Rheumatoid arthritis

....

Follow up studies in rheumatoid arthritis

R Landewé, D van der Heijde

....

Do we really need follow up studies of randomised clinical trials in RA?

In rheumatology we have been told for years that the results of randomised controlled trials may not be extrapolated to the long term, and that a gain in the beginning may not necessarily mean a gain in what we solemnly call the "final outcome". In almost every review on disease modifying antirheumatic drugs (DMARDs) in rheumatoid arthritis (RA), long term efficacy and long term toxicity are always considered. The importance of long term data is repeatedly stated, without, however, explaining what information is really necessary. And regulatory authorities nowadays require long term studies before approving drugs which claim, for example, to preserve function.

Probably as a consequence of this, we are now facing an increasing number of published follow up reports of clinical trials—for example, references 1–6a.[1–6a] The conclusion is always that some drug is still as effective and as toxic as its comparator, x years after the start of the trial, or that drug A "maintains its efficacy or its superiority over drug B" over time.

REASONS TO PERFORM FOLLOW UP STUDIES OF CLINICAL TRIALS

A large number of arguments are mentioned by different authors to justify follow up studies. These arguments—more or less valid—include doubts about long term clinical efficacy, effects on radiological progression and function, long term toxicity, "drug survival time", cost effectiveness, and quality of life. Intuitively, follow ups of randomised controlled trials (RCTs) enjoy a greater respect than observational studies, because patients were allocated to groups by chance, and most authors assume they can make an appropriate between-group comparison at the end.

The questions to be asked in this editorial are:

- Does the methodology used in follow up studies of RCTs provide data which are valid?
- Do these studies really increase our knowledge of particular drugs and their long term benefits and harms?

METHODOLOGICAL ISSUES

Randomisation is the best insurance for prognostic similarity of the treatment groups.[7] In other words, patients in both groups differ from each other in nothing but their trial treatment. Ideally, this prognostic similarity is kept intact during the study; this is why in most RCTs co-interventions are forbidden or at least strictly regulated, and why RCTs are of limited duration.

> "If there are no differences in treatment effects during an RCT, they are unlikely to appear at follow up."

It has become common practice that at the end of a formal RCT patients are asked—and sometimes encouraged—to continue their trial drug indefinitely. Often, for ethical reasons, statements about this request are included in the original trial protocol. Obviously, only the completers of a comparative clinical trial will continue the study treatment, which may give rise to serious, sometimes interrelated, problems of prognostic similarity.

Confounding by indication

Confounding by indication is a potential pitfall of follow ups of RCTs and observational studies. Most follow up studies do not define which criteria should be met in order for the study drugs to be continued after the end of the trial, or which strategy should be followed if the drug fails. Treatment choice is "left to the opinion of the patient and the doctor", which implies that patients with more severe disease and a worse prognosis will probably be treated with—or switched earlier to—more intensive treatment than patients with less severe disease. In an RCT with evidence that drug A is better than drug B at the end of the trial, such a policy will ultimately reduce prognostic similarity, leading to confounding by indication. As a result of confounding by indication, treatment effects that are divergent at the end of the RCT may converge during follow up (see below).

Cross over of effective treatment[8]

In comparative drug trials, in which one drug is more effective than the other, there is a considerable chance that patients treated with the less effective drug will be changed to the more effective drug after the end of the trial. Sometimes, patients consent to take part in the study only when they receive a promise that they will be treated with the most effective drug after the end of the trial. We have seen protocols in which a statement about cross over of the effective drug has been included, often at the instigation of ethical committees. Cross over of effective treatment differs from confounding by indication in that it is not the (perceived) severity of the disease that determines the choice of drug, but the consequences are similar: loss of prognostic similarity. Clinical trials with significant cross over of treatment are in our opinion not suitable for follow up of efficacy.

Bias of completers

In trial designs, clinical investigators, including the authors of this editorial, pharmaceutical companies, and regulatory authorities are convinced that the intention to treat (ITT) approach is the most robust type of analysis.[9–11] ITT reasonably assures prognostic similarity at baseline (but does not protect against confounding by indication in follow up studies). In follow ups of clinical trials there is often some resistance against ITT analysis, especially if the follow up is long. This feeling is understandable, because true ITT requires follow up of all patients, including those who have taken "only one drug dose" or those who have withdrawn consent during the trial. It is difficult to imagine how patients withdrawing early may contribute to assessing the long term efficacy of a drug that they have hardly taken, and these patients are often reluctant to cooperate years after they have left the study, and may not even be traced. However, results are almost certainly biased if only those patients that continue their study drug for some years are included in the analysis. A completers-only analysis is extremely biased towards good clinical efficacy or good tolerability, or both.[11] In a completers-only analysis with drug A which is relatively toxic and drug B which is tolerated better, for example, the make-up of group A at the end will be determined largely by the absence of toxicity and that of group B largely by acceptable efficacy: failure of prognostic similarity. It is obvious that an appropriate comparison about the efficacy outcome cannot be made in such a case.

STATISTICAL ISSUES

In follow up studies, people have a tendency to assess and report the same

outcome measures as those used in the original RCT. Most often, a mixture of process and outcome variables (for example, WHO/ILAR core set), composite indices (for example, DAS28), and response proportions (for example, ACR20) is presented as in the original trial report. It is questionable whether all these measures are appropriate for assessing efficacy in follow up studies.

Process variables that are appropriate for comparing disease activity in well designed RCTs (including patients with "active disease" at the start) are less suitable for assessing efficacy of a drug during follow up. If there is no "obligatory" treatment protocol, patients and doctors will be inclined to alter the intensity of their treatment to a level of "acceptable" disease activity. In the follow up of the COBRA trial, we found that the DAS28—as well as other process variables—converged towards a mean level of four points over time in both groups, compatible with low to moderate disease activity, and apparently this was the level that was accepted by patient and doctor.[12]

To overcome these shortcomings of process variables, an increasing number of investigators propose time averaged values for disease activity parameters, by using an area under the curve method. There is some rationale for this because time averaged estimates for disease activity correlate better with radiological progression than point estimates.[13 14] But again, time averaged estimates as outcome measures for efficacy in comparative follow up studies are of limited value, if one realises that any contrast between treatment groups will certainly be due to differences appearing during the RCT—not during the follow up. Mutatis mutandis: if there is no significant contrast in the effect of treatment at the end of an RCT, it is most unlikely that such a contrast will appear during follow up.

We have seen follow up studies reporting proportions of patients that "still meet" the ACR criteria for improvement over time (see for example, Scott *et al*[5]).The ACR improvement criteria have been developed to measure improvement, not to measure maintenance of low disease activity, which is an entirely different concept.[15] Repeatedly using ACR response proportions or change from baseline scores at consecutive times overemphasises baseline values, without taking into account actual disease activity.

Many authors analyse measures at different times as if these repeated measurements are independent of each other. Repeated measurements of any kind of variable, however, are characterised by a high intra-patient correlation; it makes the best predictor of the DAS28 at

t_x the DAS28 at t_{x-1}. Neglecting this intra-patient correlation may lead to overinterpretation of contrasts between treatment groups, because a statistically significant difference at t_{x-1} will have a higher probability of existing also at t_x, irrespective of treatment.

ARE FOLLOW UP STUDIES OF RANDOMISED CONTROLLED TRIALS VALUABLE?

Based on the considerations mentioned here, we doubt whether many of the follow up studies of RCTs now being published teach us more about a particular drug. It has been argued already that any contrast between two drugs is most likely to arise under the ideal experimental conditions of prognostic similarity— namely, the RCT. Probably, ideal experimental conditions are increasingly weakened towards the end of an RCT, so that an ideal RCT is of limited duration. In RA trials we need 6–12 months, because this is the time frame at least required to demonstrate differences in radiological progression in well designed, appropriately powered clinical trials.

> "Mortality, joint prostheses, withdrawal from the labour force, and radiological damage are real outcome measures in a follow up trial"

If in an RCT of appropriate duration no contrast in the efficacy of different treatments can be found, it is meaningless to follow up those patients under far from ideal experimental conditions for many years.

If there is a treatment contrast at the end of an RCT, it is of little value to follow up process variables as primary outcome measures; the results—either positive or negative—will not be interpretable because of too many interrelating factors. One should look for real outcome measures, such as mortality, joint prostheses, withdrawal from the labour force, and radiological damage. A radiological damage score is the sum of all the damage that has occurred over time, and correlates well, although incompletely, with disease activity and, in the long term, with function.[16 17] Radiological damage thus represents historic disease activity, and is not subject to fluctuations of process variables. But again, if there is no difference at the end of the trial, it is useless to follow up the patients in a further study.

Some people advocate functional ability as the best outcome measure for the long term. In theory, functional ability is

a better outcome than disease activity, but the instruments to measure function, such as the Health Assessment Questionnaire, are so greatly influenced by disease activity that the component of function that is not subject to fluctuation cannot easily be derived.[18]

Of course there are valid reasons for performing follow up studies of RCTs. The most important reason might be long term toxicity, especially if there are theoretical arguments suggesting that specific adverse events may occur some time after the exposition to the drug— for example, the fear of malignancies after tumour necrosis factor α blocking treatment, or after cyclosporin A.[19 20]

If long term studies are performed, it is wise to use appropriate statistical techniques. The recent development of longitudinal regression techniques, such as generalised estimating equations (GEE), has made it possible to study longitudinal relationships between variables.[21 22] GEE have other advantages in follow up studies: they adjust for unequal time intervals and missing values, two common occurrences in follow up studies. We have not often seen the application of GEE in rheumatology, probably because they are not available in common software packages, and their real merit should be established experimentally. In our experience, however, GEE are a powerful technique, particularly for analysing data with some longitudinal development, such as radiological damage.[12]

CONCLUSIONS

The number of published follow up studies of randomised clinical trials increases. These studies are often open label extensions of the formal trial, without a specified treatment protocol. Analyses often are repeats of the primary analysis of the RCT, whereas sophisticated longitudinal techniques would be more appropriate. Reports lack a predefined study question, and confounding by indication or biases introduced by some kind of skewed withdrawal of patients is neglected. All too easily, conclusions are overinterpretations of the findings.

We recommend that follow up studies should only assess data that can be reliably obtained and analysed, and that provide better insight into "real outcome". For example, a follow up study of an RCT with two DMARDs in patients with RA should include annual assessments of mortality, malignancy, and/or other comorbidity (for example, joint prostheses), radiological damage, and labour participation.

In all other cases, the rheumatological community might be served better with data that answer the question, "Which patients should be treated intensively

(expensively) and which not?", rather than "Is drug A still effective after four years?" These data can be obtained within the original RCT, and follow up is not necessary.

Ann Rheum Dis 2002;**61**:479–481

....

Authors' affiliations
R Landewé, D van der Heijde, University Hospital, Department of Internal Medicine/Rheumatology, Maastricht, The Netherlands

Correspondence to: Dr R B M Landewé, Department of Rheumatology, University Hospital, PO Box 5800, 6202 AZ Maastricht, The Netherlands; rlan@sint.azm.nl

REFERENCES

1 **Jessop JD**, O'Sullivan MM, Lewis PA, Williams LA, Camilleri JP, Plant MJ, et al. A long-term five-year randomized controlled trial of hydroxychloroquine, sodium aurothiomalate, auranofin and penicillamine in the treatment of patients with rheumatoid arthritis. Br J Rheumatol 1998;37:992–1002.
2 **Sander O**, Herborn G, Bock E, Rau R. Prospective six year follow up of patients withdrawn from a randomised study comparing parenteral gold salt and methotrexate. Ann Rheum Dis 1999;58:281–7.
3 **Weinblatt ME**, Kaplan H, Germain BF, Block S, Solomon SD, Merriman RC, et al. Methotrexate in rheumatoid arthritis. A five-year prospective multicenter study. Arthritis Rheum 1994;37:1492–8.
4 **Gonzalez-Gay MA**, Blanco R, Garcia-Porrua C, Ibanez D, Vazquez-Caruncho M, Sanchez-Burson J, et al. Long-term followup of patients receiving combined therapy with cyclosporine and methotrexate. Arthritis Rheum 1998;41:1703–4.
5 **Scott DL**, Smolen JS, Kalden JR, van de Putte LB, Larsen A, Kvien TK, et al. Treatment of active rheumatoid arthritis with leflunomide: two year follow up of a double blind, placebo controlled trial versus sulfasalazine. Ann Rheum Dis 2001;60:913–23.
6 **Stein CM**, Pincus T, Yocum D, Tugwell P, Wells G, Gluck O, et al. Combination treatment of severe rheumatoid arthritis with cyclosporine and methotrexate for forty-eight weeks: an open-label extension study. The Methotrexate-Cyclosporine Combination Study Group. Arthritis Rheum 1997;40:1843–51.
6a Kvien TK, Zeidler HK, Hannonen P, Wollheim FA, Førre Ø, Hafstrom I, et al. Long term efficacy and safety of cyclosporin versus parenteral gold in early rheumatoid arthritis: a three year study of radiographic progression, renal function, and arterial hypertension. Ann Rheum Dis 2002;61:xxx–xx.
7 **Pocock SJ**. *Clinical trials, a practical approach.* Chichester: Wiley, 1983.
8 **Sim J**, Wright C. *Research in health care: concepts, designs and methods.* Cheltenham: Stanley Thornes; 2000.
9 **Lee YJ**, Ellenberg JH, Hirtz DG, Nelson KB. Analysis of clinical trials by treatment actually received: is it really an option? Stat Med 1991;10:1595–605.
10 **Newell DJ**. Intention-to-treat analysis: implications for quantitative and qualitative research. Int J Epidemiol 1992;21:837–41.
11 **Heitjan DF**. Causal inference in a clinical trial: a comparative example. Control Clin Trials 1999;20:309–18.
12 **Landewe RB**, Boers M, Verhoeven AC, Westhovens R, van De Laar MA, Markusse HM, et al. COBRA combination therapy in patients with early rheumatoid arthritis: long-term structural benefits of a brief intervention. Arthritis Rheum 2002;46:347–56.
13 **Plant MJ**, Williams AL, O'Sullivan MM, Lewis PA, Coles EC, Jessop JD. Relationship between time-integrated C-reactive protein levels and radiologic progression in patients with rheumatoid arthritis. Arthritis Rheum 2000;43:1473–7.
14 **van Leeuwen MA**, van Rijswijk MH, Sluiter WJ, van Riel PL, Kuper IH, van de Putte LB, et al. Individual relationship between progression of radiological damage and the acute phase response in early rheumatoid arthritis. Towards development of a decision support system. J Rheumatol 1997;24:20–7.
15 **Felson DT**, Anderson JJ, Boers M, Bombardier C, Furst D, Goldsmith C, et al. American College of Rheumatology. Preliminary definition of improvement in rheumatoid arthritis. Arthritis Rheum 1995;38:727–35.
16 **Welsing PM**, van Gestel AM, Swinkels HL, Kiemeney LA, van Riel PL. The relationship between disease activity, joint destruction, and functional capacity over the course of rheumatoid arthritis. Arthritis Rheum 2001;44:2009–17.
17 **Drossaers-Bakker KW**, de Buck M, van Zeben D, Zwinderman AH, Breedveld FC, Hazes JM. Long-term course and outcome of functional capacity in rheumatoid arthritis: the effect of disease activity and radiologic damage over time. Arthritis Rheum 1999;42:1854–60.
18 **Wolfe F**. A reappraisal of HAQ disability in rheumatoid arthritis. Arthritis Rheum 2000;43:2751–61.
19 **van den Borne BE**, Landewe RB, Houkes I, Schild F, van der Heyden PC, Hazes JM, et al. No increased risk of malignancies and mortality in cyclosporin A- treated patients with rheumatoid arthritis. Arthritis Rheum 1998;41:1930–7.
20 **Landewe RB**, van den Borne BE, Breedveld FC, Dijkmans BA. Does cyclosporin A cause cancer? Nat Med 1999;5:714.
21 **Zeger SL**, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. Biometrics 1986;42:121–30.
22 **Twisk JW**. Different statistical models to analyze epidemiological observational longitudinal data: an example from the Amsterdam Growth and Health Study. Int J Sports Med 1997;18(suppl 3):S216–24.