CLINICAL DATA GAP BETWEEN PHASE III CLINICAL TRIALS (PRE-MARKETING) AND PHASE IV (POST-MARKETING) STUDIES: EVALUATION OF ETANERCEPT IN RHEUMATOID ARTHRITIS

Pendar Farahani[1,3], Mitchell Levine[1,2], Kathryn Gaebel[1], Lehana Thabane[1,2]

[1]Centre for Evaluation of Medicines, St. Joseph's Healthcare, Hamilton, Ontario, [2]Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, [3]Health Research Methodology (HRM) Graduate Programme, Faculty of Health Sciences, McMaster University

Corresponding author: LEVINEM@mcmaster.ca

ABSTRACT

Background
There are fundamental differences in design between phase III clinical trials and phase IV post-marketing studies that involve patient characteristics, the clinical setting (environment) and the manner of drug use. As well, many phase IV studies are extensions of randomized clinical trials (RCTs) and suffer from selection bias.

Objective
To determine if the data obtained from RCTs of etanercept (Enbrel®) in the treatment of rheumatoid arthritis would be representative of the effects attainable in community practice.

Method
An analysis was conducted comparing data from published RCTs of etanercept use in rheumatoid arthritis patients with data collected in a community based cohort study that was not an extension of an RCT.

Results
Baseline clinical data, such as tender or painful joint count, patient's global assessment, the heath assessment questionnaire, physical and mental component summary of the SF-36, and rheumatoid arthritis drug profile were significantly different between the patients receiving etanercept in the phase IV community cohort study and the patients enrolled in the RCTs. Differences in the baseline data for the control patients were also noted amongst the RCT studies. The *treatment outcome*, American College of Rheumatology (ACR) response rate of 20%, 50% and 70% at 6 month, was the same between the cohort study and the RCTs, but at 12 months the clinical response was less for the community based patients than for the RCT patients. At 6 months there were fewer withdrawals involving community-based patients than RCT patients due to less frequent withdrawals associated with lack of efficacy. At 12 months the withdrawal rate due to either a lack of efficacy or from adverse events was similar between data sets.

Conclusion
The data from the etanercept phase III RCTs may not reflect the characteristics of patients using etanercept in community practice, nor the clinical outcomes observed by RA patients at 12 months. These discrepancies may be derived from methodological differences in study design and patient selection. On the other hand, outcomes such as withdrawal rates at 12 months appear comparable between the two types of populations.

Key Words: *Rheumatoid Arthritis, etanercept, post-marketing study, comparative study*

In the development of a new drug several established assessment steps must be completed prior to approval by the regulatory authorities for use in clinical practice. In phase III of this process, it is common to conduct double blind, randomized clinical trials designed to evaluate the benefits and harms of the drug. After this pivotal stage the drug may be approved for marketing. In the phase IV or post-marketing studies that follow, further evaluation of the drug is conducted in circumstances of routine use.[1-3] It is also evident that data gathered during phase III might not be sufficient to determine the value of a drug in clinical practice. The assessment of a drug during routine use can uncover evidence of additional side effects or reduced efficacy, and has contributed to the withdrawal of some drugs from the market or resulted in restrictions for their use[4].

From a practical sense there are some fundamental differences between the design and conduct of phase III clinical trials and phase IV post-marketing studies. During phase III studies the efficacy of a drug is measured in an idealized and somewhat artificial clinical environment. There are narrow inclusion criteria and patients with co-morbidities (multi-morbidity), children, elderly patients and pregnant women are often excluded. The treatment strategies are fixed by design, drug doses are fixed and drug combinations are protocol defined. In contrast, phase IV studies can be conducted in a "real life" situation that involves a wide spectrum of patients with broad inclusion criteria and few if any exclusion criteria. Because of the "real life" setting the combinations of drugs are typical, the drug doses are fluctuant and the treatment strategies are flexible and dependent upon the course of the illness.[5-11]

The objective of this study was to determine if the data obtained through randomized clinical trials of etanercept (Enbrel®) use in the treatment of rheumatoid arthritis would be representative of the effects anticipated in real use. For this purpose, data from a clinical practice cohort study (phase IV) were compared with the published data from randomized clinical trials.

METHODS

Data were obtained from the self-reported community-based cohort study conducted at the Centre for Evaluation of Medicines (CEM), St. Joseph's Healthcare-McMaster University, Hamilton.[12]

In this cohort study, patients requesting etanercept therapy were stratified into treatment and control arms based upon their individual accessibility to obtain the drug. Because the drug was in short supply patient allocation to the treatment or the control group was based upon the temporal availability of the drug from the manufacturer, and therefore quasi-randomization could be assumed. Patients were interviewed serially during a 12-month period of monitoring. The exclusion or inclusion of data from patients who withdrew from the treatment group is reflected in the on-treatment analysis and intention-to-treat analysis respectively.

An extensive systematic literature search, through Medline, EMBASE and Cochrane Library for the period of January 1, 1990 to January 1, 2004, was conducted to identify phase III randomized clinical trials (RCTs) of etanercept in the treatment of rheumatoid arthritis (RA) patients, meeting the following selection criteria:

1. The duration of the clinical study was at least 6 months,
2. There was at least one treatment arm with etanercept 25 mg twice a week in the clinical study, and,
3. The clinical study included some of the variables and outcomes that had been measured in the cohort study.

The text articles were identified using several combinations of the MESH terms including 'rheumatoid arthritis', 'Drug Therapy', 'etanercept', and 'tumor necrosis factor inhibitor' limiting results to clinical trials or randomized controlled trials and English-language publications on human adults. During initial search 11 text articles were found. Five of the articles did not measure the same variables that were measured in the cohort study and three articles were replicate of other original reports. `Finally, three RCTs (referred to as A, B, C) met these criteria and were compared with the data from the cohort study:

A) Moreland et al. evaluated etanercept therapy in rheumatoid arthritis (RA) patients for a duration of 6 months. This randomized study included three groups: placebo, etanercept 10 mg twice a week, or etanercept 25 mg twice a week.[13]

B) Weinblatt et al. randomly allocated RA patients who had been receiving methotrexate (MTX) into two groups (etanercept 25 mg twice a week plus MTX compared to placebo plus MTX) and followed them for 6 month.[14]

C) Bathon et al. chose recent onset RA patients and randomized them into three arms: MTX, etanercept 10 mg twice a week, or etanercept 25 mg twice a week. Patients were followed for 12 months.[15]

Statistical analysis

Baseline and outcome data from the cohort study were compared with the RCT data using chi-squared test of association for binary variables and student t-test for continuous variables. Because of multiple comparisons, alpha = 0.01 was considered the threshold for statistical significance. Data in the tables are presented as mean (SD = Standard Deviation) for continuous variables and number (percentage) for binary or categorical variables.

The measured outcomes in these studies include painful or tender joint count, swollen joint count, early morning stiffness (minutes), pain severity (range 0 to 10, 0 being least pain), overall well being or patient's global assessment (range 0 to 10, 10 being best overall well being), quality of life measured by SF-36 (range 0 to 100, 100 being the best)[16], functional (disability) assessment measured by Health Assessment Questionnaire-HAQ (range 0 to 3, 3 being the best)[17] and the American College of Rheumatology response rate[18,19] (i.e. ACR20, ACR50 and ACR70). For example, the ACR20 means a 20% improvement in the tender and swollen joint count, as well as a 20% improvement in 3 of the following 5 parameters: patient's global assessment, physician's global assessment, patient's assessment of pain, degree of disability, and level of acute-phase reactant. (Appendix 1)

RESULTS

Baseline demographic and clinical data for *treatment* groups (Table 1) demonstrated significant differences between the cohort study and the RCTs. Swollen Joint Count (SJC) and Tender Joint Count (TJC), Patient's Global Assessment, Physical Component Summary of the SF-36 (PCS) and Mental Component Summary of the SF-36 (MCS) were significantly different between cohort study and at least two of the three RCTs.

The proportion of patients with concomitant drug use at baseline for the cohort study and RCTs were comparable with respect to NSAIDs, but not for DMARDs and Corticosteroids. Similar results were observed for *control* group data at *baseline* (Table 2).

Outcomes

Response rates using modified American College of Rheumatology criteria for 20%, 50% and 70% were calculated for the cohort study (Appendix 1) using on-treatment analysis and intention-to-treat analysis. A comparison between the cohort study data using the on-treatment analysis method (Table 3) and the RCTs data did not demonstrate any significant differences at 6 months, but there were significant dissimilarities for ACR 20, 50 and 70 at 12 months. Using the intention-to-treat analysis (Table 4) ACR outcomes for the cohort study data were significantly lower values than the RCTs data for ACR 20 at 6 months, and at 12 months ACR 20, 50 and 70 outcomes were significantly lower in the cohort study than the RCTs. The only consistent difference in the results obtained between the on-treatment analysis and the intention-to-treat analysis, in the comparison between cohort and RCTs data, was the ACR 20 data at 6 months.

Withdrawals (Table 5) were significantly diverse between the cohort study and the RCTs at 6 months, with one of the RCTs primarily having an increase in withdrawal due to a lack of efficacy. At 12 months the withdrawal data were comparable between cohort study and RCT C.

TABLE 1 Demographic and clinical data at Baseline- Treatment Groups

	Cohort	RCT A	p-value	RCT B	p-value	RCT C	p-value
Sample Size	223	78	---	59	---	207	---
Age (years)	53.5 (22.5)	53	NS	48	NS	51 (13)	NS
Women	74%	74%	NS	90%	<0.01	74%	NS
Swollen Joint Count	17.7 (15)	25	<0.001	20	NS	24 (11.9)	<0.001
Tender Joint Count	23.6 (15)	33	<0.001	28	NS	31 (15.8)	<0.001
Early Morning Stiffness (minutes)	102 (65)	300	<0.001	90	NS	---	---
Patient Global Assess (0-10, 10 is best)	4.9 (2)	7	<0.001	6	<0.001	---	---
Pain Severity (score 0-10, 10 is best)	6.25 (2.25)	6.7	NS	5	<0.001	---	---
HAQ (score 0-3, 0 is best)	1.7 (0.7)	1.63 (0.53)	NS	1.5	NS	---	---
PCS (SF-36) (score 0-100, 100 is best)	25.5 (9)	66	<0.001	---	---	28 (7)	0.001
MCS (SF-36) (score 0-100, 100 is best)	50.5 (12)	42	<0.001	---	---	46 (10)	<0.001
Duration of Disease in Year	12.5 (9.2)	11	NS	13	NS	1 (0.9)	<0.001
DMARDs	86%	87%	NS	100%	<0.001	23%	<0.001
Corticosteroids	47%	81%	<0.001	53%	NS	39%	NS
NSAIDs	77%	84%	NS	75%	NS	86%	NS

Cohort: self reported cohort study in a community based setting, duration 12 months
RCT A: comparing etanercept with placebo, duration 6 months
RCT B: comparing etanercept + methotrexate with placebo + methotrexate, duration 6 months
RCT C: comparing etanercept with methotrexate in early RA patients, duration 12 months
---: No data were available through literatures
Data are either presented as Mean (SD); or Mean, where the SD was not available through literatures
Abbreviations: Physical Summary Score (PSC) and Mental Summary Score (MSC) of the health-related quality of life questionnaire SF-36, Health Assessment Questionnaire (HAQ), Disease-Modifying Anti-rheumatic Drugs (DMARDs), Nonsteroidal Anti-inflammatory Drugs (NSAIDs)

TABLE 2 Demographic and clinical data at baseline- Control Groups

	Cohort	RCT A	p-value	RCT B	p-value	RCT C	p-value
Sample Size	208	80	---	30	---	217	---
Age (years)	52.7 (11.3)	51	NS	53	NS	49 (13)	0.001
Women	68%	76%	NS	73%	NS	75%	NS
Swollen Joint Count	19.5 (15)	25	<0.01	17	NS	24 (11.9)	<0.001
Tender Joint Count	26 (15)	35	<0.001	28	NS	30 (16.1)	<0.01
Early Morning Stiffness (minutes)	110 (64)	288	<0.001	120	NS	---	---
Patient Global Assess (0-10, 10 is best)	5.1 (2)	6.9	<0.001	6	NS	---	---
Pain Severity (score 0-10, 10 is best)	6.25 (2.25)	6.5	NS	5.6	NS	---	---
HAQ (score 0-3, 0 is best)	1.8 (0.7)	1.7	NS	1.5	NS	---	---
PCS (SF-36) (score 0-100, 100 is best)	26.5 (9)	69	<0.001	---	---	---	---
MCS (SF-36) (score 0-100, 100 is best)	48 (12)	42	<0.001	---	---	---	---
Duration of Disease (years)	12.3 (9.7)	12	NS	13	NS	1 (0.9)	<0.001
DMARDs	92%	90%	NS	100%	<0.01	24%	<0.001
Corticosteroids	49%	58%	NS	70%	<0.01	41%	NS
NSAIDs	73%	67%	NS	80%	NS	80%	NS

Cohort: self reported cohort study in a community based setting, duration 12 months
RCT A: comparing etanercept with placebo, duration 6 months
RCT B: comparing etanercept + methotrexate with placebo + methotrexate, duration 6 months
RCT C: comparing etanercept with methotrexate in early RA patients, duration 12 months
---: No data were available through literatures
Data are either presented as Mean (SD); or Mean, where the SD was not available through literatures
Abbreviations: Physical Summary Score (PSC) and Mental Summary Score (MSC) of the health-related quality of life questionnaire SF-36, Health Assessment Questionnaire (HAQ), Disease-Modifying Anti-rheumatic Drugs (DMARDs), Nonsteroidal Anti-inflammatory Drugs (NSAIDs)

TABLE 3 Outcome (response rate)*****

At 6 months	Cohort	RCT A	p-value	RCT B	p-value	RCT C	p-value
ACR20	56%*	59%	**NS**	71%	NS	65%	**NS**
ACR50	33%*	40%	**NS**	39%	NS	40%	**NS**
ACR70	16%*	15%	**NS**	15%	NS	21%	**NS**
At 12 months							
ACR20	50%*					72%	**<0.001**
ACR50	28%*					49%	**<0.001**
ACR70	11%*					25%	**0.001**

Cohort: self reported cohort study in a community based setting, duration 12 months
RCT A: comparing etanercept with placebo, duration 6 months
RCT B: comparing etanercept + methotrexate with placebo + methotrexate, duration 6 months
RCT C: comparing etanercept with methotrexate in early RA patients, duration 12 months
* On treatment analysis

TABLE 4 Outcome (response rate)*

At 6 months	**Cohort**	**RCT A**	**p-value**	**RCT B**	**p-value**	**RCT C**	**p-value**
ACR20	41%*	59%	**<0.01**	71%	**<0.001**	65%	**<0.001**
ACR50	24%*	40%	**NS**	39%	**NS**	40%	**0.001**
ACR70	12%*	15%	**NS**	15%	**NS**	21%	**NS**
At 12 months							
ACR20	30%*					72%	**<0.001**
ACR50	17%*					49%	**<0.001**
ACR70	7%*					25%	**<0.001**

Cohort: self reported cohort study in a community based setting, duration 12 months
RCT A: comparing etanercept with placebo, duration 6 months
RCT B: comparing etanercept + methotrexate with placebo + methotrexate, duration 6 months
RCT C: comparing etanercept with methotrexate in early RA patients, duration 12 months
* Intention to treat analysis

TABLE 5 Outcome (withdrawal)

	Cohort 6 months	RCT A	p-value	RCT B	p-value	Cohort 12 months	RCT C	p-value
Withdrawal (drop-out)	12%	24%	**NS**	3%	**0.01**	19%	15%	**NS**
Withdrawal (adverse events)	8%	3%	**NS**	3%	**NS**	9%	5%	**NS**
Withdrawal (lack of efficacy)	3%	15%	**<0.01**	0%	**NS**	6%	5%	**NS**
Withdrawal (other)	0.5%	6%	**0.01**	0%	**NS**	2%	5%	**NS**
Withdrawal (Cost)	0.5%	0%	**NS**	0%	**NS**	2%	0%	**NS**
Number of patients	223	78	**---**	89	**---**	223	207	**---**

Cohort: self reported cohort study in a community based setting, duration 12 months
RCT A: comparing etanercept with placebo, duration 6 months
RCT B: comparing etanercept + methotrexate with placebo + methotrexate, duration 6 months
RCT C: comparing etanercept with methotrexate in early RA patients, duration 12 months

DISCUSSION

The results of this analysis have demonstrated some differences in the baseline clinical and demographic data between phase III RCTs and a phase IV cohort study. As well, there were differences in data amongst the three individual RCTs (A, B and C), and this finding applies to both the treatment and the control arms in the RCT studies. Important patient characteristics differed with respect to disease activity, duration and concomitant drug use at baseline. Some of these differences were by design (i.e., RCT B patients were required to be using methotrexate prior to entering the study, RCT C only included patients with new onset RA, etc). Not all baseline differences that were statistically significant were necessarily clinically meaningful. The baseline differences between the cohort study and RCTs regarding the MCS and PCS of the SF-36 were less than 5, and differences of this small magnitude are considered of little clinical significance. Baseline clinical data (joint count, pain severity, and patient's global assessment) demonstrated that the patients who participated in RCTs had more active RA disease compared to the patients in the cohort study. Although the joint count in the cohort study is self-reported, there is evidence that self-reported joint counts are a reliable and responsive measure that agrees highly with the observer-assessed joint count and is significantly associated to the health-related quality of life of patients with RA.[20]

The fact that the baseline data from the RCTs differed from the community based cohort study suggests that data from RCTs may not thoroughly reflect patients' characteristics in the real world. In addition, with considerable observed variability amongst the RCTs, extrapolations of phase III results to clinical practice would be quite dependent upon which specific phase III RCT the data were derived from. This heterogeneity, which even transcends the distinction between phase III and phase IV and occurs amongst phase III trials, has implications for the selectivity of trial data used in generating practice guidelines and other materials used to guide or influence prescribing in community-based practices.

Thus this study reinforces the need for community-based (effectiveness) clinical trials to

include patients from a real world experience. For this particular clinical population (RA patients using etanercept) the RCT that would most reflect patient use of etanercept in community-based clinical practice is the study where patients were concomitantly using a DMARD such as methotrexate (i.e., the study that evaluates the adjunctive therapeutic effect of etanercept, rather than etanercept as a replacement for DMARD therapy).

The ACR response rates for the cohort study were calculated using both an on-treatment and an intention-to-treat analysis. Because the RCTs reported the ACR results with an intention-to-treat analysis, the intention-to-treat ACR response rates for the cohort study should be used in the comparison. In the cohort study using the intention to treat analysis is likely to produce an under estimation of ACR because all dropouts would be assumed to be non-responders. To assess if this occurred, the ACR results using on-treatment analysis were also compared with the RCT data. Since the only consistent difference in the results between the two types of analyses was the ACR 20 at 6 months, the intention-to-treat analysis appears to be acceptable.

The ACR response rates at 6 months are comparable between the RCTs and the community based cohort study, but the data at 12 months demonstrate discrepancies between the two types of studies, with the cohort results showing less benefit than what was observed in the RCT. Patients' selection criteria may explain the discrepancy in ACR results between the studies at 12 months. In RCT C (the only RCT with 12 months data) the patients had early RA whereas the patients in the cohort study clearly did not have early RA (mean duration of disease of 1 year versus 12.5 years respectively). Years of RA can lead to joint destruction that may result in a blunted response to the therapy due to the irreversible progression.

A potential limitation of this inter-study comparison of efficacy is the difference in methods used to calculate the ACR. The ACR for the cohort study was calculated using a modified method that might have provided a conservative ACR estimate (Appendix 1) but when we adjusted the parameters in a sensitivity analysis to provide a more lenient threshold for the ACR outcome the results did not change in any of the comparisons with the RCT data. In addition, if the modified ACR outcomes were a low estimate of the traditional ACR we would have expected to observe a difference between the cohort and RCT data at both 6 and 12 months, but our results only showed a consistent difference between cohort and RCT data at 12 months. Nevertheless, extensive evidence has demonstrated that measures on a patient self-report method of physical function (i.e. HAQ), pain, and global status, are as informative as joint counts, radiographic scores, laboratory tests, or any measure by a health professional to document status, estimate prognosis, and monitor responses to therapies.[21-23]

The withdrawals at 12 months for the different studies (RCT C vs. the cohort study) were similar but the withdrawal rates at 6 months were different amongst the three RCTs, and between the RCTs and the cohort study. The withdrawal difference between RCT A and the cohort study was mostly the consequence of a difference in efficacy failure. This latter comparison emphasizes the difference between the adjunct (additive) effects of etanercept as opposed to its use as an alternative mono-therapy. In addition, withdrawal because of the cost of the etanercept was an issue in the cohort study but would not have been a factor in the treatment-sponsored RCTs, although this difference did not achieve statistical significance in our analysis.

CONCLUSION

There is considerable variability amongst RCT studies, and the data derived from some RCTs may not reflect the patients' characteristics in the real world. These discrepancies may be derived from methodological differences in study design and patient selection. This study emphasizes the necessity for community-based clinical trials to select the patients from a real world perspective in order to evaluate the effectiveness (as opposed to efficacy) of new drugs and to determine the long-term safety of drugs in a practical setting.

Acknowledgement

The project was supported by a grant from Wyeth-Ayerst Canada Inc.

Appendix 1

Calculation of modified ACR for self-reported cohort study

Due to the limitations in the clinical data that could be collected from the patients by telephone (e.g. patient assessed tender or painful joint count, patient's global assessment, pain severity and HAQ were measured in this study) a modified ACR was constructed. However, there is evidence that the self-reported joint counts is a reliable and responsive measure that agrees highly with the observer-assessed joint count and is significantly associated to the health-related quality of life of patients with RA.[20] For estimation of swollen joint count, the measured tender joint count was used. In the literature from previous studies on etanercept the tender (T) joint count is always higher than swollen (S) joint count and S/T is equal to 0.75 [S = 0.75 T] for baseline data and 0.83 [S = 0.83 T] after treatment. In addition, the impact of the unmeasured variables (doctor's global assessment and CRP or ESR) on the ACR proportions was extrapolated from the data. Nevertheless, extensive evidence has demonstrated that measures on a patient self-report method of physical function (i.e. HAQ), pain, and global status, are as informative as joint counts, radiographic scores, laboratory tests, or any measure by a health professional to document status, estimate prognosis, and monitor responses to therapies.[21-23] Patients who improved (20%, 50%, 70%) in tender or swollen joint count and had the requisite amount of impact in at least two of the measured variables (pain severity, HAQ or patient's global assessment) were considered ACR responders. The probability of improvement in at least one of the unmeasured variables (doctor's global assessment and CRP / ESR) would be 75%. Patients with sufficient improvement in only one of the three measured variables were all considered non responders even though there would be a crude (or independent) probability of 25% that they could have had both of the unmeasured variable sufficiently improved, and would have been defined as ACR responders. Using these assumptions, the modified ACR would be a reasonable approximation for the traditional ACR.

REFERENCES

1. Kuhlmann J. Drug research: from the idea to the product. Int J Clin Pharmacol Ther. 997;35:541-52.
2. Peck CC, Barr WH, Benet LZ, Collins J, Desjardins RE, Furst DE et al. Opportunities for integration of pharmacokinetics, pharmacodynamics, and toxicokinetics in rational drug development. J Pharm Sci. 1992;81:605-10.
3. Weissinger J. Pharmacology and toxicology of novel drug delivery systems. Regulatory issues. Drug Saf. 1990;5:107-13.
4. Lexchin J. Drug withdrawals from the Canadian market for safety reasons, 1963-2004. CMAJ. 2005;172:765-67.
5. Streiner DL. The 2 "Es" of research: Efficacy and effectiveness trials. Canadian Journal of Psychiatry-Revue Canadienne de Psychiatrie 2002;47:552-56.
6. Wells KB. Treatment research at the crossroads: The scientific interface of clinical trials and effectiveness research. American Journal of Psychiatry 1999;156:5-10.
7. Corrigan OP. A risky business: the detection of adverse drug reactions in clinical trials and post-marketing exercises. Social Science & Medicine 2002;55:497-507.
8. Linden M. Controlled clinical trials and postmarketing experience. Psychopharmakotherapie 1997;4:26-31.
9. Linden M. Differences in adverse drug reactions in phase III and phase IV of the drug evaluation process. Psychopharmacol Bull. 1993;29:51-56.
10. Lawson DH. Postmarketing surveillance versus clinical trials: which benefits the patient? Cardiology 1994;85:18-23.
11. Linden M. Phase-IV research: specifics, objectives and methodology. Pharmacopsychiatry 1984;17:140-42.
12. Community-Based Evaluation of Etanercept in Rheumatoid Arthritis Patients. Farahani P, Levine M, Gaebel K, Wang E, Khalidi N. J Rheumatology *(in press)*
13. Moreland LW, Schiff MH, Baumgartner SW, Tindall EA, Fleischmann RM, Bulpitt KJ et al. Etanercept therapy in rheumatoid arthritis - A randomized, controlled trial. Annals of Internal Medicine 1999;130:478-486.
14. Weinblatt ME, Kremer JM, Bankhurst AD, Bulpitt KJ, Fleischmann RM, Fox RI et al. A trial of etanercept, a recombinant tumor necrosis factor receptor: Fc fusion protein, in patients with rheumatoid arthritis receiving methotrexate. New England Journal of Medicine 1999;340:253-59.

15. Bathon JM, Martin RW, Fleischmann RM, Tesser JR, Schiff MH, Keystone EC et al. A comparison of etanercept and methotrexate in patients with early rheumatoid arthritis. New England Journal of Medicine 2000;343:1586-93.
16. Ware J. SF-36 Health Survey. Manual and Interpretation Guide. The Health Institute, New England Medical Centre.Boston 1993.
17. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. Arthr Rheum 1980;23:137-45.
18. Felson DT, Anderson JJ, Boers M, Bombardier C, Chernoff M, Fried B et al. The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. The Committee on Outcome Measures in Rheumatoid Arthritis Clinical Trials. Arthritis Rheum. 1993;36:729-40.
19. 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.
20. Escalante A. What do self-administered joint counts tell us about patients with rheumatoid arthritis? Arthritis Care Res. 1998;11:280-90.
21. Pincus T, Callahan LF, Brooks RH, Fuchs HA, Olsen NJ, Kaye JJ. Self-report questionnaire scores in rheumatoid arthritis compared with traditional physical, radiographic, and laboratory measures. Ann.Intern.Med 1989;110:259-66.
22. Deyo RA, Inui TS, Leininger JD, Overman SS. Measuring functional outcomes in chronic disease: a comparison of traditional scales and a self-administered health status questionnaire in patients with rheumatoid arthritis. Med Care 1983;21:180-92.
23. Pincus T, Sokka T, Kavanaugh A. Quantitative documentation of benefit/risk of new therapies for rheumatoid arthritis: patient questionnaires as an optimal measure in standard care. Clin Exp.Rheumatol. 2004;22:S26-S33.