 5 cross-sectional studies) and 4 were *post hoc* analyses of randomized controlled trials or real-world studies.

Evaluation of study quality using the CASP checklist revealed low-to-moderate quality across most studies, as depicted in [Supplementary Table 5](#).

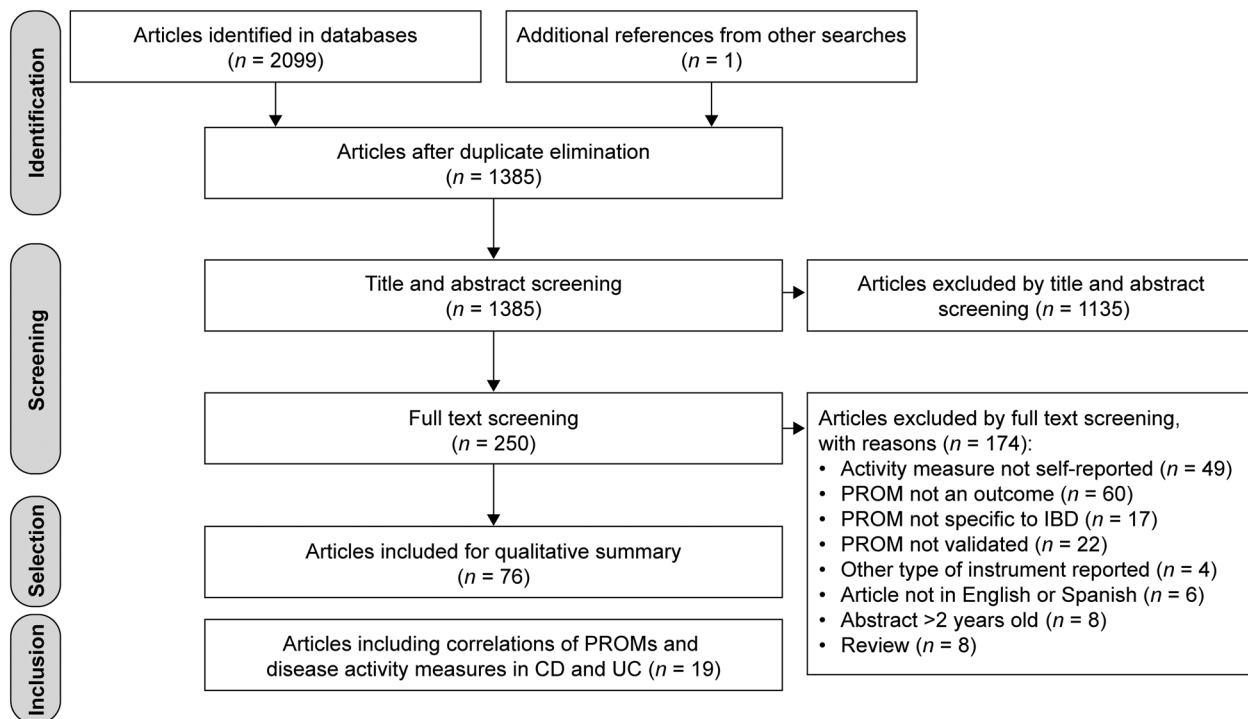


**Figure 1.** PRISMA flowchart: results from systematic search and article selection. CD, Crohn's disease; IBD, inflammatory bowel disease; PROM, patient-reported outcome measure; UC, ulcerative colitis.

### 3.3 Characteristics of included PROMs

Among the 19 included articles, 16 different validated IBD-specific PROMs were identified (Supplementary Table 6). Seven of these were utilized across IBD populations, whereas 3 and 6 were specific for CD and UC, respectively. The PROMs included the assessment of various signs and symptoms of IBD activity, of which the most reported signs and symptoms were stool frequency (SF; day and night), abdominal pain (AP) or discomfort, frequency of loose stools/diarrhea, rectal bleeding (RB), urgency of defecation, general well-being, and fatigue. Additional aspects assessed were anal fissures, active fistulas and/or abscesses, and extracolonic manifestations, including arthralgia, red or swollen joints, uveitis, erythema nodosum, pyoderma gangrenosum, and aphthous ulcers. Additionally, certain PROMs also included aspects of health-related quality of life, such as the impact on emotional components and daily activities.

### 3.4 Concordance between PROMs and endoscopic measures of disease activity

For CD, 6 studies reported data on the correlation between PROMs and endoscopic activity measures and/or the diagnostic accuracy of PROMs for detecting disease activity (Table 1).<sup>28–33</sup> A single study published by Dragasevic et al. reported a strong inverse correlation between the PRO-2 and the SES-CD in 63 patients with CD and colonic involvement.<sup>29</sup> Most of the remaining studies described a weak-to-moderate positive correlation between PROMs and endoscopic scores, ranging from 0.29<sup>32</sup> to 0.36.<sup>30</sup> Sensitivity to detect endoscopic activity ranged from 23%<sup>29</sup> to 68%<sup>28</sup>; specificity ranged from 81%<sup>28</sup> to 98%.<sup>29</sup> Finally, AUROCs ranged from 0.54 for the HBI to 0.70 for the patient-reported opinion on whether IBD is active.<sup>33</sup> In addition, Zittan et al. demonstrated that the combination of HBI

with other outcome measures was not a useful surrogate to detect endoscopic remission.<sup>33</sup>

For UC, 6 studies reported data on the correlation between a PROM and endoscopic activity measures, and/or the diagnostic accuracy of PROMs for detecting endoscopic activity.<sup>28,29,32,38,39,41</sup> Correlations between PROMs and the MES were positive and moderate to strong, ranging from 0.34<sup>41</sup> to 0.84.<sup>29</sup> PROMs generally showed a moderate-to-high sensitivity and specificity to detect endoscopic activity: sensitivity ranged from 45%<sup>29</sup> to 98%<sup>39</sup> and specificity from 45%<sup>39</sup> to 99%.<sup>29</sup> Finally, AUROCs ranged from 0.82<sup>32</sup> to 0.93.<sup>39</sup> Regarding individual items, RB subscores generally demonstrated a noticeably higher accuracy when compared to SF.<sup>38,39,41</sup>

### 3.5 Concordance between PROMs and laboratory measures of disease activity

Six studies evaluated the concordance between PROMs and different laboratory measures,<sup>35,36,39,40,43,44</sup> mainly CRP and FCP (Table 2). Concordance was weak-to-moderate for PROMs and biological markers of activity. In CD, neither the IBD-10 nor the HBI effectively distinguished patients with increased FCP levels (>250 µg/g) from those with normal levels.<sup>43</sup> The Numerical Rating Scale (NRS) did not correlate with CRP, erythrocyte sedimentation rate (ESR), or FCP (Table 2).<sup>44</sup> Similar limitations were observed in UC patients. Neither the IBD-10 nor the SCCAI satisfactorily discriminated patients with normal vs. elevated FCP levels.<sup>43</sup> CRP showed fair inter-rater agreement with the Patient Simple Clinical Colitis Activity Index (P-SCCAI) and the SCCAI, while a substantial agreement was found between the Physician Global Assessment (PGA) and both P-SCCAI and SCCAI.<sup>35</sup> FCP correlated moderately with the 8-item questionnaire, IBD-Control-8, and the VAS, IBD-Control-VAS (Table 2).<sup>36</sup>

**Table 1.** Descriptive analysis of articles evaluating the correlation and/or association/predictive accuracy of PROMs with endoscopic scores in IBD

| CD                                     |                        |                     |                                                                                                                                                                                                          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|----------------------------------------|------------------------|---------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Article                                | Number of patients (N) | PROM                | Correlation/association/predictive accuracy measures                                                                                                                                                     | Correlation with SES-CD <sup>a</sup> Association and predictive accuracy with respect to SES-CD                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| de Jong MJ et al. 2019 <sup>28</sup>   | 135                    | MIAH <sup>c</sup>   | Validation phase: MIAH-CD questionnaires' sensitivity, specificity, PPV, and NPV                                                                                                                         | <ul style="list-style-type: none"> <li>• Sensitivity = 67.6%</li> <li>• Specificity = 81.3%</li> <li>• PPV = 69.3% (with cut-off at 3.6)</li> <li>• NPV = 80.0% (with cut-off at 3.6)</li> </ul>                                                                                                                                                                                                                                                                                                                                                                       |
| Dragasevic S et al. 2020 <sup>29</sup> | 63                     | PRO-2 <sup>b</sup>  | Spearman's correlation coefficients. Sensitivity and specificity were also measured against endoscopy                                                                                                    | <ul style="list-style-type: none"> <li>• <math>r_s = -0.67</math>, <math>p &lt; 0.05</math></li> <li>• Sensitivity = 22.5% (95% CI: 10.8-38.4%)</li> <li>• Specificity = 97.7% (95% CI: 71.9-98.9%)</li> <li>• PPV = 81.8% (95% CI: 51.5-95.0%)</li> <li>• NPV = 40.4% (95% CI: 35.5-45.5%)</li> <li>• NA</li> </ul>                                                                                                                                                                                                                                                   |
| Lewis JD et al. 2020 <sup>30</sup>     | 353                    | PRO-2 <sup>b</sup>  | Pearson correlation coefficients                                                                                                                                                                         | <ul style="list-style-type: none"> <li>• PRO-2: <math>r = 0.36</math> (95% CI: 0.27-0.45), <math>p &lt; 0.001</math></li> <li>• NA</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                          |
| Morris MW et al. 2018 <sup>31</sup>    | 164                    | PRO-2<br>PRO-3      | AUROC curves to evaluate active disease using SES-CD at 2 levels: (SES-CD $\geq 7$ and SES-CD $\geq 2$ for moderate disease). Sensitivity, specificity                                                   | <ul style="list-style-type: none"> <li>• PRO-3: SES-CD <math>\geq 7</math>; AUROC = 0.56 (95% CI: 0.55-0.57); sensitivity = 63%; specificity = 88%</li> <li>• PRO-2: SES-CD <math>\geq 7</math>; AUROC = 0.56 (95% CI: 0.55-0.57); sensitivity = 61%; specificity = 55%</li> <li>• PRO-3: SES-CD <math>\geq 2</math>; AUROC = 0.66 (95% CI: 0.65-0.67); sensitivity = 34%; specificity = 94%</li> <li>• PRO-2: SES-CD <math>\geq 2</math>; AUROC = 0.58 (95% CI: 0.57-0.59); sensitivity = 38%; specificity = 85%</li> <li>• AUROC: 0.632 (CI not provided)</li> </ul> |
| Van Deen WK et al. 2016 <sup>32</sup>  | 301                    | mHI-CD <sup>d</sup> | Spearman's correlation coefficient, ROC curves                                                                                                                                                           | <ul style="list-style-type: none"> <li>• <math>r_s = 0.29</math>, <math>p &lt; 0.039</math></li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| Zittan E et al. 2017 <sup>33</sup>     | 88                     | HBI-PRO<br>PROp     | ROC curves                                                                                                                                                                                               | <ul style="list-style-type: none"> <li>• NA</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| UC                                     |                        |                     |                                                                                                                                                                                                          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| Article                                | Number of patients (N) | PROM                | Correlation/association/predictive accuracy measures                                                                                                                                                     | Correlation with MES <sup>a</sup> Association and predictive accuracy for remission (MES=0 or 1)                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| Colombel JF et al. 2017 <sup>38</sup>  | 103                    | PRO-2               | PPV, NPV, sensitivity, and specificity of RB and SF subscores for endoscopic remission (MES=0/1) were calculated for each definition separately for RB subscore=0, SF subscore=0, and RB + SF subscore=0 | <ul style="list-style-type: none"> <li>• NA</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| (Continued)                            |                        |                     |                                                                                                                                                                                                          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |

Table 1. Continued.

| UC                                     |                                  |                                 |                                                                                                                                                                                                                                                                |                                                                                                                                                                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|----------------------------------------|----------------------------------|---------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Article                                | Number of patients (N)           | PROM                            | Correlation/association/predictive accuracy measures                                                                                                                                                                                                           | Correlation with MES <sup>a</sup>                                                                                                                               | Association and predictive accuracy for remission (MES=0 or 1)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| de Jong MJ et al. 2019 <sup>28</sup>   | Development: 80, Validation: 131 | SF and RB (MIAH- <sup>c</sup> ) | Spearman correlation coefficients were calculated between individual items and endoscopic disease activity in the development phase. Sensitivity, specificity, PPV, and NPV for MIAH-UC were calculated in the validation phase, defining remission as MES 0/1 | <ul style="list-style-type: none"><li><math>r_s</math> (RB)=0.58, <math>p&lt;0.001</math></li><li><math>r_s</math> (SF)=0.51, <math>p&lt;0.001</math></li></ul> | <ul style="list-style-type: none"><li>Sensitivity = 68.2%</li><li>Specificity = 80.5%</li><li>PPV = 63.8% (with cut-off at 3.5)</li><li>NPV = 83.3% (with cut-off at 3.5)</li></ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| Dragasevic S et al. 2020 <sup>29</sup> | 96                               | PRO-2 <sup>b</sup>              | Spearman's correlation coefficient. Sensitivity, specificity, PPV, and NPV for endoscopic remission                                                                                                                                                            | <ul style="list-style-type: none"><li><math>r_s</math> =0.84, <math>p&lt;0.05</math></li></ul>                                                                  | <ul style="list-style-type: none"><li>Specificity = 99%</li><li>Sensitivity = 45%</li><li>PPV = 96%</li><li>NPV = 59%</li></ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
| Golovics PA et al. 2021 <sup>39</sup>  | 171                              | PRO-2                           | AUROC, sensitivity, specificity, PPV, NPV to assess remission (MSC=0/1)                                                                                                                                                                                        | <ul style="list-style-type: none"><li>NA</li></ul>                                                                                                              | AUROC: <ul style="list-style-type: none"><li>PRO-2 =0.868</li><li>SF=0.842</li><li>RB =0.814</li><li>pMCS=0.927</li></ul> PRO SF= 0 versus MES ≤1 <ul style="list-style-type: none"><li>Sensitivity = 89.0%</li><li>Specificity = 74.4%</li><li>PPV = 91.1%</li><li>NPV = 69.6%</li></ul> PRO RB= 0 versus MES= 0 <ul style="list-style-type: none"><li>Sensitivity = 97.9%</li><li>Specificity = 45.2%</li><li>PPV = 70.4%</li><li>NPV = 94.3%</li></ul> PRO RB= 0 versus MES ≤1 <ul style="list-style-type: none"><li>Sensitivity = 95.3%</li><li>Specificity = 67.4%</li><li>PPV = 89.6%</li><li>NPV = 82.9%</li></ul> PRO-2 remission <ul style="list-style-type: none"><li>Sensitivity = 92.1%</li><li>Specificity = 76.7%</li><li>PPV = 92.1%</li><li>NPV = 76.7%</li></ul> pMCS <ul style="list-style-type: none"><li>Sensitivity = 97.6%</li><li>Specificity = 74.4%</li><li>PPV = 91.9%</li><li>NPV = 91.4%</li></ul> |

(Continued)

(Continued)

**Table 1.** Continued.

| UC                                    |                        |                     |                                                                                                                                                                                                                                          |                                                                                                                                                                                                                     |                                                                                                                                                                                                                                                  |
|---------------------------------------|------------------------|---------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Article                               | Number of patients (N) | PROM                | Correlation/association/predictive accuracy measures                                                                                                                                                                                     | Correlation with MES <sup>a</sup>                                                                                                                                                                                   | Association and predictive accuracy for remission (MES=0 or 1)                                                                                                                                                                                   |
| Ma C et al. 2020 <sup>41</sup>        | 817                    | PRO-2 <sup>b</sup>  | Spearman's rank correlation coefficient with 95% CI. Sensitivity, specificity, PPV, and NPV of RB and SF for endoscopic remission (MES = 0/1) at week 8 and week 38 separately for PROs with RB = 0, SF = 0, and RB + SF = 0 as criteria | <ul style="list-style-type: none"><li>• <math>r_s</math> (RB) = 0.39 (95% CI: 0.32-0.45), <math>p &lt; 0.001</math></li><li>• <math>r_s</math> (SF) = 0.34 (95% CI: 0.27-0.40), <math>p &lt; 0.001</math></li></ul> | Remission (MES = 0 or 1):<br>RB+SF (week 38): <ul style="list-style-type: none"><li>• Sensitivity = 59.5%</li><li>• Specificity = 82.2%</li><li>• PPV = 72.7%</li><li>• NPV = 71.7%</li></ul>                                                    |
| Van Deen WK et al. 2016 <sup>32</sup> | 265                    | mHI-UC <sup>d</sup> | Spearman's correlation coefficient, AUROC, sensitivity, specificity, PPV, and NPV to predict endoscopic remission (MES = 0/1). ROC analysis                                                                                              | <ul style="list-style-type: none"><li>• <math>r_s = 0.60</math>, <math>p &lt; 0.0001</math></li></ul>                                                                                                               | <ul style="list-style-type: none"><li>• AUC = 0.82 (with a 3.2 cut-off)</li><li>• Sensitivity = 85% (with a 3.2 cut-off)</li><li>• Specificity = 80% (with a 3.2 cut-off)</li><li>• NPV = 95%</li><li>• PPV = 55% (with a 3.2 cut-off)</li></ul> |

<sup>a</sup>The correlations provided were performed using an MES of 0 or 1 as the definition for remission. Scales vary across the studies that use different coefficients and should not be compared directly. Correlations were considered strong when the absolute Pearson ( $r$ ) or Spearman's ( $r_s$ ) correlation coefficient lay between 0.70 and 1.00, moderate when the coefficient lay between 0.30 and 0.69, and weak when the coefficient was less than or equal to 0.29. <sup>b</sup>PRO-2 was based on RB and SF for UC and abdominal pain and number of liquid or very loose stools and abdominal pain for CD. <sup>c</sup>The MIAH questionnaire assessed IBD-specific items of abdominal pain, fatigue, fever, frequency of liquid or soft stools, general well-being, mucus loss, patient-reported disease activity, perceived stress, RB, SF (day), SF (night), urgency of defecation, and weight loss. <sup>d</sup>The mHI-UC and mHI-UC assessed abdominal pain, anorexia, disease activity, fever, general well-being, nausea/vomiting, SF stool consistency, RB, and urgency. Abbreviations: CD, Crohn's disease; CI, confidence interval; IBD, inflammatory bowel disease; MES, Mayo endoscopic subscore; mHI-UC, mobile Health Index for UC; mHI-UC, mobile Health Index for UC; MIAH, Monitor IBD at Home; pMCS, partial Mayo Clinic Score; PRO-2, 2 item patient-reported outcome; PRO, patient-reported opinion on whether CD is active; PROM, patient-reported outcome measure;  $r$ , Pearson correlation coefficient; RB, rectal bleeding;  $r_s$ , Spearman's correlation coefficient; SES-CD, Simple Endoscopic Score for CD; SF, stool frequency; UC, ulcerative colitis.

**Table 2.** Concordance between PROMs and clinical and/or laboratory measures of disease activity.

| CD                                              |                             |                                         |                                                                                     |                                                                                                                                                                                                                                                                  |                                                                                                                                                                                                                                                                                                                                                                                                                   |                                                                                                                                                                                                                                     |
|-------------------------------------------------|-----------------------------|-----------------------------------------|-------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Article                                         | Number of patients (N)      | PROM                                    | Disease activity measure                                                            | Correlation/association/predictive accuracy measures                                                                                                                                                                                                             | Correlation                                                                                                                                                                                                                                                                                                                                                                                                       | Association and predictive accuracy                                                                                                                                                                                                 |
| Bennebroek Evertsz KA et al. 2013 <sup>34</sup> | 181                         | P-HBI                                   | HBI                                                                                 | Agreement between HBI and P-HBI was calculated with Spearman's $r$ . Agreement regarding active disease versus remission was calculated by percent agreement and Cohen's $\kappa$ . Sensitivity, specificity, PPV, and NPV were calculated from the article data | <ul style="list-style-type: none"> <li><math>r_s = 0.82</math></li> </ul>                                                                                                                                                                                                                                                                                                                                         | <ul style="list-style-type: none"> <li>Cohen's <math>\kappa = 0.52</math></li> <li>Sensitivity = 74%</li> <li>Specificity = 91%</li> <li>PPV = 96%</li> <li>NPV = 53%</li> </ul>                                                    |
| Bodger K et al. 2013 <sup>36</sup>              | 160                         | IBD-Control                             | HBI Physician global assessment (PGA). Composite evaluation for "quiescent disease" | Correlation was expressed as Spearman's $r_s$ values. A negative correlation was expected (the higher the IBD-c, the lower the symptoms score). AUROC, sensitivity, specificity, PPV, and NPV to predict "quiescent diseases" pooled for UC and CD               | IBD-Control-8: <ul style="list-style-type: none"> <li>vs. HBI: <math>r_s = -0.68</math>, <math>p &lt; 0.01</math></li> <li>vs. PGA: <math>r_s = -0.45</math>; <math>p &lt; 0.01</math></li> </ul> IBD-Control-VAS: <ul style="list-style-type: none"> <li>vs. HBI: <math>r_s = -0.60</math>, <math>p &lt; 0.01</math></li> <li>vs. PGA: <math>r_s = -0.47</math>, <math>p &lt; 0.01</math></li> <li>NA</li> </ul> | <ul style="list-style-type: none"> <li>NA</li> </ul>                                                                                                                                                                                |
| Clara I et al. 2009 <sup>37</sup>               | 184                         | MIBDI                                   | HBI                                                                                 | Sensitivity and specificity for detecting activity (HBI $\geq 4$ ) were provided at 3 time periods (0, 12, and 24 months). PPV and NPV were calculated using the data provided in the study                                                                      | <ul style="list-style-type: none"> <li>NA</li> </ul>                                                                                                                                                                                                                                                                                                                                                              | At 0, 12, and 24 months, respectively: <ul style="list-style-type: none"> <li>Sensitivity = 88%, 85%, and 84%</li> <li>Specificity = 51%, 57%, and 66%</li> <li>PPV = 78%, 75%, and 77%</li> <li>NPV = 67%, 72%, and 75%</li> </ul> |
| Sexton KA et al. 2019 <sup>42</sup>             | 124                         | IBDSI (long form, LF) Self-reported HBI | Nurse/clinician-administered HBI, PGA                                               | Pearson correlation analyses, AUROC. Sensitivity, specificity, PPV, and NPV using HBI $\geq 5$ or PGA                                                                                                                                                            | <ul style="list-style-type: none"> <li>IBDSI (LF) vs. PI: <math>r = 0.86</math></li> <li>IBDSI (LF) vs. PGA: <math>r = 0.70</math></li> <li>Self-reported HBI vs. nurse/clinician administered HBI: <math>r = 0.86</math></li> </ul>                                                                                                                                                                              | <ul style="list-style-type: none"> <li>IBDSI (LF) vs. HBI</li> <li>AUROC = 0.93</li> <li>Sensitivity = 84%</li> <li>Specificity = 89%</li> </ul>                                                                                    |
| Subramanian S et al. 2016 <sup>43</sup>         | 142                         | IBD-10 NRS                              | HBI CRP FCP                                                                         | Spearman's rank correlation coefficients. AUROC. Sensitivity, specificity for HBI-defined remission ( $\leq 4$ ) Mann-Whitney U test for comparing values of IBD-10 between patients with CRP $> 5$ vs. normal and FC $> 250$ vs. normal                         | <ul style="list-style-type: none"> <li>IBD-10 vs. HBI: <math>r_s = -0.69</math>, <math>p &lt; 0.001</math></li> <li>IBD-10 scores were similar in patients with normal vs. elevated FC and normal vs. elevated CPR</li> </ul>                                                                                                                                                                                     | <ul style="list-style-type: none"> <li>AUROC = 0.86 (95% CI: 0.81-0.91; <math>p = 0.035</math>)</li> <li>Sensitivity = 95% (95% CI: 86-94)</li> <li>Specificity = 75% (95% CI: 67-82)</li> </ul>                                    |
| Surti B et al. 2013 <sup>44</sup>               | Study A: 46<br>Study B: 135 | IBD NRS                                 | CDAI HBI CRP ESR FCP                                                                | Pearson's correlation coefficient                                                                                                                                                                                                                                | Study A: <ul style="list-style-type: none"> <li>vs. CDAI: <math>r = 0.59</math>, <math>p &lt; 0.0001</math></li> <li>vs. HBI: <math>r = 0.32</math>, <math>p &lt; 0.0001</math></li> </ul> Study B: <ul style="list-style-type: none"> <li>vs. HBI: <math>r = 0.25</math>, <math>p &lt; 0.0001</math></li> <li>vs. CRP no significant correlation</li> </ul>                                                      | <ul style="list-style-type: none"> <li>NA</li> </ul>                                                                                                                                                                                |
| Tow KE et al. 2019 <sup>45</sup>                | 48                          | PRO-2, PRO-3                            | MHBI                                                                                | Spearman's rho                                                                                                                                                                                                                                                   | <ul style="list-style-type: none"> <li>PRO-2: <math>r = 0.95</math>, <math>p &lt; 0.001</math></li> <li>PRO-3-CD: <math>r = 0.88</math>, <math>p &lt; 0.001</math></li> <li>PRO-2-CDSR: <math>r = 0.81</math>, <math>p &lt; 0.001</math></li> <li>PRO-3-CDSR: <math>r = 0.79</math>, <math>p &lt; 0.001</math></li> </ul>                                                                                         | <ul style="list-style-type: none"> <li>NA</li> </ul>                                                                                                                                                                                |

(Continued)

Table 2. Continued.

| Article                                        | Number of patients (N) | PROM        | Disease activity measure                            | Correlation/association/predictive accuracy measures                                                                                                                                                                                                                                                                    | Correlation                                                                                                                                                                                                                                                                                                                                                    | Association and predictive accuracy                                                                                                                                                                                                                                                                                                                         |
|------------------------------------------------|------------------------|-------------|-----------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Van Deen WK et al. 2016 <sup>32</sup>          | 110                    | mHI-CD      | HBI                                                 | Spearman's correlation coefficient, AUROC analysis for detecting active disease (HBI >4)                                                                                                                                                                                                                                | • $r_s = 0.75, p < 0.0001$                                                                                                                                                                                                                                                                                                                                     | <ul style="list-style-type: none"> <li>• AUROC = 0.90</li> <li>• Sensitivity = 94%</li> <li>• Specificity = 67%</li> </ul>                                                                                                                                                                                                                                  |
| Vicente Lidón et al. 2021 <sup>46</sup>        | 177                    | IBD-Control | HBI<br>Composite evaluation for "quiescent disease" | Spearman's correlation coefficients. AUROC. Sensitivity and specificity for "quiescent disease" (data pooled, separate data for CD and UC were not provided)                                                                                                                                                            | <ul style="list-style-type: none"> <li>• Moderate negative correlation with the HBI: <math>r_s = -0.524, p &lt; 0.001</math></li> </ul>                                                                                                                                                                                                                        | <ul style="list-style-type: none"> <li>• AUROC for "quiescent disease" = 0.864 (95% CI: 0.811-0.916)</li> <li>• Sensitivity = 70%</li> <li>• Specificity = 90%</li> </ul>                                                                                                                                                                                   |
| UC                                             |                        |             |                                                     |                                                                                                                                                                                                                                                                                                                         |                                                                                                                                                                                                                                                                                                                                                                |                                                                                                                                                                                                                                                                                                                                                             |
| Article                                        | Number of patients (N) | PROM        | Disease activity measure                            | Correlation/association/predictive accuracy measures                                                                                                                                                                                                                                                                    | Correlation                                                                                                                                                                                                                                                                                                                                                    | Association and predictive accuracy                                                                                                                                                                                                                                                                                                                         |
| Bennebroek Evertsz F et al. 2013 <sup>35</sup> | 149                    | P-SCCAI     | SCCAI<br>CRP<br>PGA of disease activity             | Agreement between SCCAI and P-SCCAI was calculated using Spearman's rho. Agreement between P-SCCAI and SCCAI and between CRP and PGA vs. P-SCCAI to detect remission (SCCAI >5) was calculated by percent agreement and Cohen's $\kappa$ . Sensitivity, specificity, PPV, and NPV were calculated from the article data | <ul style="list-style-type: none"> <li>• P-SCCAI vs. SCCAI: <math>r_s = 0.79</math></li> </ul>                                                                                                                                                                                                                                                                 | <ul style="list-style-type: none"> <li>• P-SCCAI vs. SCCAI: 87% agreement (<math>\kappa = 0.66</math>)</li> <li>• Sensitivity = 91%</li> <li>• Specificity = 79%</li> <li>• PPV = 94%</li> <li>• NPV = 69%</li> <li>• P-SCCAI vs. CRP: <math>\kappa = 0.32, p = 0.04</math></li> <li>• P-SCCAI vs. PGA: <math>\kappa = 0.73, p &lt; 0.001</math></li> </ul> |
| Bodger K et al. 2014 <sup>36</sup>             | 139                    | IBD-Control | SCCAI                                               | Correlation was expressed as Spearman's $r_s$ values. Negative correlation expected (the higher the IBD-c, the lower the symptoms score). AUROC, sensitivity, specificity, PPV, and NPV to predict "quiescent diseases" pooled for UC and CD                                                                            | <ul style="list-style-type: none"> <li>• IBD-Control-8 vs. SCCAI: <math>r_s = -0.72, p &lt; 0.01</math></li> <li>• IBD-Control-8 vs. PGA: <math>r_s = -0.67, p &lt; 0.01</math></li> <li>• IBD-Control-VAS vs. SCCAI: <math>r_s = -0.75, p &lt; 0.01</math></li> <li>• IBD-Control-VAS vs. PGA: <math>r_s = -0.65, p &lt; 0.01</math></li> <li>• NA</li> </ul> | <ul style="list-style-type: none"> <li>• NA</li> </ul>                                                                                                                                                                                                                                                                                                      |
| Clara I et al. 2009 <sup>37</sup>              | 169                    | MIBDI       | PTI                                                 | Sensitivity and specificity values to detect activity (PTI $\geq 4$ ) were provided at 3 time periods (0, 12, and 24 months). PPV and NPV were calculated using the data provided in the study                                                                                                                          | <ul style="list-style-type: none"> <li>• NA</li> </ul>                                                                                                                                                                                                                                                                                                         | <p>At 0, 12, and 24 months, respectively:</p> <ul style="list-style-type: none"> <li>• Sensitivity = 84%, 82%, 66%</li> <li>• Specificity = 61%, 68%, 61%</li> <li>• PPV = 75%, 79%, 68%</li> <li>• NPV = 74%, 72%, 59%</li> <li>• NA</li> </ul>                                                                                                            |
| Golovics PA et al. 2021 <sup>39</sup>          | 171                    | PRO-2, pMCS | CRP<br>FCP                                          | Kappa and Chi-2 test for correlation by PROMs-defined remission and elevated CRP or FCP                                                                                                                                                                                                                                 | <p>k agreement:</p> <ul style="list-style-type: none"> <li>• PRO-2 vs. CRP: <math>\kappa \sim 0.2</math></li> <li>• pMCS vs. CRP: <math>\kappa \sim 0.2</math></li> <li>• PRO-2 vs. FCP: <math>\kappa = 0.6</math></li> <li>• pMCS vs. FCP: <math>\kappa = 0.59</math></li> </ul>                                                                              | <ul style="list-style-type: none"> <li>• NA</li> </ul>                                                                                                                                                                                                                                                                                                      |

(Continued)

Table 2. Continued.

| Article                                   | Number of patients (N) | PROM                             | Disease activity measure | Correlation/association/predictive accuracy measures                                                                                                 | Correlation                                                                                                                                                                                                 | Association and predictive accuracy                                                                                                                                                                                                                                                                                      |
|-------------------------------------------|------------------------|----------------------------------|--------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Kamat N et al. 2022 <sup>40</sup>         | 57                     | IBD-Control-8, IBD-Control-VAS   | SCCAI, FCP               | Spearman's rank correlation coefficient. A negative correlation was expected. AUROC, sensitivity, and specificity to detect remission (SCCAI ≤2)     | IBD-Control-8<br>• vs. SCCAI: $r_s = -0.59$ , $p < 0.001$<br>• vs. FCP: $r_s = 0.57$ , $p < 0.001$<br>IBD-Control-VAS<br>• vs. SCCAI: $r_s = -0.60$ , $p < 0.001$<br>• vs. FCP: $r_s = -0.40$ , $p < 0.001$ | IBD-Control-8:<br>• AUROC = 0.87 (95% CI: 0.78-0.97)<br>• Sensitivity = 71.8%<br>• Specificity = 88%<br>IBD-Control-VAS:<br>• AUROC = 0.83 (95% CI: 0.72-0.94)<br>• Sensitivity = 34.4%<br>• Specificity = 92%<br>FCP:<br>• AUROC: 0.88 (95% CI: 0.77-0.97), $p < 0.001$<br>• Sensitivity = 81.3%<br>• Specificity = 88% |
| Sexton KA et al. 2019 <sup>42</sup>       | 125                    | IBDSI (long form, LF [38 items]) | Nurse-administered PTI   | Pearson correlation, AUROC, sensitivity, specificity, PPV, and NPV using PTI ≤5 or PGA                                                               | • IBDSI vs. PTI: $r = 0.85$ , $p < 0.001$<br>• IBDSI vs. PGA: $r = 0.70$ , $p < 0.0001$                                                                                                                     | IBDSI (LF) vs. PTI<br>• AUROC: 0.95 (95% CI: 0.92-0.99)<br>• Sensitivity: 91%<br>• Specificity: 87%                                                                                                                                                                                                                      |
| Subramanian S et al. 2016 <sup>43</sup>   | 196                    | IBD-10 NRS                       | SCCAI CRP FCP            | Spearman's rank correlation coefficient (Spearman's rho). AUROC, sensitivity, and specificity were calculated for remission (SCCAI ≤2)               | • IBD-10 vs. SCCAI: $r_s = -0.79$ , $p < 0.001$                                                                                                                                                             | • AUROC = 0.94 (95% CI: 0.89-0.97)<br>• Sensitivity = 90% (95% CI: 86-94)<br>• Sensitivity = 75% (95% CI: 67-82)<br>• Patients with CRP <5 mg/dL had a significantly higher IBD-10 score than patients with CRP ≥5 ( $p = 0.002$ )                                                                                       |
| Surti B et al. 2013 <sup>44</sup>         | 81                     | IBD NRS                          | SCCAI CRP ESR FCP        | Pearson correlation coefficient ( $r$ )                                                                                                              | • NRS vs. SCCAI: $r = 0.25$ , $p < 0.0001$<br>• vs. CRP, ESR, or FCP: no correlation                                                                                                                        | • NA                                                                                                                                                                                                                                                                                                                     |
| Tow KE et al. 2019 <sup>45</sup>          | 33                     | PRO-2-UC                         | Clinician-verified pMCS  | Spearman's correlation coefficient                                                                                                                   | • PRO-2-UC (paper-based) vs. pMCS: $r_s = 0.99$ , $p < 0.001$<br>• PRO-2-UC (tablet-based) vs. pMCS: $r_s = 0.56$ , $p < 0.01$<br>• $r_s = 0.72$ , $p < 0.0001$                                             | • NA                                                                                                                                                                                                                                                                                                                     |
| Van Deen WK et al. 2016 <sup>32</sup>     | 109                    | mHI-UC                           | pMCS                     | Spearman's correlation coefficient, ROC analysis to predict disease activity                                                                         | • $r_s = -0.524$ , $p < 0.001$                                                                                                                                                                              | • AUROC = 0.91<br>• Sensitivity = 73%<br>• Specificity = 90%                                                                                                                                                                                                                                                             |
| Vicente Lidón R et al. 2021 <sup>46</sup> | 71                     | IBD-Control                      | SCCAI                    | Spearman's correlation coefficient. Sensitivity and specificity for "quiescent disease" (data pooled, separate data for CD and UC were not provided) |                                                                                                                                                                                                             | • AUROC for "quiescent disease" = 0.864 (95% CI: 0.811-0.916)<br>• Sensitivity = 70%<br>• Specificity = 90%                                                                                                                                                                                                              |

Some studies analyzed the correlation with disease activity in both CD and UC.

Abbreviations: AP, abdominal pain; AUROC, area under the receiver operating characteristic curve; CD, Crohn's disease; CDAI, Crohn's Disease Activity Index; CDEIS, Crohn's Disease Endoscopic Index of Severity; CI, confidence interval; CRP, C-reactive protein; ESR, erythrocyte sedimentation rate; FCP, fecal calprotectin; GELS, global evaluation of lesion severity; HBI, Harvey-Bradshaw Index; IBD, inflammatory bowel disease; IBD-Control-8, inflammatory bowel disease control-8 items; IBD-Control-VAS, inflammatory bowel disease control-visual analogue scale; IBDQ, Inflammatory Bowel Disease Questionnaire; IBDSI, Inflammatory Bowel Disease Symptom Inventory; MES, Mayo endoscopic subscore; MHBI, modified Harvey-Bradshaw Index; mHI-CD, mobile Health Index for Crohn's disease; mHI-UC, mobile Health Index for ulcerative colitis; MIAH, Monitor IBD at Home; MIBDI, Manitoba IBD Index; NA, not available; NPV, negative predictive value; NRS, Numerical Rating Scale; PGA, Physician's Global Assessment; p-HBI, patient-based Harvey-Bradshaw Index; pMCS, partial Mayo Clinic Score; PPV, positive predictive value; PRO-2, 2 item patient-reported outcome; PRO-3, 3 item patient-reported outcome; PROM, patient-reported outcome measure;  $r$ , Pearson correlation coefficient; RB, rectal bleeding; ROC, receiver operating characteristic;  $r_s$ , Spearman's correlation coefficient; SCCAI, Simple Clinical Colitis Activity Index; SES-CD, Simple Endoscopic Score for CD; SF, stool frequency; UC, ulcerative colitis; VAS, visual analogue scale.

### 3.6 Correlation and predictive performance between PROMs and clinical measures of disease activity

Correlations between PROMs and disease activity evaluated by healthcare professionals were moderate-to-strong both in UC and CD in most of the studies (Table 2).<sup>32,34–37,40,42–46</sup> In CD,  $r$  values ranged from 0.25 for the correlation between IBD NRS and Harvey-Bradshaw Index<sup>44</sup> to 0.95 for those between PRO2 and a modified Harvey-Bradshaw Index.<sup>45</sup> Most correlations were strong (0.50 to 0.90).<sup>32,34,36,42–46</sup> Sensitivity and specificity values spanned from 70%<sup>46</sup> to 95%<sup>43</sup> and from 51%<sup>37</sup> to 91%,<sup>35</sup> respectively. AUROCs ranged from 0.86<sup>43</sup> to 0.93.<sup>42</sup>

In patients with UC, the values for correlations were between 0.25<sup>44</sup> and 0.99,<sup>45</sup> with most studies reporting values above 0.50.<sup>32,34,36,40,42–46</sup> Sensitivity ranged from 34%<sup>40</sup> to 91%<sup>35,42</sup> and specificity from 61%<sup>37</sup> to 92%.<sup>40</sup> AUROCs were reported in 2 studies, with values from 0.83<sup>40</sup> to 0.95.<sup>42</sup>

## 4. Discussion

Our systematic review shows a poor correlation of PROMs and more objective endoscopic measures of disease activity especially for CD. Correlations were also very poor for laboratory-based activity measures, whereas, as expected, correlations were adequate for some of the clinical activity measures. Although PROMs are valuable tools for capturing the patient's perspective, their ability to detect inflammatory activity in CD seems limited. For UC, predictive value is better, although most of the predicted value comes from self-reporting rectal bleeding. Finally, our data do not suggest that a particular PROM set might perform much better than the others for detecting endoscopic activity.

Most studies we analyzed had a real-world observational design, underscoring the growing recognition of the value of PROs in understanding and managing IBD beyond controlled clinical trials. The most common symptoms in the identified PROMs were SF, RB, and AP. Some tools also considered extraintestinal/systemic manifestations, general well-being, and emotional/self-perception aspects providing a more comprehensive overview of the patient's condition. Incorporating these aspects did not seem, however, to improve the ability to detect active inflammation. Interestingly, none of the identified PROMs fully evaluated all relevant signs and symptoms of the diseases. Significant symptoms, such as tenesmus, are absent from all evaluated PROMs, highlighting gaps in the current tools available for IBD assessment.

In CD, correlations between PROMs and endoscopic measures were uniformly poor. For instance, studies assessing the commonly used PRO-2 reported that it was weakly correlated with the SES-CD.<sup>47</sup> Regarding individual symptoms, the SF item of the MIAH questionnaire showed the highest correlation with SES-CD; however, it was still a modest association.

In UC, correlations were generally stronger than those observed in CD. PROMs, like PRO-2, the MIAH RB item, and the mHI for UC, showed weak-to-moderate correlations with the MES. The rectal bleeding component of MIAH had a stronger correlation with endoscopic findings, suggesting that self-reported bleeding may rather accurately reflect UC disease activity. Indeed, in a meta-analysis of 5 studies, the authors found that most patients with UC who had normal RB and SF subscores had attained endoscopic remission.<sup>48</sup>

Our results agree with prior studies showing that PROMs are not standalone markers of disease activity in IBD. For example, the correlations between PROMs and endoscopic scores in CD were also weak according to Khanna et al.<sup>49</sup> On the other hand, the study performed by Ma et al. in UC<sup>41</sup> showed that the correlations of PRO-2 with endoscopic scores were better when determining the value of the RB, which is in line with our findings. This all underlines the importance of the specificity of symptoms in the utility of PROMs.

The reviewed studies also analyzed the correlations between PROMs and laboratory parameters such as CRP and FCP. Neither the IBD-10 nor the HBI showed a good association with high levels of FCP in CD. In the same way, the NRS did not show any correlation with either CRP, ESR, or FCP. In UC, associations were slightly better but still indicated limitations. For instance, IBD-Control-8 and IBD-Control-VAS showed only a moderate correlation with FCP. Previous studies have found similar results, showing that while FCP is a reliable marker for detecting mucosal inflammation, its correlation with PROMs like the IBD-10 and HBI in IBD is generally weak, likely due to the nonspecific nature of the symptoms of IBD that are measured by PROMs.<sup>40</sup>

Regarding the correlation between PROMs and clinically assessed scores, we found heterogeneous results, but, as expected, the correlation between different symptom-based scores was acceptable, with some exceptions. In CD, the IBD-10 showed a strong negative correlation with the HBI and the SCCAI indices. In UC, tools like the 6-point Mayo Clinic Score and SCCAI exhibited high sensitivity and specificity in predicting clinical remission measured by SCCAI. Therefore, their alignment with clinical activity indexes is generally better than those with endoscopic measures but can still differ significantly depending on the disease context and specific PROM used.

After closing the review period, a new study reported the correlation between RB, SF, PRO-2, partial Mayo Clinic Score, and SCCAI and endoscopic healing in UC. Correlation ranged from moderate to strong, with no significant differences in predictive accuracy between these measures.<sup>50</sup> Another investigation recently published in IBD indicated that urgency for bowel movements was consistently associated with disease activity and that fatigue is the most prevalent symptom, irrespective of the disease activity measure used;<sup>51</sup> Sninsky et al. also supported this finding.<sup>52</sup> However, it is noteworthy to mention that fecal urgency and fatigue were included in only 5 and 4 out of the 17 PROMs that we identified, respectively. Similarly, Moon et al.<sup>53</sup> emphasized the importance of considering fecal urgency, tenesmus, mucoid stool, and night-time defecation to improve the prediction of endoscopic mucosal healing status with respect to RB and SF in UC.<sup>53</sup>

The main limitation of this work arises from the challenge of precisely categorizing PROMs based on their intended purpose. Many instruments mix various patient-reported aspects such as general health-related quality of life, disease burden, disability, and disease activity.<sup>54</sup> Moreover, the considerable heterogeneity observed among the selected studies, derived from differences in study design and comparator selection, complicates the formulation of definitive conclusions based on the results. An additional critical limitation is the inherent bias in patient-reported outcomes, which can be influenced by various factors, including psychological state, socio-economic status, and the patient-clinician relationship. This subjectivity,

while valuable in understanding the patient's perspective, necessitates cautious interpretation of PROMs in isolation from clinical assessments. Furthermore, it should be noted that the quality of conclusions made based on a systematic review largely depends on the biases that may affect the gathering of the original literature, such as the language bias (a type of selection bias) and publication bias toward positive results, which is independent of our methodology. We have included only studies reported in English or Spanish, and have attempted to mitigate possible biases by searching within 3 of the main publicly accessible databases of medical literature, as well as performing manual searches and reviewing the bibliographies of the studies found. We acknowledge that the quality of the available and included studies was generally low-to-moderate. We hope this review will raise awareness about the need to standardize how PROMs are assessed and validated. Finally, the high heterogeneity in the activity measure used as standard, the methods of correlation and prediction estimates, as well as cut-offs used for calculating sensitivity and specificity make it difficult to quantitatively compare between studies. Despite this heterogeneity, the conclusions on the poor correlation between PROMs and endoscopy were consistent, suggesting that the findings of our review are likely to be generalizable.

Nonetheless, our investigation also has some strengths. Our analysis provides a global review of the available validated IBD-specific PROMs, offering valuable insights into their utility and limitations. Moreover, the utilization of a rigorous systematic review methodology, including a detailed literature search and the quality assessment of the included studies, enhances the reliability of our findings.

In conclusion, our review suggests that PROMs cannot substitute objective measures of disease activity in IBD, such as endoscopic and laboratory results. This limitation may be attributed to the fact that IBD symptoms can overlap with those of other gastrointestinal disorders, particularly in CD. However, these results must be considered with caution due to the generally high methodological heterogeneity between studies and low-to-moderate quality of the original studies. Homogenization of methods and standards is warranted for future validation studies. Nonetheless, PROMs represent a valuable complementary instrument to describe disease control comprehensively by capturing the patient's perspective. Further studies are needed to develop more refined PROMs for IBD, ensuring they capture relevant symptoms and correlate more closely with objective disease measures. Moreover, developing composite indices integrating PROMs with clinical, endoscopic, and laboratory measures could enhance the accuracy and utility of patient-reported data in clinical practice.

## Acknowledgments

Medical writing support was provided by Medical Statistics Consulting SL (Valencia, Spain), and was funded by Galapagos Biopharma Spain SL (Madrid, Spain). Editorial and publications management support was provided by PharmaGenesis London, London, UK, and was funded by Galapagos NV.

## Author contributions

All authors were involved in the conception and design of the study, the acquisition or analysis of data, and the interpretation

of data. All authors contributed to drafting the manuscript or revising it critically for important intellectual content and approved the final version.

## Supplementary material

[Supplementary material](#) is available at *ECCO-JCC* online.

## Funding

This study was funded by Galapagos NV (Mechelen, Belgium).

## Conflicts of interest

X.C. reports advisory board and speaker fees from AbbVie, Allergan, Eli Lilly, MSD, Shire, Takeda, and Vifor Pharma; and research funding from AbbVie, Galapagos, Janssen Pharmaceuticals, MSD, and Vifor Pharma.

M.G.F. and V.M. are employees of Medical Statistics Consulting SL (Valencia, Spain).

S.A. was an employee of Galapagos Biopharma Spain SL (Madrid, Spain) at the time of this work, and is now an employee of Alfasigma S.p.A.

M.B.-d.A. reports consulting, advisory board, and speaker fees, or research funding from AbbVie, Adacety Therapeutics, Celltrion, Eli Lilly, Fresenius, Galapagos, Janssen Pharmaceuticals, MSD, Pfizer, and Takeda.

## Data availability

The data underlying this article were obtained from company-sponsored research and cannot be shared publicly due to company policy. The data may be shared upon reasonable request to the authors.

## References

1. Cosnes J, Gower-Rousseau C, Seksik P, Cortot A. Epidemiology and natural history of inflammatory bowel diseases. *Gastroenterology* 2011;140:1785-1794.
2. Gajendran M, Loganathan P, Jimenez G, et al. A comprehensive review and update on ulcerative colitis. *Dis Mon* 2019;65:100851.
3. Baumgart DC, Sandborn WJ. Crohn's disease. *Lancet* 2012;380:1590-1605.
4. Wong E, Mavondo F, Fisher J. Patient feedback to improve quality of patient-centred care in public hospitals: A systematic review of the evidence. *BMC Health Serv Res* 2020;20:530.
5. Kim EJ, Nam IC, Koo YR. Reframing patient experience approaches and methods to achieve patient-centeredness in healthcare: Scoping review. *Int J Environ Res Public Health* 2022;19.
6. Louis E, Ramos-Goni JM, Cuervo J, et al. A qualitative research for defining meaningful attributes for the treatment of inflammatory bowel disease from the patient perspective. *Patient* 2020;13:317-325.
7. Siegel CA. Refocusing IBD patient management: Personalized, proactive, and patient-centered care. *Am J Gastroenterol* 2018;113:1440-1443.
8. Ye BD, Travis S. Improving the quality of care for inflammatory bowel disease. *Intest Res* 2019;17:45-53.
9. Russell G. Have we forgotten the moral justification for patient-centred care? *BMJ Qual Saf* 2022;31:172-174.

10. Kuntz JL, Safford MM, Singh JA, et al. Patient-centered interventions to improve medication management and adherence: A qualitative review of research findings. *Patient Educ Couns* 2014;97:310-326.
11. Bosworth HB, Fortmann SP, Kuntz J, et al. Recommendations for providers on person-centered approaches to assess and improve medication adherence. *J Gen Intern Med* 2017;32:93-100.
12. Bjorner JB, Gandek B, Cole J, et al. Response to the FDA draft guidance for industry document: Patient-reported outcome measures: Use in medical product development to support labeling claims (docket 2006d-0044). 2006.
13. Churrua K, Pomare C, Ellis LA, et al. Patient-reported outcome measures (PROMs): A review of generic and condition-specific measures and a discussion of trends and issues. *Health Expect* 2021; 24:1015-1024.
14. El-Matary W. Patient-reported outcome measures in inflammatory bowel disease. *Can J Gastroenterol Hepatol* 2014;28: 536-542.
15. de Jong MJ, Huijbregtse R, Masclee AAM, Jonkers D, Pierik MJ. Patient-reported outcome measures for use in clinical trials and clinical practice in inflammatory bowel diseases: A systematic review. *Clin Gastroenterol Hepatol* 2018;16:648-663.e3.
16. Kim AH, Roberts C, Feagan BG, et al. Developing a standard set of patient-centred outcomes for inflammatory bowel disease-an international, cross-disciplinary consensus. *J Crohns Colitis* 2018; 12:408-418.
17. Tran F, Schirmer JH, Ratjen I, et al. Patient reported outcomes in chronic inflammatory diseases: Current state, limitations and perspectives. *Front Immunol* 2021;12:614653.
18. Daperno M, D'Haens G, Van Assche G, et al. Development and validation of a new, simplified endoscopic activity score for Crohn's disease: The SES-CD. *Gastrointestinal Endoscopy* 2004;60: 505-512.
19. Mary JY, Modigliani R. Development and validation of an endoscopic index of the severity for Crohn's disease: A prospective multicentre study. Groupe d'études thérapeutiques des affections inflammatoires du tube digestif (GETAID). *Gut* 1989;30:983-989.
20. Schroeder KW, Tremaine WJ, Ilstrup DM. Coated oral 5-aminosalicylic acid therapy for mildly to moderately active ulcerative colitis. A randomized study. *N Engl J Med* 1987;317:1625-1629.
21. Peyrin-Biroulet L, Panes J, Sandborn WJ, et al. Defining disease severity in inflammatory bowel diseases: Current and future directions. *Clin Gastroenterol Hepatol* 2016;14:348-354.e17.
22. Harvey RF, Bradshaw JM. A simple index of Crohn's-disease activity. *Lancet* 1980;1:514.
23. Best WR, Becktel JM, Singleton JW, Kern Jr. F., Development of a Crohn's disease activity index. National cooperative Crohn's disease study. *Gastroenterology* 1976;70:439-444.
24. Walmsley RS, Ayres RC, Pounder RE, Allan RN. A simple clinical colitis activity index. *Gut* 1998;43:29-32.
25. Programme CAS. Critical appraisal skills programme checklists <https://casp-uk.net/casp-tools-checklists/>, 2021.
26. Armenteros SC, Calvet X, Barreiro M. Correlation between patient-reported outcome measures and disease activity in inflammatory bowel disease: A systematic review. [https://www.crd.york.ac.uk/prospetro/display\\_record.php?RecordID=383899](https://www.crd.york.ac.uk/prospetro/display_record.php?RecordID=383899), 2022.
27. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Int J Surg* 2021;88:105906.
28. de Jong MJ, Roosen D, Degens J, et al. Development and validation of a patient-reported score to screen for mucosal inflammation in inflammatory bowel disease. *J Crohns Colitis* 2019;13:555-563.
29. Dragasevic S, Sokic-Milutinovic A, Lalošević MS, et al. Correlation of patient-reported outcome (PRO-2) with endoscopic and histological features in ulcerative colitis and Crohn's disease patients. *Gastroenterol Res Pract* 2020;2020:2065383.
30. Lewis JD, Rutgeerts P, Feagan BG, et al. Correlation of stool frequency and abdominal pain measures with simple endoscopic score for Crohn's disease. *Inflamm Bowel Dis* 2020;26:304-313.
31. Morris MW, Stewart SA, Heisler C, et al. Biomarker-based models outperform patient-reported scores in predicting endoscopic inflammatory disease activity. *Inflamm Bowel Dis* 2018;24:277-285.
32. Van Deen WK, van der Meulen-de Jong AE, Parekh NK, et al. Development and validation of an inflammatory bowel diseases monitoring index for use with mobile health technologies. *Clin Gastroenterol Hepatol* 2016;14:1742-1750.e7.
33. Zittan E, Kabakchiev B, Kelly OB, et al. Development of the Harvey-Bradshaw index-pro (HBI-PRO) score to assess endoscopic disease activity in Crohn's disease. *J Crohns Colitis* 2017;11: 543-548.
34. Bennebroek Evertsz F, Hoeks CC, Nieuwkerk PT, et al. Development of the patient Harvey Bradshaw index and a comparison with a clinician-based Harvey Bradshaw Index assessment of Crohn's disease activity. *J Clin Gastroenterol* 2013;47:850-856.
35. Bennebroek Evertsz F, Nieuwkerk PT, Stokkers PC, et al. The patient simple clinical colitis activity index (P-SCCAI) can detect ulcerative colitis (UC) disease activity in remission: A comparison of the P-SCCAI with clinician-based SCCAI and biological markers. *J Crohns Colitis* 2013;7:890-900.
36. Bodger K, Ormerod C, Shackcloth D, Harrison M, IBD Control Collaborative. Development and validation of a rapid, generic measure of disease control from the patient's perspective: The IBD-control questionnaire. *Gut* 2014;63:1092-1102.
37. Clara I, Lix LM, Walker JR, et al. The Manitoba IBD index: Evidence for a new and simple indicator of IBD activity. *Am J Gastroenterol* 2009;104:1754-1763.
38. Colombel JF, Keir ME, Scherl A, et al. Discrepancies between patient-reported outcomes, and endoscopic and histological appearance in UC. *Gut* 2017;66:2063-2068.
39. Golovics PA, Gonczi L, Reinglas J, et al. Patient-reported outcome and clinical scores are equally accurate in predicting mucosal healing in ulcerative colitis: A prospective study. *Dig Dis Sci* 2022;67: 3089-3095.
40. Kamat N, Vuyyuru SK, Kedia S, et al. Correlation of fecal calprotectin and patient-reported outcome measures in patients with ulcerative colitis. *Intest Res* 2022;20:269-273.
41. Ma C, Sandborn WJ, D'Haens GR, et al. Discordance between patient-reported outcomes and mucosal inflammation in patients with mild to moderate ulcerative colitis. *Clin Gastroenterol Hepatol* 2020;18:1760-1768.e1.
42. Sexton KA, Walker JR, Targownik LE, et al. The inflammatory bowel disease symptom inventory: A patient-report scale for research and clinical application. *Inflamm Bowel Dis* 2019;25: 1277-1290.
43. Subramanian S, Asher R, Weston W, et al. Validation of a simple 0 to 10 numerical score (IBD-10) of patient-reported inflammatory bowel disease activity for routine clinical use. *Inflamm Bowel Dis* 2016;22:1902-1907.
44. Surti B, Spiegel B, Ippoliti A, et al. Assessing health status in inflammatory bowel disease using a novel single-item numeric rating scale. *Dig Dis Sci* 2013;58:1313-1321.
45. Tow KE, Rogge C, Lee T, Caputi P, Knowles SR. Validation of a digital support app to assess inflammatory disease activity and

- mental health patient-reported outcomes (PROs): A pilot investigation. *Gastroenterol Res Pract* 2019;2019:7618468.
46. Vicente Lidón R, García López S, Corsino Roche P, et al. Translation into Spanish and validation of a short questionnaire to measure the control of inflammatory bowel disease from the patient's perspective: IBD-control, EII-control. *Gastroenterologia y Hepatologia* 2021;45:524-534.
  47. Ma C, Panaccione R, Fedorak RN, et al. Heterogeneity in definitions of endpoints for clinical trials of ulcerative colitis: A systematic review for development of a core outcome set. *Clin Gastroenterol Hepatol* 2018;16:637-647.e13.
  48. Narula N, Alshahrani AA, Yuan Y, Reinisch W, Colombel JF. Patient-reported outcomes and endoscopic appearance of ulcerative colitis: A systematic review and meta-analysis. *Clin Gastroenterol Hepatol* 2019;17:411-418.e3.
  49. Khanna R, Zou G, Stitt L, et al. Responsiveness of endoscopic indices of disease activity for Crohn's disease. *Am J Gastroenterol* 2017;112:1584-1592.
  50. Tse CS, Nguyen HP, Singh S, et al. Correlations between gastrointestinal symptoms and endoscopic-histologic disease activity in adults with ulcerative colitis. *Dig Dis Sci* 2023;68:3254-3258.
  51. Kulyk A, Shafer LA, Graff LA, et al. Urgency for bowel movements is a highly discriminatory symptom of active disease in persons with IBD (the Manitoba Living with IBD study). *Aliment Pharmacol Ther* 2022;56:1570-1580.
  52. Sninsky JA, Barnes EL, Zhang X, Long MD. Urgency and its association with quality of life and clinical outcomes in patients with ulcerative colitis. *Am J Gastroenterol* 2022;117:769-776.
  53. Moon JM, Yoon H, Park J, et al. Usefulness of examining various patient-reported outcomes in predicting endoscopic mucosal healing in ulcerative colitis. *Turk J Gastroenterol* 2022;33: 682-688.
  54. van Andel EM, Koopmann BDM, Crouwel F, et al. Systematic review of development and content validity of patient-reported outcome measures in inflammatory bowel disease: Do we measure what we measure? *J Crohns Colitis* 2020;14:1299-1315.