Feasibility and Validity of the Pediatric Ulcerative Colitis Activity Index in Routine Clinical Practice

Jennifer L. Dotson, MD, MPH,
Division of Pediatric Gastroenterology, Hepatology and Nutrition, Nationwide Children's Hospital, 700 Children's Drive, Columbus, OH 43205

Wallace V. Crandall, MD,
Division of Pediatric Gastroenterology, Hepatology and Nutrition, Nationwide Children's Hospital, 700 Children's Drive, Columbus, OH 43205

Peixin Zhang, PhD,
Children's Hospital of Philadelphia, 34th Street and Civic Center Boulevard, Philadelphia, PA 19104

Christopher B Forrest, MD, PhD,
Children's Hospital of Philadelphia, 34th Street and Civic Center Boulevard, Philadelphia, PA 19104

L. Charles Bailey, MD, PhD,
Children's Hospital of Philadelphia, 34th Street and Civic Center Boulevard, Philadelphia, PA 19104

Richard B. Colletti, MD, and
University of Vermont, Department of Pediatrics, 89 Beaumont Avenue, Burlington, VT 05405

Michael D. Kappelman, MD, MPH
University of North Carolina at Chapel Hill, Department of Pediatrics, Division of Gastroenterology and Hepatology, 260 MacNider Building, Chapel Hill, NC 27599

Jennifer L. Dotson: jennifer.dotson@nationwidechildrens.org; Wallace V. Crandall: wallace.crandall@nationwidechildrens.org; Peixin Zhang: zpx8069@hotmail.com; Christopher B Forrest: FORRESTC@email.chop.edu; L. Charles Bailey: BAILEYC@email.chop.edu; Richard B. Colletti: richard.colletti@med.uvm.edu; Michael D. Kappelman: michael_kappelman@med.unc.edu

Abstract

Objectives—The Pediatric Ulcerative Colitis Activity Index (PUCAI) is a non-invasive disease activity index developed as a clinical trial endpoint. More recently, practice guidelines have recommended the use of PUCAI in routine clinical care. We therefore sought to evaluate the

feasibility, validity and responsiveness of PUCAI in a large, diverse collection of pediatric gastroenterology practices.

Methods—We extracted data from the two most recent encounters for patients with ulcerative colitis in the ImproveCareNow registry. Feasibility was determined by the percent of patients for whom all PUCAI components were recorded, validity by correlation of PUCAI scores across Physician Global Assessment (PGA) categories, and responsiveness to change by the correlation between the change in PUCAI and PGA scores between visits.

Results—2503 patients were included (49.5% male, age 15.2±4.1 years, disease duration 3.7±3.2 years). All items in the PUCAI were completed for 96% of visits. PUCAI demonstrated excellent discriminatory ability between remission, mild and moderate disease; discrimination between moderate and severe disease was less robust. There was good correlation with PGA [r=0.76 (p<0.001), weighted kappa k=0.73 (p<0.001)]. The PUCAI change scores correlated well with PGA change scores (p<0.001). Test-retest reliability of the PUCAI was good (intra-class correlation coefficient=0.72 [95% CI 0.70–0.75], p<0.001). Guyatt's responsiveness statistic was 1.18 and the correlation of ΔPUCAI with ΔPGA was 0.69 (p<0.001).

Conclusions—The PUCAI is feasible to use in routine clinical settings. Evidence of its validity and responsiveness support its use as a clinical tool for monitoring disease activity for patients with ulcerative colitis.

Introduction

Ulcerative colitis (UC) is a chronic, relapsing inflammatory disorder affecting the colon. In routine clinical practice, gastroenterologists often use the Physician Global Assessment (PGA) as a way to classify disease activity. PGA is based on thephysician's subjective clinical impression of the patient, taking into account factors such as abdominal pain, stool characteristics, fatigue, abdominal tenderness, and available lab tests. PGA is classified as either remission (inactive), mild, moderate or severe. The Pediatric Ulcerative Colitis Activity Index (PUCAI)[1] was developed to provide a more objective assessment of disease activity in ulcerative colitis (the PUCAI components and corresponding scoring is illustrated in Table 1). The PUCAI underwent a rigorous development process, and has been shown to be a valid, responsive tool. Indeed, the PUCAI has been widely-adopted by clinical researchers as a non-invasive measure of disease activity[2–7].

Use of the PUCAI has also been recommended in the routine clinical management of pediatric patients with UC, and has been incorporated into recent clinical guidelines[2,8]. Although the use of PUCAI has been evaluated in single-center and small multi-center research studies[1,5,9–12], little is known about the feasibility and performance of PUCAI when used in routine clinical practice. We therefore sought to evaluate the feasibility, validity and responsiveness to clinical change of the PUCAI in a large, diverse collection of pediatric gastroenterology practices. The specific objectives of this study were to: 1.)

evaluate the feasibility of using PUCAI in routine clinical practice settings, 2.) evaluate the cross-sectional and longitudinal associations between PUCAI and PGA.

Materials and Methods

ImproveCareNow (ICN) is a network of pediatric gastroenterology practices established in 2007 to improve the health and healthcare of children and adolescents with IBD. The ICN registry contains demographic, disease and treatment data collected prospectively and longitudinally during outpatient encounters. Data collection at each center is embedded into the clinical workflow, generally through the use of a paper-based or electronic health record (EHR) clinic template. During the time of this study, a study coordinator at most sites subsequently entered this data into a web-based registry form. At a small number of centers, data from the EHR template was Extracted, Transformed, and Loaded (ETL) into the registry. Patients were diagnosed and managed according to the usual practice of the primary gastroenterologist and quality improvement methodologies were applied within centers.

In this retrospective analysis, we extracted data from the two most recent encounters for all patients with ulcerative colitis in the ICN registry (September 2006 to December 2012). Data elements used in this study included basic demographics, disease duration, disease extent as defined by Paris classification,[13] laboratory studies (hematocrit, erythrocyte sedimentation rate (ESR), c-reactive protein (CRP), albumin), PGA, medications and PUCAI measurement components (abdominal pain, rectal bleeding, consistency of most stools, the total number of stools, nocturnal stools, and activity level).

Feasibility

The feasibility is a measure used to determine the practicality of collecting the PUCAI components. It was determined by the percent of patients for whom all PUCAI components were recorded at the most recent visit.

Validity

The validity is the extent to which a tool measures what it claims to measure in a real world setting. To assess validity, we first evaluated the distribution of PUCAI scores across categories (remission, mild, moderate, severe) of the Physician Global Assessment (PGA) using boxplots and then compared the differences using the Kruskal-Wallis test. We then calculated Pearson's correlation coefficient to evaluate the correlation between PUCAI and PGA. Finally, to measure the agreement between the disease severity captured by PUCAI and that captured by PGA, we used the percentage of agreement and Cohen's Kappa coefficients, based on the following published PUCAI cut-points: remission <10, mild 10–34, moderate 35–64, and severe 65–85.[1,9] Specifically, the absolute agreement was assessed using the unweighted Kappa; whereas, a Kappa coefficient with quadratic weighting was calculated for the relative concordance between the two measures.

Responsiveness

The responsiveness of an instrument is its ability to detect minimal clinically important differences and is directly related to the magnitude of that change[14–16]. The responsiveness in this study reflects the extent to which PUCAI changes in relation to a corresponding change in PGA (reference measure) over time. A successful approach to defining the change in PGA for longitudinal analysis has been previously documented[9,17,18]. A <u>small change</u> in PGA between baseline and repeated visits was defined by a change in one category (e.g. severe to moderate, mild to moderate, mild to remission). A <u>moderate change</u> in PGA was defined as a change in 2 categories (e.g. moderate to remission, mild to severe), and a <u>large change</u> as 3 categories (e.g. severe to remission).[18]

We evaluated the distribution of the change in PUCAI according to the change in PGA using boxplots, and compared PUCAI across change in PGA subgroups using the Kruskal-Wallis test. We then assessed the test-retest reliability of the assigned PUCAI values of each visit for patients whose disease assessment (PGA) remained unchanged[18]. This was assessed using the intra-class correlation coefficient (ICC) using ANOVA. The PUCAI change score was determined by subtracting the follow-up PUCAI score from the previous visit PUCAI score.

Different statistical methods were used to evaluate the responsiveness of PUCAI. Following the distributional-based approach, we calculated Guyatt's responsiveness statistic with the minimal important difference (MID). The MID was set as the PUCAI change score associated with the highest sensitivity and specificity to distinguish changed versus unchanged patients.[10,14] Additionally, we used Pearson's correlation coefficients to evaluate associations between change in PUCAI score and change in PGA.

A *p*-value of 0.05 was used to determine significance. Analyses were performed using R version 2.14.1. The analysis of this data set was considered exempt by the Institutional Review Board at Children's Hospital of Philadelphia.

Results

Our study population included 2503 pediatric patients with ulcerative colitis from 35 centers. The population was composed of 49.4% males (n = 1237), with a mean age of 15.2 ± 4.1years (median 15.9 years), and mean disease duration of 3.7 ± 3.2 years. Additional demographic and clinical details are provided in Table 2.

Feasibility

Over 96% of visits contained all six required components to calculate a PUCAI, with approximately 2% missing only 1 component and 2% missing 2 or more components. (Supplemental Table 1).

Validity

The study sample that contributed to the validity analysis consisted of patients with both a recorded PUCAI and PGA for the visit (n=2355, 94%). Of these patients, 1654 (70%) had a PUCAI <10 (remission), 496 (21%) had a PUCAI between 10–34 (mild disease activity),

176 (8%) had a PUCAI between 35–64 (moderate disease activity), and 29 (1%) had a PUCAI between 65–85 (severe disease activity). The PUCAI demonstrated discriminatory ability between disease classification based on PGA (Figure 1, $p < 0.001$). There was good correlation with PGA [$r = 0.76$ ($p < 0.001$), agreement 77%, Cohen's unweighted kappa statistic $k = 0.53$ ($p < 0.001$) and weighted kappa $k = 0.73$ ($p < 0.001$)].

Responsiveness

A total of 1874 patients had both PUCAI and PGA recorded for the two most recent visits (of the 2355 patients, 481 (20%) did not have a prior visit). The mean number of days between the current and previous visit was 166 days (SD ± 149, median = 133, IQR 75 to 198, range 1 to 1846). The change in PUCAI scores differentiated well the different categories of change in PGA (Figure 2; Table 3; Kruskal-Wallis with degree of freedom 6, $p < 0.001$).

There were 1236 patients whose disease severity (judged by PGA) at the most recent and the previous visit remained unchanged. For these 1236 patients, the distribution of the PGA-based disease severity was 1040 (84%) in remission, 145 (12%) with mild disease activity, 44 (4%) with moderate disease activity and 7 (<1%) with severe disease activity. The test-retest reliability of the PUCAI (median change in PUCAI is 0) was good (ICC = 0.72 [95% CI 0.70–0.75], $p < 0.001$). We defined improved and worsened disease activity as at least a moderate change. Stable was defined as a small change or no change. The MID of small change in PUCAI, which was used to calculate the responsiveness statistic (Table 3), was 10 points (for improvement: sensitivity 78.4%, specificity 77.7%; area under the ROC curve 0.819 [95% CI, 0.790 – 0.848]; for worsening: sensitivity 82.7%, specificity 80.3% and area under the ROC curve 0.852 [95% CI, 0.822 – 0.882]). The correlation of PUCAI change scores with PGA change scores was 0.69 ($p < 0.001$).

Discussion

This is the first large-scale, multicenter evaluation of the feasibility, validity and responsiveness to change of the PUCAI in relation to PGA in a simulated-routine clinical practice setting. The feasibility of using PUCAI in an outpatient clinical practice setting was excellent. Over 96% of visits contained all six required components to calculate a PUCAI. Via several analytic methods, we were able to demonstrate high correlation of the PUCAI with PGA. The test-retest reliability of the PUCAI was also quite good. This study extends the foundational work completed by Turner, et al by including a large sample size and diversity of practice sites (approximately 2000 patients from 35 centers).[18] The PUCAI differentiated very well among the four PGA-based disease severity categories, with fairly clear separation between disease categories. The PUCAI change scores also differentiated well among different categories (no change, small, moderate, large) of change in PGA. A small change in PUCAI (indicated with a 10 point change) gave a sensitivity and specificity of approximately 80%.

The moderately strong associations between PGA and PUCAI observed in our cross-sectional and longitudinal analyses are in line with prior work, and not unexpected given that many clinicians base their PGA on a careful history, including many of the components

of the PUCAI. The weighted kappa, k = 0.733, observed here is consistent with previously published results by Lee et al (k = 0.71).[12] The ICC observed here was slightly lower than that reported by Turner et al[18] (72% versus 89%), likely owing to the diversity of practices and physicians included in this network, along with variable degrees of IBD expertise.

These findings have several important implications for clinical care. The high feasibility and validity of the PUCAI in routine clinical practice provide strong support for the use of this instrument as a clinical tool, including serving as a basis for inpatient and outpatient care algorithms. It is worth emphasizing that while PUCAI is not an objective measure of inflammation, like colonoscopy or non-invasive biochemical markers (i.e. calprotectin), it is a more standardized approach to measuring disease activity than PGA. Indeed, the wide variability (distribution) of PUCAI scores within categories of PGA depicted in the box plot figures illustrates the subjectivity of the PGA and underscores the importance of a more standardized, yet practical and feasible, measure such as PUCAI. The role of PUCAI in clinical care has been recently highlighted in study of hospitalized patients demonstrating that PUCAI scores on hospital days 3 and 5 of corticosteroid therapy can predict patients that are likely to fail corticosteroid therapy, and thus should proceed with salvage therapy (in this case infliximab) provides additional support for its use in clinical care.[19]

There are several limitations of our study. First, ImproveCareNow sites are a self-selected group of practices who participate in a voluntary quality improvement network and collect standardized data in the context of routine clinical care. Additionally, all centers receive data quality reports, including metrics on data completeness, and sites are encouraged to improve the quality of both data collection and clinical care. Hence, our findings may not be applicable to all pediatric GI practices. However, as mentioned above, the diversity of practices and IBD volume/expertise in the network supports the generalizability of our study. Furthermore, the paper-based or EHR clinic templates used in the network to can be easily embedded into the clinical workflow of any pediatric GI practice. Another limitation is the small sample size at the periphery of the distribution of the change in PGA categories (both large worsening and large improvement categories each had 5 patients). This is because these data were derived from an outpatient database, and thus few patients were classified as severe. (Patients with severe disease are more likely to be hospitalized). Thus, caution should be exercised when interpreting the results pertaining to these large changes. An additional limitation is the possibility of data entry errors, as would be expected in a clinical registry comprised of 2503 patients with UC and 11,734visits, which might account for some of the extreme outliers seen in our figures. Finally, it is possible that since the PUCAI items were recorded by the same physician that assigned the PGA, we may have overestimated the association between PUCAI and PGA. However, the PUCAI score itself was not routinely calculated by providers and therefore this should not affect the assignment of a PGA any more than taking a good patient history.

In this study, we did not evaluate associations between laboratory values and PUCAI, as the associations between these measures have been previously described as part of the original validation study by Turner et. al.[18] Similarly, we did not evaluate associations between PUCAI and non-invasive markers of inflammation such as colonoscopy and calprotectin, given that the fundamental advantage of the PUCAI is that its non-invasive nature makes it

suitable for the repeated, often frequent, clinical assessments which are a necessary component of clinical care. Furthermore, in the PUCAI validation study, the observed correlation of the PUCAI with macroscopic mucosal inflammation was sufficiently strong to allow measurement of disease activity without endoscopic assessment.[12,17] This is important, as obtaining frequent endoscopic assessment and/or measurement of calprotectin is neither practical nor cost-efficient.

In summary, this is the first large-scale, multicenter evaluation of the PUCAI used by hundreds of pediatric gastroenterologists among dozens of centers in a simulated routine (based on registry data) clinical practice setting in which we conclude that the PUCAI is highly feasible, valid and responsive to change. This study expands prior work as it examines the use of PUCAI in a much more diverse group of patients, providers and centers. This study demonstrated additional evidence to support the generalizability and ease of use of the PUCAI. These findings support the use of PUCAI as a clinical tool, as well as a research tool.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Turner D, Otley AR, Mack D, et al. Development, validation, and evaluation of a pediatric ulcerative colitis activity index: a prospective multicenter study. Gastroenterol. 2007; 133(2):423–432.

2. Turner D, Travis SP, Griffiths AM, et al. Consensus for managing acute severe ulcerative colitis in children: a systematic review and joint statement from ECCO, ESPGHAN, and the Porto IBD Working Group of ESPGHAN. Am J Gastroenterol. 2011; 106(4):574–588. [PubMed: 21224839]

3. Gray FL, Turner CG, Zurakowski D, et al. Predictive value of the Pediatric Ulcerative Colitis Activity Index in the surgical management of ulcerative colitis. J Pediatr Surg. 2013; 48(7):1540–1545. [PubMed: 23895969]

4. Teitelbaum JE, Rajaraman RR, Jaeger J, et al. Correlation of health related quality of life in children with inflammatory bowel disease, their parents and physician as measured by a visual analog scale. J Pediatr Gastroenterol Nutr. 2013; 57(5):594–597. [PubMed: 23752079]

5. Turner D, Griffiths AM, Veerman G, et al. Endoscopic and clinical variables that predict sustained remission in children with ulcerative colitis treated with infliximab. Clin Gastroenterol Hepatol. 2013; 11(11):1460–1465. [PubMed: 23672831]

6. Watson S, Pensabene L, Mitchell P, et al. Outcomes and adverse events in children and young adults undergoing tacrolimus therapy for steroid-refractory colitis. Inflamm Bowel Dis. 2011; 17(1):22–29. [PubMed: 20722055]

7. Turner D, Griffiths AM. Acute severe ulcerative colitis in children: a systematic review. Inflamm Bowel Dis. 2011; 17(1):440–449. [PubMed: 20645317]

8. Turner D, Levine A, Escher JC, et al. Management of pediatric ulcerative colitis: joint ECCO and ESPGHAN evidence-based consensus guidelines. J Pediatr Gastroenterol Nutr. 2012; 55(3):340–361. [PubMed: 22773060]

9. Turner D, Griffiths AM, Walters TD, et al. Mathematical weighting of the pediatric Crohn's disease activity index (PCDAI) and comparison with its other short versions. Inflamm Bowel Dis. 2012; 18(1):55–62. [PubMed: 21351206]

10. Turner D, Schunemann HJ, Griffith LE, et al. Using the entire cohort in the receiver operating characteristic analysis maximizes precision of the minimal important difference. J Clin Epidemiol. 2009; 62(4):374–379. [PubMed: 19013766]

11. Turner D, Griffiths AM, Mack D, et al. Assessing disease activity in ulcerative colitis: patients or their physicians? Inflamm Bowel Dis. 2010; 16(4):651–656. [PubMed: 19708058]

12. Lee JJ, Colman RJ, Mitchell PD, et al. Agreement between patient- and physician-completed Pediatric Ulcerative Colitis Activity Index scores. J Pediatr Gastroenterol Nutr. 2011; 52(6):708–713. [PubMed: 21593644]

13. Levine A, Griffiths A, Markowitz J, et al. Pediatric modification of the Montreal classification for inflammatory bowel disease: the Paris classification. Inflamm Bowel Dis. 2011; 17(6):1314–1321. [PubMed: 21560194]

14. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. J Chronic Dis. 1987; 40(2):171–178. [PubMed: 3818871]

15. Husted JA, Cook RJ, Farewell VT, et al. Methods for assessing responsiveness: a critical review and recommendations. J Clin Epidemiol. 2000; 53(5):459–468. [PubMed: 10812317]

16. Guyatt GH, Deyo RA, Charlson M, et al. Responsiveness and validity in health status measurement: a clarification. J Clin Epidemiol. 1989; 42(5):403–408. [PubMed: 2659745]

17. Turner D, Griffiths AM, Walters TD, et al. Appraisal of the pediatric Crohn's disease activity index on four prospectively collected datasets: recommended cutoff values and clinimetric properties. Am J Gastroenterol. 2010; 105(9):2085–2092. [PubMed: 20372111]

18. Turner D, Hyams J, Markowitz J, et al. Appraisal of the pediatric ulcerative colitis activity index (PUCAI). Inflamm Bowel Dis. 2009; 15(8):1218–1223. [PubMed: 19161178]

19. Turner D, Mack D, Leleiko N, et al. Severe pediatric ulcerative colitis: a prospective multicenter study of outcomes and predictors of response. Gastroenterol. 2010; 138(7):2282–2291.
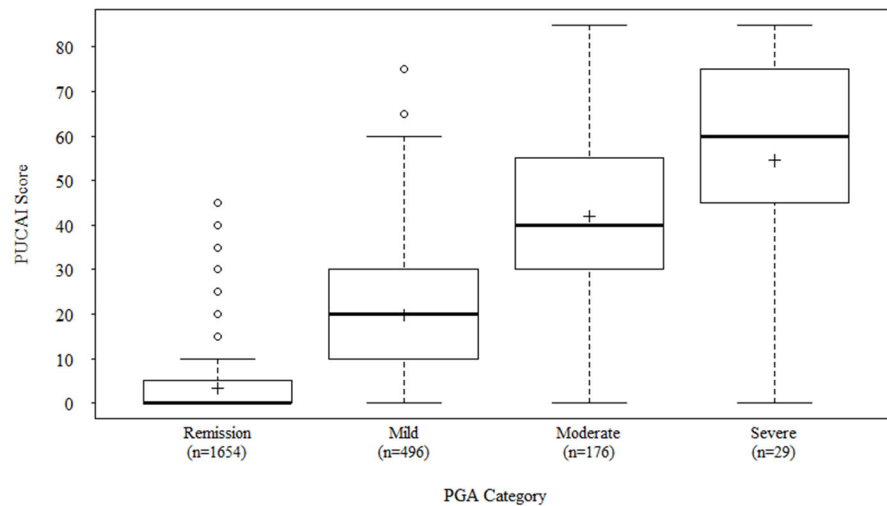
Figure 1.
Boxplot depicting the distribution of Pediatric Ulcerative Colitis Activity Index (PUCAI) scores according to physician global assessment (PGA) categories. Within the boxplots, the "+" symbol denotes the mean activity index score for each category of disease severity, the vertical boxes depict the range of scores from the 25th to 75th percentile, and the horizontal hash mark in the middle of the vertical box represents median score. The mean PUCAI($\pm$ 1 SD) for all patients whose PGA is in remission was 3.23 ± 6.52, 19.66 ± 14.32 for mild disease, 41.93 ± 19.52 for moderate disease, and 54.48 ± 23.92 for severe disease.
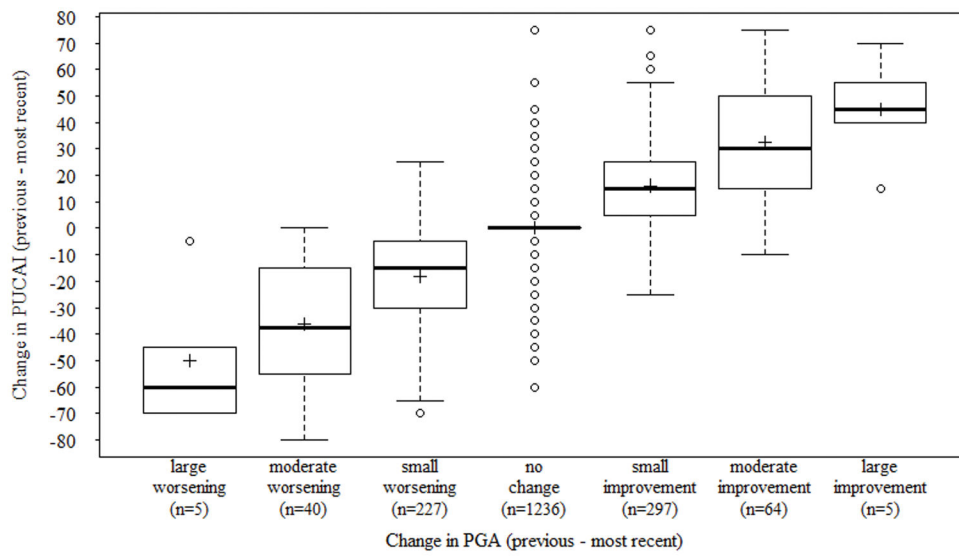
Figure 2.
Boxplot depicting the distribution of change in Pediatric Ulcerative Colitis Activity Index (PUCAI) according to the degree of change in Physician Global Assessment (PGA) categories.

Table 1

The Pediatric Ulcerative Colitis Activity Index Components[1]

Item	Points
1. Abdominal pain	
No pain	0
Pain can be ignored	5
Pain cannot be ignored	10
2. Rectal bleeding	
None	0
Small amount only, in <50% of stools	10
Small amount with most stools	20
Large amount (>50% of stool content)	30
3. Stool consistency of most stools	
Formed	0
Partially formed	5
Completely unformed	10
4. Number of stools per 24 hours	
0–2	0
3–5	5
6–8	10
>8	15
5. Nocturnal stools (any episode causing wakening)	
No	0
Yes	10
6. Activity level	
No limitation of activity	0
Occasional limitation of activity	5
Severe restricted activity	10
Sum of PUCAI	**0–85**

Table 2

Patient Demographics at the Most Recent Visit

Variable	n (%)
Total number of patients	2503
Gender	
Male	1237 (49.4)
Age (mean ± SD)	15.2 ± 4.1 years
Race/Ethnicity	
White	1920 (81.9)
Black	199 (8.5)
Hispanic or Latino	87 (3.7)
Asian	42 (1.8)
Other	97 (4.1)
Disease duration	3.7 ± 3.2 years
Paris Classification (n=1773 (70.8%))	
E1: ulcerative proctitis	154 (8.7)
E2: left-sided (distal to splenic flexure)	330 (18.6)
E3: extensive (hepatic flexure distally)	135 (7.6)
E4: pancolitis (proximal to hepatic flexure)	1154 (65.1)
Laboratory studies	
Mean Hematocrit (n=1983)	38.4 ± 4.5 in %
Mean ESR (n=1752)	14.0 ± 14.2 mm/h
Mean CRP (n=1527)	1.3 ± 5.6 mg/dL
Mean Albumin (n=1900)	4.3 ± 0.9 g/dL
PGA	
Remission	1703 (70.0)
Mild	518 (21.3)
Moderate	183 (7.5)
Severe	30 (1.2)
Medications	
5-ASA	1776 (71.1)
Prednisone	411 (16.5)
6-MP/AZA	850 (34.0)
Methotrexate	91 (3.7)
Anti-TNFα	400 (16.0)

PGA = physician global assessment; 5-ASA = 5-aminosalicylates; 6-MP=6-mercaptopurine; AZA=azathioprine, biologics=infliximab, adalimumab, certolizumab

Table 3

PUCAI Responsiveness Analysis

Median change in PUCAI (IQR)	
[a] Improved (n = 69)	35 (−10 to 75)
[a] Stable (n = 1760)	0 (−5 to 5)
[a] Worsened (n = 45)	−40 (−80 to 0)
[b] Responsiveness statistic [MID/SD$_{unchanged}$]	1.18
Correlation with change in PGA	0.69 (p < 0.001)

[a]Groups categorized based on change in PGA and the corresponding statistics represent the median with IQR in parentheses

[b]SD$_{unchanged}$ is the standard deviation of change in PUCAI for all the patients whose PGA is unchanged; Effect sizes are defined as: 0.2–0.5 = small effect, 0.5–0.8 = moderate effect, and >0.8 = large effect

PUCAI = Pediatric Ulcerative Colitis Activity Index; IQR = interquartile range; MID=minimal important difference; SD = standard deviation; PGA = Physician Global Assessment