Appraisal of the Pediatric Ulcerative Colitis Activity Index (PUCAI)

Dan Turner, MD, PhD,[1,2] Jeffrey Hyams, MD,[3] James Markowitz, MD,[4] Trudy Lerer, MS,[3] David R. Mack, MD,[5] Jonathan Evans, MD,[6] Marian Pfefferkorn, MD,[7] Joel Rosh, MD,[8] Marsha Kay, MD,[9] Wallace Crandall, MD,[10] David Keljo, MD,[11] Anthony R. Otley, MD, MSc,[12] Subra Kugathasan, MD,[13] Ryan Carvalho, MD,[14] Maria Oliva-Hemker, MD,[15] Christine Langton, MSW, MPH,[3] Petar Mamula, MD,[16] Athos Bousvaros, MD, MPH,[17] Neal LeLeiko, MD, PhD,[18] and Anne M. Griffiths, MD,[2] for the Pediatric IBD Collaborative Research Group

Background: We evaluated the psychometric performance of the Pediatric Ulcerative Colitis Activity Index (PUCAI) in a real-life cohort from the Pediatric IBD Collaborative Research Group.

Methods: Two consecutive visits of 215 children with ulcerative colitis (UC) were included (mean age 11.2 ± 3.6 years; 112 (52%) males; 63 (29%) newly diagnosed and the others after disease duration of 24 ± 15.6 months). Validity was assessed using several constructs of disease activity. Distributional and anchor-based strategies were used to assess the responsiveness of the PUCAI to change over time following treatment.

Results: Reflecting feasibility, 97.6% of 770 eligible registry visits had a completed PUCAI score versus only 47.6% for a contemporaneously collected Pediatric Crohn's Disease Activity Index (odds ratio = 45.8, 95% confidence interval [CI] 28.6–73.5) obtained for children with Crohn's disease accessioned into the same database. The PUCAI score was significantly higher in patients requiring escalation of medical therapy (45 points [interquartile range, IQR, 30–60]) versus those who did not, (0 points [IQR 0–10]; $P < 0.001$), and was highly correlated with physician's global assessment of disease activity ($r = 0.9$, $P < 0.001$). The best cutoff to differentiate remission from active disease was 10 points (area under receiver operating characteristic curve [AUC] 0.94; 95% CI 0.90–0.97). Test–retest reliability was excellent (intraclass correlation coefficient = 0.89; 95% CI 0.84–0.92, $P < 0.001$) as well as responsiveness to change (AUC 0.96 [0.92–0.99]; standardized response mean 2.66).

Conclusion: This study on real-life, prospectively obtained data confirms that the PUCAI is highly feasible by virtue of the noninvasiveness, valid, and responsive index. The PUCAI can be used as a primary outcome measure to reflect disease activity in pediatric UC.

(*Inflamm Bowel Dis 2009;15:1218–1223)*

Key Words: PUCAI, pediatrics, ulcerative colitis, validity, responsiveness, feasibility

It is mandated that the rapidly emerging novel therapies for inflammatory bowel diseases (IBDs) be evaluated in the clinical trial setting. Robust outcome measures are of utmost importance in determining the outcome of these trials. No single clinical or biochemical parameter consistently reflects activity of intestinal inflammation and, thus, multi-attribute measures of disease activity have been developed. Although other Crohn's disease (CD) activity indices have been used,[1] the Crohn's Disease Activity Index (CDAI) is generally accepted as the standard clinical outcome measure in adult CD trials.[2] The Pediatric Crohn's Disease Activity Index (PCDAI) has become the accepted disease activity measure in childhood CD.[3] In contrast to CD and until recently, no single multi-attribute measure has been consistently employed in either adult or pediatric ulcerative colitis (UC).[4]

In 2007 a Pediatric UC Activity Index (the PUCAI) was developed and validated using prospectively enrolled cohorts

Received for publication November 20, 2008; Accepted December 1, 2008.

From the [1]Shaare Zedek Medical Center, Jerusalem, Israel, [2]Hospital for Sick Children, Toronto, Ontario, Canada, [3]Connecticut Children's Medical Center, Hartford, Connecticut, [4]North Shore-Long Island Jewish Health System, New Hyde Park, New York, [5]Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada, [6]Nemours Children's Clinic, Jacksonville, Florida, [7]Riley Hospital for Children, Indianapolis, Indiana, [8]Goryeb Children's Hospital/Atlantic Health, Morristown, New Jersey, [9]Children's' Hospital, Cleveland Clinic, Cleveland, Ohio, [10]Nationwide Children's Hospital, Columbus, Ohio, [11]Children's Hospital of Pittsburgh, Pittsburgh, Pennsylvania, [12]IWK Health Centre, Halifax, Nova Scotia, Canada, [13]Medical College of Wisconsin, Milwaukee, Wisconsin, [14]Childrens Medical Center of Dayton, Ohio, [15]Johns Hopkins Hospital, Baltimore, Maryland, [16]Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, [17]Children's Hospital Boston, Boston, Massachusetts, [18]Hasbro Children's Hospital, Providence, Rhode Island.

The registry is maintained with support from Centocor, Inc., AstraZeneca, PLC, and the participating institutions.

Reprints: Dan Turner, MD, PhD, Pediatric Gastroenterology Unit, Shaare Zedek Medical Center, POB 3235, Jerusalem 91031, Israel (e-mail: turnerd@szmc.org.il).

DOI 10.1002/ibd.20867

Published online 22 January 2009 in Wiley InterScience (www.interscience.wiley.com).

of children with UC (see Appendix).[5] The PUCAI, lacking invasive items, is suitable for longitudinal use in clinical trials and for determining timely introduction of second-line therapy in severe acute UC.[6] Recently, the FDA endorsed the PUCAI as a substitute to endoscopic evaluation for the primary outcome measure in a pediatric clinical trial evaluating a 5-aminosalicylate (5-ASA) regimen.

Despite the encouraging initial results, the original report of an outcome measure often presents an optimistic reflection of its performance,[7] and thus, duplication of the results in independent populations is necessary. Moreover, validity is not a property of the instrument but a property of how it is used, so it is important to reassess validity in different populations and setups.[8] We aimed to use a prospective multicenter pediatric IBD registry to determine the feasibility of the PUCAI in real life and to assess its validity and responsiveness for measuring disease activity in pediatric UC.

MATERIALS AND METHODS

Patients

Data generated from the database of the Pediatric Inflammatory Bowel Disease Collaborative Research Group Registry were used. This registry was initiated in January 2002 by 18 pediatric gastroenterology centers in North America as a method of describing the natural history of IBD in newly diagnosed children. Data are recorded at the time of initial diagnosis, 30 days after diagnosis, and quarterly thereafter on standardized forms and transmitted to a central data repository. Patients are managed according to the discretion of their physicians. Approval for the registry has been received from the research ethics committee at each participating institution and informed consent was obtained from all families.

Recorded data include basic and demographic characteristics, physician global assessment (PGA) of disease activity (inactive, mild, moderate, and severe), blood test results, therapies, the need for change in therapy, and admissions. Since November 2006 the attending physician was asked to score the items of the PUCAI on all UC patients. The PUCAI is a 6-item disease activity index intended for use in pediatric UC clinical trials with a score range of 0–85 (see Appendix). Development, weighting, and validation were performed using combined judgmental and mathematical strategies utilizing a Delphi group and prospective cohort of 205 children.[5] For this study, data of children (0–18 years of age) with a confirmed diagnosis of UC collected at first visit after November 2006 were included. Responsiveness was assessed using data collected at the subsequent visit.

Analytic Approach and Statistics

To assess the feasibility of the PUCAI we calculated the rate of complete PUCAI assessments of all eligible visits recorded by the registry from January 1, 2007 to February 29, 2008. This was compared with the corresponding completion rate of the PCDAI in contemporaneously evaluated children with CD accessioned into the same pediatric IBD database (total of 770 eligible UC visits and 1961 CD visits, including repeated measures).

Validity is the degree to which the instrument measures the concept that it purports to measure and includes construct, predictive, and criterion validity.[9] In this study we used construct validation based on association between the PUCAI score and PGA and the need for change in medical therapy. Correlation of the PUCAI with erythrocyte sedimentation rate (ESR), C-reactive protein (CRP), hemoglobin, and albumin was expected to be only fair since laboratory tests do not reflect well disease activity in UC.[10] Spearman's or Pearson's correlations were used as appropriate for the distribution normality.

Agreement between baseline and follow-up PUCAI values of patients who remained unchanged (test–retest reliability) was assessed using the intraclass correlation coefficient (ICC, using Shrout and Fleiss's 2-way random analysis of variance [ANOVA] model ± 95% confidence interval [CI][11] reporting the "average measures" value in the SPSS [Chicago, IL] output).

Responsiveness is the ability of an instrument to accurately detect change in disease activity over time, when it occurs. The short-term responsiveness of the PUCAI was assessed by comparing the PUCAI scores at the baseline visit and the subsequent one (1–3 months apart). A change in the PUCAI scores was determined by subtracting the follow-up score from the initial score (labeled ΔPUCAI). Patients were classified as those having an important change in disease activity or being stable in 2 different ways: 1) *Anchor-based approach*: patients who changed by 1 category on the PGA (e.g., from "moderate disease activity" at the first visit to "mild" at the second visit) were considered as experiencing small change, whereas a change in 2 categories as a moderate change, and at least 3 categories as a large change; 2) *Data-driven approach:* all patients commenced on therapy of proven efficacy (i.e., new or increased dose of steroids, anti-TNF therapy, or oral 5-ASA) were considered as changed and those with unchanged medical therapy were considered unchanged, regardless of the actual change judged by the PGA score.

Responsiveness was assessed using distributional, correlational, and anchor-based approaches using several different statistical methods.[12]

(1) *Distributional-based approach:* Effect size statistics represent the magnitude of change in the evaluated index (ratio of signal [i.e., observed change] to noise [i.e., some measure of variance]). A higher effect size statistic indicates a greater change effect, and as a general rule >0.8 is considered a large effect, 0.5–0.8 moderate, and 0.2–0.5 small (13).

The following statistical tests were used (Table 2): i) standardized response mean (SRM)[14]; ii) standardized effect size (SES)[15]; and iii) Guyatt et al's responsiveness statistic[16] substituting 20 for the minimal important difference (MID).[5]

(2) *Anchor-based approach*: To differentiate patients who experienced at least moderate change from those who remained unchanged, receiver operating characteristic (ROC) curves, sensitivity, and specificity were used.[12]

(3) *Correlational:* Correlation between the ΔPUCAI with ΔPGA.

The minimal important difference (MID) is the smallest change in a health-related outcome measure that, in association with minimal toxicity and cost, is large enough to trigger a change in management.[17,18] The MID determines the change required to define improvement in clinical trials. Following the anchor-based approach,[19] the MID was set as the ΔPUCAI associated with the highest sensitivity and specificity to differentiate changed versus unchanged patients, using the ROC curve approach.[13,20,21]

Serial ROC curves were used to define cutoff scores for categorical disease activity (none, mild, moderate, and severe). Area under the ROC curve (±95% CI) of over 0.7 was considered fair, 0.8 good, and at least 0.9 excellent discriminative ability.

The minimal detectable change (MDC), sometimes also termed "smallest detectable difference," is the smallest change in score that can be detected beyond random error[22–24]; anything lower than that could be the same that one would observe in stable patients. To ascertain the MDC, Jacobson's Reliable Change Index was used (RCI = $1.96*SD_{baseline}*(SQRT[2*(1-r)])$ where r is the test–retest reliability coefficient).[25,26] As previously found, RCI is the preferred way to ascertain the MDC for a clinimetric index such as the PUCAI.[19,27] The RCI has the same units as the instrument under study and it is based on the standard error of measurement.[28,29] To supplement approximation of the MDC, we also reported Norman et al's 0.5 standard deviation of the total change score.[30]

Data are presented as means ± standard deviation, or medians (interquartile range, IQR). Comparisons between PUCAI subgroups were made using the nonparametric Wilcoxon rank sum test or Kruskal–Wallis on ranks. Proportions were compared using the χ^2 or Fisher's exact as appropriate. All comparisons were made using 2-sided significance levels of $P < 0.05$ and performed using SPSS v.15.0.

RESULTS

A total of 215 children with UC were included in this study, of which 63 (29%) were enrolled at the time of diagnosis and the others after disease duration of 24 ± 15.6 months (Table 1).

TABLE 1. Patient's Characteristics at First Included Visit

	Entire Cohort (N=215)
Males	112 (52%)
Age (years)	11.2 ± 3.6
Range (years)	1.4–16.2
Race	
White	183 (85%)
Black	14 (7%)
Asian	3 (1%)
Hispanic	6 (3%)
Others	9 (4%)
Disease duration (years)	1 (0–2.5)
At disease onset	64 (30%)
Exacerbation	151 (70%)
Disease extent	
Proctitis	2 (1%)
Left-sided	40 (19%)
Extensive[a]	173 (80%)
Disease activity[b]	
Remission	100 (46%)
Mild	57 (27%)
Moderate	45 (21%)
Severe	13 (6%)
Concurrent medications[c]	
Oral 5-ASA preparation	157 (73%)
Thiopurines	80 (37%)
Anti-TNFα	17 (8%)

Medians (interquartile range) or mean (± SD) are presented as appropriate for the data distribution.
[a]Beyond the splenic flexure according to the Montreal classification (Ref. 38).
[b]As per physicians' global assessment.
[c]Started prior to or at visit date.

Feasibility

Of the 770 eligible UC visits included in this analysis, 752 (97.6%) provided complete data on all items of the PUCAI with a valid total score. In comparison, only 935 of 1961 CD visits (47.6%) had complete PCDAI data, mainly due to lack of blood tests and/or height velocity data (odds ratio [OR] = 45.8; 95% CI 28.6–73.5).

Validity

The PUCAI was highly correlated with PGA ($r = 0.90$; $P < 0.001$). The correlation did not differ between left-sided and extensive colitis ($r = 0.88$ and 0.91; $P > 0.05$) and between different age groups ($r = 0.9$ and 0.92 for those over and under 8 years of age, respectively; $P > 0.05$). As expected, the PUCAI was only fairly correlated with common blood tests including ESR ($r = 0.39$, $P < 0.001$), CRP (r

TABLE 2. Responsiveness Analysis of the PUCAI

	PUCAI
Median change (IQR)	
Improved[a]	55 (45–65)
Stable	0 (0–10)
Worsened[a]	−42.5 (−75–10)
Deemed to have improved	32.5 (5–50)
SES[b] [change/$SD_{baseline}$]	
Changed	2.60
Stable	0.36
Deemed to have changed	1.45
SRM[b] [change/$SD_{changed}$]	
Changed	2.66
Stable	0.46
Deemed to have changed	1.20
Responsiveness statistics[b] [MID/ $SD_{unchanged}$]	1.32
Correlation with change in PGA	0.84 ($P<0.001$)
AUC of ROC between changed and unchanged (95% CI)	0.96 (0.92–0.99)

SES, standardized effect size; SD, standard deviation; SRM, standardized response mean; MID, minimal important difference; PGA, physician global assessment; AUC, area under the curve; ROC, receiver operating characteristic

[a]Defined as at least moderately changed.

[b]Effect size statistics (SES, SRM, and RS) should be interpreted as follows: 0.2–0.5 small effect, 0.5–0.8 moderate effect, >0.8 large effect.

= 0.37, $P = 0.001$), hemoglobin ($r = -0.22$; $P = 0.007$), and albumin ($r = -0.57$; $P < 0.001$).

Median PUCAI scores were significantly higher in patients whose therapy escalated during that visit than those whose medical therapy decreased or remained unchanged (0 [IQR 0–10] versus 45 [30–60]; Wilcoxon Rank Sum Test, $P < 0.001$). In fact, only 4% of patients with PUCAI <10 (i.e., remission) had any escalation of therapy during that visit versus 50% of patients with mild disease (PUCAI 10–34), and 89% of patients with moderate or severe disease (PUCAI ≥35, $P < 0.001$). Overall, the PUCAI score highly predicted the need for escalating medical therapy (area under the ROC curve 0.94; 95% CI 0.90–0.98).

Cutoffs of Disease Activity

The PUCAI differentiated very well the different categories of disease activity captured by the PGA (Fig. 1; Kruskal–$Wallis_3$ $P < 0.001$; Spearman $r = 0.87$, $P < 0.001$). The best cutoff to differentiate remission from active disease was <10 points (sensitivity 89%, specificity 89%, area under ROC curve 0.94; 95% CI 0.90–0.97), to differentiate mild from moderate disease activity at least 30 points (sensitivity 95%, specificity 95%, AUC 0.98; 95% CI 0.97–1), and moderate disease activity from severe at least 65 points (sensitivity 92%, specificity 94%, AUC 0.99; 95% CI 0.97–1).

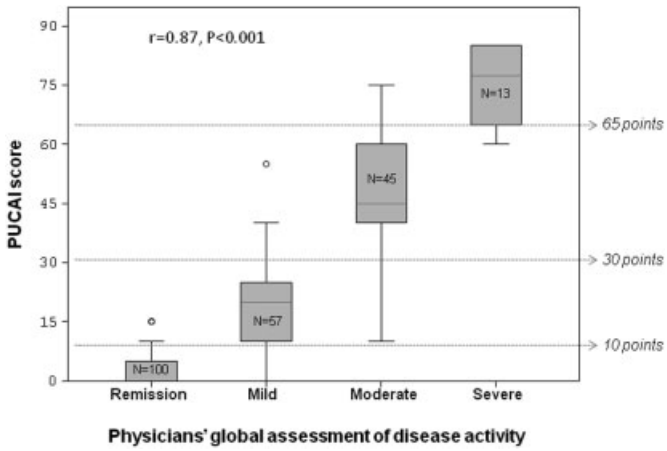

FIGURE 1. Distribution of the PUCAI according to disease activity as measured by physicians' global assessment (PGA). The numbers on the right denote the best cutoff scores to differentiate disease activity, obtained by serial ROC curves.

Test–Retest Reliability

Median PUCAI values at baseline and follow-up of 121 children who remained unchanged (judged by PGA of disease activity) were 0 points (IQR 0-10) at both times. Accordingly, ICC analysis of these patients showed excellent test–rest reliability (ICC = 0.89 [95% CI 0.84–0.92], $P < 0.001$). Subgroup analysis of 2 age groups did not differ substantially (ICC = 0.89 [95% CI 0.83–0.92] for <8 years versus 0.85 [95% CI 0.67–0.83] for >8 years of age).

Responsiveness, MDC, and the MID

A total of 213 children (99%) had a follow-up visit recorded and thus contributed to the responsiveness analysis. The ΔPUCAI differentiated well the different groups of improvement, unchanged and deterioration, as judged by the change in PGA (Fig. 2; $r = 0.87$; $Kruskal–Wallis_5$ $P < 0.001$). The 3 conceptual methods to evaluate responsiveness (i.e., distributional, anchor-based, and correlational) indicated high responsiveness of the PUCAI across all statistical methods (Table 2).

The MID of small change in PUCAI (worsening or improvement) was 10 points (for improvement: sensitivity 81% specificity 86%, area under the ROC curve 0.92 [95% CI 0.88–0.96]; for deterioration: sensitivity 91%, specificity 80% and AUC 0.93 [95% CI 0.88–0.98]). The MID for moderate improvement was at least 30 points (sensitivity 92%, specificity 90%; area under the ROC curve 0.96 [95% CI 0.92–0.99]). It was impossible to calculate the MID for moderate worsening since there were only 2 patients in this subgroup.

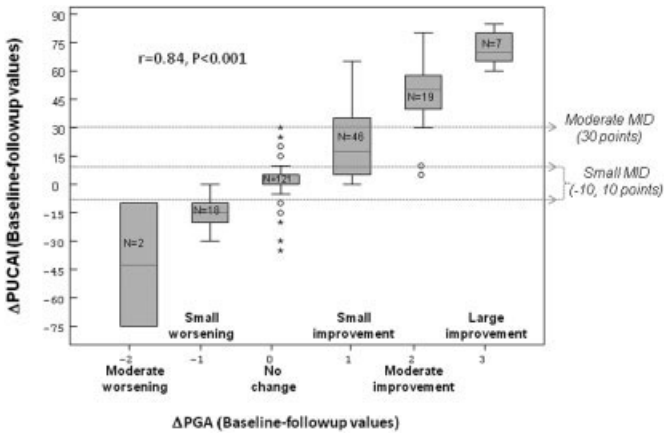

FIGURE 2. Distribution of the ΔPUCAI according to the change in physicians' global assessment (PGA). The minimal important difference (MID) was determined by ROC curve, finding the best cutoff on the PUCAI to differentiate change: small change denotes a change of 1 category on the PGA, moderate change denotes a change in 2 categories, and large change, 3 categories.

The reliable change index, reflecting the MDC, was 12.2 points (4.4*1.96*[13.2*SQRT((2*1−0.89))]). In agreement, 0.5 standard deviation of ΔPUCAI was 11.9 points. These data imply that the PUCAI unit to be considered real beyond interrater variability is 12 points, similar to the MID found for small change.

DISCUSSION

This study provided novel data indicating the high feasibility of the PUCAI and the good correlation with treatment decision-making. This study also replicated the cutoff score that defines remission (PUCAI < 10 points), and severe disease (at least 65 points). The cutoff for moderate disease was at least 30 points, similar to the 35 point value found in the original study.[5] Small variations such as this one are anticipated from inherent interpopulation variability.

It is difficult to compare the MID obtained here (10 points for small change and 30 for moderate change) with those obtained previously (10 and 20 points, respectively), since the 2 studies utilized different anchors to reflect change. In the previous study a 7-point global rating of change was used (−3 to +3 where 0 means no change), whereas here we calculated the difference of a 4-point PGA of disease activity. Global rating of change has several limitations.[31–34] Recall biases exist in comparing the current state with a previous one and retrospective estimates of change are highly correlated with the present state.[32] The rater may be influenced by recent good or bad events that falsely alter the judgment of change for the entire period. The current study used a combined estimate of both timepoints, avoiding the need for retrospective evaluation. The higher value obtained here for moderate improvement is interesting and will facilitate a final MID of the PUCAI based on more required data. In the interim we suggest leaving the current recommendation of the MID for the PUCAI at 20 points as previously described.[5]

Psychologically, patients admit to deterioration only when a substantial worsening occurs and are often looking for the smallest difference to reassure themselves of improvement. Indeed, asymmetry in MID for improvement and deterioration have been previously reported.[35] The data presented here on worsening are novel, as we did not have enough patients who deteriorated in previous studies. It seems that, for the PUCAI, MID of worsening patients mirrors the one of improvement. This is important to determine what increment in the PUCAI score reflects treatment failure and to guide, for instance, an increase in standardized steroid dose in clinical trials.

Although the data were prospectively collected, it was not collected specifically for this study. Therefore, some of the limitations inherent to a retrospective design apply also here. Only 1 physician scored the PGA and it was the same 1 who also decided on therapy. This may have biased the correlation with therapy escalation. Colonoscopic evaluation

APPENDIX. Pediatric Ulcerative Colitis Activity Index (PUCAI)

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 less than 50% of stools	10
Small amount with most stools	20
Large amount (>50% of the 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)	

For User's Guide and cutoff of disease activity, refer to the original study (Ref. 5).

is a very important variable to reflect disease activity, but it cannot be considered a gold standard. Some clinical symptoms are not directly related to the mucosal inflammation, and, on the other hand, macroscopic assessment of the degree of inflammation is subjective and endoscopic healing tends to lag behind symptom improvement.[36,37] The registry does not record colonoscopic data and, thus, this was not included here. This study emphasizes, yet again, the rarity of disease limited to the rectum in children; it is still unknown how well the PUCAI performs in patients with proctitis. Nonetheless, the results here were consistent across different statistical strategies and various constructs, using data obtained during routine clinical practice, thus reflecting "real life." Most important, this study supplements the original prospective evaluation that presented colonoscopic data.[5] Future research may include the feasibility of using the PUCAI as a routine clinical tool in a community-based setting.

The results of this study are consistent with those previously presented.[5] Replication of the psychometric properties of the PUCAI on real-life data is of utmost importance for using the PUCAI with confidence as the sole outcome measure in pediatric UC without the need for endoscopy.

REFERENCES

1. Noble A, Turner D. Clinical outcomes in pediatric inflammatory bowel disease. In: Mamula P, Markowitz J, Baldassano R, eds. Springer, NY. Pediatric Inflammatory Bowel Disease, 1st ed. 2008. p 744.
2. Best WR, Becktel JM, Singleton JW, et al. Development of a Crohn's disease activity index. National Cooperative Crohn's Disease Study. *Gastroenterology.* 1976;70:439–444.
3. Hyams JS, Ferry GD, Mandel FS, et al. Development and validation of a pediatric Crohn's disease activity index. *J Pediatr Gastroenterol Nutr.* 1991;12:439–447.
4. D'Haens G, Sandborn WJ, Feagan BG, et al. A review of activity indices and efficacy end points for clinical trials of medical therapy in adults with ulcerative colitis. *Gastroenterology.* 2007;132:763–786.
5. Turner D, Otley AR, Mack D, et al. Development and evaluation of a Pediatric Ulcerative Colitis Activity Index (PUCAI): a prospective multicenter study. *Gastroenterology.* 2007;133:423–432.
6. Turner D, Walsh CM, Benchimol EI, et al. Severe paediatric ulcerative colitis: incidence, outcomes and optimal timing for second-line therapy. *Gut.* 2008;57:331–338.
7. McDowell I, Jenkinson C. Development standards for health measures. *J Health Services Res Policy.* 1996;1:238–246.
8. DeVellis RF. A consumer's guide to finding, evaluating, and reporting on measurement instruments. *Arthritis Care Res.* 1996;9:239–245.
9. Streiner D, Norman G. Validity. Health Measurement Scales. A Practical Guide to Their Development and Use, 2nd ed. Oxford: Oxford University Press; 1995.
10. Mack DR, Langton C, Markowitz J, et al. Laboratory values for children with newly diagnosed inflammatory bowel disease. *Pediatrics.* 2007; 119:1113–1119.
11. Shrout P, Fleiss J. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull.* 1979;86:420–428.
12. Deyo RA, Centor RM. Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. *J Chronic Dis.* 1986;39:897–906.
13. Kundhal PS, Critch JN, Zachos M, et al. Pediatric Crohn Disease Activity Index: responsive to short-term change. *J Pediatr Gastroenterol Nutr.* 2003;36:83–89.
14. Liang MH, Fossel AH, Larson MG. Comparisons of five health status instruments for orthopedic evaluation. *Med Care.* 1990;28:632–642.
15. Stratford PW, Binkley FM, Riddle DL. Health status measures: strategies and analytic methods for assessing change scores. *Phys Ther.* 1996;76:1109–1123.
16. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chronic Dis.* 1987;40:171–178.
17. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials.* 1989;10:407–415.
18. Schunemann HJ, Puhan M, Goldstein R, et al. Measurement properties and interpretability of the chronic respiratory disease questionnaire (CRQ). *COPD.* 2005;2:81–89.
19. Turner D, Schünemann HJ, Griffith LE, et al. Using the entire cohort in the ROC analysis maximizes precision of the minimal important difference. *J Clin Epidemiol.* 2008 (in press).
20. Hays RD, Woolley JM. The concept of clinically meaningful difference in health-related quality-of-life research. How meaningful is it? *Pharmacoeconomics.* 2000;18:419–423.
21. Beaton DE, Boers M, Wells GA. Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. *Curr Opin Rheumatol.* 2002;14:109–114.
22. Lassere MN, van der Heijde D, Johnson KR. Foundations of the minimal clinically important difference for imaging. *J Rheumatol.* 2001;28:890–891.
23. Lassere MN, van der Heijde D, Johnson K, et al. Robustness and generalizability of smallest detectable difference in radiological progression. *J Rheumatol.* 2001;28:911–913.
24. Beaton DE. Understanding the relevance of measured change through studies of responsiveness. *Spine.* 2000;25:3192–3199.
25. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol.* 1991;59:12–19.
26. Jacobson NS, Roberts LJ, Berns SB, et al. Methods for defining and determining the clinical significance of treatment effects: description, application, and alternatives. *J Consult Clin Psychol.* 1999;67:300–307.
27. Turner D, Schünemann HJ, Griffith LE, et al. The minimal detectable change cannot reliably replace the minimal important difference. *J Clin Epidemiol.* 2008 (in press).
28. Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. *J Clin Epidemiol.* 1999;52:861–873.
29. Wyrwich KW, Nienaber NA, Tierney WM, et al. Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. *Med Care.* 1999;37:469–478.
30. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care.* 2003;41:582–592.
31. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol.* 2003;56:395–407.
32. Norman GR, Stratford P, Regehr G. Methodological problems in the retrospective computation of responsiveness to change: the lesson of Cronbach. *J Clin Epidemiol.* 1997;50:869–879.
33. Wyrwich KW, Metz SM, Kroenke K, et al. Measuring patient and clinician perspectives to evaluate change in health-related quality of life among patients with chronic obstructive pulmonary disease. *J Gen Intern Med.* 2007;22:161–170.
34. Guyatt GH, Norman GR, Juniper EF, et al. A critical look at transition ratings. *J Clin Epidemiol.* 2002;55:900–908.
35. Cella D, Hahn EA, Dineen K. Meaningful change in cancer-specific quality of life scores: differences between improvement and worsening. *Qual Life Res.* 2002;11:207–221.
36. Irvine E. Assessing outcomes in clinical trials. In: Satsangi JSL, ed. Inflammatory Bowel Disease. Philadelphia: Churchill Livingstone; 2003. p 319–333.
37. Beattie RM, Nicholls SW, Domizio P, et al. Endoscopic assessment of the colonic response to corticosteroids in children with ulcerative colitis. *J Pediatr Gastroenterol Nutr.* 1996;22:373–379.
38. Silverberg MS, Satsangi J, Ahmad T, et al. Toward an integrated clinical, molecular and serological classification of inflammatory bowel disease: report of a Working Party of the 2005 Montreal World Congress of Gastroenterology. *Can J Gastroenterol.* 2005;19(suppl A):5–36.