

# Assessing outcome in randomized clinical trials: Inflammatory bowel disease

E JAN IRVINE, MD, FRCPC, MSc

**EJ IRVINE.** Assessing outcome in randomized clinical trials: Inflammatory bowel disease. *Can J Gastroenterol* 1993;7(7):561-567. Methodological principles or standards to assess the quality of a clinical trial must be simple and user-friendly if we expect busy clinicians to adopt them. When assessing the success or failure of a new treatment in inflammatory bowel disease, the patient characteristics, therapeutic strategy, outcome assessment and interpretation of results are unique to the disease and the study. Because inflammatory bowel disease rarely is fatal, we use surrogate markers, such as disease activity index, endoscopic appearance, histology, blood tests, tissue markers, need for other medications and quality of life scores, to assess treatment efficacy. Primary and secondary outcomes must be identified and uniformly evaluated to ensure unbiased objective assessment. Careful a priori definition of outcome events is essential. The statistical analysis depends on the data type which comprise the outcome events and the study design. Both intention to treat as well as efficacy analyses should be performed. Interpretation of results should address both clinical and statistical importance.

**Key Words:** Crohn's disease, Outcome assessment, Randomization, Study design, Ulcerative colitis

## Évaluation de l'issue d'essais cliniques randomisés sur la maladie intestinale inflammatoire

**RÉSUMÉ:** Les principes méthodologiques ou les normes en vue d'évaluer la qualité d'un essai clinique doivent être simples et conviviaux si nous voulons que les médecins très occupés s'en servent. Si l'on évalue le succès ou l'échec d'une nouvelle thérapeutique dans la maladie intestinale inflammatoire, les caractéristiques relatives au patient, la stratégie thérapeutique, l'évaluation de l'issue et l'interprétation des résultats sont uniques à la maladie et à l'étude. Parce que la maladie intestinale inflammatoire est rarement fatale, nous utilisons des marqueurs substitués comme l'indice d'activité, l'aspect endoscopique, l'histologie, les analyses sanguines, les marqueurs tissulaires, la nécessité d'administrer d'autres médicaments et les scores de qualité de vie pour évaluer l'efficacité du traitement. Les paramètres primaires et secondaires doivent être identifiés et mesurés unifor-

**T**HE METHODOLOGICAL PRINCIPLES used to design or interpret clinical trials, whether for patients with inflammatory bowel disease (IBD) or other disorders, are surprisingly similar. Strategies for patient selection, recruitment, randomization, blinding, assessing compliance, data analysis and communication of results are not disease-specific. Distinctiveness lies in the features of the diseased populations, such as age, sex or genetics, the unique characteristics of the disease, therapeutic approaches, outcomes of importance and interpretation of results. The general and unique features of assessing outcome or treatment efficacy, with relevant examples in Crohn's disease (CD) or ulcerative colitis (UC), will be discussed.

## GOALS OF THERAPY

As responsible clinicians with a general objective to improve our patients' health state, we read relevant literature to optimize therapy for different groups of patients and to individualize therapy for our IBD patients.

The questions we pose which are germane to patients with CD or UC are the following: How can we prevent a disease exacerbation? How can we induce a remission? How do we prevent or minimize the complications of disease, such as cancer in UC or obstruction in patients with CD? Can we avoid treatment toxicity, such as moon facies,

Department of Medicine, Division of Gastroenterology, McMaster University, Hamilton, Ontario

Correspondence: Dr E Jan Irvine, Associate Professor, Department of Medicine, Division of Gastroenterology, Room 4W8-HSC, McMaster University, 1200 Main Street West, Hamilton, Ontario L8N 3Z5. Telephone (416) 521-2100 ext 6402

mément pour assurer une évaluation objective et impartiale. La définition attentive à priori des incidents à survenir est essentielle. L'analyse statistique dépend du type de données, qui englobe la survenue d'incidents et le modèle de l'étude. L'intention de traiter et les analyses d'efficacité doivent être identifiées. L'interprétation des résultats doit porter sur la portée clinique et statistique.

acne, hirsutism, weight gain or other unpleasant effects of corticosteroids? How can we improve this patient's quality of life? What is the cost to this patient, to the hospital, to the pharmaceutical or insurance company or to society as a whole for a particular treatment? What are the costs and benefits of one treatment strategy over another?

### DEFINING AN OUTCOME EVENT

To answer these questions, we must more closely examine outcome measurement. The modalities to assess success or failure of a new therapy in IBD are unique to the disease, while the statistical methods, interpretation and distribution of the results to the medical community, are not.

A generalist might consider a treatment 'success' or 'failure' as the sum of events occurring after a 'treatment intervention'. A purist would demand more careful dissection of phenomena into those that were directly attributable to the treatment and those which were likely coincidental. As well, quantification of the efficacy or lack of efficacy of the new treatment would enhance the meaning of the results. A rudimentary approach to measuring outcome has been to examine the number of fatal versus nonfatal events (mortality) after an intervention. However, as IBD rarely is fatal, staggering numbers of patients would be required for such a clinical trial, and the potential benefit (or harm) of the drug would be dissipated if it failed to reduce or worsen mortality.

### SURROGATE OUTCOME MEASUREMENT

Outcome assessment in IBD, as for other disease entities, has been replaced by 'surrogate markers', ie, symptoms, signs or laboratory test results which are important measurements of health status for that particular disease.

For IBD, these could include stool frequency, presence of blood in stools, fever, extraintestinal complications, abdominal tenderness, mass or blood parameters, such as erythrocyte sedimentation rate or orosomucoid. More recently, functional change, such as the ability to attend work or school, or changes in quality of life index are becoming more widely accepted.

A fundamental trait of an outcome measurement is that it reflect the proficiency of the intervention, ie, the new drug or surgical technique. To decide the best outcome, it is helpful to know something of the disease biology and the expected mechanism of the drug or intervention. Although it was early appreciated that sulfasalazine was an 'anti-inflammatory drug' with potent benefits in the treatment of active UC, it was only later discovered that the active moiety was aminosalicylic acid (5-ASA) (1), and later still that some of its effect was due to inhibition of prostaglandin and leukotriene production (2,3). Thus, early outcomes or measurements of health status changes primarily included clinical disease activity index; later, when the mechanisms of drug action were the focus of research questions, blood and tissue inflammatory mediators were more appropriate.

Other features of outcome measurement are variably present. Less commonly, an outcome measurement may represent a single state which was not present at the outset of the trial. An example of this would be a clinical trial testing a new drug to prevent disease exacerbation in UC. Although exacerbation would not have been present in patients at entry, careful definition of remission and 'exacerbation' at entry would be essential to permit recognition of an 'outcome event' and subsequent assessment of the drug efficacy.

In many studies, a comparison is made between the health status at the

beginning of the trial (or baseline) and at the end, in which we look for 'improvement' or 'worsening'. This often is referred to as a 'transition event'. Most clinical trials are of a pre-specified duration with repeated measurements over time, sometimes called 'time series' or 'time dependent' measurements.

Other general descriptors for surrogate outcomes reflect our common sense aspirations to prove the intervention beneficial, such as a desirable outcome (benefit) in which the patient improves, or an undesirable outcome (risk) in which the patient worsens or suffers a side effect. These effects may be anticipated further if we have prior knowledge of the drug mechanism or toxicity, or be unanticipated if they are unexpected or idiosyncratic.

### SELECTING AN OUTCOME EVENT

Selecting the tools to measure study outcome (Table 1) is crucial to the scientific rigour and results of a clinical trial, and must be decided before a study begins. The criteria must be objective and interpretable by all participating investigators (or readers) in the same fashion. Thus, clearly defined criteria are necessary. There must be adequate patient follow-up and uniform application of outcome assessment. Subjects dropping out of trials prematurely often do so because a drug is ineffective or has an unpleasant side effect. The importance of blinding of outcome assessment warrants reiteration to assure the objectivity upon which a clinical trial result may hang perilously. Finally, this assessment often is undertaken by an adjudication process or a committee to ensure fairness and standardized application of outcome criteria.

When assessing results of (or participation in) a clinical trial, how do we recognize that the outcome events selected were the most appropriate? Is a decrease in the CD Activity Index (CDAI) of 50 points clinically important? Does such a result suggest we should advocate or avoid the treatment?

Consider the following hypothetical questions which may have been posed for two clinical trials in IBD.

**Question one:** Does drug 'X', at a dose of 1 g tid orally, induce remission within eight weeks in patients who have a non-severe exacerbation of extensive UC? We can discern that the study patients have UC and are experiencing a mild or moderate exacerbation. Drug X will be given orally for eight weeks at 1 g tid. What should be our outcome events?

**Question two:** Does drug 'Y', at 500 mg bid given orally for one year, prevent exacerbation of CD? This study will examine patients with CD who are in remission and who will receive drug Y for one year. Thus, outcome assessment in the two trials will necessarily be different.

In the first study, we expect the patients with UC exacerbation at baseline to be completely well or improved after eight weeks. Thus, we must define 'remission', 'improvement' and 'exacerbation' for these time-dependent events, using the same yardstick. Such a trial would likely assess several outcome events, such as a disease activity index of symptoms and physical signs, endoscopic appearance, possibly histology and adverse drug effects. It must be decided beforehand which combination of these four gauges is the most important and worthy of the title 'primary outcome'. Other secondary outcomes might lend support to the primary measurements but in the event of a discrepancy between primary and secondary outcomes would not sway beliefs about the drug efficacy. These may be softer measures, such as quality of life index, biochemical test results, or number of hospitalizations or surgeries for UC. We might arbitrarily define improvement as a "decrease in the disease activity index (4) of, say, five points together with improvement in endoscopic appearance of one grade (5)". Remission could be defined as a score of '0' on the clinical index, normal endoscopic appearance and no histological activity.

Turning to the second question, let us examine the outcome measurements used in the Canadian Crohn's Relapse Prevention Trial of chronic low dose cyclosporine administration to prevent disease worsening (6). The primary outcome measurements in that trial

were the median time to exacerbation, the proportion of patients having an exacerbation, and the frequency and severity of adverse events over the 18 months of the trial. Exacerbation was defined as an increase in the CDAI (7) of 100 points above baseline. Patients at entry were stratified by activity index (CDAI  $\leq 150$ , CDAI  $> 150$ ). Adverse events were a descriptive list of events classed as requiring premature withdrawal or not. Secondary outcome measurements included quality of life index (8,9), mean dose of corticosteroids or 5-ASA, drugs and the frequency of hospitalization, surgery or death. These outcomes permitted us to assess the impact of low dose cyclosporine use and to demonstrate that it provided no additional benefit in patients already stable at study entry (6).

The outcome events must therefore be linked to the questions of the study, be objective and preferably be quantitative.

#### THE IMPORTANCE OF COMPLIANCE

Compliance occurs at several levels in clinical trials. It includes adherence to the protocol by clinicians and patients, especially the taking of study medication but also keeping appointments and avoiding other nonstudy interventions. Compliance may be considered an outcome measurement in its own rite as it may completely explain the results of a study. Thus, it must also be measured objectively when possible, for example by pill count, blood or urine testing, to prevent bias. If drug A appears to be no better than placebo, it is possible that patients presumed to be taking the drug were noncompliant or failed to continue in the trial due to an undesirable side effect, or that those who were 'on placebo' bought or somehow procured and took the active drug.

#### EXAMPLES OF OUTCOME EVENTS IN IBD

**Activity measurement in UC:** Activity indices, whether for UC or CD, are clusters of symptoms, signs and test results which suggest a particular health state, ie, remission, or activity of a cer-

**TABLE 1**  
**Outcome measurement**

- Selection of outcome events should capture and quantify the efficacy of the new therapy and answer the research questions posed
- A priori definition of objective criteria for primary and secondary outcomes must be undertaken
- There must be a standardized frequency of assessments
- Assessment of outcome events must be blinded
- Adjudication process by a committee should be considered to ensure objective assessment

tain severity. Disease activity indices tend not to be useful in daily practice when assessing individual patients on repeated occasions before and after usual therapy. In practice, clinicians perform a quick 'gestalt' activity index with a few basic criteria that are rechecked at follow-up visits for improvement or worsening. During clinical trials, however, objective reproducible criteria are necessary to ensure that there is standardization among many clinicians or centres involving large groups of patients.

In the early 1960s, Truelove and Witts (10) developed criteria for mild and severe UC. However, as with most categorical or 'ordinal' indices, moderate exacerbation was ill-defined, lying somewhere between 'mild' and 'severe', and patients often had some, but not all, features in a single category. Thus, defining improvement or worsening based on such indices was difficult except when patients were in complete remission with absence of all features of active disease. A second generation index was the St Marks index for extensive UC (Table 2) which graded 11 different features of the disease, including sigmoidoscopy from 0 to 3, with a full score range from 0 to 22 (4). This permitted a clearer quantitative definition of disease activity or change in activity, ie, improvement or worsening. However, most features of this index are not pertinent for patients with limited colitis or proctitis since few such patients experience anorexia, abdominal tenderness, fever, nausea or signifi-

**TABLE 2**  
St Marks ulcerative colitis activity index

1	<b>General health</b>
	0 good
	1 slightly impaired
	2 activities reduced
	3 unable to work
2	<b>Abdominal pain</b>
	0 none
	1 with bowel actions
	2 prolonged
3	<b>Bowel frequency</b>
	0 less than three times a day
	1 three to six times a day
	2 more than six times a day
4	<b>Stool consistency</b>
	0 formed
	1 semiformed
	2 liquid
5	<b>Blood in stool</b>
	0 none
	1 trace
	2 more than a trace
6	<b>Anorexia</b>
	0 no
	1 yes
7	<b>Nausea/vomiting</b>
	0 no
	1 yes
8	<b>Abdominal tenderness</b>
	0 none
	1 mild
	2 marked
	3 rebound
9	Choose 1 if only one of the following is present and if is mild; choose 2 if one of the following is severe (or if more than one of the following was chosen):
	eye inflammation
	arthralgia
	oral ulceration
	related skin lesions
10	<b>Body temperature</b>
	0 less than 37.1°C
	1 37.1°C
	2 greater than 38°C
11	<b>Sigmoidoscopy</b>
	0 nonhemorrhagic
	1 friable
	2 spontaneous bleeding
<b>Total score:</b> 0 to 29	

cant impairment of daily activities. Thus, an index (11) was developed for patients with limited colitis to provide a quantitative index (range 0 to 12) which included rectal bleeding, stool frequency, sigmoidoscopic appearance and global physician assessment.

**TABLE 3**  
Crohn's Disease Activity Index

Daily diary for seven days prior to visit for 1 to 3	
1	Number of liquid/very soft stools: total for one week times 2
2	Abdominal pain: total score for week times 5 (0 none; 1 mild; 2 moderate; 3 severe)
3	Well being: total score for week times 7 (0 normal, under par; 1 slightly under par; 2 poor; 3 very poor; 4 terrible)
4	Number of extraintestinal problems times 20 arthritis/arthralgia iritis/uveitis, erythema nodosum, pyoderma gangrenosum, aphthous ulcers anal fissure, fistula, abscess other fistula temperature >37.5°C
5	Taking antidiarrheals times 30 (0 no; 1 yes)
6	Abdominal mass times 10 (0 no; 2 dubious; 5 definite)
7	Hematocrit males (47 minus hematocrit) times 6; females (42 minus hematocrit) times 6 include $\pm$ sign
8	% ideal body weight ((ideal - measured)/ideal) times 100
<b>Total score:</b> Inactive disease <150, severe activity >450	

Based on reference 7 (Best WR, et al. *Gastroenterology* 1979;77:843-6)

The CDAI (7), although not used by most of us in clinical practice, is the one most familiar from reported studies and includes eight features of disease activity which are variably weighted to yield a score range from 0 to about 700 (Table 3). Active disease is considered greater than 150 and severe disease greater than 450. The simple index (12) has reduced the number of features from eight to five, and eliminated the weighting of scores and the need for patients to keep a diary (score range 0 to 29). Both indices have been shown to be reliable in clinical trials (as have other similar indices). Garrett and Drossman (13) have summarized the limitations of these and other scales, which have included lack of validation in some cases, substantial interobserver variability and inability to detect functional disability (13). Some scales show poor responsiveness to change in clinical status or fail to gauge symptoms such as intestinal obstruction or perianal disease (13). The author's group (14) recently developed a simple five-point index to evaluate change in perianal disease activity.

**Radiological assessments:** Radiological assessment, such as double contrast barium enema in IBD, undoubtedly is

more suited as a diagnostic – rather than an evaluative – tool. For safety reasons, it is undesirable to undertake repeated studies which expose patients to ionizing radiation. Barium studies are useful to evaluate small bowel CD because of the relative insensitivity and inaccessibility to small bowel with other techniques. Most radiological features, however, are nonresponsive to change except in long term studies of many months (15). The computed tomography (CT) scan is expensive and requires considerable radiation exposure. CT, ultrasound or magnetic resonance scanning might reveal nonspecific colonic wall or mesenteric thickening in IBD, or the presence of abscess or fistula in the abdomen, pelvis or perianal regions. For patients with colonic disease, fistula or abscess, <sup>111</sup>indium scanning or fecal indium quantitation are useful for disease localization or severity but, like x-rays, are poorly conducive to repeated measurements (16). Intestinal permeability studies using oral, intestinal or colonic administration of probe molecules, like <sup>51</sup>CrEDTA, polyethylene glycols, mono- and/or disaccharides with quantification of urinary recovery, also yield a measure of disease activity but which may be affected by the use of nonsteroidal or other drugs (17).

**Endoscopic assessment:** Endoscopic appearance of rectal mucosa seen at sigmoidoscopy has changed little in UC since the Baron score (Table 4) (5), which is graded on a four-point scale from 0 (inactive) to 3 (severely active appearance). Although minor variations have been used in recent studies, most clinicians and investigators still employ this index to assess activity in UC. The Crohn's Disease Endoscopic Index of Severity (18), an endoscopic classification of CD useful in diagnosis and grading severity of both upper and lower intestinal CD, like the CDAI, is not practical for routine application, but is quantitative and reliable for use in clinical trials. It was used to demonstrate that prolonged steroid therapy failed to improve endoscopic appearance significantly (19). This index does not appear to correlate closely with clinical activity.

Endoscopic assessment also permits procurement of intestinal mucosal biopsies for histological and inflammatory mediator measurements. Histological evaluation in UC has been standardized and although minor adaptations have been reported, the one used by Riley et al (Table 5) (20) – which was validated during their study – has appeared frequently in the literature. Histology has played a lesser role in the evaluation of change in CD activity since the site of disease limits the availability of tissue, endoscopic biopsy evaluates only the mucosa and not the full wall thickness, and is complicated by the focality of the disease and sampling bias.

**Biochemical assessments:** Biochemical surrogate marker levels in both CD and UC often are increased due to other conditions, but because these processes are automated, there is less bias and random error in their measurement (21,22). For example, low hemoglobin may be due to blood loss, type of drug therapy or nutritional parameters, and depends somewhat on disease extent. Erythrocyte sedimentation rate has a moderate correlation with disease activity except in proctitis, while platelet count tends to be elevated in more severe colonic disease. The seromucoids, such as the acid alpha-glycoprotein

**TABLE 4**  
**Baron score: ulcerative colitis endoscopic score**

0	None (normal, noninflamed mucosa)
1	Mild (abnormal vascular pattern, edema, granularity, hyperemia)
2	Moderate (friability, petechiae, mucopus, plus all of grade 1)
3	Severe (spontaneous bleeding, erosions, ulceration, and all of grades 1 and 2)

Based on reference 5 (Baron JH, et al. *Br Med J* 1964;1:89-92)

orosomucoid or C-reactive protein, rise in response to certain cytokines during moderate and severe activity, and tend to fall quickly after initiation of steroid treatment. Albumin, when abnormal, reflects nutritional state and disease extent as well as severity. Other markers not available and impractical for clinicians include other blood, urine, fecal, and tissue inflammatory mediators, enzymes, cytokine adhesion molecules, receptors or specific mRNAs, which reflect both disease activity and mechanisms of inflammation.

#### QUALITY OF LIFE MEASUREMENT IN IBD

Because of limitations of disease activity indices and the chronicity of IBD, several disease-specific quality of life indices have been developed (9,22,23). Although several general instruments, like the Sickness Impact Profile and other disease-specific instruments, have been assessed, only the McMaster inflammatory bowel disease quality (IBDQ) study (9) was developed specifically as an outcome measurement for clinical trials. Quality of life instruments are generally patient-reported (subjective) and quantitative instruments which evaluate patient satisfaction and function in the physical (eg, pain, diarrhea), social (ability to work, attend social engagements) and emotional (anger, irritability) spheres.

The IBDQ score (range 32 to 224, higher score indicates better quality of life) in a group of healthy volunteers was  $212 \pm 12$  (unpublished data). In one clinical trial, patients 'in remission' at study entry had mean scores of  $168 \pm 31$ , which remained stable in patients who

**TABLE 5**  
**Histological activity in ulcerative colitis**

0	Normal
1	Mild increase in chronic inflammatory cell infiltrate, no tissue destruction
2	Moderate increase in chronic inflammatory cell infiltrate, no tissue destruction
3	Marked increase in chronic inflammatory cell infiltrate, mild tissue destruction
4	Marked increase in chronic inflammatory cell infiltrate, obvious tissue destruction

Based on reference 20 (Riley SA, et al. *Gut* 1988;29:669-74)

did not experience a flare-up over 18 months (24). In a second study, patients experiencing a mild to moderate exacerbation of CD had a significant improvement of 34% ( $P < 0.05$ ) in the total IBDQ score after two weeks of prednisone therapy (25).

#### ADVERSE TREATMENT EFFECTS

Clearly, the appraisal of adverse effects of therapeutic interventions in the existent literature lacks sophistication and is, perhaps, at a level more rudimentary than even the early activity indices. Most often, adverse effects are described as 'deaths', 'life-threatening adverse events' or a frequency distribution of self-reported symptoms displayed as occurring in patients taking either active or placebo therapy. Although this is a practical foundation on which to build, the full spectrum of toxicity, as for efficacy, must be evaluated. Additional features should include frequency and cause of fatal or nonfatal incidents, frequency and grading of severity of the latter, whether they occur early or late after drug initiation and whether they appear to be dose-related or idiosyncratic. Further elucidation of requirements for complete discontinuation of therapy, dose reduction, the degree of reversibility and whether there are potential interactions with other drugs or implications for child-bearing women are essential. Standardized instruments, like the quantitative activity or quality of

**TABLE 6**  
**Statistical guidelines**

Study design	Data type	Statistical test
Two independent groups	Proportions (nominal), rank ordered (ordinal), measured quantities ie, mean (interval/ratio), Survival data	Chi-square; Fisher's exact, Mann-Whitney 'U'; Unpaired 'T'; Mantel Haensel; log rank
Two related groups	Proportions Rank ordered Measured quantities	McNemars Chi-square; binomial sign test; Wilcoxon signed rank; Paired 't' test
More than two independent groups	Proportions Rank ordered Measured quantities Survival data	Chi-square Kruskal-Wallis Analysis of variance Log rank
More than two related groups	Proportions Rank ordered Measured quantities	Cochran Q, Friedman, Analysis of variance
Multivariate	Proportions Measured quantity	Log linear models Multiple regression, discriminant function
Cohort design/ case control	Rates exposed/unexposed, Rates exposed/unexposed	Relative risk Odds ratio

life indices, should be developed for adverse effect application since clinicians often are faced with choosing between a highly efficacious drug with serious side effects and a drug of both lesser efficacy and lesser toxicity.

### CLINICAL VERSUS STATISTICAL SIGNIFICANCE

As clinicians, we must decide when to adopt a particular management strategy for an individual patient. As investigators, we must decide whether study results support the success or failure of a certain therapy. Statistical analysis is a practical mathematical means to assess success or failure objectively. Nevertheless, it is the clinician who must decide whether 40% of patients in remission taking treatment A versus 35% taking treatment B or a difference in CDAI of 30 points between two treatment strategies is sufficient to justify the associated adverse effects or to choose one or the other treatment, despite  $P < 0.05$ . Clinical significance reflects the relevance of the primary and secondary outcome measurements to the research question, while statistical significance evaluates the number of patients, type of data collected, choice of statistical test, arbitrary predefined test parameters (like types I and II er-

ror) as well as the success or failure of the treatment.

Guidelines for the use of statistical tests are presented in Table 6. Selection of statistical test is dependent on the study design and the type of data which constitute the outcome events. There are four different types of data. Nominal data, such as sex, site of disease or smoking status, describe features that are absent or present. Ordinal or categorical data, such as mild, moderate or severe disease, categorize a particular feature and imply a direction of change, ie, mild is better than moderate or severe. Interval data, such as the difference in CDAI or UC activity score between the end of the study and baseline, tell us the amount of rise or decrease in the activity while ratio data have well-defined anchor points, such as the normal hemoglobin ( $\geq 150$  g/L) or absolute CDAI score (normal is 0,  $>150$  implies active disease). For each data type and study design, there are one or more appropriate summary statistics best suited to the analysis.

Data analysis and presentation of results must also answer questions of clinical importance (Table 7). Readers or investigators must be able to tell how missing data, patients lost to follow-up, dying of unrelated causes or taking ad-

**TABLE 7**  
**Statistical analysis**

1	The test statistics should relate to the research question and outcome assessment
2	Techniques should be used to handle missing data
3	The statistical software programs used should be stated
4	If an interim analysis was done, it should be recorded and whether appropriate adjustments were made
5	The appropriate alpha, beta, power levels and confidence intervals should be reported

ditional treatments were handled in the analysis. It also is important to correct for multiple statistical testing to minimize the risk of a type I error. Omission of such information from study results can alter the interpretation of outcome dramatically. Reporting the balance of prognostic variables between treatment groups, eg, number of patients who are steroid-dependent, or taking 5-ASA versus nonsteroid dependent or not on 5-ASA also has important effects on the final result. Statistical evaluation should assess both intention to treat and efficacy analysis. The former assesses whether patients are better because of the original treatment group assignment (and thus accounts for patients dropping out due to lack of drug efficacy or side effects) while the latter tests whether the drug worked in those patients who actually took it.

While clinicians may not appreciate the subtle differences in the more complex statistical tests, they can at least recognize the common and notable flaws in data analysis.

### INTERPRETATION OF RESULTS

Finally, the study results should be interpreted objectively. The discussion section of a manuscript should identify the important findings – but also the limitations – of the study, and why these results are consistent or incongruent with findings of others. They should be presented clearly for the clinician as well as the investigator. The reader (investigator) must question

the validity of the conclusions. Problems or new questions to be addressed by future research should be identified. When tying together the results of a study, as clinicians, we ought to be able to predict the impact of the results on state-of-the-art clinical practice. Alternatively, are we satisfied that a negative study has been adequately explained?

**ACKNOWLEDGEMENTS:** Thanks to Professor DW Taylor of the Department of Clinical Epidemiology and Biostatistics for preparing Table 6.

#### REFERENCES

1. Khan A, Piris J, Truelove S. An experiment to determine the active therapeutic moiety of suphasalazine. *Lancet* 1977;ii:892-5.
2. Ligumsky M, Karmeli F, Sharon P, et al. Enhanced thromboxane A2 and prostacyclin production by cultured rectal mucosa in ulcerative colitis and its inhibition by steroids and sulfasalazine. *Gastroenterology* 1981;81:443-9.
3. Stenson WF, Lobos E. Sulfasalazine inhibits the synthesis of chemotactic lipids by neutrophils. *J Clin Invest* 1982;69:494-7.
4. Powell-Tuck J, Brown RL, Lennard-Jones JE. A comparison of oral prednisone given as single or multiple daily doses for active procto colitis. *Scand J Gastroenterol* 1978;13:833-7.
5. Baron JH, Connell AM, Lennard-Jones JE. Variation between observers in describing mucosal appearances in procto colitis. *Br Med J* 1964;1:89-92.
6. Archambault A, Feagan B, Fedorak R, et al. The Canadian Crohn's Relapse Prevention Trial (CCRPT). *Gastroenterology* 1991;100:A193. (Abst)
7. Best WR, Becktel JM, Singleton JW. Rederived values of the eight coefficients of the Crohn's disease activity index (CDAI). *Gastroenterology* 1979;77:843-6.
8. Mitchell A, Guyatt G, Singer J, et al. Quality of life in patients with inflammatory bowel disease. *J Clin Gastroenterol* 1988;10:306-10.
9. Guyatt GH, Mitchell A, Irvine EJ, et al. A new measure of health status for clinical trials in inflammatory bowel disease. *Gastroenterology* 1989;96:804-10.
10. Truelove SC, Witts LJ. Cortisone in ulcerative colitis: Final report on a therapeutic trial. *Br Med J* 1955;2:1041-4.
11. Sutherland LR, Martin F, Greer S, et al. 5-Aminosalicylic acid enema in the treatment of distal ulcerative colitis, proctosigmoiditis and proctitis. *Gastroenterology* 1987;92:1894-8.
12. Harvey RF, Bradshaw JM. A simple index of Crohn's disease activity. *Lancet* 1980;i:514-5.
13. Garrett JW, Drossman DA. Health status in inflammatory bowel disease: Biological and behavioural considerations. *Gastroenterology* 1990;99:90-6.
14. Irvine EJ, Stoskopf B, Donnelly M. A disease activity index for patients with perianal Crohn's disease (CD). *Gastroenterology* 1990;98:A177. (Abst)
15. Goldberg H, Caruthers SB, Nelson JA, Singleton JW. Radiological findings of the National Cooperative Crohn's Disease Study. *Gastroenterology* 1979;77:925-37.
16. Saverymuttu SH, Camilleri M, Rees H, Lavender JP, Hodgson HJF, Chadwick VS. Indium<sup>111</sup> granulocyte scanning in the assessment of disease extent and disease activity in inflammatory bowel disease. *Gastroenterology* 1986;90:1121-8.
17. Peters TJ, Bjarnason I. Uses and abuses of intestinal permeability measurement. *Can J Gastroenterol* 1988;2:127-32.
18. Mary JY, Modigliani R. Development and validation of a Crohn's disease endoscopic index: A prospective multicenter study. *Gut* 1989;30:983-9.
19. Modigliani R, Mary JY, Simon JF, et al. Clinical biological and endoscopic picture of attacks of Crohn's disease: Evolution on prednisolone. Groupe d'etudes therapeutiques des affections inflammatoires digestives. *Gastroenterology* 1990;98:811-8.
20. Riley SA, Mani V, Goodman MJ, Herd ME, Dutt S, Turnberg LA. Comparison of delayed release 5 aminosalicylic acid (mesalazine) and sulphasalazine in the treatment of mild to moderate ulcerative colitis relapse. *Gut* 1988;29:669-74.
21. Hodgson HJF, Mazlam MA. Assessment of drug therapy in IBD. *Aliment Pharmacol Ther* 1991;5:555-84.
22. Drossman DA, Patrick DL, Mitchell CM, Murp EA, Zagomi EA, Appelbaum MI. HRQOL in inflammatory bowel disease: Functional status and patient worries and concerns. *Dig Dis Sci* 1989;34:1379-86.
23. Farmer RG, Easley KA, Farmer JM. Quality of life assessment by patients with inflammatory bowel disease. *Cleve Clin J Med* 1992;59:35-52.
24. Irvine EJ, Feagan B, Rochon J, et al and the CCRPT Study Group. Quality of life adds dimension to clinical trials in IBD. *Gastroenterology* 1992;102:A640.
25. Martin F, Sutherland L, Beck IT, et al. Oral 5-ASA versus prednisone in short term treatment of Crohn's disease: A multicenter controlled trial. *Can J Gastroenterol* 1990;4:452-7.



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

