Mathematical Weighting of the Pediatric Crohn's Disease Activity Index (PCDAI) and Comparison with Its Other Short Versions

Dan Turner, MD, PhD, Anne M. Griffiths, MD,† Thomas D. Walters, MD,† Tong Seah,† James Markowitz, MD,‡ Marian Pfefferkorn, MD,§ David Keljo, MD,¶ Jacob Waxman,* Anthony Otley, MD, MSc,‖ Neal S. LeLeiko, MD, PhD,** David Mack, MD,†† Jeffrey Hyams, MD,‡‡ and Arie Levine, MD§§*

Background: The Pediatric Crohn's Disease Activity Index (PCDAI) has become the standard outcome measure in pediatric Crohn's disease (CD) clinical research. Other versions have been proposed but without systematic evaluation. The aim was to assess validity and responsiveness of the abbreviated PCDAI (abbrPCDAI), short PCDAI (shPCDAI), and modified PCDAI (modPCDAI) as measures of disease activity and to compare these with a mathematically weighted version developed here (wPCDAI).

Methods: The raw data from four prospectively collected datasets were used, totaling 437 children with CD (including two clinical trials). Discriminant validity utilized physician global assessment of disease activity (PGA), and construct validity the correlation with PGA and laboratory results. Feasibility and face validity were ascertained by a survey of 33 experts in pediatric CD.

Results: The wPCDAI had better performance than the PCDAI in construct validity and responsiveness and it discriminated better between the disease activity categories (area under the receiver operator characteristic [ROC] 0.97; 95% confidence interval [CI]: 0.95–0.99). In comparison to the original PCDAI, the noninvasive versions (abbrPCDAI and shPCDAI) had lower face, construct, and discriminant validity but were judged to be significantly more feasible. The modPCDAI performed well in the construct validation but was consistently inferior in all other parameters. Cutoffs that correspond to remission, response, and gradations of disease activity were determined for each index.

Conclusions: The newly weighted wPCDAI performed better than the original PCDAI and is more feasible. The noninvasive versions (shPCDAI and abbrPCDAI) are inferior to the full PCDAI, but when needed in retrospective studies either may be equally used.

(Inflamm Bowel Dis 2012;18:55–62)

Key Words: Pediatric Crohn's Disease Activity Index, PCDAI, inflammatory bowel disease, Crohn's disease, wPCDAI, Pediatrics, clinimetrics

New interventions in inflammatory bowel disease (IBD) should be tested in the clinical trial setting, where validated multiitem measures of the disease activity such as the Pediatric Crohn's Disease Activity Index (PCDAI)[1–5] should be used to assess response.

During the last 20 years, the PCDAI has become the standard of measuring disease activity in pediatric Crohn's disease (CD) but it is not without limitations. First, the inclusion of laboratory results, perianal examination, and height velocity in the PCDAI reduces its feasibility especially for retrospective review of patients' health records. Even in a prospectively collected "real-life" registry cohort, the PCDAI was scored in only 48% of eligible visits compared with 98% for the Pediatric Ulcerative Colitis Activity Index (PUCAI), which requires no laboratory values.[6] A recent study has found that data to complete the PCDAI retrospectively were available in the charts of only 20% of 3643 clinical visits.[7] Second, although the height item undoubtedly is a very important marker of disease activity in children, it is relevant only to young children in the Tanner-growing stages (until stages 2–3) and its calculation over many months reduces short-term responsiveness and discriminant validity. Acknowledging that fact, we

Additional Supporting Information may be found in the online version of this article.

Received for publication November 28, 2010; Accepted December 27, 2010.

From the *Pediatric Gastroenterology Unit, Shaare Zedek Medical Center, Hebrew University of Jerusalem, Israel, †Hospital for Sick Children, Toronto, Ontario, Canada, ‡North Shore-Long Island Jewish Health System, New Hyde Park, New York, §Riley Hospital for Children, Indianapolis, Indiana, ¶Children's Hospital of Pittsburgh, Pennsylvania, ‖IWK Health Centre, Halifax, Nova Scotia, Canada, **Hasbro Children's Hospital, Providence, Rhode Island, ††Children's Hospital of Eastern Ontario, Ottawa, Canada, ‡‡Connecticut Children's Medical Center, Hartford, Connecticut, §§Pediatric Gastroenterology Unit, Wolfson Medical Center, Tel Aviv University, Israel.

Reprints: Dan Turner, MD, PhD, Pediatric Gastroenterology and Nutrition Unit, Shaare Zedek Medical Center, The Hebrew University, P.O.B 3235, Jerusalem 91031, Israel (e-mail: turnerd@szmc.org.il)

DOI 10.1002/ibd.21649

Published online 23 February 2011 in Wiley Online Library (wileyonlinelibrary.com).

TABLE 1. Studies of the Different PCDAI Later Versions

Version	Study	Population	Validity	Suggested Cutoffs
Abbreviated PCDAI (abbrPCDAI)	Loonen 2003 (9)	n=71, data from previously reported prospective cohort (2)	AUC of ROC 0.93 to discriminated between remission from active disease	Remission <10 points
	Shepanski 2004 (10)	n=40, prospective single center cohort (5-24 years)	abbrPCDAI to PCDAI r=0.85 abbrPCDAI to IMPACT r=-0.58	—
	Kappelman 2010 (7)	n=2815 visits of approximately 600 children from a registry	abbrPCDAI to PCDAI r=0.68 abbrPCDAI to PGA r=0.64 AUC of ROC 0.82 to discriminate remission from active disease	—
Short PCDAI (shPCDAI)	Kappelman 2010 (7)	n=4241 visits of approximately 900 children from a registry	shPCDAI to PCDAI r=0.66 shPCDAI to PGA r=0.60 AUC of ROC 0.80 to discriminate remission from active disease	Remission <15 points Mild 15-20 points Moderate-severe >20 points
Modified PCDAI (modPCDAI)	Leach 2010 (11)	n=100 visits of 62 children	modPCDAI to PCDAI r=0.66 modPCDAI to PGA r=0.79 modPCDAI to calprotectin r=0.48	Remission <7.5 Mild 7.5 – 10 Moderate 12.5-17.5 Severe >17.5

AUC, area under the curve; ROC, receiver operator characteristic; PGA, physician global assessment.

recently determined that remission of the PCDAI should be defined as <10 points, or <7.5 points without the height item.[5] Third, the inclusion of perianal item is debated as it reflects a different concept than luminal disease activity. Finally, the PCDAI cannot differentiate well the moderate from the severe end of disease activity.[3,8]

To address the poor feasibility of the PCDAI, two shorter versions of the index have been published but with limited evaluation and validation. Two groups proposed an abbreviated PCDAI (abbrPCDAI), removing the height, extraintestinal manifestation and the three laboratory items (Appendix A1), Table 1 (see Supporting Information).[9,10] The remaining items of the abbrPCDAI were not reweighted. Recently, a larger study presented a short version of the PCDAI (shPCDAI), excluding items with a low frequency of completion in a patient registry (Appendix B1, Table 1) (see Supporting Information).[7] The difference between the shPCDAI from the abbrPCDAI is that the extraintestinal manifestation item has replaced the perianal item, and that new weights have been mathematically assigned to each item by multivariate modeling, reflecting their relative importance to physician global assessment (PGA) of disease activity.

A third version, a modified PCDAI (modPCDAI), was recently proposed by Leach et al[11] based on the three laboratory items of the PCDAI (i.e., hematocrit, erythrocyte sedimentation rate [ESR], and albumin) with an addition of C-reactive protein (CRP) (Appendix C1, Table 1) (see Supporting Information). It aimed to overcome the ambiguity of the subjective and anthropometric components of the full index.

In the four aforementioned studies, only limited analyses of the clinimetric properties have been performed. In addition, scarce data are available on the cutoff values that should be used to define remission and response and other gradations of disease activity (Table 1). Finally, the PCDAI itself has hitherto never been subjected to multivariate mathematical weighting and item reduction, likely due to the insufficient sample size in previous studies for this purpose. We have previously shown that mathematical weighting of a disease activity index is superior to the judgemental approach.[15]

We therefore aimed to use the raw data from four prospectively collected datasets of pediatric CD to mathematically weight items in the PCDAI. We then systematically compared this mathematically weighted PCDAI (wPCDAI) with the original PCDAI, abbrPCDAI, shPCDAI, and modPCDAI with respect to feasibility, validity, and responsiveness as measures of disease activity in pediatric CD. Cutoffs that correspond to remission, response, and gradations of disease activity were determined for each version.

MATERIALS AND METHODS

Setting and Patients

Four prospectively collected datasets of children with CD, previously used to establish the best cutoff values of the PCDAI,[5] were utilized in this study including the REACH infliximab trial, a multicenter controlled ileal release

budesonide trial, the North American Pediatric IBD Collaborative Research Group Registry, and a cohort in which linear growth is being evaluated longitudinally. The details of each datasets were previously described[5] and are mentioned in brief. The REACH study was a 1-year randomized controlled trial of induction and maintenance of remission of infliximab in children with CD aged 6–17 years with baseline PCDAI>30.[12] For this study, data generated at baseline and at week 10 were used. As agreed with the owner of the data (Centocor Ortho Biotech, Horsham, PA), in order to receive the raw data we used a random sample of 90% of the entire REACH cohort. The second cohort came from the Pediatric IBD Collaborative Research Group Registry in North America that includes data on newly diagnosed IBD children under the age of 16 years. The current study includes data from June 2005 until September 2009. The first two visits from diagnosis were included, typically at baseline and 30 days thereafter, but when data were missing we used the subsequent two visits performed quarterly. The third dataset was from an ongoing prospective inception cohort study examining linear growth of children with CD (Hospital for Sick Children, Toronto). We used data from the first two visits from diagnosis on subjects enrolled until September 2009. The last study was a randomized, double-blind, controlled study of controlled ileal release budesonide in children ages 10–19 years with mild to moderate CD.[13] Data from baseline and the 7-week visit were included.

We obtained from the owners of the datasets the raw data to allow robust pooling of the data. Moreover, we previously performed subgroup and sensitivity analyses of these datasets and found no significant variations with regard to the PCDAI outcome.[5]

Analysis

The raw data of each of the PCDAI items were available for all four datasets at both the baseline and follow-up visits, thus enabling calculation of the different abbreviated versions of the index. All analyses were performed using SPSS v. 16 (Chicago, IL).

Derivation of the Mathematically Weighted Version (wPCDAI)

In order to mathematically derive weights to the PCDAI, the cohort was randomly split to derivation (2/3 of the cohort, $n = 291$) and validation cohorts ($n = 140$ children), using the automatic function in the SPSS software. A multivariate regression model was constructed for the derivation cohort, wherein the PGA was utilized as dependent variable and the PCDAI items (entered unweighted) as the exploratory variables. The standardized regression coefficient values (i.e., β scores) guided the weight of each item. Items with a β score not significantly different than zero (taking a conservative P-value threshold of 0.1) were excluded from the model since they did not explain any of the variance of the dependent variable (i.e., global assessment of disease activity). Change in the model's $R^2 < 10\%$ further justified excluding the insignificant items. Governed by our aim to optimize the clinimetric performance of the PCDAI (and not to produce the most feasible index), significant items with low feasibility were not excluded.

Validation

Validity is the degree to which the instrument measures the concept that it purports to measure.[14] In this study we used construct validation based on Spearman's correlation between the indices and constructs of disease activity: PGA, CRP, ESR, hemoglobin, platelets, and albumin. These constructs were used based on availability within the existing datasets. Correlation r of 0–0.25 was considered a priori as lack of correlation, 0.25–0.5 poor, 0.5–0.75 fair to good, and >0.75 very good to excellent correlation. Testing for the wPCDAI was performed first on the validation set and then on the entire set, without any difference.

For discriminant validity, sensitivity, specificity, and area under the receiver operator characteristic (ROC) curve were used to express the ability of the PCDAI versions to differentiate patients in remission from those with active disease and from the different disease activity states (mild, moderate, and severe). Only one visit per patient (decided a priori as the first one) was used for the validity analyses to avoid repeated measures bias.

"Face validity" addresses whether, on the face of it, the index makes sense and includes also "content validity," which is a subjective assessment whether the measure left out any items that most experts would agree that they are important to the measure. "Feasibility" encompasses both respondent and administrative burden. An instrument is feasible if the participant and researcher report that the instrument is completed within reasonable limits of participant discomfort and both participant and researcher time constraints.

In order to assess face validity and feasibility of the contending PCDAI versions, we contacted 40 experts in pediatric IBD from four continents. The selection was based on IBD working groups in North America and Europe and through personal acquaintance. A questionnaire was sent via email with a reminder 1 week later. Participants were asked to score feasibility and face validity of the four versions on a 7-point Likert scale (from "very feasible" and "very good face validity" to "very much not feasible" and "very poor face validity"). The questionnaire described the study and its aims and provided the different versions with the weighting themes.

Longitudinal Analysis

A change of at least one category in the PGA between the baseline and repeated visits was considered a small change (e.g., from severe to moderate disease activity). A large change was defined as a change of at least two categories (e.g., from moderate disease activity to remission). This approach has been successfully used previously.[5,6] The sensitivity of the PCDAI versions to detect any change was assessed using area under the ROC curve considering the change score of the index (follow-up score minus baseline

TABLE 2. Patient Characteristics

	All Patients ($n = 437$)	Registry Data ($n = 179$)	REACH Trial ($n = 101$)	Budesonide Trial ($n = 71$)[b]	Toronto Data ($n = 86$)
Males (%)	268 (61%)	100 (56%)	60 (59%)	45 (63%)	63 (73%)
Age (years)	12.9 ± 2.6	11.7 ± 2.5	13.3 ± 2.5	14 ± 2.7	11.7 ± 2.2
Range (years)	4-19	4.2-15.9	6-17	9-19	4-16
Disease duration (months)	13 (5-27)	0.7 (0-1.1)[c]	19 (12-32)	11 (4-26)	5 (0-17.4)
Disease location					
Ileum	84 (19%)	15 (8%)	14 (14%)	25 (35%)	30 (35%)
Ileocolonic	264 (60%)	129 (72%)	55 (55%)	43 (61%)	37 (43%)
Colonic	85 (20%)	32 (18%)	31 (31%)	3 (4%)	19 (22%)
Upper GI[a]	146 (33%)	93 (52%)	37 (37%)	1 (1%)	15 (17%)
Treatment at baseline					
Steroids/budesonide	191 (44%)	116 (65%)	32 (32%)	0 (0%)	43 (40%)
AZA/6MP/MTX	174 (40%)	42 (25%)	100 (99.9%)	8 (11%)	24 (28%)
Antibiotics	77 (23%)	38 (21%)	NA	5 (7%)	34 (40%)
Anti-TNF treatment	21 (5%)	9 (5%)	0 (0%)	0 (0%)	12 (14%)
Enteral therapy	11 (3%)	7 (4%)	NA	0 (0%)	4 (5%)
5-ASA regimens	170 (39%)	92 (51%)	54 (54%)	11 (16%)	13 (15%)
Baseline PCDAI score	32 ± 15	27 ± 15	41 ± 8.5	28 ± 7.3	24 ± 19.7
Follow-up PCDAI score	13 ± 12	10.7 ± 9.3	9.7 ± 9.6	15 ± 12	15 ± 12
Baseline disease activity					
Remission (%)	46 (13%)	14 (8%)	1 (1%)	—	31 (36%)
Mild (%)	81 (22%)	51 (29%)	8 (8%)	—	22 (26%)
Moderate (%)	190 (52%)	97 (54%)	76 (75%)	—	17 (20%)
Severe (%)	49 (13%)	17 (10%)	16 (16%)	—	16 (19%)
Follow-up disease activity					
Remission (%)	163 (45%)	74 (41%)	44 (44%)	—	45 (52%)
Mild (%)	159 (43%)	83 (46%)	47 (47%)	—	29 (34%)
Moderate (%)	36 (10%)	19 (11%)	6 (6%)	—	11 (13%)
Severe (%)	8 (2%)	3 (2%)	4 (4%)	—	1 (1%)
Improvement					
None	—	—	—	3 (4%)	—
Poor	—	—	—	25 (35%)	—
Good	—	—	—	19 (27%)	—
Very good	—	—	—	23 (32%)	—

Proportions, medians (interquartile range) and means ± SD are presented as appropriate for the data distribution.
[a]Not mutually exclusive with other distributions as per the Montreal classification.[22]
[b]The only global assessment collected in this study was longitudinal effectiveness since initiation of the study.
[c]A total of 138 (77%) were enrolled at diagnosis, 23 (13%) one month after diagnosis and 18 (10%) during the following 5 months.
AZA, azathioprine; 6MP, 6-mercaptopurine; GI, gastrointestinal; MTX, methotrexate; NA, not applicable; FU, follow-up.

score). The correlation between the change score with the change in PGA was also calculated.

Cutoff Scores of Disease Activity

Serial ROC curves (±95% confidence interval [CI]) were used to define the cutoff scores for categorical disease activity (remission, mild, moderate, and severe) as judged by the PGA of the baseline visit, and the change score that defines response. The best cutoff score was selected as the point in which the sensitivity and specificity were maximized (typically the value corresponding to the most upper left shoulder on the ROC curve).

RESULTS

The raw data of 437 children were included in this study ($n = 101$ in the REACH dataset, $n = 86$ in the Toronto growth data, $n = 179$ in the registry data, and $n = 71$ in the budesonide trial) (Table 2). The calculation of the modPCDAI required CRP values, available on 285

TABLE 3. Results of the β Coefficients in the Regression Model of the PCDAI

Item	#$\beta-$ Coefficient[a]	t	*P*-value	Frequency of Endorsement[b]
Abdominal pain	.209	4.532	<0.001	159 (36%)
Stool frequency	.146	3.938	<0.001	65 (15%)
General well-being	.268	5.916	<0.001	100 (23%)
Abdominal examination	.060	1.576	0.116	19 (4%)
Perirectal disease	.152	4.490	<0.001	24 (6%)
EIM	.106	3.028	0.003	5 (1%)
Hematocrit	.033	0.858	0.391	35 (8%)
ESR	.153	3.909	<0.001	92 (21%)
Albumin	.194	5.063	<0.001	84 (19%)
Height velocity	−.047	−1.419	0.157	94 (22%)
Weight	.116	2.982	0.003	61 (14%)

EIM, extraintestinal manifestations; ESR, erythrocyte sedimentation rate.
[a]The standardized β-coefficients represent the score of the item, entered as continuous variable.
[b]Number of children (%) of the full cohort (n=437) who scored the maximum points of the item at baseline.

patients from the entire cohort. Univariate analysis of all major variables in the randomized datasets showed that the derivation and validation sets were similar (all $P > 0.1$; data not shown).

Mathematical Weighting of the PCDAI

In the validation set the regression analysis indicated that three items of the PCDAI are redundant: height, hematocrit, and abdominal examination (Table 3). This does not necessarily imply that these items do not reflect disease activity independently, but rather that the other items together explain also the contribution of these three items. Indeed, the adjusted R^2 of the model remained unchanged after excluding these three items (a decrease from 0.604 to 0.601). Four items had low frequency of endorsement when considering the worse scoring option, two of which were already excluded based on low weights (hematocrit and abdominal examination) (Table 3). Only five children received the highest score in the "extraintestinal manifestations" item and it was, therefore, reduced from three to two response options (i.e., no extraintestinal manifestations and one or more manifestations). The fourth item, perirectal disease, was left unchanged since many such children were excluded from the original studies. We weighted each item by multiplying the corresponding β-coefficient by 100 and rounding to the nearest 2.5. We penalized the item "well being" (judgmentally assigning 20 instead of 27 points) since this item is closely associated with the dependent variable, used to guide the weighting in the model (i.e., PGA). The final weighted index is presented in Appendix D1 (see Supporting Information).

Validity

Construct Validity

The modPCDAI, followed by the newly weighted wPCDAI, had the highest validity of the different PCDAI versions in the validation set, while the original PCDAI came only third (Table 4). The same rank order of the Spearman's *r* was obtained across the different PCDAI versions when using the entire cohort. Both versions without laboratory values and height items (abbrPCDAI and shPCDAI) had similar degree of construct validity.

Discriminative Validity and Recommended Cutoff Values (Fig. 1, Table 5)

The newly weighted wPCDAI differentiated best those in remission versus those with active disease (area under the ROC curve 0.97 [95% CI: 0.95–0.99]) followed

TABLE 4. Construct Validity of the Contending PCDAI Versions Using the 1/3 Random Validation Cohort

	modPCDAI	wPCDAI	PCDAI	shPCDAI	abbrPCDAI
PCDAI	0.61	0.92	—	0.86	0.86
PGA	0.57	0.75	0.67	0.66	0.65
C-reactive protein	0.48	0.27	0.26	0.18	0.17*
ESR	0.65	0.58	0.49	0.38	0.35
Albumin	−0.82	−0.46	−0.37	−0.18	−0.12*
Hemoglobin	−0.67	−0.39	−0.40	−0.25	−0.18
Platelets	0.55	0.51	0.58	0.48	0.45
SUM[a]	**0.62**	**0.55**	**0.46**	**0.43**	**0.40**

[a]Sum represents the average of the absolute values above.
Numbers represent Spearman's correlation.
*$P > 0.05$ (all other $P < 0.05$).
ESR, erythrocyte sedimentation rate; PGA, physician global assessment.

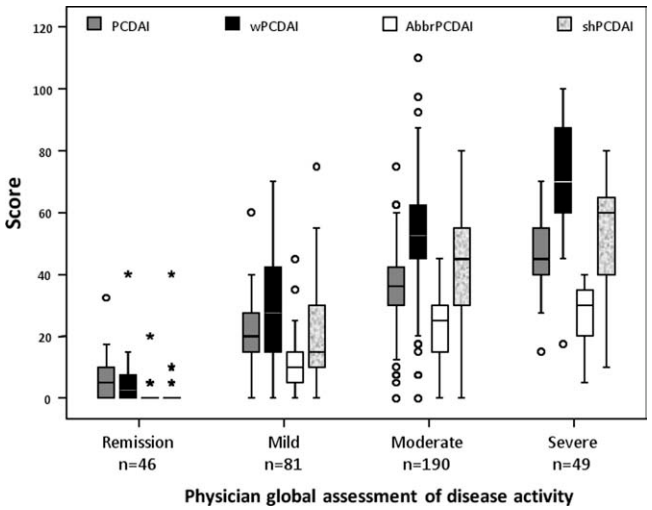

FIGURE 1. Discrimination of the different disease activity states of the different PCDAI versions. The wPCDAI discriminates best the categories. The figure excluded the budesonide trial data for which no physician global assessment was collected.

closely by the original PCDAI and abbrPCDAI (both 0.95 [0.93–0.98]), and shPCDAI (0.94 [0.92–0.97]) and finally the modPCDAI (0.88 [0.83–0.93]).

The superiority of the wPCDAI was pronounced in differentiating the moderate from the severe disease activity category (area under the ROC curve 0.87 [95% CI: 0.82–0.92]), followed by the original PCDAI (0.83 [0.77–0.88]), the shPCDAI (0.79 [0.72–0.86]), and lastly the abbrPCDAI and modPCDAI (both 0.78 [0.72–0.84]). These differences are independent of any given cutoff value as ROC curves are composed of all different possible cutoff values of the continuous measure.

The best cutoff values to differentiate the different disease activity states are presented in Table 5.

Face Validity and Feasibility

Response rate of the electronic survey was 83% (33 of 40 experts). One who stated that he rarely use the PCDAI was excluded; all others used the PCDAI sometimes ($n = 3$), often ($n = 11$) or very often ($n = 18$). The mean age of 32 participants was 47 ± 10 years (range 33–60) and mean time caring for IBD patients was 14.8 ± 7.5 years (range 3–30). The modPCDAI was not included in the survey as it was published after the survey was mailed out.

According to this survey, the feasibility of the original PCDAI was below the "somewhat feasible" (mean 3.4 points of 7), and the face validity of the short and abbreviated PCDAI versions was below the "slight face validity" mark (mean 3.7–3.8 of 7).

The most feasible versions were the shPCDAI and abbrPCDAI (84% scored them as "feasible" or" very feasible") followed by the wPCDAI (66%) and then the original PCDAI (31% only) ($P < 0.01$). The rank order was reciprocal for face validity: 91% scored the PCDAI as having "good" or "very good face validity," followed by the wPCDAI (59%), shPCDAI (22%), and finally the abbrPCDAI (19%) ($P < 0.01$).

The wPCDAI had the highest combined product of feasibility and validity, which added up to 5.17 (lower values reflect better performance) versus 5.4 for the original PCDAI, 5.33 for the shPCDAI, and lastly the abbrPCDAI (5.6).

Longitudinal Assessment (Fig. 2)

All five versions detected change over the follow-up period (median 10 weeks; interquartile range [IQR] 7–13 weeks), but the wPCDAI was the only one that was able to differentiate moderate from large improvement (Fig. 2). Similarly, all versions were responsive to change except for the modPCDAI (area under the ROC curve to differentiate changed from unchanged patients for the wPCDAI 0.83 [95% CI: 0.79–0.87], for the PCDAI 0.82 [0.78–0.86], for the shPCDAI and abbrPCDAI 0.81 [0.76–0.86], and for the modPCDAI 0.72 [0.64–0.79]).

TABLE 5. Recommended Cutoff Values of the Different PCDAI Versions

	Remission	Mild	Moderate	Severe	Small Improvement	Moderate Improvement
PCDAI[a]	<10 or <7.5 excluding the height item (85%/67%)	↔	>27.5 (81%/83%)	>37.5 (78%/75%)	>12.5 (80%/80%)	>22.5 (73%/76%)
wPCDAI	<12.5 (94%/93%)	↔	>40 (82%/86%)	>57.5 (82%/78%)	>17.5 (86%/76%)	>37.5 (71%/77%)
shPCDAI	<10 (87%/91%)	↔	>25 (80%/82%)	>40 (69%/67%)	>10 (86%/75%)	>30 (68%/67%)
abbrPCDAI	<10 (98%/88%)	↔	>15 (82%/83%)	>25 (61%/73%)	>5 (85%/74%)	>15 (71%/74%)
modPCDAI	<7.5 (79%/81%)	↔	>7.5[b] (68%, 72%)	>12.5 (73%, 74%)	>2.5 (64%/79%)	>2.5[a] (72%, 63%)

Corresponding sensitivity/specificity in parentheses.
[a]As established previously.[8]
[b]The modPCDAI did not differentiate between small and moderate change.

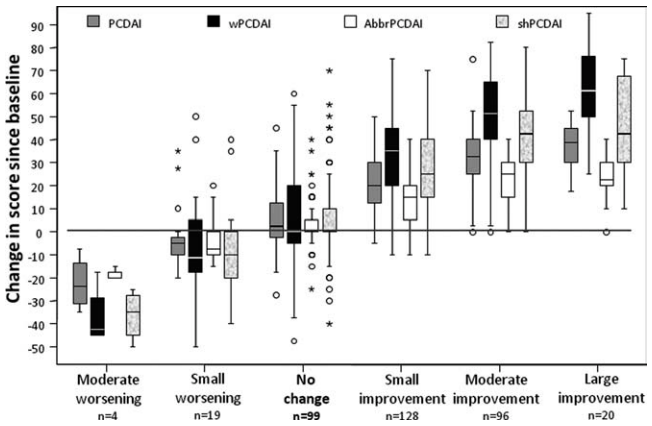

FIGURE 2. Discrimination of change for the different PCDAI versions, stratified by longitudinal global assessment of change. The wPCDAI discriminates best the categories.

The same rank order was achieved when expressing responsiveness by the Spearman's correlation between the change in the index and change in PGA (Fig. 2): $r = 0.72$ for the wPCDAI, $r = 0.70$ for the PCDAI, $r = 0.67$ for both the shPCDAI and the abbrPCDAI, and only $r = 0.52$ for modPCDAI (all $P < 0.001$).

DISCUSSION

We systematically compared, for the first time, the clinimetric properties of the different PCDAI versions and determined the best cutoff scores that correspond to remission, active disease, and response. We also weighted the PCDAI mathematically, hitherto not done, thereby excluding three, statistically redundant items, and producing a modified index with better performance.

Despite the several concerns related to the original PCDAI outlined above, it has performed well in multiple studies over the years. Our survey showed that 91% of experts think that the index has good to very good face validity but it was found to be inferior to the wPCDAI in the construct validity, discriminant validity, and responsiveness. Our survey is likely affected by response bias (i.e., the tendency to respond in a particular way that leads to systematic bias), in which the participants assumed that the longer the version the more valid it is and less feasible. The high face validity among experts who use the current PCDAI frequently may also simply express comfort with a well-known and frequently used tool. This may explain the difference in the validity obtained in the analysis versus the findings of the survey.

The wPCDAI was obtained by mathematical evaluation of the weights of the PCDAI items, originally determined judgmentally. We previously compared the judgmental and the mathematical strategies in weighting the items of the Pediatric Ulcerative Colitis Activity Index (PUCAI), showing that assigning weights mathematically yielded an index that performed just as well as the judgmental one but without the need for laboratory tests, a major advantage in pediatrics.[15] Similarly, the newly weighted wPCDAI performed just as well as the full index but without items of low feasibility. Indeed, evidence from cognitive psychology suggests that humans perform poorly in discriminating between important and less important items.[16] Two rheumatologists were asked in a clinical judgment analysis to provide a PGA of disease activity on patients, and then to state how much emphasis they placed on specific items when providing that assessment.[17] Both physicians placed comparable weighting across five items, but multivariate modeling showed that, in practice, the decision relied on only part of the items that the physicians stressed as important. In a different study, multivariate analyses calculated from clinical judgments in rheumatoid arthritis explained 88% of the variance of the model, whereas rheumatologist's specified judgment policies could explain only 34%.[18] The evidence, thus, convincingly support our finding that mathematical modeling for assigning weights to the PCDAI yields a more valid index than the original judgmental weighting.

The primary aim of developing the abbrPCDAI and shPCDAI was to maximize feasibility, even in the expense of validity, and as such no blood tests are included. The results of our survey reflected that concept; both versions had the highest feasibility but with an associated low face validity. All in all, the two versions performed similarly in most of the evaluated clinimetric categories. Despite the lower overall performance of these indices compared with the PCDAI and the wPCDAI, their utility is inimitable when a more feasible index is needed (such as in retrospective chart review when lab tests are not always available). Based on the results of this study, both the shPCDAI and the abbrPCDAI may be used and the availability of specific items should dictate the use of either version.

The modPCDAI is a version comprised of only laboratory items that were developed with the aim of producing an objective measure of disease activity. Its responsiveness and discriminant validity proved significantly inferior to the PCDAI and wPCDAI. Indeed, laboratory tests have at most fair correlation with intestinal inflammation in CD.[19] Consistent with the original study, the modPCDAI had only moderate correlation with PGA.[11] Nonetheless, overall it performed well in the construct validation, likely since we chose constructs that were largely the very same blood test that constructed the index.

This highlights the most significant limitation of our study, which is the ambiguity in defining disease activity for validation. The latter is a concept for which no gold standard exists; it is the combined constellation of clinical, laboratory, endoscopic, and radiographic parameters that

best define disease activity. Thus, validity is a process of continuous learning about the measure in different scenarios and using different constructs. Our study was limited to the constructs collected in the original studies, although other constructs (e.g., fecal calprotectin and ileocolonoscopy) are also important. Similarly, the PGA used to weight the wPCDAI may not necessarily reflect a "true" estimate of disease activity. Nonetheless, there are multiple precedents in using the PGA in this way including the PCDAI, the PUCAI, and disease activity measures used in rheumatologic diseases.[20,21] Another limitation of this study is the lack of reliability testing, which is the last aspect in clinimetric evaluation. The major strengths of this study is in including several constructs, a very large sample size that allowed statistical manipulations, and robust methodological techniques, including a large survey among experts in the field.

There is no perfect tool that combines responsiveness, discriminative and construct validity, and high feasibility. The performance of the modPCDAI was inferior to the other versions. When a very feasible index is needed, the shPCDAI or the abbrPCDAI have sufficient and similar validity and responsiveness. However, their overall performance was inferior compared with the full indices (i.e., PCDAI and wPCDAI), which should be preferred, most certainly in prospective studies. The newly weighted wPCDAI had the highest overall performance despite, or as a consequence of, the exclusion of three items. Two of the three (height velocity and abdominal examination) have low feasibility and two (abdominal examination and hematocrit) had a low frequency of endorsement in the datasets studied. Despite these encouraging results, the wPCDAI cannot yet replace the full version, which has gained extensive credibility through 20 years of successful experience. More comparative studies are necessary in the different scenarios to grasp the utility of the different versions.

ACKNOWLEDGMENTS

The authors thank Centocor Ortho Biotech for providing the raw REACH data and Falk for providing the raw Budesonide trial data.

REFERENCES

1. 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.
2. Otley A, Loonen H, Parekh N, et al. Assessing activity of pediatric Crohn's disease: which index to use? *Gastroenterology*. 1999;116: 527–531.
3. Hyams J, Markowitz J, Otley A, et al. Evaluation of the pediatric Crohn's disease activity index: a prospective multicenter experience. *J Pediatr Gastroenterol Nutr*. 2005;41:416–421.
4. Kundhal PS, Critch JN, Zachos M, et al. Pediatric Crohn's Disease Activity Index: responsive to short-term change. *J Pediatr Gastroenterol Nutr*. 2003;36:83–89.
5. 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;105:2085–2092.
6. Turner D, Hyams J, Markowitz J, et al. Appraisal of the pediatric ulcerative colitis activity index (PUCAI). *Inflamm Bowel Dis*. 2009;15: 1218–1223.
7. Kappelman M, Crandall W, Colletti R, et al. A short pediatric Crohn's disease activity index for quality improvement and observational research. *Inflamm Bowel Dis*. 2009;15:S18.
8. 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;105:2085–2092.
9. Loonen HJ, Griffiths AM, Merkus MP, et al. A critical assessment of items on the Pediatric Crohn's Disease Activity Index. *J Pediatr Gastroenterol Nutr*. 2003;36:90–95.
10. Shepanski MA, Markowitz JE, Mamula P, et al. Is an abbreviated Pediatric Crohn's Disease Activity Index better than the original? *J Pediatr Gastroenterol Nutr*. 2004;39:68–72.
11. Leach ST, Nahidi L, Tilakaratne S, et al. Development and assessment of a modified pediatric Crohn's disease activity index. *J Pediatr Gastroenterol Nutr*. 2010;51:126–127.
12. Hyams J, Crandall W, Kugathasan S, et al. Induction and maintenance infliximab therapy for the treatment of moderate-to-severe Crohn's disease in children. *Gastroenterology*. 2007;132:863–873; quiz 1165–1166.
13. Levine A, Kori M, Dinari G, et al. Comparison of two dosing methods for induction of response and remission with oral budesonide in active pediatric Crohn's disease: a randomized placebo-controlled trial. *Inflamm Bowel Dis*. 2009;15:1055–1061.
14. Streiner D, Norman G. Validity. Health measurement scales A practical guide to their development and use, 2nd ed. Oxford: Oxford University Press; 1995.
15. Turner D, Griffiths AM, Steinhart AH, et al. Mathematical weighting of a clinimetric index (Pediatric Ulcerative Colitis Activity Index) was superior to the judgmental approach. *J Clin Epidemiol*. 2009;62: 738–744.
16. Dawes RM, Faust D, Meehl PE. Clinical versus actuarial judgment. *Science*. 1989;243:1668–1674.
17. Kirwan JR, Chaput de Saintonge DM, Joyce CR, et al. Clinical judgment in rheumatoid arthritis. II. Judging 'current disease activity' in clinical practice. *Ann Rheum Dis*. 1983;42:648–651.
18. Kirwan JR, Chaput de Saintonge DM, Joyce CR, et al. Inability of rheumatologists to describe their true policies for assessing rheumatoid arthritis. *Ann Rheum Dis*. 1986;45:156–161.
19. Mack DR, Langton C, Markowitz J, et al. Laboratory values for children with newly diagnosed inflammatory bowel disease. *Pediatrics*. 2007;119:1113–1119.
20. 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.
21. 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.
22. 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.