

The Crohn's Disease Activity Index, its derivatives and the Inflammatory Bowel Disease Questionnaire:

A review of instruments to assess Crohn's disease

Eric M Yoshida MD FRCPC

EM Yoshida. The Crohn's Disease Activity Index, its derivatives and the Inflammatory Bowel Disease Questionnaire: A review of instruments to assess Crohn's disease. *Can J Gastroenterol* 1999;13(1):65-73. The Crohn's Disease Activity Index (CDAI) has been used to measure Crohn's activity for over a quarter of a century. The development of the CDAI is reviewed and its reliability and validity are examined. Instruments used to assess Crohn's disease that were developed subsequent to the CDAI, including the Harvey-Bradshaw Index, the Cape Town Index and a three-variable version of the CDAI (modified for survey research), are similarly reviewed. The most recent instrument to assess Crohn's disease, the Inflammatory Bowel Disease Questionnaire, which assesses patients in the domains of bowel, systemic, emotional and social function, is also discussed.

Key Words: *Cape Town Index, Crohn's Disease Activity Index, Harvey-Bradshaw Index, Inflammatory Bowel Disease Questionnaire, Reliability, Validity, Van Hees Index*

L'Indice d'activité de la maladie de Crohn, ses dérivés et le Questionnaire sur la maladie inflammatoire de l'intestin : aperçu des outils d'évaluation de la maladie de Crohn

RÉSUMÉ : Depuis un quart de siècle maintenant, on utilise l'Indice d'activité de la maladie de Crohn (ou CDAI, pour *Crohn's Disease Activity Index*) pour mesurer l'évolution de la maladie. Le développement de cet outil est passé en revue et on se penche sur sa fiabilité et sa validité. Les outils utilisés pour évaluer la maladie de Crohn qui ont été développés par la suite, notamment l'index Harvey-Bradshaw, l'index de Cape Town et une version à trois variables du CDAI (modifié pour les enquêtes) ont également été passés en revue. On présente en outre le plus récent outil à avoir été créé pour évaluer la maladie de Crohn, soit le Questionnaire sur la maladie inflammatoire de l'intestin, qui évalue l'état des patients sur les plans intestinal, systémique, émotionnel et sur le plan de la fonction sociale.

The two main diseases that comprise inflammatory bowel disease (IBD) are ulcerative colitis (UC) and Crohn's disease (CD). Both conditions are chronic illnesses associated with significant patient morbidity and have natural histories characterized by exacerbations of disease activity and

episodes of quiescence. The etiology of both conditions is unknown ('idiopathic'), although it is widely accepted that the pathogenesis of both conditions is on the basis of immunodysregulation (1). As a consequence of the similarities in inflammatory mucosal damage, medications that typically

Department of Medicine, Division of Gastroenterology, University of British Columbia, Vancouver, British Columbia
Correspondence: Dr EM Yoshida, Vancouver Hospital and Health Sciences Centre, 3300-950 West 12th Avenue, Vancouver, British Columbia V5Z 4E3. Telephone 604-875-5862, fax 604-875-5447
Received for publication April 9, 1998. Accepted June 10, 1998

TABLE 1
The original 18 items selected in the development of the Crohn's Disease Activity Index and their rating scale

1. Number of liquid or very soft stools in one week
2. Daily abdominal pain ratings score total in one week (rating from 0 to 3)
3. Daily well-being score total in one week (rating 0 to 4 – 0 = well to 4 = terrible)
4. Symptoms or signs of Crohn's disease (present = 1, absent = 0):
a) arthritis/arthralgia
b) mucocutaneous lesions (eg, erythema nodosum)
c) ocular inflammation (iritis, uveitis)
d) perianal disease including fissures, fistulas or perirectal abscess
e) extra-anal fistulae (enterocutaneous, enterovaginal, etc)
f) fever >37.8°C (100°F) during week
5. Antidiarrheal usage (eg, diphenoxylate hydrochloride) (yes=1, no=0)
6. Abdominal mass (absent = 0, equivocal = 0.4, present = 1)
7. 47 (males) or 43 (females) minus hematocrit
8. $100 \times (1 - [\text{body weight}/\text{standard weight}])$.
9. Mean daily peak body temperature in previous week
10. Mean daily periods of pain in hours
11. Number of days in a week with bloody stools
12. Nausea/vomiting (absent = 0, occasionally = 1, frequently = 3, daily = 5)
13. Appetite (good = 0, mild = 1, moderate = 3, severe = 5)
14. Sleep interrupted by bowel movements (yes = 1, no = 0)
15. Tenderness on abdominal examination (none = 0, mild = 1, moderate = 3, severe = 5)
16. Serum albumin (3.9 – albumin in g/dL)
17. Pain that awakens from sleep (yes = 1, no = 0)
18. Location of Crohn's disease (any small bowel involvement = 0, colonic involvement = 1)

Each item (1-18) was an independent variable 'x' in the general multiple regression equation, $y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n$. Adapted from reference 5

target the immune system either systemically or locally (mesalamine, corticosteroids, azathioprine, etc) are used to treat both conditions.

UC and CD, however, are distinct disease entities (2). Inflammation in UC is confined to the mucosa, is limited to the colon and appears distally in the rectum, extending more proximally in a confluent pattern. CD, on the other hand, may affect any part of the gastrointestinal tract in a patchy distribution, although it commonly affects the small bowel (usually distal ileum) and colon. Unlike inflammation in UC, inflammation in CD can be transmural, resulting in intestinal fibrosis, luminal stenosis and fistula formation between adjacent hollow organs (including skin) and other loops of intestine. Moreover, in UC, definitive surgical re-

section is curative whereas in CD, inflammatory recurrence is common. Both diseases often present with systemic symptoms (fever, fatigue, weight loss, etc) and both can have extra-intestinal manifestations (eg, cutaneous ulcers, arthritis and ocular inflammation) (3).

Given the heterogeneous clinical features present in any cohort of CD patients, as well as the many possible surrogate laboratory markers of disease activity, it became necessary to develop an instrument that would accurately reflect the activity of CD. This instrument would have to incorporate the clinical features of the disease and any of its known laboratory markers. The instrument would have to demonstrate reliability and possess validity (both content and criterion), as well as demonstrate responsiveness to change in disease activity, in order to be suitable for use in clinical trials. Moreover, for such an instrument (and the clinical trials that employed such an instrument) to be accepted by experienced clinicians, it would have to possess face validity, ie, appear to reflect a bedside global assessment of CD activity.

In 1970, the National Cooperative Crohn's Disease Study (NCCDS) was initiated to test the efficacy and safety of prednisone, sulfasalazine and azathioprine in controlled clinical trials (4). For this study, the investigators developed an instrument to assess CD severity and response to therapy. This instrument, the Crohn's Disease Activity Index (CDAI) (5), was first used in clinical trials in 1972 and was the instrument of evaluation in two large trials that resulted from the NCCDS (6,7). Today, the CDAI remains widely used in clinical research. A random overview of important published clinical trials of the past four years (8-13) reveals that more than 25 years after the CDAI was developed, it remains a popular instrument of evaluation of CD.

This article reviews the CDAI, and its development, reliability and validity. Other instruments developed subsequent to the CDAI, including quality of life indexes, are also reviewed.

DEVELOPMENT OF THE CDAI

The CDAI is a disease-specific index, ie, its use is restricted to CD and it is not applicable to IBD in general. The CDAI attempts to evaluate the activity of CD in several domains, each of which evaluates a specific aspect of CD. The CDAI sums up the weighted value of each item of the domain and quantifies the global disease severity in a final numerical score (5).

The CDAI was not the first instrument that attempted to provide a global numerical score of disease severity using this approach. An earlier instrument, adapted from a medical doctoral thesis, was developed at St Bartholomew's Hospital (14). In this earlier instrument, physicians assigned a grade of 0 to 3 to 11 features of CD ranging from lassitude, nausea and pain to ocular inflammation and fistulas. The unweighted sum of the index provided a global rating score of disease activity. Any reliability assessment of this instrument before its use in a clinical trial is unpublished. Likewise, the only validity assessment that can be ascertained from the brief description of the instrument is the reader's own face

validity. The development of the CDAI by the investigators of the NCCDS, in contrast to that of the St Bartholomew's Index, was well documented (5).

In the development of any new health questionnaire or scale, there must be some method of selecting the instrument's items. Two techniques that have been described in devising items in the health sciences are the use of focus groups and clinical observation. Focus groups comprise small groups of 'informants' who have insight into the condition that is to be assessed (usually those affected by the condition) (15). They generate general themes from which items can be derived.

Items derived from clinical observation are based on the opinions of experienced clinicians obtained at 'key informant interviews' (15). Items in the CDAI were devised by combining both methods. Gastroenterologists with expertise in CD from 13 participating university centres (analogous to the focus group of health sciences research) drew upon their clinical experience and agreed upon a list of 18 parameters (Table 1) to be used (4,5). As can be seen in Table 1, the original list of items from the physician focus group sampled several different domains of CD, including subjective patient symptoms and need for symptomatic medications (eg, well-being, abdominal pain that wakes patient up, taking diphenoxylate hydrochloride, etc); objective clinical findings on physical examination (abdominal mass and tenderness, arthritis, etc); extra-intestinal manifestations of CD (eg, iritis and erythema nodosum); complications of CD (eg, fistulas); radiologic and endoscopic examinations (site of disease, distal ileal involvement, colitis, etc); and laboratory parameters (eg, serum albumin and hematocrit).

After the list of items and rating scales were devised, 112 patients with known CD were then studied prospectively. At each visit, patients were assessed by a gastroenterologist, and one-week diaries were completed. These 112 patients made a total of 186 visits. At each patient visit, the gastroenterologist recorded a global assessment score of their impression of the patient's disease activity (rating: "very well" = 1, "fair to good" = 3, "poor" = 5, "very poor" = 7). In the analysis of the CDAI variables, the principles of multiple regression analysis were applied to determine which of the 18 items correlated with the physicians' global assessment.

In a multiple regression model, the dependent variable (y) is related to the independent variables (x) in such a way that the relationship fits the general equation:

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n$$

where b_0 is a constant (the y intercept) and $b_{(1-n)}$ are constant coefficients (slope of the line) of the independent variables (the x 's). In determining the regression equation for a data set, the method of least squares (16) is employed. This least squares method determines the line of best fit between the observed value y (the points scattered around the line) and the y value predicted by equation (the line). In developing the CDAI, the physicians' global assessments of each of the 112 patients became the dependent variable (y) and the items selected by the focus group of gastroenterologists be-

TABLE 2
Final items included in the Crohn's Disease Activity Index and their weights

Item (daily sum per week)	Weight
Number of liquid or very soft stools	2
Abdominal pain score in one week (rating: 0-3)	5
General well-being (rating: 1-4)	7
Sum of findings per week:	20
Arthritis/arthralgia	
Mucocutaneous lesions (eg, erythema nodosum aphthous ulcers)	
Iritis/uveitis	
Anal disease (fissure, fistula, etc)	
External fistula (enterocutaneous/vesicle/vaginal, etc)	
Fever >37.8°C	
Antidiarrheal use (eg, diphenoxylate hydrochloride)	30
Abdominal mass (none = 0, equivocal = 2, present = 5)	10
47 minus hematocrit (males) or 42 minus hematocrit (females)	6
100 x (1 - [body weight divided by standard weight])	1

Adapted from references 4 and 5

came the independent variables ($x_{[1-18]}$). In this analysis, the y intercept (b_0) was set to zero. In deciding which items contributed to the multiple regression model, a method known as backwards step-wise was chosen. This method begins with the full nested model (ie, the initial model including all 18 independent variables). Each independent variable is then dropped from the equation and the remaining equation analyzed to determine whether the deletion of this variable had a significant effect on the remaining model (ie, the variable's effect on the model was more than that of chance as determined by an alpha level of significance such as $P < 0.05$) (16). This stepwise deletion is continued until only independent variables with significant coefficients remain in the model.

Using this method, the number of items of the CDAI were reduced from 18 to eight (Table 2). Body weight was removed from the CDAI in the early computational analysis, however, because of its intuitive importance; it was reinserted only to have gained strength in subsequent stepwise analysis (5).

THE CDAI: THE FINAL INSTRUMENT AND ITS APPLICATION

One of the a priori requirements of the CDAI stipulated by the NCCDS was that the instrument be relatively simple to use. In order to convert a complicated multiple independent variable regression equation into an index that clinical investigators could use, each of the final coefficients ($b_{(1-8)}$) of the equation was divided by the smallest coefficient, which was item 8 (body weight) (5). Each of the adjusted coefficients was then rounded to the nearest integer. From this mathematical manipulation, a weighting scale ranging from 1 to 30 was produced (Table 2).

The actual mechanics of applying the CDAI to a study subject requires that the patient complete a three-item patient diary recording the number of stools per day over a one-week period, the degree of abdominal pain over a week (on a scale of 0 to 3 – “mild” to “severe”) and daily general well-being (on a scale of 0 to 3 – “generally well” to “terrible”). Similarly, at each patient visit, the clinical findings on physician history and physical examination, as well as the relevant laboratory values, were recorded. The individual item scores were then multiplied by the item’s weighting factor and the final score totalled (5). In determining threshold cut-off values of the CDAI that indicated quiescent disease from active, the CDAI scores for the cohort of study patients were compared with the physician’s global categorical assessment (ie, “very well” to “very poor”) (5). A value less than 150 was defined as quiescent CD (ie, nonactive disease, clinical remission) whereas values greater than 450 were associated with extremely severe disease (5).

RELIABILITY OF THE CDAI

The reliability of a test or instrument in a general sense can be considered to represent the degree of consistency of the instrument’s measurement. Reliability can be further defined as the proportion of subject variability to total variability (which includes both subject variability and, importantly, the variability introduced by the instrument’s inherent error of measurement) (17). The kappa statistic – which is well-known to clinicians as a measure of interobserver agreement in diagnostic test interpretation beyond mere chance (eg, comparing chest radiograph interpretations between two radiologists) (18,19) – is a form of reliability testing. Statistical tests of a health care instrument’s reliability are the intraclass correlation coefficient (ICC) and the Pearson correlation coefficient (product moment correlation).

Best et al’s (5) original report of the CDAI did not include formal tests of reliability. The original report did, however, examine the consistency of the instrument over two successive visits (ie, test-retest reliability) in 32 patients. The investigators noted a positive association between changes in the physician global assessment and the CDAI between visits with some overlap. The differential between patients considered slightly improved/worse versus the same was 50 index points. The investigators attempted to describe the within patient variation under relatively stable conditions using the pooled standard deviation of replication (ie, between assessments of patients unchanged between two visits), which was reported to be 45.1 CDAI units (5). The co-efficient of correlation (which explains the degree of variability between the observed data points scattered around the line of regression [20]) between the clinical assessment of disease activity and CDAI was not formally reported in the original report but was subsequently cited as 0.7 (21). As a general rule of thumb, correlations of 0.5 to 0.75 indicate a moderate to good relationship and correlations in excess of 0.75 indicate a very good to excellent rela-

tionship (20). The reliability of the CDAI by this rule of thumb, falls within the good to very good range.

The original CDAI was developed from 112 patients; the CDAI was later re-analyzed after it had been applied to 1058 patients from two studies (22). This was a study of test-retest reliability and the investigators re-derived the regression coefficients ($b_{[1-8]}$) of the original eight independent item variables). A multiplication factor was then used to arrive at a new weighting factor for the eight items. The recalculated CDAI was found to be very similar to the original because the correlation co-efficient determinations between the original CDAI and the recalculated derivative for the four patient groupings of “very well” to “very poor” (based on physician global assessment) ranged from 0.969 to 0.994 (22). Because the rederived CDAI did not differ significantly from the original, the investigators did not recommend any changes to the CDAI.

Intra-class correlation is another test of an instrument’s reliability. Unlike Pearson’s correlation co-efficient, which is derived from regression analysis, the ICC is determined from an ANOVA between total variance and components of this variance (17,23). Mathematically, a significant difference between the ICC versus the correlation co-efficient is that the y intercept of the correlation coefficient may not be zero for the correlation co-efficient, whereas it is zero for the ICC (17). Nevertheless, comparability between the two tests is reported to be good (17). Over 20 years after the development of the CDAI, Irvine et al (24), in a study of the Inflammatory Bowel Disease Questionnaire (IBDQ) (a quality of life questionnaire), examined the ICC of several CD indexes including the CDAI. The analysis was based on 305 patients enrolled in a clinical trial and compared baseline scores with follow-up scores. The ICC of the CDAI in this case was 0.66, which is similar to the correlation coefficient of 0.7 from the original CDAI data (5).

Despite relatively good reliability of the CDAI in the hands of experienced investigators, considerable interobserver variation in scoring individual cases has been demonstrated prospectively (25). When clinicians attending a conference were asked to calculate activity scores for several indexes, including the CDAI, based on individual case history presentation, significant differences in CDAI scores were obtained. Formal tests of reliability were not performed; however, wide ranges in CDAI scores of several hundred points were obtained. The range of CDAI scores became narrower on retesting after discussion with the study participants. This experience demonstrates that the reliability of the CDAI – like any diagnostic skill – is, first, operator-dependent with a learning curve and, second, improves as users become more experienced.

VALIDITY OF THE CDAI

For a medical or health care instrument or test to be clinically useful, it must possess validity. Although the term validity is used frequently and the meaning appears to be obvious, it is not a single concept. Conceptually, validity can

take several forms. Classically, these forms of validity are grouped into construct validity, content validity and criterion validity (26).

Construct validity: Construct validity refers to the creation of an instrument based on known or postulated features (ie, the theoretical constructs) of a disease or condition. The selection of the original items of the CDAI by the focus group of expert gastroenterologists based on accepted features of the disease can be considered an example of construct validity – the items represent accepted features of CD and, therefore, their inclusion is valid.

Content validity: Content validity refers to the concept that the instrument accurately reflects the disease that it purports to measure. An example of content validity is the assumption that a candidate who passes the Royal College certification examinations in gastroenterology, in fact, has the breadth of knowledge expected of a competent gastroenterologist. The requirement of content validity for the CDAI appears to be satisfied by the ability of the instrument to be responsive to changes in clinical disease severity (5,22) as reflected by the derivation of threshold CDAI scores that allow categorization of patients with quiescent or active disease.

As mentioned in the preceding section on reliability, the correlation coefficient of this relationship between CDAI scores and clinical disease severity was 0.7 (21). Although the correlation coefficient indicates the variability of the observed data points around the line of regression and is an indicator of reliability, it is also an indicator of validity. If the null hypothesis – that there was no relationship between CDAI scores and disease severity – was true, then the correlation coefficient value would be expected to be closer to zero than unity.

Since the CDAI has been used as an instrument of evaluation in clinical trials, and, more importantly, that the findings of those trials have been accepted by practicing gastroenterologists for over 25 years, indicate that the CDAI possesses content validity.

While the CDAI appears to possess content validity in many circumstances, there are aspects of CD in which the CDAI does not reflect the effects of the disease. For instance, the CDAI is not suitable for pediatric patients with CD. This is hardly surprising because the CDAI was developed by adult gastroenterologists, and the instrument, therefore, fails to possess construct validity for children whose disease manifestation/presentation may differ from that of adults. A pediatric CDAI has been developed (27).

The domains of the CDAI also may not adequately cover all aspects of adult CD that are important to patients and their families. In terms of quality of life, including psychological, social, sexual and occupational functioning, the CDAI does not possess content validity. As Garrett and Drossman (28) point out, a patient with a low CDAI score may, in fact, be severely limited by the disease in these areas. In order to cover these qualitative aspects of CD in which the CDAI has poor content validity, quality of life instruments have been developed for IBD (these will be discussed later).

Criterion validity: Criterion validity refers to the ability of an instrument to correlate positively and significantly with an established or accepted gold standard test (26). The heterogeneous manifestations of CD and the many possible diagnostic studies available – including endoscopic, radiologic, nuclear and laboratory studies – make choosing a gold standard test difficult. In fact, over the past 25 years, the CDAI itself become the gold standard instrument, with other diagnostic tests compared with it. Nonetheless, because the CDAI is based on mostly clinical features, including those self-reported by the patient, it would be of interest to assess the comparativeness of the CDAI with more objective diagnostic modalities that assess mucosal inflammation.

Studies comparing markers of mucosal inflammation with the CDAI generally have not demonstrated significant correlations and have sometimes yielded conflicting results. It is important to realize a priori that many of the studies have not involved large numbers of patients, that the diagnostic tests themselves are only surrogate markers of inflammation and that, in many cases, the CDAI was the gold standard of comparison. There has been only poor or weak correlation between the CDAI and direct endoscopic or microscopic mucosal examination in Crohn's colitis (29); conflicting results with fecal ¹¹¹In excretion, with reports of both no correlation (30,31) and positive correlation (32); positive correlation with computerized ^{99m}Tc hexamethyl propylene amine oxime nuclear scanning (33); and reports of falsely positive high CDAI scores compared with gut lavage measuring protein loss (34).

The lack of a clearly positive correlation between mucosal inflammation and CDAI does not necessarily mean that the CDAI lacks criterion validity. Rather, this lack suggests that mucosal inflammation as detected by diagnostic testing is only one domain from which to assess CD severity.

Another aspect of criterion validity is the predictability of the instrument, ie, can the instrument predict a disease outcome? Wright et al (35) assessed several CD indexes as well as individual laboratory parameters, and noted that a rise in CDAI from baseline predicted an acute exacerbation of CD two months before the actual attack. Conversely, the CDAI in this study demonstrated a decrease towards baseline one month after the acute exacerbation, further suggesting that the CDAI possesses validity.

OTHER CD INDEXES POST-CDAI

Despite the intentions of the NCCDS to develop the CDAI with simplicity of application and computation in mind, the CDAI has been reported to be "cumbersome" (36) to use, and attempts have been made to simplify it. The most popular simplified derivative of the CDAI is the Harvey-Bradshaw Index (HBI) (37), also known as the Simple Index and the Modified CDAI. Unlike the CDAI, which requires completion of a seven-day patient diary, the HBI looks only at symptoms and signs over the preceding 24 h. The HBI also reduced the original eight items of the CDAI to five, removing anti-diarrheal use, hematocrit level and body weight, and eliminated the weighting factor. The HBI had a 93% corre-

TABLE 3
The Cape Town Index (South African Index)

Item/score	0	1	2	3
Diarrhea (episodes/day)	None	≤4	5	≥6
Abdominal pain	None	Mild	Moderate	Severe
Well-being	Normal	Below par	Unwell	Terrible
Complications				
Local	None	Skin tag	Sinus	Fistula
Systemic	None	Stomatitis	Arthralgia	A/I/ER
Fever	Normal (37°C)	≤38°C	≤39°C	>39°C
Weight versus previous weight	No change	No change	<95%	<90%
Abdominal examination				
Mass	None	None	Indefinite	Definite
Tenderness	None	Mild	Moderate	Severe
Hemoglobin (g/L)	≥120	≤120	≤110	≤100
ESR* (mm/first hour)	≤15	>15	>25	>40

*Erythrocyte sedimentation rate (ESR) was not included in the original index but was subsequently added as a modification: the Cape Town Index + ESR. A/I/ER Arthritis/iritis/erythema nodosum. Adapted from reference 38

TABLE 4
The Van Hees Index

Regression variable $x_{(1-9)}$	Index item
1	Serum albumin (g/L)
2	Erythrocyte sedimentation rate (mm) after 1 h
3	Quetelet index*
4	Abdominal mass (grade 1-5) [†]
5	Sex (male = 1, female = 2)
6	Temperature (°C) daily average in 1 week
7	Stool consistency (rating: 1-3) [‡]
8	Intestinal resection (no = 1, yes = 2)
9	Extra-intestinal lesions

The index score is obtained by inserting each dependent variable value (x_n) into the equation:

$$-209 [the constant, b_0] - 5.48 x_1 + 0.29 x_2 - 0.22 x_3 + 7.83 x_4 - 12.3 x_5 + 16.4 x_6 + 8.46 x_7 - 9.17 x_8 + 10.7 x_9$$

Adapted from reference 41. *Weight (kg) x 10/height (cm)²; [†]No mass = 1, dubious = 2, diameter <6 cm = 3, diameter 6-12 cm = 4, diameter >12 cm = 5; [‡]Well-formed = 1, soft = 2, watery = 3

lation with the CDAI in the original prospective study of 112 patients (37). This high correlation has been duplicated in later studies (29,35). The HBI clearly possesses easier practical application and is amenable to same-day clinic visits. The previous day nature of the HBI, however, introduces the possibility of less reliability versus the CDAI's seven-day patient diary. The ICC of the HBI in Irvine et al's study (24) was 0.55 (versus 0.66 for the CDAI).

It has been suggested that the frequency of diarrhea item introduces the strong possibility of single-parameter bias within the HBI (38). In an effort to avoid this bias, the Cape Town Index (CTI) (also known as the South African Index) (38) was developed. Similar to the HBI, the CTI is also a derivative of the CDAI. Parameters from the domains of patient subjective symptoms, physician clinical findings and laboratory data (Table 3) are included. The CTI was studied on 86 patients who made 1068 visits (38). The correlation coefficient of the CTI and CDAI was 0.761, and was 0.810 of the CTI and HBI (the correlation between the HBI and CDAI in this study was 0.745).

Sandler et al (21) have pointed out that a significant shortcoming of the CDAI, as well as the other clinical indexes of CD, is the unsuitability for population survey research. In an effort to derive a new instrument from the CDAI that was suitable for survey research, these investigators reanalyzed data from a previously published study (7). Using multiple regression analysis with the physician global assessment as the dependent variable and transforming the new index into CDAI equivalents, they derived a three-variable index from the original CDAI (21). Scoring for this new index was calculated from the formula:

$$(3 \times \text{number of stools}) + (10 \times \text{abdominal pain}) + (8 \times \text{well-being})$$

The correlation coefficient for this new index compared with the original CDAI was 0.866. When the dataset of the new index and CDAI were divided into quartiles (ie, one quartile spans the 25th to 75th percentile), there was agreement in 61% of cases and 97% agreement within one quartile. Although the three-variable derivation of the CDAI seems to be very amenable to survey research with possible further conversion of the three variables to Likert or visual analogue scales, this derivation of the CDAI has not been used to any great extent in clinical research. Such a three-parameter instrument may not be of great appeal to clinicians conducting clinical trials because disease assessment is restricted to only one domain: patient subjective symptoms. Such a single domain index would be perceived as lacking face validity.

A total of 39% of the CDAI is derived from the domain of the patient's symptoms (eg, general well-being and pain). Strong objections to the subjective nature of the CDAI were raised a few years after publication of the NCCDS, and attempts have been made to determine CD activity on objective grounds only. The use of laboratory-based indexes (36,39,40) as a replacement for, or an adjuvant to, a clinical instrument has been suggested. One instrument that eliminated subjective criteria is the Van Hees Index (VHI) (41) (also referred to as the Dutch Index). The derivation of the VHI, akin to that of the CDAI, employed a multiple regression analysis with a step-wise deletion using a physician global assessment as the dependent variable. The final items included in the index were: sex, two laboratory parameters (serum albumin and erythrocyte sedimentation rate) and seven clinical features obtained through physician history or

physical examination (a weight:height index [the Quetelet index], abdominal mass, temperature, stool consistency, previous intestinal resection and extraintestinal manifestations). In determining the VHI score for the individual patient, the VHI items are not multiplied by a weighting factor to obtain a final score; instead, the VHI score is obtained by directly inserting the item value into the multiple regression equation (Table 4). VHI scores less than 100 imply quiescent disease, less than 150 imply mild disease, less than 210 imply moderate disease and greater than 210 imply severe disease. The correlation coefficient of the VHI regression analysis was 0.95. Comparing the CDAI with the VHI, Van Hees et al (41) reported a correlation of only 0.57 – a marginal correlation that has been subsequently noted by others (35,38). Of interest, although Wright et al (35) found a poor correlation between the VHI and CDAI, both had predictive ability regarding CD exacerbations, although the CDAI reportedly demonstrated greater sensitivity. The poor correlation between CDAI and VHI may suggest that the VHI differs in reliability although it may also suggest that the two instruments do not measure the same features of CD.

QUALITY OF LIFE IN CD AND THE IBDQ

As discussed in the previous section, some clinical investigators who raised objections to the inclusion of subjective patient-reported symptoms in the CDAI attempted to create indexes without them. Regardless, there has been an increasing awareness of the importance of quality of life issues surrounding CD in the past 10 years. The content of previous instruments of CD activity have generally reflected a physician's perspective of disease activity. Quality of life studies, however, have suggested that functional status from the patient's viewpoint may be more indicative of health care utilization than the physicians' (42), and that patients are less likely to volunteer information regarding psychosocial impairment (43). Drossman et al (42) used a generic health-related quality of life questionnaire, the Sickness Impact Profile – which examines 12 areas of daily living – and an IBD-specific instrument, the Rating Form of IBD Patient Concerns – which assesses patient concerns and worries – to assess a mixed UC/CD cohort of patients (43 UC, 54 CD). These investigators reported that IBD patients are more likely to indicate functional impairment in psychosocial aspects of health, such as emotional behaviour, sleep, rest and social interactions, than control patients without IBD. Moreover, the IBD patients in their study indicated that functional impairment resided more in the psychosocial domains than in the physical. CD patients also indicated greater impairment than UC patients.

In the assessment of quality of life in CD, however, not all instruments may be sensitive enough to detect changes in functional impairment and may not be appropriate. For instance, results from a time trade-off questionnaire, in which patients indicate the total number of life-years of poor health they would be willing to trade for a lesser number of life-years of good health, indicated that almost half of the IBD patient sample (n=93) were unwilling to trade-off (43).

TABLE 5
An example of questions from the Inflammatory Bowel Disease Questionnaire

Question #	Symptom domain	Question
5	Bowel	How much of the time during the last two weeks have your bowel movements been loose?
6	Systemic	How much energy have you had during the last two weeks?
15	Emotional	How often during the last two weeks have you felt depressed or discouraged?
28	Social	To what extent has your bowel problem limited sexual activity during the last two weeks?

Adapted from reference 45

The assessment of quality of life in IBD has led to the development of a disease-specific instrument: the IBDQ (44,45). In contrast to the VHI (41) (which intentionally excludes patient-self reporting), the IBDQ does not include physician-derived physical findings or laboratory data. The IBDQ is an interviewer-administered questionnaire, although a revised, self-administered, mail-in version has also been developed (46). Unlike the proposed modified three-variable survey version of the CDAI (23), the IBDQ has gained a degree of acceptance and has been used as an adjunct outcome measure in several studies (8,10,11). The IBDQ, in its final form, consists of 32 questions that cover four domains: bowel symptoms, systemic systems, emotional function and social function (44,45) (Table 5 includes an example of questions from each domain). Each question response is in the form of a seven-point scale.

The selection of items in the IBDQ came from an open questionnaire of both gastroenterologists and patients with IBD (n=77) (45). The administered open questionnaire was derived from a literature review of both existing disease-specific and -nonspecific instruments. A group of patients with IBD then assessed the items, and the 32 most important items were included in the final survey. The involvement of patients in the creation of the instrument differs from the development of the previously described CD instruments, which were entirely clinician-focused. The IBDQ was formally evaluated on 63 patients on two occasions, one month apart. IBD activity was also assessed with the VHI (41) for CD and the St Mark's index (47) for UC. Test-retest reliability (referred to as reproducibility) in stable patients revealed a small but statistically significant difference (paired Student's *t* test) in two of the four domains: bowel (5.5% change from baseline) and systemic symptoms (12.1% change from baseline), a finding that suggests a lesser degree of reliability in these domains.

Responsiveness of the IBDQ was suggested by the consistency of domain scores with changes in patient self-assessed global ratings (significance of score differences from baseline

versus follow-up assessed by paired Student's *t* test) and by the ratio of standard deviations between stable and clinically changed patient scores in excess of unity in all domains. Construct validity and content validity were reflected by the significant correlation among patient's global domain assessment, physician assessment and domain scores for bowel/systemic domains. For the most part, these correlations achieved statistical significance but were typically less than 0.5. Criterion validity testing assessing the IBDQ versus the VHI did not reveal a statistically significant correlation with bowel symptoms ($r=0.33$, $P=0.08$) or systemic symptoms ($r=0.036$, $P=0.442$), suggesting that in these domains, the IBDQ may not provide an adequate assessment compared with a gold standard.

The VHI, as previously discussed, makes no attempt to assess subjective patient symptoms; therefore, the emotional domain of the IBDQ was assessed against the emotional function dimension of the Rand questionnaire. A significant correlation of 0.76 was determined, suggesting criterion validity in this domain.

The IBDQ was subsequently assessed in later studies. Love et al (46) compared a modified IBDQ in a cohort of IBD patients ($n=182$) with a sex- and age-matched control group ($n=48$). Comparison of scores between IBD and controls in each domain revealed statistically significant differences suggesting content validity (ie, the instrument distinguished IBD patients from healthy patients). Irvine et al (24) reassessed the IBDQ in a larger group of patients ($n=305$) with CD only (previously studied groups included both CD and UC). Reliability testing revealed an ICC among the four domains ranging from 0.65 to 0.67. (As previously mentioned, the ICC for the CDAI in this study was 0.66.) Criterion validity testing of the IBDQ versus the CDAI revealed baseline correlation coefficients from -0.5 for emotional function, -0.53 for systemic function, -0.56 for social function and -0.71 for bowel function ($r=-0.67$ for overall IBDQ score). The negative correlation merely denotes that higher IBDQ scores reflect less functional impairment whereas higher CDAI scores reflect worse disease activity. The

eight-week follow-up correlation was lower for all domains except systemic function, which was unchanged ($r=-0.37$ for social function, $r=-0.40$ for emotional function and $r=-0.60$ for bowel function). The better correlation with bowel function and decreased correlation with emotional and social function reflects the differing degree of overlap between the CDAI and the IBDQ domains, as opposed to lack of criterion validity. The correlation with the CDAI in this study differed markedly from that of the original study (45), which compared the VHI and found lower coefficients of correlation. This finding suggests less overlap in IBDQ domains and the VHI compared with the IBDQ and the CDAI. As noted, the correlation between the CDAI and the VHI is modest (35,38,41), so one may infer that these instruments also tap into different domains of CD activity.

CONCLUSIONS

Since 1972, when the CDAI was first used in clinical research in the NCCDS, the CDAI has become the gold standard with which other instruments of disease assessment and diagnostic tests are compared. The CDAI demonstrates acceptable reliability, although the range of variability in scoring can be considerable with a learning curve effect (25). The CDAI appears to be a valid instrument of assessment although it is not appropriate for all patient subgroups with CD (eg, children) and does not cover all domains of the CD spectrum (eg, psychosocial). The correlation of the CDAI with diagnostic markers of mucosal inflammation is inconsistent, which may reflect the lack of correlation with these markers and the global effects of CD on patients. Since the introduction of the CDAI, derivations of the index have been developed to facilitate its clinical administration (eg, the HBI [37] or the CTI [38]) or to make the index more objective (ie, the VHI [41]). Quality of life, from a patient perspective, including social and emotional function, has been poorly assessed by these instruments. The IBDQ (24,44,45) possesses both reliability and validity in these domains and can be used in combination with a traditional CD activity index in clinical research.

REFERENCES

- Sartor RB. Current concepts of the etiology and pathogenesis of ulcerative colitis and Crohn's disease. *Gastroenterol Clin North Am* 1995;24:475-507.
- Ogorek CP, Fisher RS. Differentiation between Crohn's disease and ulcerative colitis. *Med Clin North Am* 1994;78:1249-58.
- Levine JB, Lukawski-Trubish D. Extraintestinal considerations in inflammatory bowel disease. *Gastroenterol Clin North Am* 1995;24:633-46.
- Winship DH, Summers RW, Singleton JW, et al. National Cooperative Crohn's Disease Study: study design and conduct of the study. *Gastroenterology* 1979;77:829-42.
- Best WR, Beckett JM, Singleton JW, Kern F Jr. Development of a Crohn's disease activity index. *Gastroenterology* 1976;70:439-44.
- Summers RW, Switz DM, Sessions JT Jr, et al. National Cooperative Crohn's Disease Study: results of drug treatment. *Gastroenterology* 1979;77:847-69.
- Singleton JW, Summers RW, Kern F Jr, et al. A trial of sulfasalazine as adjunctive therapy in Crohn's disease. *Gastroenterology* 1979;77:887-97.
- Greenberg GR, Feagan BG, Martin F, et al. Oral budesonide for active Crohn's disease. *N Engl J Med* 1994;331:836-41.
- Rutgeerts P, Lofberg R, Malchow H, et al. A comparison of budesonide with prednisolone for active Crohn's disease. *N Engl J Med* 1994;331:842-5.
- Feagan BG, McDonald JWD, Rochon J, et al. Low-dose cyclosporine for the treatment of Crohn's disease. *N Engl J Med* 1994;330:1846-51.
- Feagan BG, Rochon J, Fedorak RN, et al. Methotrexate for the treatment of Crohn's disease. *N Engl J Med* 1995;332:292-7.
- Targan SR, Hanauer SB, van Deventer SJ, et al. A short-term study of chimeric monoclonal antibody cA2 to tumor necrosis factor alpha for Crohn's disease. Crohn's Disease CA2 Study Group. *N Engl J Med* 1997;337:1029-35.
- Stack WA, Mann SD, Roy AJ, et al. Randomized controlled trial of CDP571 antibody to tumor necrosis factor-alpha in Crohn's disease. *Lancet* 1997;349:521-4.
- Willoughby JMT, Beckett J, Kumar PJ, Dawson AM. Controlled trial of azathioprine in Crohn's disease. *Lancet* 1971;ii:944-7.
- Streiner DL, Norman GR. Devising the items. In: *Health Measurement Scales. A Practical Guide to their Development and Use*, 2nd edn. New York: Oxford University Press, 1995:16-27.
- Pagano M, Gauvreau K. Multiple regression. In: *Principles of Biostatistics*. Belmont: Duxbury Press, 1992:409-25.

17. Streiner DL, Norman GR. Reliability. In: *Health Measurement Scales. A Practical Guide to their Development and Use*, 2nd edn. New York: Oxford University Press, 1995:104-27.
18. Sackett DL. Clinical disagreement: I. How often it occurs and why. *Can Med Assoc J* 1980;20:499-504.
19. Ker M. Issues in the use of kappa. *Invest Radiol* 1991;26:78-83.
20. Colton T. Regression and correlation. In: *Statistics in Medicine*. Boston: Little Brown and Company, 1974:189-917.
21. Sandler RS, Jordan MD, Kupper LL. Development of a Crohn's index for survey research. *J Clin Epidemiol* 1988;41:451-8.
22. Best WR, Beckett JM, Singleton JW. Rederived values of the eight coefficients of the Crohn's disease activity index (CDAI). *Gastroenterology* 1979;77:843-6.
23. Selvin S. General concepts. In: *Practical Biostatistical Methods*. Belmont: Duxbury Press, 1995:1-30.
24. Irvine EJ, Feagan B, Rochon J, et al. Quality of life: a valid and reliable measure of therapeutic efficacy in the treatment of inflammatory bowel disease. *Gastroenterology* 1994;106:287-96.
25. De Dombal FT, Softley A. IOIBD report no 1: observer variation in calculating indices of severity and activity in Crohn's disease. *Gut* 1987;28:474-81.
26. Streiner DL, Norman GR. Validity. In: *Health Measurement Scales. A Practical Guide to their Development and Use*, 2nd edn. New York: Oxford University Press, 1995:144-62.
27. Harms HK, Blomer R, Bertele-Harms RM, et al. A paediatric Crohn's disease activity index (PCDAI). Is it useful? *Acta Paediatr* 1994;395(Suppl):22-6.
28. Garrett JW, Drossman DA. Health status in inflammatory bowel disease. *Gastroenterology* 1990;99:90-6.
29. Gomes P, Du Boulay C, Smith CL, Holdstock G. Relationship between disease activity indices and colonoscopic findings in patients with colonic inflammatory bowel disease. *Gut* 1986;27:92-5.
30. Park RHR, McKillop JH, Duncan A, et al. Can ¹¹¹indium autologous mixed leucocyte scanning accurately assess disease extent and activity in Crohn's disease? *Gut* 1988;29:821-5.
31. Crama-Bohbouth G, Pena AS, Biemond I, et al. Are activity indices helpful in assessing active intestinal inflammation in Crohn's disease? *Gut* 1989;30:1236-40.
32. Saverymuttu SH, Peters AM, Lavender JP, et al. Quantitative fecal indium 111-labeled leukocyte excretion in the assessment of disease in Crohn's disease. *Gastroenterology* 1983;85:1333-9.
33. Gjaffer MH, Tindale WB, Senior S, et al. Quantification of disease activity in Crohn's disease by computer analysis of Tc-99m hexamethyl propylene amine oxime (HMPAO) labelled leucocyte images. *Gut* 1993;34:68-74.
34. Acciuffi S, Ghosh S, Ferguson A. Strengths and limitations of the Crohn's disease activity index, revealed by an objective gut lavage test of gastrointestinal protein loss. *Aliment Pharmacol Ther* 1996;10:321-6.
35. Wright JP, Young GO, Tigler-Wybrandi N. Predictors of acute relapse of Crohn's disease. A laboratory and clinical study. *Dig Dis Sci* 1987;32:164-70.
36. Mee AS, Brown DJ, Jewell DP. Crohn's disease activity index – is it useful? *Gut* 1978;19:A990.
37. Harvey RF, Bradshaw JM. A simple index of Crohn's disease activity. *Lancet* 1980;i:514.
38. Wright JP, Marks IN, Parfitt A. A simple clinical index of Crohn's disease activity – the Cape Town index. *S Afr Med J* 1985;68:502-3.
39. Andre C, Descos L, Landais P, Fermanian J. Assessment of appropriate laboratory measurements to supplement the Crohn's disease activity index. *Gut* 1981;22:571-4.
40. Cooke WT, Prior P. Determining disease activity in inflammatory bowel disease. *J Clin Gastroenterol* 1984;6:17-25.
41. Van Hees PAM, Van Elteren PH, Van Lier HJJ, Van Tongeren JHM. An index of inflammatory activity in patients with Crohn's disease. *Gut* 1980;21:279-86.
42. Drossman DA, Patrick DL, Mitchell CM, et al. Health-related quality of life in inflammatory bowel disease. Functional status and patient worries and concerns. *Dig Dis Sci* 1989;34:1379-86.
43. Mitchell A, Guyatt G, Singer J, et al. Quality of life in patients with inflammatory bowel disease. *J Clin Gastroenterol* 1988;10:306-10.
44. Guyatt GH, Deyo RA, Charlson M, et al. Responsiveness and validity in health status measurement: a clarification. *J Clin Epidemiol* 1989;42:403-8.
45. Guyatt G, Mitchell A, Irvine EJ, et al. A new measure of health status for clinical trials in inflammatory bowel disease. *Gastroenterology* 1989;96:804-10.
46. Love JR, Irvine EJ, Fedorak RN. Quality of life in inflammatory bowel disease. *J Clin Gastroenterol* 1992;14:15-9.
47. Powell-Tuck J, Brown RL, Lennard-Jones JE. A comparison of oral prednisone given as single or multiple daily doses for active proctocolitis. *Scand J Gastroenterol* 1978;13:833-7.



Hindawi
Submit your manuscripts at
<http://www.hindawi.com>

