Published in final edited form as:

Ann Rheum Dis. 2008 May ; 67(5): 580–583. doi:10.1136/ard.2007.079632.

Trials in rheumatoid arthritis: choosing the right outcome measure when minimal disease is achievable

David T Felson¹, Bin Zhang¹, and Jeffrey N Siegel²

¹ Clinical Epidemiology Unit, Boston University School of Medicine, Boston, Massachusetts, USA

² Food and Drug Administration, Silver Springs, Maryland, USA

Exciting advances in rheumatoid arthritis therapy have improved the prospects for patients to live for many years with minimal, if any, detectable disease activity. Recent trials have not only provided evidence on specific medications that might be used to improve outcome but also on therapeutic strategies that might optimise that outcome.

Obviously, the elimination of disease activity is the goal of treating patients with rheumatoid arthritis (RA), or any rheumatic disease. The goal of this editorial is not to question this ultimate aim of treatment or to discuss the design of new treatment strategies or the promise of specific treatments. We propose rather a discussion of how best to measure outcomes in RA trials to optimise our likelihood of achieving this aim of low disease activity at a time when low disease activity is an attainable goal for many patients.

Since they are clinically achievable goals, one reasonable approach to trial outcome measurement in RA might be to focus on low disease activity and, as much as possible, on remission or cure. As suggested by a recent editorial, “Successful therapeutic strategies of the next decade will be measured by the percentage of patients able to achieve remission or at least achieve a very low disease activity state and not by how many patients improved by a certain amount”.¹ We shall contend that RA trial outcomes should not focus on remission or even low disease activity, but rather on outcomes that optimise the detection of treatment efficacy. This depends not just on the rate of remission (or low disease activity) in the active treatment group, but on the difference in rates between the treatments being compared. If low disease activity is used as a primary endpoint in a trial in which low disease activity is not attained more often with a new than with a comparator agent, a trial of a demonstrably efficacious agent will fail to show efficacy.

Outcome Measurement in RA Trials: The TICORA Trial as an Outlier

One reason for the focus on remission or low disease activity in RA trials has been the success of the treatment regimen tested in the TICORA (for “tight control of rheumatoid arthritis”) trial. TICORA tested intensive vs conventional treatment for recent onset RA, and showed a dramatic difference between these groups.² Response rates in TICORA were extraordinary, perhaps the highest ever seen. For example, in RA trials published of anti-tumour necrosis factor α (TNF α) and other treatments published in 2006 (table 1), the proportion of patients achieving 70% improvement in symptoms as per American College of Rheumatology criteria (ACR70) was consistently 46% or less for patients on active treatment, whereas in TICORA, it was 71%. In the BeSt trial,³ published before our 2006 year of focus and designed similarly

Correspondence to: David T Felson, Room X200, Boston University School of Medicine, 650 Albany Street, Boston, MA 02118, USA; dfelson@bu.edu.

Competing interests: None declared.

to TICORA with aggressive combination therapy for early disease patients, the ACR70 rate was 40%.³ In BeSt, the European League Against Rheumatism (EULAR) remission rate was roughly 36% in the early combination treatment group (in BeSt, this rate was not significantly different across treatment groups). One reason for the high response rates may be that in TICORA, treatment naïve patients with RA received effective therapy for the first time. This is shown by the high rate of ACR70 response in the control group (table 1). Also, TICORA patients had early disease and response rates to all treatments are systemically higher in early than late RA.⁴

Perhaps the most striking finding in the TICORA results was the dramatic difference between the active treatment and control groups, a far higher difference than has been seen in other trials (table 1). For example, 65% of those in the intensive group went into remission vs 16% of those in the control group. In the BeSt Study, a study of similar design to TICORA, there were no significant differences in EULAR remission and ACR70 rates between treatment groups, and differences between treatment groups in these rates did not exceed 10%. If we classify TICORA as a placebo trial, the difference in ACR70 response rates ranged from 15–23% in 2006 trials whereas in TICORA, it was 53%. Similar responses can be seen for EULAR remission rates.

Those who study clinical trials¹⁴ have noted that initial trials testing novel therapies or therapeutic approaches have often reported impressive results that are difficult to replicate in subsequent trials. While a crystal ball is not available and while the TICORA intensive treatment group may represent an outstanding model for RA treatment, we suggest that a replication of similar impressive results as the TICORA trial will be difficult.

Small Effects of Treatment as the Rule

Current treatment regimens in RA have proven their success in randomised trials. While dramatic efficacy of a new compared to conventional treatment is a fervent desire, as can be seen for trials published in 2006 (table 1), small improvements in treatment conferred by new regimens or treatments compared with the existing standard are likely to be the rule. Note that many patients may achieve low disease activity or even remission, but a trial tests the difference in response between two treatment groups and unless this difference is substantial, the trial may fail.

Rheumatoid arthritis treatment success parallels the successful treatment of other chronic diseases where improved disease control and enhanced survival have been the products of incremental gains in therapy. For example, in coronary heart disease, a series of modest improvements in the efficacy of treatments have summed to produce a major drop in mortality.¹⁵ Peto¹⁶ has suggested that since new treatments are likely to provide only small incremental benefits over available treatments, very large trials will be needed in the future; only such “megatrials” will have enough statistical power to be likely to detect such small differences in efficacy. One example is aspirin therapy, which, for those who sustain a myocardial infarction, has lowered mortality from 13% to 10%, a small effect that has required trials of over 10 000 persons to detect. In the GISSI 1 trial, over 11 000 patients were randomised to prove the marginal efficacy (mortality difference of 10.7 vs 13%) of thrombolytics for prevention of mortality in MI.¹⁷

Rheumatoid arthritis obviously is not so common, and megatrials in RA would be prohibitively difficult to carry out. Nonetheless, the message of Peto and of the experience in cardiology and other chronic disease areas such as oncology, asthma, osteoporosis and diabetes that incremental overall improvements in treatment produce better noticeable overall outcomes is highly relevant to RA. And in RA trials, unlike those in cardiology, outcomes are not dichotomous (death/no death). Continuous outcomes like those available in RA can provide

more outcome information per subject than dichotomising a subject's outcome experience. Using continuously measured outcomes, definitive evidence on treatment efficacy can be secured with fewer patients in trials,^{18 19} and megatrials are unnecessary.

For most patients with RA, outcomes have undergone a secular improvement,²⁰ and patients being considered for trials start out with milder disease, in part, because they have already been treated more effectively than patients in an earlier era. Even if the primary outcome of all trials was remission, a dichotomous and uncommon outcome, this outcome would represent only an incremental improvement for some patients. Thus, like other chronic diseases, the overall effect of RA treatment is large, and many patients attain low disease activity or even remission. However, the marginal (or additional) effect of a new treatment over a traditional treatment is small or incremental. Incremental improvements in disease are likely to be the rule in the future.

One Way to Detect Small Effects: The ACR Hybrid

Previous work¹⁸ showed that dichotomous measures of response (improved or not improved or low disease activity vs not low disease activity) sacrificed statistical power compared to continuously defined outcomes and that plus the growing obsolescence of the ACR20 in the face of improving RA outcomes led the ACR to re-examine the ACR20. As part of an ACR funded effort to reevaluate the ACR20, the ACR Committee to Reevaluate Improvement Criteria tested the performance of 135 candidate measures of response or state in RA in 11 large recently completed RA multi-centre trials. The committee created and selected a new measure, the ACR Hybrid, a measure which combines a continuous scale of percentage improvement with the well known ACR20, 50, and 70 (see Appendix). The ACR Hybrid was developed with the assumption that outcome measures for future RA trials would need to detect small treatment effects. Among the most statistically powerful of all measures tested, the ACR Hybrid provides much more statistical power than the ACR20 and low disease activity measures. RA trials using it will require many fewer subjects to detect the same treatments as statistically significant better than placebo. In addition, the ACR Hybrid permits the detection of small treatment effects. (eg, when one treatment is only modestly more efficacious than its comparator). We should note that the ACR Hybrid represents only one of many possible examples of continuous outcome measures. While in recent trials, the Disease Activity Score (DAS) has been dichotomised to examine thresholds for remission and low disease, it could also be used as a continuous outcome measure and would detect smaller treatment differences.

Among the candidates evaluated by the ACR committee were two different definitions of low disease activity, which, among 135 measures tested, were among the five least sensitive to change in all of the trials. The low disease activity measures were so insensitive to change that they would have produced non-significant trial findings for treatments commonly accepted as efficacious for rheumatoid arthritis. These include methotrexate vs placebo, and cyclosporine vs placebo.

What are the consequences on trial results of using a dichotomous measure of low disease activity/response or of remission as the primary outcome? Also, what are the consequences when, like the TICORA trial, the achievement of “good” response according to EULAR criteria, is used as the primary outcome measure? Using data from nine large multi-centre randomised RA trials, we tested “good response”, a dichotomous definition of improvement, for its sensitivity to change vs the ACR Hybrid and the ACR20 (table 2). We also tested EULAR remission, which was reported in TICORA and other recent trials. The “good” response and EULAR remission scored as less sensitive to change than the ACR20 and much less sensitive to change than the ACR Hybrid. In fact, for trials of known efficacious disease-modifying antirheumatic drugs (DMARDs; ie, methotrexate vs placebo; leflunomide vs placebo), neither the “good” EULAR response nor the EULAR remission outcome would have shown efficacy

vs placebo ($p>0.05$). Both these drugs easily demonstrated significant efficacy vs placebo using either ACR20 or ACR Hybrid.

Thus, if low disease activity or remission is used as the primary outcome measure in future RA trials, we will fail to detect the efficacy of treatments that work. Failure to detect efficacy would deprive our patients of effective treatments.

Achieving Low Disease Activity without Using it as the Primary Outcome of Trials

In addition to its poor sensitivity to change, there is one other potential drawback to using either low disease activity or remission as a primary outcome in RA trials. If remission (or low disease activity) is the primary outcome of a trial, persons close to that state may be enrolled because they are more likely to attain the outcome and show success of the treatment. This may exclude both persons already below the definition of low disease activity used as the primary outcome and persons with highly active disease who are unlikely to achieve either remission or low disease activity. Treatments might be tested on a select subset of patients. Given recruitment challenges already being faced, a study design decision with adverse effects on likely recruitment may be inadvisable.

At any rate, the goal of treatment remains to cure patients or at least effectively to suppress most disease activity. How can we accommodate these goals in rheumatoid arthritis yet not use these measures as our primary outcomes in RA trials?

First, we should design trials so that new treatments have the best opportunity to demonstrate efficacy by employing an outcome measure likely to detect efficacy such as the ACR Hybrid. Low disease activity or remission can be used as a secondary outcome in many trials.

Second, to enhance the sensitivity to change of low disease activity, we can measure the proportion of visits during the trial in which the patient meets the criteria for low disease activity, so that the outcome becomes more of a continuous measure (eg, 10 of 12 visits vs 6 of 12 visits) than a dichotomous one (eg, patient met criteria for low diseases activity at study's end only). Because its sensitivity to change is poor, low disease activity is unlikely ever to achieve the same sensitivity to change as more continuously measured outcomes such as the ACR Hybrid or even the DAS assessed continuously, especially if the continuous measures are also tested as an average over time.

Ultimately, the goal of RA treatment is to achieve remission or extreme low disease activity. Ironically, expecting much higher rates of remission or even low disease activity for each new treatment compared to traditional treatments will not produce major gains in treatment for our patients.

Acknowledgments

Funding: The views presented in this article do not necessarily reflect those of the FDA. Supported by National Institutes of Health grant NIH AR47785.

APPENDIX: How to score the ACR Hybrid Response Measure

- Calculate the average percentage change in core set measures: for each core set measure, subtract score after treatment from baseline score and determine percentage improvement in each measure. Next, if a core set measure worsened by $>100\%$, limit that percentage changes to 100% (a -100% bound). Then, average percent changes for all core set measures.

- Determine whether the patient has achieved ARC20, ACR50 or ACR70.
- Using the following table (table A1), obtain the ACR Hybrid response measure. To use the table, take the ACR20, 50 or 70 status of the patient (left column) and the mean percentage improvement in core set items (other columns). Where these intersect in the table is the ACR Hybrid score for the patient.

Table A1
Hybrid scoring table

	Mean % change in core set measures			
	<20	≥20 but <50	≥50 but <70	≥70
Not ACR20	–	19.99	19.99	19.99
ACR20 but not ACR50	20	–	49.99	49.99
ACR50 but not ACR70	50	50	–	69.99
ACR70	70	70	70	–

References

1. O'Dell JR. The best way to treat early rheumatoid arthritis. *Ann Intern Med* 2007;146:459–60. [PubMed: 17371891]
2. Grigor C, Capell H, Stirling A, McMahon AD, Lock P, Vallance R, et al. Effect of a treatment strategy of tight control for rheumatic arthritis (the TICORA study): a single-blind randomised controlled trial. *Lancet* 2004;364:263–9. [PubMed: 15262104]
3. Goekoop-Ruiterman Y, de Vries-Bouwstra J, Allaart CF, Van Zeben D, Kerstens P, Hazes MW, et al. Clinical and radiographic outcomes of four different treatment strategies in patients with early rheumatoid arthritis (the BeSt Study). A randomized, controlled trial. *Arthritis Rheum* 2005;52:3381–90. [PubMed: 16258899]
4. Anderson JJ, Bolognese JA, Felson DT. Effects of disease duration and outcome definition on the power of rheumatoid arthritis clinical trials. *Arthritis Rheum* 2002;46:S520.
5. Vanags D, Williams B, Johnson B, Hall S, Nash P, Taylor A, et al. Therapeutic efficacy and safety of chaperonin 10 in patients with rheumatoid arthritis: a double-blind randomised trial. *Lancet* 2006;368:855–63. [PubMed: 16950363]
6. Maini RN, Taylor PC, Szechinski J, Pavelka K, Broll J, Balint G, Emery P, Raemen F, Petersen J, Smolen J, Thomson D, Kishimoto T. Double blind randomized controlled clinical trial of the interleukin-6 receptor, antagonist, tocilizumab, in European patients with rheumatoid arthritis who had an incomplete response to methotrexate. *Arthritis Rheum* 2006;54:2817–29. [PubMed: 16947782]
7. Kremer JM, Genant HK, Moreland LW, Russell AS, Emery P, Abud-Mendoza C, et al. Effects of abatacept in patients with methotrexate resistant active rheumatoid arthritis. *Ann Int Med* 2006;144:865–76. [PubMed: 16785475]
8. Van der Heijde D, Klareskog L, Rodriguez-Valverde V, Codreanu C, Bolosiu H, Melo-Gomes J, et al. Comparison of etanercept and methotrexate, alone and combined, in the treatment of rheumatoid arthritis. *Arthritis Rheum* 2006;54:1063–74. [PubMed: 16572441]
9. Emery P, Fleischman R, Filipowicz-Sosnowska A, Schechtman J, Szczepanski L, Kavanaugh A, et al. The efficacy and safety of rituximab in patients with active rheumatoid arthritis despite methotrexate treatment. *Arthritis Rheum* 2006;54:1390–400. [PubMed: 16649186]
10. Hetland ML, Stengaard-Pedersen K, Junker P, Lottenburger T, Ellingsen T, Andersen LS, et al. Combination treatment with methotrexate, cyclosporine and intraarticular betamethasone compared with methotrexate and intraarticular betamethasone in early active rheumatoid arthritis. *Arthritis Rheum* 2006;54:1401–9. [PubMed: 16645967]

11. Breedveld FC, Weisman MH, Kavanaugh AF, Cohen SB, Pevelka K, Van Vollenhoven R, et al. A multicenter, randomized, double-blinded clinical trial of combination therapy with adalimumab plus methotrexate versus methotrexate alone or adalimumab alone in patients with early, aggressive rheumatoid arthritis who had not had previous methotrexate treatment. *Arthritis Rheum* 2006;54:26–37. [PubMed: 16385520]
12. Van Riel PL, Taggart AJ, Sany J, Gaubitz M, Nab HW, Pedersen R, et al. Efficacy and safety of combination etanercept and methotrexate versus etanercept alone in patients with rheumatoid arthritis with an inadequate response to methotrexate: the ADORE study. *Ann Rheum Dis* 2006;65:1478–83. [PubMed: 16464988]
13. Combe B, Codreanu C, Fiocco U, Gaubitz M, Geusens PP, Kvien TK, et al. Etanercept and sulfasalazine, alone and combined, in patients with active rheumatoid arthritis despite receiving sulfasalazine: a double-blind comparison. *Ann Rheum Dis* 2006;65:1357–62. [PubMed: 16606651]
14. Lau J, Ioannidis JPA, Schmid C. Summing up evidence: one answer is not always enough. *Lancet* 1998;351:123–7. [PubMed: 9439507]
15. Cardiovascular Disease Statistics. [28 December 2007].
<http://www.americanheart.org/presenter.jhtml?identifier=4478>
16. Peto R, Baigent C. Trials: the next 50 years. Large scale randomised evidence of moderate benefits. *BMJ* 1988;317:1170–1. [PubMed: 9794846]
17. Gruppo Italiano per lo Studio della Streptochinasi nell'Infarto Miocardico (GISSI). Effectiveness of intravenous thrombolytic treatment in acute myocardial infarction. *Lancet* 1986;1:397–402. [PubMed: 2868337]
18. Anderson JJ, Bolognese JA, Felson DT. Comparison of rheumatoid arthritis clinical trial outcome measures: a simulation study. *Arthritis Rheum* 2003;48:3031–8. [PubMed: 14613263]
19. Pincus T, Amara I, Koch GG. Continuous indices of core data set measures in rheumatoid arthritis clinical trials. Lower responses to placebo than seen with categorical responses with the American College of Rheumatology 20% criteria. *Arthritis Rheum* 2005;52:1031–6. [PubMed: 15818698]
20. Pincus T, Sokka T, Kautiainen H. Patients seen for rheumatoid arthritis care have significantly better articular, radiographic, laboratory, and functional status in 2000 than in 1985. *Arthritis Rheum* 2005;54:1009–19. [PubMed: 15818706]

Table 1
Rates of ACR70 and EULAR remission among subjects in RA trials published in 2006

Trial by first author (name of trial or primary treatment)	Percentage attaining ACR70		Percentage attaining EULAR remission		Difference between active and control group (%)
	Active treatment (%)	Control (placebo, active, or routine care) (%)	Active treatment (%)	Control (%)	
Vanags <i>et al</i> (Chaperonin) ⁵	29	NR [*] (active)	NR [*]	NR [*]	
Maini <i>et al</i> (CHARISMA) ⁶	37	16 (placebo)	34	8	26
Kremer <i>et al</i> (Abatacept) ⁷	29	6 (placebo)	15	3	12
Van der Heijde <i>et al</i> (TEMPO) ⁸	49	21 (active)	30	14	16
Emery <i>et al</i> (DANCER) ⁹	20	5 (placebo)	NR [*]	NR [*]	
Hetland <i>et al</i> (CIMESTRA) ¹⁰	40	36 (active)	43	34	9
Breedveld <i>et al</i> (PREMIER) ¹¹	46	28 (active)	43	21	22
Van Riel <i>et al</i> (ADORE) ¹²	18	17 (probably active) [†]	17	15	2
Combe <i>et al</i> (SSZ Etanercept) ¹³	25	2 (placebo)	NR [*]	NR [*]	
Grigor <i>et al</i> (TICORA) ²	71	18 (routine care)	65	16	49

* Not reported in trial publication.

[†] Comparison group discontinued MTX to which they had not responded adequately and were started on etanercept).

ACR70, 70% improvement in symptoms as per American College of Rheumatology criteria; EULAR, European League Against Rheumatism; MTX, methotrexate.

Sensitivity to change of different outcomes measures tested in 9 recent large scale RA trials presenting 11 active-control comparisons

Table 2

Definition of response	TNF inhibitor			DMARD			Combination TNF/ DMARD vs DMARD alone
ACR Hybrid	<0.001	<0.001	<0.001	<0.001	0.0016	<0.001	<0.001
ACR20	<0.001	<0.001	0.0045	<0.001	0.0018	0.0017	0.0021
EULAR good response	0.0011	<0.001	0.0186	<0.001	0.8297	0.0581	<0.001
EULAR remission	0.2377	0.0418	0.1412	0.0259	0.4768	0.5083	0.1393
					0.5537		0.2363

p Values are shown testing efficacy of treatment vs placebo. Numbers shown in table are p values comparing active treatment vs control (in most trials, control is placebo).

*Combination: these are trials testing a combination of second-line drugs vs one second-line drug.

ACR, American College of Rheumatology; DMARD, disease-modifying antirheumatic drug; EULAR, European League Against Rheumatism; TNF, tumour necrosis factor.