Methodological Problems in RCTs on IBD

Cottone M[#], Criscuoli V[#,*], Modesto I[#] and Orlando A[#]

[#]*Dibimis (Biomedic Department of Internal and Specialistic Medicine), "V.Cervello" Hospital Palermo, Medicine Department, University of Palermo, Italy*

Abstract: Randomized controlled trials (RCTs) are the gold standard method for developing evidence-based medicine in inflammatory bowel disease (IBD).

Methodological problems in RCTs in IBD concern different aspects such as the definition of the study population due to the extreme variability of patients with IBD, the indices of disease activity, a clearly defined outcome, the environmental risk factors (i.e smoking behaviour) that may influence the randomization, the heterogeneous placebo rate of remission and the different statistical methods used to analyze the results.

It is important that trials are designed efficiently, done well and complement clinical practice with a careful subject selection, standardization of disease activity indices, and precise outcome measurement in order to continue the improvement of the IBD research process.

Keywords: IBD, RCTs.

Randomized controlled trials (RCTs) provide a gold standard evidence on which to base treatment decisions. They give the strongest data on the efficacy of both pharmacologic and non-pharmacologic interventions in inflammatory bowel disease (IBD) [1].

The evaluation of trial design and outcome is an indispensable exercise for correct interpretation of results and for planning future studies. The ideal clinical study should answer to a single primary research question in a well-defined study population and produce outcomes that are both statistically and clinically sound. Common features of RCTs include: definition of a study population, randomized treatment assignment, standardized and well-defined interventions and clearly stipulated outcome measures. Practical considerations in trial design require evaluation of compliance, reproducibility in clinical practice, and fulfilment of regulatory requirements.

Key disease characteristics that should be considered in targeting patients (pts) with IBD include: a) different genetic patterns b) wide variability of clinical presentation c) different extension of the disease d) presence of complications like stenosis or perianal disease e) serologic markers such as C-reactive protein f) natural history of disease, characterized by period of activity and spontaneous remission g) placebo effect.

Subgroup analyses are commonly performed to identify pts who are likely to derive the maximum benefit from a treatment. These analyses are only valid if they are planned before knowledge of the study results by the investigators.

RANDOMIZATION

Purpose of randomization is to obtain comparable groups for prognostic features so that the only difference between them is the assigned treatment. A benefit of randomization is to minimize the potential for bias (a systematic deviation from the truth) [2].

In the RCTs in which many pts are enrolled, we can assume that main demographic and clinical characteristics of pts are well distributed. The clinical trial with a reduced number of pts instead can have groups not completely superimposable as in that it can happen that some pts with the same characteristics are assigned to the same group.

In IBD the randomization does not necessarily produce groups that are similar in important prognostic factors because many environmental risk factors may influence the similarity of baseline characteristics, as listed in Table **1**. 2:1 allocation is better to reduce the number of patients on placebo. The prognostic stratification by severity is essential in the planning of the trials considered the different baseline risk. Rotwell [3] demonstrated how the baseline risk may change the application of clinical trials to all pts.

The consecutiveness is crucial to ensure that the clinician will be objective during the inclusions of pts. In a RCT of Lochs [4] on mesalamine for prevention of postoperative recurrence in Crohn's Disease (CD), the authors show a significant benefit on clinical relapse after resective surgery but only a small favourable effect mainly in pts with isolated small bowel disease. In this trial the consecutiveness has strictly been respected: in the methods are described in detail pts consecutively observed (617), those who did not meet the inclusion criteria and those who did not give consent, so in the end only 324 patients were enrolled in the study.

*Address correspondence to this author at the Medicine Department "Ospedali Riuniti V.Sofia-Cervello" Hospital, Via Trabucco 180, 90146, Palermo, Italy; Tel: +39 091 6802966; Fax: +39 091 6885111; E-mail: vale.cic@tiscali.it

Table 1. Environmental Risk Factor for IBD

	CD	UC
Smoking	+	-
Appendectomy	+	-
Oral contraceptive	+ (?)	+ (?)
Breastfeeding	-	-
NSAIDs	+	+
Genetic factors	+	+
Diet	+	+
• Western stile diet (total fat polyunsaturated fatty acid)	+	+
• Intake vegetable and fiber	?	-

- protective effect; + increased risk, ? inconclusive findings

CD: Crohn's Disease; UC: Ulcerative Colitis

BLINDNESS

The blindness of treatment allocation is important to assure that the trial is unbiased. The principle of blindness wants to avoid both the patient and the physician investigator being influenced in the evaluation of efficacy of a drug by the fact of knowing who was given the trial drug. There are studies in which blindness can not be met: blinding is easy to perform if two medications are being compared while it is more difficult if two different intervention are being compared (for example it is difficult to compare total parenteral nutrition with corticosteroid therapy: in this case you can only have the blindness of the physician evaluating the results of the study). Blinding of the patient and physician may be impossible if a surgical therapy is being compared with medical therapy; it is less difficult if two surgical interventions are being compared. If blinding does not occur, there may be a bias in favour of one of the treatment arm. In a metanalysis [5] of trials for maintainance of remission in CD with mesalamine, the result show a benefit in the post surgical setting; when the study of Caprilli [6], who reported the highest benefit of mesalamine were excluded, a statistically significant effect for maintenance therapy was observed. The trial of Caprilli did not provide for the blindness of the patient and the physician, therefore its result should be viewed with caution.

CONTROL

The control is the medication with which the new drug is being compared. Studies that do not include a control intervention do not have the same power as those which include controls. Any intervention may have a placebo effect and thus placebo medication is administered to the control group in medical trials. Understanding the outcomes among patients with IBD receiving placebo is important for conducting clinical trials because the use of placebo remains the gold standard for testing the efficacy of a new therapy. But while it is justified the use of placebo in a mild disease, the use of placebo as a control arm in pts with moderate-severe disease remains questionable (instead of unethical). In these cases, when an effective conventional therapy exists, the patient is being denied a treatment of proven benefit and thus may experience some additional degree of morbidity. The analysis of the placebo response shows that in every setting there is a high heterogeneity. This demonstrates mainly that the selection of pts is not often homogeneous and that the outcome criteria used in the trials are not highly reliable [7]. A meta-analysis [8] of the placebo rate of remission and response in clinical trial of active CD showed that remission rates for pts receiving placebo are heterogeneous ranging from 0% to 50%. The authors identified several factors that may influence the likelihood to achieve remission in pts receiving placebo: most of the heterogeneity appears to occur among studies with a duration < 2 months, it was related to CDAI score on enrolment and to the number of study visits. Although shorter study duration is desiderable to minimize the placebo remission rate (because pts are more likely to enter spontaneous remission with time), the optimal study duration for an individual clinical trial also depends on the pharmacodinamics of the drug under study, including its time to onset of action. In another study the placebo response resulted to be greater in trials with more frequent study visits (more than three) [9]. An accurate estimation of rates of remission and response among patients receiving placebo is essential for both calculating sample size and interpreting results of clinical trials. Recently a small study evaluated the role of scheduled Infliximab in reducing 12 months severe endoscopic recurrence showing a rate of recurrence significantly lower in the infliximab group (1 of 11 patients; 9.1%) compared with the placebo group (11 of 13 patients; 84.6%) (P =0.0006). This means a 75% absolute risk reduction of severe endoscopic recurrence and a NNT of 1.3. Although these results are impressive, the small sample size of the study and the unexpected recurrence rate of the placebo arm weakens the conclusion [10]. It is important to emphasize that the comparison drug may be the true measure of efficacy of study medication.

SAMPLE SIZE

Pocock found that sample size was based on prior statistical power calculations [11].

Determining sample size is a key design issue; it is the number of events (endpoints) that determines power. If the event rate is lower than predicted, there will be fewer events (thus less power at the planned analysis time). A smaller difference between treatment requires more events to determine the true effect size. If there are fewer events than planned, one should consider how much follow-up information is available. Follow-up assessments and rates of missing data should be similar on each arm. Studies that use small patients groups must detect very large differences in the outcome. If the results of an open study have demonstrated a clear efficacy of a new drug, it is acceptable to design a trials in which the sample is not too large. In a randomized, double-blind, controlled trial [12] on cyclosporine on severe Ulcerative Colitis (UC) refractory to steroid therapy 82% of pts treated with cyclosporine had a response to therapy compared to none pts who received placebo. After that 20 pts were enrolled the study monitor agreed that was unethical to continue to enter pts into placebo arm so the trial was interrupted due to the evident efficacy of the treatment. Trials on anti-TNF molecules [13] have been conducted on small samples of patients showing however, a clear efficacy. On the contrary, to demonstrate the efficacy of 5-ASA in maintaining remission in CD, trials with a large sample size were necessary for demonstrating a marginal clinical effect. However, a large sample size may detect a statistically significant test difference that is not clinically meaningful (α error), while a sample size that is too small may yield inconclusive results due to low power (β error). An example of alpha error is the result of the study on the effectiveness of antisense molecules of adhesion molecules in the resistant CD [14]. This study showed that 47% (CI 20-70) of steroid-dependent CD pts had achieved a response versus 20% (CI 7-70) of placebo showing an error in the calculation of the confidence intervals.

OUTCOME

In pts with IBD, a wide spectrum of potential outcomes are available to assess efficacy. Interpretation of research studies conducted in pts with IBD is complicated by the differentiation of the goals of therapy (e.g., induce and maintain remission, minimize complications, and maximize quality of life) and the multitude of patient subsets (e.g., mild-to-moderate, moderately active, severely active, steroid-dependent, and steroid-refractory).

The main outcomes assessed in IBD clinical trials are: remission, recurrence, surgery, closure of fistulas, steroid withdrawal, quality of life. In the last year endoscopic remission has been considered by many authors a main outcome which is debatable until a clear demonstration that this outcome correlates with a better prognosis.

Remission is the best outcome when evaluating a drug on active disease. In the PRECISE 1 trial [15] for the approval of certolizumab pegol in moderate to severe CD who had an inadequate response to conventional therapy, the primary endpoints were induction of a response at week 6 and a response at both weeks 6 and 26 but not the remission. This is why the EMEA has not authorized the use of the drug in Europe. As for the recurrence there are three typed to consider such as clinical, surgical and endoscopical recurrence.

Two distinct approaches are possible in clinical practice for postoperative CD management: starting prophylactic medical therapy after surgery, which aims to avoid or postpone the recurrence; or waiting for endoscopic recurrence, and eventually identifying pts with postoperative recurrence [16], starting or adapting medical therapy before the patient becomes symptomatic. The severity of recurrence assessed by endoscopy is the best predictor of clinical outcome in postoperative CD [17]. Rutgeerts *et al.* have developed a validated endoscopic score to measure the postoperative recurrence and stratify pts, according to their endoscopic findings, in five categories. Endoscopic signs of recurrence appear before symptomatic disease, which allows early treatment that may be effective in modifying the natural course of the disease, preventing clinical recurrence and possibly, the need for new surgery.

A fistula drainage assessment has been used to classify entero-cutaneous or perianal fistulas as being either open and actively draining, or closed [18]. A fistula is open if the investigator can express purulent material from the fistula with the application of gentle pressure. This instrument served as the primary endpoint for a placebo-controlled trial of infliximab for abdominal and perianal enterocutaneous fistulas that led to regulatory approval of the drug in the United States and Europe. Additional placebo-controlled trials with infliximab, tacrolimus, and the humanized anti-tumor necrosis factor antibody CDP571 have used the fistula drainage assessment as a primary measure. Recently, in studies on fistulizing CD, radiological outcomes have been considered. Interesting data have been reported regarding the utility of MRI in following the radiologic course of fistulous tracts in response to infliximab therapy [19,20]. In one of these studies clinical and radiologic response were evaluated in 18 pts with fistulizing CD, before and after treatment with infliximab, given at 5 or 10 mg/kg at weeks 0, 2, and 6 and compared with placebo. MRI was performed at inclusion and at week 6. All 18 pts had signs of active inflammation at baseline. After infliximab treatment, 11 pts had a cessation of fistula drainage, but 8 of these 11 had persistent tracts on MRI. The role of EUS in following of the response to medical therapies such as infliximab is unclear [21].

Many disease activity indexes [22-34] (Table **2**) have been generated to permit assessment of therapy efficacy or to predict clinical course of disease in individuals or groups of patients. The ideal index should be simple to administer and quantitative and it should be composite of subjective symptoms, objective findings and laboratory markers of inflammation. No single clinical or biochemical parameter consistently reflects IBD disease activity or intestinal inflammation.

Ideally, an instrument designed to measure disease activity should be reproducible, valid, and responsive to change. It should also be simple to administer and follow accepted standards of reporting. In pediatric population there has been the need to create specific indexes because of the problem of growth and maturation, which are poorly assessed by adult indices [35]. No single index can satisfy the needs of all

Table 2. Principal Indexes for Measuring Clinical Disease Activity in IBD

Index for Measuring Clinical Disease Activity in CD	Index for Measuring Clinical Disease Activity in UC
• Crohn's Disease Activity Index [21]	• Truelove and Witt's Score [27]
• Harvey Bradshaw Index or ''simple index''2 [22]	• Lichtiger Score [28]
• Van Hess or Dutch Index [23]	• Powell Tuck Index [29]
• Therapeutic Goals Score (Present/Korelitz Index) [24]	• Clinical Activity Index (CAI; Rachmilewitz Index) [30]
• International Organization of Inflammatory Bowel-Disease-Oxford [25]	• Mayo Score [31]
• Cape Town Index [26]	• Sutherland Index [32]
	• Physician's Global Assessment (PGA) [33]

Table 3. Methodological Aspects of RCTs in IBD: Main Problems

Randomization	Blindness	Control/Placebo	Sample Size	Outcome	Statistical Methods
In IBD environmental risk factors may influence the similarity of baseline characteristics

2:1 allocation is better to reduce the number of patients on placebo | Indispensable considering that CDAI and Mayoclinic score are based on symptoms

Impossible if comparing medical vs surgical treatment | The placebo rate of remission and response in CD is heterogeneous

Ethical problems when evaluation a new biologics in moderate severe disease | Calculated on the basis of placebo rate of remission

The heterogeneity of the placebo remission rate make difficult to calculate the sample size | Remission is the more relevant outcome

Problems linked to the scores adopted (CDAI and Mayo-Score) | Intention to treat analysis(ITT): necessary to ensure correct randomization

Not often present in trials on IBD |

trials so the definition of remission, improvement or worsening, the important outcome events in any clinical trial, must be defined by the investigators based on the study objective and the expected mechanism of treatment.

STATISTICAL METHODS

The best way to review the results of a clinical trial is evaluating the statistical method. In the "intention to treat analysis" all randomized pts who took even one tablet are included in the analysis; this leads to a conservative estimate of the effect size and tends to give results that are less biased toward showing efficacy, including also the treatment failure. Not often is present in trials on IBD. The "per protocol analysis" is confined to pts who have complied with all aspects of the protocol, by excluding pts who stopped during the study. This type of analysis leads to biased estimates. Confidence intervals provide the clinicians with the range of possible results if the experiment was repeated, using the same number of patients 95 out of 100 times. The value are sensitive to the number of patients enrolled in the trial. If the intervals include the unity or a range of positive and negative values, the results are not statistically significant.

Meta-analysis can combine the data of trials sufficiently homogeneous into an overall estimate of the size of the treatment effect, overcoming the risk of false negative results of undersized trials, and providing a more precise and stable measure of the treatment effect. The main problems are the quality of trials included: even if the search for relevant evidence in the literature has been well conducted, heterogeneity among the included studies may alter the result. An emerging problem is the "publication bias", a systematic preference for publication of studies that demonstrate statistically significant and positive findings compared to studies exhibiting negative or indeterminate findings. Timmer *et al.* [36] examined publication bias in gastroenterology in the early 1990s. The tables and the plot displaying the results of trials and the pooled estimate of meta-analyses are well standardized and easily understandable but we need to look the distribution of the treatment effect of the individual trial. If there is substantial qualitative heterogeneity (several trials showed results in favour of the control treatment) any pooled estimate of the treatment must be considered with great caution, even more if the heterogeneity is statistically significant. Kottachachi [37] evaluated all abstracts of phase III randomized controlled trials in IBD accepted at Digestive Disease Week from 1998 to 2003 and assessed the presence of publication bias for negative abstract and evaluate publication rates of complete articles after abstract presentation and quality of information. They showed that the 5-year cumulative rate of publication was significantly higher for abstract with positive compared to negative findings (91% ver-

sus 50% P< 0.001) and reveals the use of data from abstract as a potential means of mitigating publication bias in IBD literature. Moreover a small but clinically relevant 6.3% of fully published abstract changed the overall significance and conclusion of their initial study.

Multivariate modeling is an essential part of many clinical studies; it is the only feasible way to deal with many variables at one time during the analysis phase of a study. It is used to adjust simultaneously for the effects of many variables to determine the independent effect of one. The method can select from a large set of variables a smaller subset that independently and significantly contributes to the overall variation in outcome and should be performed on a large sample of patients. Simpler methods, such as stratification or matching, can only consider a few variables at a time and the only by sacrificing statistical power.

CONCLUSION

Given the large amount of evidence coming from the medical literature, strategies are necessary to limit the attention on reports that have the potential to change the clinical management of the pts (Table **3**). Critical evaluation of literature can help the clinician to consider useful data only those based on properly executed trials of high quality.

CONFLICT OF INTEREST

The author(s) confirm that this article content has no con-flicts of interest.

ACKNOWLEDGEMENT

REFERENCES

[1]   Cottone M, Martorana G. Methodology of controlled trials in IBDs. Medic 2001; 9: 25-31.
[2]   Feagan B, McDonald J, Koval J. Therapeutics and inflammatory bowel disease: a guide to the interpretation of randomized controlled trials. Gastroenterology 1996; 110: 275-83.
[3]   Rotwell PM. Can overall results of clinical trials be applied to all patients? Lancet 1995; 345: 1616-9.
[4]   Lochs H Mauer M, Fleig W, *et al.* Prophylaxis of postoperative relapse in Crohn's disease: European Cooperative Crohn's Disease Study VI. Gastroenterology 2000; 118: 264-73.
[5]   Camma' C, Giunta M, Rosselli M, *et al.* Mesalamine in the treatment of Crohn's disease: a meta-analysis adjusted for confounding variables. Gastroenterology 1997; 113: 1465-73.
[6]   Caprilli R, Andreoli A, Capurso L, *et al.* Oral mesalamine for prevention of post-operative recurrence of Crohn's disease. Aliment Pharmacol Ther 1994; 8: 35-43.
[7]   Renna S Orlando A, Mocciaro F, et al. Placebo therapy in Crohn's disease. Eur J Intern Med 2009; 20: 572-8.
[8]   Su C, Lichtenstein GR, Krok K, *et al.* A meta-analysis of the placebo rates of remission and response in clinical trials of active Crohn's disease. Gastroenterology 2004; 126: 1257-69.
[9]   Ilnyckyj A, Shanahan F, Anton PA, *et al.* Quantification of the placebo response in ulcerative colitis. Gastroenterology 1997; 112: 1854-8.
[10]   Regueiro M, Schraut W, Baidoo L, *et al.* Infliximab prevents Crohn's disease recurrence after ileal resection. Gastroenterology 2009; 136: 441-50.
[11]   Pocock SJ, Hughes MD, Lee R. Statistical problems in the reporting of clinical trials: A survey of three medical journals. N Engl J Med 1987; 3: 426-32.
[12]   Lichtiger S, Present DH, Kornbluth A, *et al.* Cyclosporine in severe ulcerative colitis refractory to steroid therapy. N Engl J Med 1994; 330: 1841.
[13]   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.
[14]   Yacyshyn BR, Bowen-Yacyshyn MB, Jewell L, *et al.* A placebo-controlled trial of ICAM-1 antisense oligonucleotide in the treatment of Crohn's disease. Gastroenterology 1998; 114: 1133-42.
[15]   Sandborn WJ, Feagan BG, Stoinov S, *et al.* PRECISE 1 Study Investigators Certolizumab pegol for the treatment of Crohn's disease. N Engl J Med 2007; 357: 228.
[16]   Viscido A, Corrao G, Taddei G, *et al.* "Crohn's disease activity index" is inaccurate to detect the post-operative recurrence in Crohn's disease. A GISC study. Gruppo Italiano per lo Studio del Colon e del Retto. Ital J Gastroenterol Hepatol 1999; 31: 274-9.
[17]   Rutgeerts P, Geboes K, Vantrappen G, *et al.* Predictability of the postoperative course of Crohn's disease. Gastroenterology 1990; 99: 956-63.
[18]   Present DH, Rutgeerts P, Targan S, *et al.* Infliximab for the treatment of fistulas in patients with Crohn's disease. N Engl J Med 1999; 340: 1398-405.
[19]   Rasul I, Wilson SR, MacRae H, *et al.* Clinical and radiological responses after infliximab treatment for perianal fistulizing Crohn's disease. Am J Gastroenterol 2004; 99: 82-8.
[20]   Van Assche G, Vanbeckevoort D, Bielen D, *et al.* Magnetic resonance imaging of the effects of infliximab on perianal fistulizing Crohn's disease. Am J Gastroenterol 2003; 98: 332-339.
[21]   van Bodegraven AA, Sloots CE, Felt-Bersma RJ, *et al.* Endosonographic evidence of persistence of Crohn's disease associated fistulas after infliximab treatment, irrespective of clinical response. Dis Colon Rectum 2002; 45: 39-45.
[22]   Best WR, Beckel JN, Singleton JW, *et al.* Development of a Crohn's disease activity index: National Cooperative Crohn's Disease Study. Gastroenterology 1976; 70: 439-44.
[23]   Harvey RF, Bradshaw J. A simple index of Crohn's disease activity. Lancet 1980; 1: 514.
[24]   Van Hess PA, van Elteren PH, van Lies HJ, *et al.* An index of inflammatory activity in patients with Crohn's disease. Gut 1980; 21: 279-86.
[25]   Present DH, Korelitz BI, Wisch N, *et al.* Treatment of Crohn's disease with 6-mercaptopurine: a long –term, randomized, double blind study. N Engl J Med 1980; 302: 981-7.
[26]   Myren J, Bouchier IA, Watkinson G, *et al.* The O.M.G.E. Multinational Inflammatory Bowel Disease Survey 1976-1982: a further report on 2,657 cases. Scand J Gastroenterol Suppl 1984; 95: 1-27.
[27]   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.
[28]   Truelove SC, Witts LT. Cortisone in ulcerative colitis: final report on a therapeutic trial. BMJ 1955; 2: 1041-8.
[29]   Lichtiger S, Present DH. Preliminary report: cyclosporine in treatment of severe ulcerative colitis. Lancet 1990; 336: 16-9.
[30]   Powell-Tuck J, Day DW, Bucknell NA, *et al.* Correlations between defined sigmoidoscopic appearances and other measures of disease activity in ulcerative colitis. Dig Dis Sci 198227: 533-7.
[31]   Rachmilewitz D. Coated mesalazine (5-amynosalicyclic acid) versus sulphasalazine in the treatment of active ulcerative colitis. A randomized trial. BMJ 1989; 298: 82-6.
[32]   Schroder KW, Tremaine WJ, Ilstrup DM. Coated 5-amyno-salicyclic acid therapy for mildly to moderately active ulcerative colitis. N Engl J Med 1987; 317: 1625-9.
[33]   Sutherland LR, Martin F, Greer S, *et al.* 5-amynosalicyclic acid enema in the treatment of distal ulcerative colitis, proctosi-gmoiditis, and proctitis. Gastroenterology 1987; 92: 1894-8.
[34]   Hanauer S, Schwartz J, Robinson M, *et al.* Mesalamine capsules for treatment of active ulcerative colitis: result of a controlled trial. Am J Gastroenterol 1993; 88: 1188-97.

[35]   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: 439-47.

[36]   Timmer A, Hilsden RJ, Sutherland LR. Determinants of abstract acceptance for the Digestive Disease Week- a cross sectional study. BMC Med Res Methodol 2001; 1: 13.

[37]   Kottachchi D, Nguyen G. Quality and publication success of abstracts of randomized clinical trials in inflammatory bowel disease presented at digestive disease week. Inflamm Bowel Dis 2010; 16: 993-8.