ORIGINAL ARTICLE

Reliability among central readers in the evaluation of endoscopic disease activity in pouchitis

Mark A. Samaan, MD,[1,2] Bo Shen, MD,[3] Mahmoud H. Mosli, MD,[4] Guangyong Zou, PhD,[2,5] William J. Sandborn, MD,[2,6] Lisa M. Shackelton, PhD,[2] Sigrid Nelson, MS,[2] Larry Stitt, MS,[2] Stuart Bloom, MD,[7] Darrell S. Pardi, MD,[8] Paolo Gionchetti, MD, PhD,[9] James Lindsay, MD,[10] Simon Travis, MD, PhD,[11] Ailsa Hart, MD, PhD,[12] Mark S. Silverberg, MD, PhD,[13] Brian G. Feagan, MD,[2,5,14] Geert R. D'Haens, MD, PhD,[15] Vipul Jairath, MD, PhD[2,5,14]

London, Oxford, United Kingdom; London, Toronto, Ontario, Canada; Cleveland, Ohio; La Jolla, California; Rochester, Minnesota, USA; Jeddah, Saudi Arabia; Bologna, Italy; Amsterdam, Netherlands

Background and Aims: Pouchitis is a common adverse event after proctocolectomy with ileal pouch anal anastomosis for ulcerative colitis. Evaluation of pouchitis disease activity and response to treatment requires use of validated indices. We assessed the reliability of items evaluating endoscopic pouchitis disease activity.

Methods: Twelve panelists used a modified RAND appropriateness methodology to rate the appropriateness of items evaluating endoscopic pouchitis disease activity derived from a systematic review and also identified additional potential endoscopic items based on expert opinion. Four central readers then evaluated 50 pouchoscopy videos in triplicate, in random order. Intra- and inter-rater reliability for each item was assessed by calculating and comparing intraclass correlation coefficients (ICCs). A Delphi process identified common sources of disagreement among the readers.

Results: Ten existing endoscopic items were identified from the systematic review and an additional 7 exploratory items from the panelists. ICCs for inter-rater reliability were highest for the existing item of pouch ulceration (.72; 95% confidence interval [CI], .60-.82) and for the exploratory item of ulcerated surface in the pouch body (.67; 95% CI, .53-.75). Inter-rater reliability for all other existing and exploratory items was "moderate" (ICC < .60). The item "ulcerated surface in the pouch body" demonstrated the best correlation with a global evaluation of lesion severity (r = .80; 95% CI, .73-.85).

Conclusion: Substantial reliability was observed only for the endoscopic items of ulceration and ulcerated surface in the pouch body. Future studies should assess responsiveness to treatment in the next stage toward development of an endoscopic pouchitis disease activity index. (Gastrointest Endosc 2018;■:1-10.)

(footnotes appear on last page of article)

Surgical treatment is required in up to 30% of patients with ulcerative colitis after a decade of disease,[1-3] either as a consequence of medically refractory disease or development of dysplasia. In this situation restorative proctocolectomy with ileal pouch-anal anastomosis is usually the surgery of choice. However, pouchitis may occur within 10 years in up to 50% of patients and is associated with impaired health-related quality of life because of symptoms of diarrhea, urgency, rectal bleeding, or incontinence.[4-6] Furthermore, some patients develop chronic pouchitis and either become dependent on antibiotics for symptom relief or have symptoms refractory to conventional therapies. With no approved treatments for this condition, a large unmet medical need exists.

Importantly, several novel therapies are undergoing evaluation in clinical trials, such as the intercellular adhesion molecule-1 anti-sense oligonucleotide, alicaforsen (NCT02525523),[7] which has been granted orphan designation for this indication by the U.S. Food and Drug Administration and European Medicines Agency,[8] and vedolizumab, a monoclonal antibody to the alpha4beta7 integrin (NCT02790138). However, efficient approaches to evaluation of novel treatments in clinical trials for pouchitis require use of outcome measures with proven validity, reliability, and responsiveness.[9] Although current indices for evaluation of pouchitis typically measure a composite of clinical, endoscopic, and histologic items, none of these, including the most commonly used instrument, the Pouchitis Disease Activity Index (PDAI),[10]

was created using robust methodology for outcome measure development.

Pouchoscopy is required both for the diagnosis of pouchitis and to exclude other conditions such as Crohn's disease or structural abnormalities of the pouch. This procedure involves assessment of the rectal cuff, pouch body, and prepouch ileum. However, standardized and reliable descriptors of endoscopic disease activity do not currently exist. This study is a first step toward the development of a fully validated endoscopic instrument for the evaluation of pouchitis disease activity. Accordingly, we conducted a systematic review to identify and appraise all evaluative instruments used for the assessment of endoscopic pouchitis disease activity. We then conducted a consensus process using modified RAND appropriateness methodology[11] to combine the best available evidence and the clinical experience of experts in the field to rate appropriateness of endoscopic items. The results of blinded central review of pouchoscopy videos were then used to evaluate the reliability of these items to assess endoscopic pouchitis disease activity.

METHODS

Systematic review of literature

Search strategy. MEDLINE, EMBASE, PubMed, the Cochrane Library (CENTRAL), and abstracts presented at Digestive Disease Week and United European Gastroenterology were electronically searched without language restriction from their inception to 2014 to identify endoscopic evaluative instruments used for the assessment of endoscopic pouchitis disease activity. No restriction was placed on study design, although case studies were excluded. A summary of the specific search strategy used is detailed below and a comprehensive description is included in the Supplementary Material. Each database was searched for ("pouchitis" OR "pouch") AND ("index" OR "indice" OR "scale" OR "score" OR "grade" OR "Pouchitis Disease Activity Index" OR "PDAI" OR "Objective Pouchitis Score" OR "St. Mark's" OR "Pouchitis Activity Score").

Study selection. Eligible studies included any study design measuring pouchitis disease activity. Two reviewers (S.N. and M.A.S.) independently screened citations and abstracts before retrieving full-text publications of all potentially eligible articles. Disagreement was resolved in discussion with a third reviewer (V.J.).

The full text of eligible articles was reviewed by pairs of researchers (M.A.S.–S.N. and V.J.–M.H.M.) to extract the following prespecified variables: index used; disease activity cut-points; whether the index was used for diagnosis, measurement of disease activity, or both; and study design and number of patients. Additional variables to assess the level of index validation were also collected with regard to the index reliability, validity, and responsiveness. Disagreement was discussed among individual pairs of researchers and subsequently by all 4 if agreement was not possible.

Consensus process

Recruitment of panelists. A panel composed of 12 international gastroenterologists and a colorectal surgeon, all with a special interest in the care of patients after ileal pouch-anal anastomosis, was assembled. The panel included practicing clinicians and inflammatory bowel disease researchers from the United States, Canada, the Netherlands, the United Kingdom, and Italy who were chosen based on their recognized experience.

RAND appropriateness methodology was used to assess the face validity (the extent to which an item is subjectively viewed as addressing the concept it purports to measure) and appropriateness of items identified in the systematic review to measure endoscopic pouchitis disease activity, as well as additional items acquired from endoscopic indices used in the assessment of inflammatory bowel disease activity or considered to be of possible relevance by the experts. RAND appropriateness methodology uses a modified Delphi panel approach to combine the best available evidence with the personal clinical experience of relevant experts.[11] The use of a modified Delphi panel to facilitate decision-making is a widely accepted, iterative, evidence-based process.

First-round evaluation of appropriateness. Items measuring endoscopic pouchitis disease activity identified from the systematic review and the additional exploratory items were circulated to panelists in the form of an anonymous online survey. Definitions acquired from existing, validated scoring systems were supplied for additional clarity where possible.[12,13] Panelists rated the appropriateness of each item for the measurement of pouchitis disease activity on a scale from 1 to 9 (1 = inappropriate, 9 = highly appropriate).

Panel meeting. Results of the initial survey were distributed to and discussed with the panelists via a moderated teleconference to identify and examine areas of disagreement on appropriateness of the items and to allow panelists to explain the rationale behind their initial responses. Although this process focused on detecting consensus among panelists, in accordance with RAND appropriateness methodology, no attempt was made to force the panel to consensus.

Minor modifications were made to the questionnaire to improve the clarity of item definitions based on the outcomes of the first panel meeting. Then, the appropriateness of the modified items to assess endoscopic pouchitis disease activity was re-evaluated in a second round of panel review.

A final survey was prepared based on discussion of the results of the reliability study, with a focus to identify sources of disagreement among central readers in the interpretation or assessment of endoscopic pouchitis disease