

# ADVANCES IN IBD

Current Developments in the Treatment of Inflammatory Bowel Disease

Section Editor: Stephen B. Hanauer, MD

## The Development of Patient-Reported Outcome Measures in Inflammatory Bowel Disease



Peter D. R. Higgins, MD, PhD, MSc  
Professor  
University of Michigan School of Medicine  
Ann Arbor, Michigan

### G&H What patient-reported outcome measures have traditionally been used in inflammatory bowel disease?

**PH** To some extent, every patient visit involves a physician asking the patient about symptoms, which are patient-reported outcomes (PROs) that are not obvious to anyone but the patient. For example, only the patient knows whether he or she is feeling fatigue or pain, or has blood in his or her bowel movements. The Inflammatory Bowel Disease Questionnaire (IBDQ) is one of the oldest formal instruments that attempts to take the type of information elicited in a patient visit and arrange it into a structured format via a score. For its time, the IBDQ was a major advance. However, it had a number of flaws, which led the US Food and Drug Administration (FDA) to move away from it.

Composite measures, which mix PRO items with more objective measures, have also been used. For example, the Mayo score for ulcerative colitis uses 2 PRO items—stool frequency and stool blood—but also includes endoscopy as an objective measure and does not differentiate between the PRO and objective measures, throwing all of them together into one total score. Similarly, the Crohn's Disease Activity Index (CDAI) includes several PRO components, such as abdominal pain and the number of liquid bowel movements, but also has objective components, such as hemoglobin and weight. These composite measures have proven to be somewhat useful, but also contribute a lot of noise to measurement of disease activity because if a patient's score changes, it is unclear whether the change is due to the PRO items or to

the objective measures. When these composite measures were developed, the assumption was that all of the items of a particular measure would correlate and, thus, would increase or decrease together. However, this is not always true, resulting in noise that has become problematic in clinical trials.

### G&H What other limitations are associated with the traditional PRO or composite measures?

**PH** The traditionally used measures have important problems. For example, patients frequently report fatigue as a major issue; difficulty concentrating or focusing because of their symptoms; difficulty performing social activities, which impacts their ability to socialize; and feelings of anger, frustration, depression, and anxiety, all of which are related to inflammatory bowel disease. These symptoms are important to the patient but are not included in traditional PRO measures.

In addition, there are particular issues with the IBDQ, which was created when the science of PRO development was fairly primitive. One issue is that the IBDQ includes several double-barreled questions, such as whether the patient had anxiety about finding a bathroom. From this question, it is unclear whether the patient is mostly anxious and has a problem with anxiety or whether the patient has a problem with urgency and he or she has to rush to the bathroom very quickly. It could be one or the other, or even both.

Similarly, the Mayo score for ulcerative colitis mixes 2 components (frequency and quantity) when asking about blood. There are 4 levels for blood, from 0 (no blood

ever) to 3 (lots of blood all of the time). However, levels 1 and 2 are problematic in that 1 is traces of blood some of the time and 2 is lots of blood most of the time. Some patients have difficulty answering this question because it is combining frequency and quantity into one question. In cases where the patient has traces of blood most of the time, or lots of blood some of the time, patients do not know whether to pick a score of 1 or 2. Not only does the patient not know how to answer the question, but when the physician or clinical trialist receives the score, it is not

According to the FDA, the key is for the items of interest to come from patients, not from physicians or self-appointed experts; therefore, patient language should be used.

possible to determine whether the patient is reporting frequency or quantity.

Another problem with the IBDQ is that it was designed to measure 4 distinct domains of inflammatory bowel disease (bowel symptoms, systemic symptoms, emotional symptoms, and social function). However, factor analysis from several translations showed that there were actually 5 domains, not 4, so the intended design does not pan out in practice.

### G&H What guidance has the FDA provided for developing new PRO measures?

**PH** The FDA has put together an official guidance for developing PRO measures, which is available at <https://www.fda.gov/downloads/drugs/guidances/ucm193282.pdf>. The guidance reviews the aforementioned problems with traditional PRO measures and is fairly comprehensive on what the FDA is looking for in new measures. For example, because traditional measures miss items that are important to patients, the FDA emphasizes the need to perform a comprehensive search for everything that matters to patients and for item development to use language and terminology taken directly from patients, starting with focus groups, to elicit as many concepts as possible to capture the entire disease experience. According to the FDA, the key is for the items of interest to come from

patients, not from physicians or self-appointed experts; therefore, patient language should be used. For instance, if patients say that they are running to the bathroom instead of saying that they have urgency, the term used by the patients should be the one used in the measure. Thus, PRO instrument developers should consult many patients from different ethnic groups and races across many locations to try to obtain a broad range of descriptions of the concepts being measured and look for anything that is important to patients, is related to their disease, and that worsens when patients worsen or, conversely, improves when patients improve.

### G&H Are coprimary PRO measures necessary?

**PH** The FDA feels very strongly that PRO measures are important and necessary, and would be reluctant to approve a drug that did not show that patients felt better by patient report. One of the FDA's core missions is to approve drugs that help patients feel or function better, or have increased survival. It is essential to be able to measure that improved feeling or functioning. Although there is good evidence that objective markers such as endoscopy or fecal calprotectin have predictive value for important future clinical outcomes such as hospitalization and surgery, they do not necessarily tell physicians how a patient feels now, and what is seen on endoscopy does not always match how a patient feels. Therefore, the FDA wants drugs that produce biologic remission that will result in better clinical outcomes in the long term that will also make patients feel better in the short term.

In many ways, PRO measures and objective markers complement each other. It is not a competition in which only one can be used. Taken together, they can provide a more comprehensive understanding of patients, their disease, and their response to therapy. Some medicines will heal the bowel and result in good endoscopic appearance but make patients feel terrible. What PRO measures capture is whether patients actually feel better. For most medicines that work well, these measures and objective markers go together. It is important both to the patient and to the FDA to detect when these 2 sets of measures do not go together.

However, PRO measures and objective markers should not correlate perfectly because if they do, there would be no point in using both. Ideally, the PRO measure should correlate to some extent with an objective measure of inflammation such as endoscopy or fecal calprotectin, but the information should be orthogonal or complementary in order to collect additional, sufficiently different information that is important to the patient.

**G&H** In addition to trying to address the limitations of traditional measures, are there any other benefits to developing new measures?

**PH** If the FDA approves new PRO measures and drug manufacturers are able to use them in clinical trials, the results can be used in symptom-based labeling claims. For

... we found that, unlike with the IBDQ, Crohn's disease and ulcerative colitis were a little different, so slightly different instruments were needed for each disease.

example, currently a drug can claim that it is proven to heal the colon 30% faster. However, that does not mean much to patients. In contrast, PRO measures use patient language, which would have more meaning to patients. For example, if an urgency score can be reduced to 0, that would be important to patients and very easy for them to understand. Drug manufacturers are excited that new PRO measures may allow them to market specific symptomatic improvements.

**G&H** Is it possible to design PRO measures that are suitable for both clinical care and clinical trial?

**PH** It is possible, although to be efficient in clinical care, it will be important to leverage information technology. At my institution, my colleagues and I have been able to build PRO measures into questionnaires in our Epic electronic medical record system so that every inflammatory bowel disease patient who comes for a return visit fills out the PRO measure electronically before being seen in clinic. The PRO answers are pulled into a 4-paragraph history of present illness and a table of scores of how the patient is doing that helps the physician create a clinical note more efficiently.

**G&H** How were your PRO measures developed?

**PH** I have worked as part of a consortium with Amgen, Genentech, and Evidera on both an ulcerative colitis

PRO measure and a Crohn's disease PRO measure. We started with a framework of the concepts that patients were interested in and that would likely be in distinct domains, and then worked with patient focus groups across multiple racial and ethnic groups in multiple locations in the United States to develop patient language to generate items and response scales for each item. Pilot testing was performed to obtain feedback on content validity. These conceptual, qualitative data were submitted to and approved by the FDA almost a year ago. The PRO measures were then tested in patients participating in clinical trials sponsored by Amgen and Genentech to evaluate how the measures work compared with the IBDQ, depression scales, endoscopy, and histology, as well as to obtain a sense of how the measures perform in clinical trial patients. These developmental data have been submitted to the FDA for feedback and have been honed down from a comprehensive instrument to the items that are most likely to vary in the setting of a clinical trial. The FDA is currently reviewing the PRO measures for qualification to determine whether the clinical trial data should be accepted and whether the measures can be made publicly available for use in clinical trials with drugs of different mechanisms of action for external validation.

**G&H** Could you discuss some of the steps in the development process of these measures?

**PH** An important component was to determine the major domains or sections of the PRO measures. Both of the new PRO measures consist of 5 domains: bowel symptoms, systemic symptoms, coping behaviors (how people adapt to their inflammatory bowel disease and everyday life), impact on quality of life (including effects on work, school, social roles, and family), and emotional impact. However, we found that, unlike with the IBDQ, Crohn's disease and ulcerative colitis were a little different, so slightly different instruments were needed for each disease. In our measures, there was an overlap of 80% to 90%, but there were distinct questions for each disease. For example, urgency was more important in ulcerative colitis than in Crohn's disease.

In addition, as previously mentioned, the FDA wants measures to be defined and tested in the types of patients expected to participate in clinical trials. Thus, we had to figure out questions that make sense to patients regardless of their educational level and make sure that when patients answer questions, they are actually answering the question we think they are answering.

Administration of the instruments and collection of the data were also studied, as well as the response options. For example, patients said that they wanted to have a very broad range of options, which was challenging in

ulcerative colitis in terms of the number of bowel movements. Patients also said that they could keep count only up to approximately 1 bowel movement per hour, and the FDA specified that the recall period had to be 24 hours or less due to the fear that patients would forget what had happened during the previous day.

### G&H What are the next steps for these measures?

**PH** After they are approved by the FDA, it would be helpful to have other groups use and test the measures to see whether these are valid for therapies across multiple mechanisms of action. In addition, the PRO measures should be tested in adolescents and children, as well as in other patient subgroups, such as patients with ostomies (who are some of the sickest inflammatory bowel disease patients) and patients with perianal fistulas.

### G&H Are other groups also working on developing new PRO measures?

**PH** I certainly think that is possible, but I have not heard much about other efforts. As far as I know, the PRO measures from this consortium are the only ones that have been submitted to the FDA.

### G&H While the new PRO measures are being studied and developed, are interim measures being used?

**PH** The FDA has worked with physicians in the field to come up with compromises to use as temporary measures. For Crohn's disease, this meant focusing on the PRO items of stool frequency and stool blood in the CDAI. For ulcerative colitis, the physician global assessment (PGA) was removed, as this item was relatively easy for physicians to manipulate. In the past, if physicians wanted to get a patient into a trial, they could easily add a point or

2 to the PGA or take away a point or 2 if they wanted to show that the patient improved. Currently, the interim PRO measure for ulcerative colitis is similar to a 2-part Mayo score, with the 2 PRO items of stool frequency and stool blood.

### G&H How have these temporary measures been working out?

**PH** Because the temporary measures were derived from the measures traditionally used, the 2 sets are not that different, but I think that people are fairly comfortable with that, as the scores are fairly predictable from past data, and the temporary measures are not meant to be permanent. However, I would not be surprised if the temporary measures are used for several more years, as the FDA is taking a very measured, judicious approach to reviewing the PRO measures currently in development.

*Dr Higgins receives grant funding from the NIH, the CCF, Ascentage, Target, BioFire, Takeda, AbbVie, Shire, Seres, Eli Lilly, Pfizer, Genentech, Janssen, and UCB.*

### Suggested Reading

Bojic D, Bodger K, Travis S. Patient-reported outcome measures (PROMS) in inflammatory bowel disease: new data. *J Crohns Colitis*. 2017;11(suppl\_2):S576-S585.

Higgins PDR, Harding G, Leidy NK, et al. Development and validation of the Crohn's disease patient-reported outcomes signs and symptoms (CD-PRO/SS) diary. *J Patient Rep Outcomes*. 2017;2(1):24.

Higgins PDR, Harding G, Revicki DA, et al. Development and validation of the ulcerative colitis patient-reported outcomes signs and symptoms (UC-PRO/SS) diary. *J Patient Rep Outcomes*. 2017;2(1):26.

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(4):408-418.

US Department of Health and Human Services Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), Center for Devices and Radiological Health (CDRH). Guidance for industry. Patient-reported outcome measures: use in medical product development to support labeling claims. <https://www.fda.gov/downloads/drugs/guidances/ucm193282.pdf>. Released December 2009. Accessed October 11, 2018.