

What Are Open-Label Extension Studies For?



Open-label extension studies are reported frequently in the rheumatology literature, as successful randomized controlled trials (RCT) increase in number and maturity. One such study is reported in this month's *Journal*¹.

Despite their apparent popularity, we have to ask the question, "What are they for?" What function do open-label extension studies serve? Clearly, RCT are designed to test the hypothesis that one intervention is better than one or more comparator interventions. Great attention is paid to the design of these studies to minimize bias that might confound the interpretation of the hypothesis-testing. For example, subjects are randomized to the different treatment arms so that in all respects other than the assigned treatment, the groups are the same at baseline. Similarly, investigators and participants are blinded to the assigned treatment to minimize the possibility of unconsciously or consciously changing the way outcomes are measured on the basis of assigned treatment.

In the case of open-label extension studies, the purpose is not nearly so clear-cut. We can think of 3 possibilities. The first reason is simply to make the (now known to be) effective but as yet unlicensed drug available to participants who were randomized to placebo; this might have been a requirement of the ethics approval or a means of enhancing recruitment to the original RCT. This purpose does not require systematic data collection, and is not a sufficient reason for publishing the results of prolonged observation.

A second reason is that further, more prolonged observation may disclose adverse effects that were not observed in the original parent RCT. The likelihood of observing such events is low, since the cohorts are almost always too small to reliably detect rare events. In the case of anti-tumor necrosis factor therapies, open-label extension studies failed to detect reactivation of tuberculosis, a problem that was only identified through post-marketing surveillance and national adverse event registries. Even in the case of the early studies of prednisolone in RA, failure to identify significant steroid-induced osteoporosis was more a function of inadequate technology (lack of bone densitometry) than lack of prolonged open-label extension. For example, the study of prednisolone remained randomized for 2 years². The

safety issues do not constitute a sufficient reason for conducting open-label extension studies.

The third purpose may be to demonstrate continued efficacy of the drug over a longer period of time or to show that participants randomized to receive the active treatment during the open-label phase achieved outcomes similar to those of participants who received the drug from the beginning of the parent RCT.

This is where things become more complicated. A key reason for the ascendancy of the RCT design is control of bias. In other words, when a difference is detected, we can be confident that the reason for the difference was the drug and not some other factor. There are a number of techniques used in RCT to minimize bias. Blinding is one way of minimizing biased outcomes in RCT. It has been shown clearly that different results occur when assessment of outcome is not blinded. In open-label assessment studies, there is a significant risk of biased assessment. Analysis of all subjects who were randomized (intent to treat analysis) is another important technique, since subjects who drop out can differ in crucial ways from subjects who remain in the study. In open-label extension studies, only a proportion of the subjects continue to be available for the open-label part of the study and, critically, only these subjects are actually analyzed at the end of the observation period. This introduces bias, both in terms of analyzing only subjects who completed the randomized part of the study and also in terms of including only subjects who agreed to continue or start the therapy and be observed for a further time period. One could easily imagine that subjects originally assigned to placebo who were doing well would be less likely to agree to the open-label extension than subjects not doing well; and vice versa for subjects originally assigned to the active drug. It could be considered that such bias renders interpretation of open-label studies nearly impossible, so that trying to confirm continued or late (in the case of subjects originally assigned to placebo) efficacy is also impractical.

Now, even if we tried to account for all the sources of bias when looking at the final outcomes of each group in an open-label study, it is still necessary to make some kind of

See Continued inhibition of radiographic progression in patients with PsA following 2 years of treatment with etanercept, *page 712*

comparison between the groups to see whether there is any difference that could imply continued efficacy. Such comparisons require inferential statistical tests to allow for the possibility of sampling or random error to be the reason for any observed difference. Recently reported open-label extension studies have studiously avoided inferential statistical tests. For example, in an open-label extension study of etanercept for ankylosing spondylitis, "efficacy analyses are reported without statistical inference between treatment groups"³.

Although the CONSORT guideline has provided a framework for the design, analysis, and reporting of RCT⁴, no consensus statement has been developed for open-label extension studies. The OMERACT (Outcome Measures in Rheumatology Clinical Trials) process has developed recommendations for the standard of longitudinal observational studies in rheumatology^{5,6}, and these represent the closest published guideline for open-label extension studies that we have been able to locate. These articles describe means of identifying a number of potential biases and suggestions for analysis, in particular multivariate analyses to determine relationships between outcome and possible explanatory factors. In the case of open-label extension studies, we feel that analysis of efficacy ought to regard treatment allocation as a possible explanatory variable (in addition to others, such as baseline status and time).

The study by Mease and colleagues¹ reports the results of a 48-week open-label extension of a 24-week randomized placebo-controlled trial of etanercept for psoriatic arthritis. The rationale for conducting the study is a desire to report 2-year radiographic and clinical outcomes for people originally assigned to etanercept. There are problems with interpreting these observations.

First, the assessment of clinical outcomes could not be blinded. Radiographs were read in random order, so it is reasonable to assume some degree of blinding for radiographic assessment, even though the radiologists must have known that radiographs from at least 2 timepoints for each patient were performed while the subjects received etanercept. We assume that radiographs from the randomized portion of the study were reread for the purpose of this analysis and the previously available radiographic data (already reported) were not reused, otherwise the blind was not maintained.

Second, only subjects who completed the randomized trial were eligible to participate in the open-label portion. During this randomized phase, there were 32/104 patients randomized to placebo that dropped out before completing 24 weeks of treatment, but 23 who did not take part in the open-label extension. There were 8/101 patients randomized to etanercept who dropped out, but 14 who did not take part in the open-label extension. Final radiographic outcomes were available for 141 subjects. This represents 69% of patients originally entered and although demographic and baseline disease features were similar in this group, there must remain a significant risk of bias that cannot easily be quantified.

Third, only descriptive statistics are presented. The

authors are careful to avoid statistical comparisons, yet are quite happy to make many statements that clearly involve a comparison. For example, they conclude that "patients who switched from placebo to etanercept rapidly reached parity in all clinical response with their etanercept/etanercept treated counterparts after approximately 12 weeks of treatment..." or "The percentage of patients who showed no radiographic progression...was higher for patients originally randomized to etanercept". Such comparisons clearly infer a difference, something that is usually accompanied by a statistical test.

We also note that the 2 groups originally randomized to placebo/etanercept differed in their baseline radiographic scores (total Sharp score 18.30 vs 25.89, respectively), indicating worse radiographic appearances in the etanercept group. Since worse radiographic scores may indicate a group with a worse prognosis, it would be reasonable to adjust any observed differences in the outcome variable for such differences in baseline. In this instance, one would expect the effect of etanercept on radiographic scores to be even greater than was reported.

Notwithstanding these comments on this particular trial, we believe it is time for a wider debate upon the merits, purpose, and design of open-label extension studies. Do these studies represent research or marketing⁷?

WILLIAM J. TAYLOR, MBChB, FAFRM, FRACP,
Senior Lecturer,
Rehabilitation Teaching and Research Unit,
Wellington School of Medicine and Health Sciences,
University of Otago,
PO Box 7343,
Wellington, New Zealand;

MARK WEATHERALL, MBChB, MSc, FRACP,
Associate Professor,
Department of Medicine,
Wellington School of Medicine and Health Sciences,
University of Otago.

Address reprint requests to Dr. Taylor. E-mail: will.taylor@otago.ac.nz

REFERENCES

1. Mease PJ, Kivitz AJ, Burch FX, et al. Continued inhibition of radiographic progression in patients with psoriatic arthritis following 2 years of treatment with etanercept. *J Rheumatol* 2006;33:712-21.
2. Report by the Joint Committee of the Medical Research Council and Nuffield Foundation on clinical trials of cortisone ACTH, and other therapeutic measures in chronic rheumatic diseases. A comparison of prednisolone with aspirin or other analgesics in the treatment of rheumatoid arthritis. *Ann Rheum Dis* 1959;18:173-88.
3. Davis JC, van der Heijde DM, Braun J, et al. Sustained durability and tolerability of etanercept in ankylosing spondylitis for 96 weeks. *Ann Rheum Dis* 2005;64:1557-62.
4. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 2001;357:1191-4.
5. Silman A, Symmons D. Reporting requirements for longitudinal observational studies in rheumatology. *J Rheumatol* 1999;26:481-3.
6. Wolfe F, Lassere M, van der Heijde D, et al. Preliminary core set of domains and reporting requirements for longitudinal observational studies in rheumatology. *J Rheumatol* 1999;26:484-9.
7. Taylor GJ, Wainwright P. Open label extension studies: research or marketing? *BMJ* 2005;331:572-4.