Measure for measure. Outcome assessment of arthritis treatment in clinical practice

Gülfe, Anders

2009

Lund University
Faculty of Medicine
Department of Clinical Sciences, Lund
Section of Rheumatology

Measure for measure

Outcome assessment of arthritis treatment in clinical practice

Doctoral thesis

Anders Gülfe

LUND
UNIVERSITY

Cover photo: Peder Gülfe Örsmark

Printed in Sweden
Mediatryck, Lund 2009
ISBN 978-91-86443-12-2

I don’t know if I like it, but it is what I meant

Ralph Vaughan Williams

CONTENTS

Abstract

Objective: To investigate (i) the performance and agreement between various activity indices and response criteria in TNF-blockade of RA; (ii) the predictive ability of different response criteria and disease activity states regarding continuation of anti-TNF treatment of RA; (iii) Euro-QoL-5-dimensions utility development during TNF blockade of RA, PsA and SpA. Also, (iv) to develop a simple, utility-based outcome measure, the number needed to treat per quality adjusted life year gained (NNQ) and apply it in RA, PsA and SpA patients on anti-TNF treatment.

Methods: Data were retrieved from the South Swedish Arthritis Treatment (SSATG) register. In patients with RA, PsA and SpA commencing treatment with adalimumab, etanercept or infliximab, date of treatment start and stop, core set variables and EQ-5D were recorded, and various activity indices, responses and EQ-5D utility were calculated. Descriptive statistics and completer analysis were used. The NNQ was calculated as the inverted value of the area under the utility gain curve for one year.

Results: Agreement between RA response criteria was poor at the individual level, except at the ACR20/overall level. Disease states exhibited moderate or good agreement at all levels and for most criteria sets, except for remission. Response at ACR20/overall and ACR50/good/major level was found to significantly predict treatment continuation, for most indices already after 6 weeks. EQ-5D utilities improved rapidly (at 2 weeks in RA and PsA) and remained stable over 5 years in TNF blockade of RA, PsA and SpA. NNQ for TNF blockade of RA, PsA and SpA and was found to be 4-6, irrespective of diagnosis and treatment course order.

Conclusions: Response criteria are less suitable for use in individual patients in routine care than disease activity states in RA. By contrast, they are often useful as predictors of continued TNF blockade. EQ-5D utility rises almost instantaneously in TNF blockade and remains stable in RA, PsA and SpA patients remaining on therapy. NNQ is easy to calculate and understand and performs well across 3 diagnostic entities.

Publications

This thesis is based on the following articles, which are referred to by their roman numerals:

I Response criteria for rheumatoid arthritis in clinical practice: how useful are they?
Gülfe A, Geborek P, Saxne T
Ann Rheum Dis 2005; 64:1186—9

II Disease activity level, remission and response in established rheumatoid arthritis: Performance of various criteria sets in an observational cohort, treated with anti-TNF agents.
Gülfe A, Aletaha D, Saxne T, Geborek P
BMC Musculoskelet Disord 2009; 10:41

III Six and twelve weeks treatment response predicts continuation of TNF-blockade in rheumatoid arthritis. Observational cohort study from southern Sweden.
Gülfe A, Kristensen LE, Geborek P
J Rheumatol 2009; 36:517—21. Epub 2009 Jan 22

IV Rapid and sustained health utility gain in anti-TNF treated inflammatory arthritis. Observational data during seven years in southern Sweden.
Gülfe A, Kristensen LE, Saxne T, Jacobsson LTH, Petersson IF, Geborek P
Ann Rheum Dis 2009 Mar 27. [Epub ahead of print]

V Utility-based outcomes made easy: The Number Needed per QALY gained (NNQ). Observational cohort study from southern Sweden of TNF blockade in inflammatory arthritis.
Gülfe A, Kristensen LE, Saxne T, Jacobsson LTH, Petersson IF, Geborek P
Manuscript

Abbreviations

ACR	American College of Rheumatology
AS	ankylosing spondylitis
AUC	area under curve
BASFI	Bath ankylosing splondylitis functional index
BASDAI	Bath ankylosing splondylitis disability index
CDAI	clinical disease activity index
CI	confidence interval
CRP	C-reactive protein
CVD	cardiovascular disease
DAS	disease activity score
DFI	Dougados functional index
DMARD	disease modifying antirheumatic drug
EQ-5D	EuroQoL-5-dimensions
ESR	erythrocyte sedimentation rate
EULAR	European League Against Rheumatism
HAQ	health assessment questionnaire
HRQoL	health related quality of life
HUI	health utility index
IL-1	interleukin-1
IL-6	interleukin-6
LOCF	last observation carried forward
MCID	minimal clinically important difference
MCII	minimal clinically important improvement
NNQ	number needed (to treat) per quality adjusted life year gained
NNT	number needed to treat
NSAID	non-steroidal anti-inflammatory drug
PAS	patient activity scale
PASS	patient acceptable symptom state
PsA	psoriatic arthritis
QALY	quality adjusted life year
RA	rheumatoid arthritis
RAID	rheumatoid arthritis impact of disease score
RCT	randomised controlled trial
SSATG	South Swedish arthritis treatment group
SD	standard deviation
SDAI	simple disease activity index
SG	standard gamble
SpA	spondylarthritis
TNF	tumour necrosis factor
TTO	time trade off.
uSpA	undifferentiated spondylarthritis

Introduction

This thesis deals with some aspects of treatment effect evaluation in routine care of patients with inflammatory arthritides. The impact on the individual as well as on society of these common and chronic diseases is considerable. The burden of symptoms and functional impairment, economic loss due to the costs of care and reduced working capacity and psychosocial consequences may be hard on the affected individual. From a societal perspective, production losses as well as direct costs of drugs and other treatment are a growing concern. Reliable methods of treatment evaluation are thus increasingly important from various standpoints, such as biological and clinical efficacy, safety and economy.

Prior to marketing, new drugs, such as biologics for inflammatory arthritis, are subject to randomized controlled clinical trials (RCTs), usually involving carefully selected patients during limited periods of time. The results of these form the basis for approval by regulatory bodies. Learning more about the performance of new remedies in large patient populations over extended time periods, however, entails the systematic gathering of observational data in clinical practice. RCTs and observational studies have their strengths and weaknesses, and together they provide the best knowledge of efficacy and safety. The present work is based on observational data from the South Swedish Arthritis Treatment Group (SSATG) register, a regional database of biologic treatment of inflammatory arthritides (Geborek and Saxne 2000). The thesis deals with measures of disease activity and health related quality of life (HRQoL) as measured by various instruments, and it is both an investigation of the tools applied and the patients' response to treatment.

Arthritis

The term "arthritis" denotes joint inflammation. Traditionally, it has been used to cover a wide range of states with more or less prominent inflammation. The current work deals with rheumatoid arthritis (RA), psoriatic arthritis (PsA) and spondyloarthritis (SpA), which are considered to be autoimmune diseases with marked inflammatory features. They are prevalent worldwide, and they have profound impact on the function, quality of life and comorbidities of patients (Gabriel and Michaud 2009). Even though the diagnostic entities under consideration here are separate, they have much in common, e. g. chronicity and the potential for structural damage.

Rheumatoid arthritis

RA is an autoimmune, chronic, heterogeneous disease (Klareskog et al. 2009), affecting the synovial joints but also causing systemic inflammation, general malaise, wasting, fatigue and occasionally extra-articular manifestations, e. g. vasculitis, serositis, interstitial lung involvement and increased risk for comorbidities such as cardiovascular disease and lymphoma. The prevalence of RA is about 0.5% of the adult population (Simonsson et al. 1999) and in most cohorts about 70% of the patients are female (Lawrence et al. 1998). The ethiology and pathogenesis of RA are complex, involving both genetic and environmental factors. Furthermore, these appear to be distinct in different subsets of RA. HLA-class II alleles, especially HLA-DRB1 corresponding to the "shared epitope" are among the important genetic factors behind RA, and it has recently been shown to interact with an environmental factor, smoking, in increasing the risk for a more aggressive subset of RA (Klareskog et al. 2006).

Prominent clinical features of RA are synovitis with pannus formation and degradation of articular cartilage and bone, resulting in joint space narrowing and erosions on X-rays, respectively. Typically, there is multiple, symmetrical involvement of small joints of the hands and feet, but the presentation may be variable. There is a wide range of severity, from patients presenting with erosions shortly after the onset of symptoms and rapidly developing deformities and functional loss, to relatively mild cases with little symptoms and preserved function and quality of life. There is often acute phase reaction with elevated CRP and ESR, but this may be lacking. Since 1939, it has been known that many RA patients are positive for rheumatoid factor, an antibody to the Fc fragment of immunoglobulin G, even though this test is not specific (Waaler 1970). More specific tests are antibodies to citrullinated proteins (ACPA), one of which is known as anti-CCP (Schellekens et al. 2000). There is evidence that seropositive or anti-CCP positive RA represents a more serious subset, prone to early erosiveness and extra-articular manifestations. This may be involved in the gene-environment (smoking) interaction referred to above (Klareskog et al. 2006).

Research on RA has been much facilitated by the American College of Rheumatology 1987 classification criteria (Arnett et al. 1988). They were designed for the characterisation of groups of patients with established disease in the setting of investigations and trials, not for making the diagnosis in individuals. These criteria, however, are not sensitive to early RA (Banal et al. 2009), a serious drawback, given the importance of early treatment. Development of new classification criteria is under way in collaboration of ACR and the European League against Rheumatism (EULAR).

Psoriatic arthritis

As a rule, PsA is included among the spondyloarthritides, even though the nosology of these diagnostic entities is variable between authors and countries. The frequency of PsA in the SSATG database, as well as its distinctive features is, however, felt to merit a separate section in the current study. Psoriasis is thought to occur in 1-3% of the population, and among these 7-42% have been reported to have inflammatory arthritis; thus up to 1% of the population may have PsA (Gladman 1998). Prevalence figures vary with case definitions; estimates from the USA range from 0.1 (Gabriel and Michaud 2009) to 0.67% (Lawrence et al. 1989).

Research on PsA is hampered by a lack of universally accepted, validated classification criteria. Controversy even exists if PsA is to be regarded as a separate entity, but the weight of evidence seems to justify this (Fitzgerald and Dougados 2006). An early, frequently quoted survey (Moll and Wright 1973) defines PsA as inflammatory arthritis in the presence of psoriasis, and usually negative for rheumatoid factor. Various subtypes are described, including predominant DIP engagement, arthritis mutilans, symmetrical, RA-like polyarthritis, peripheral oligoarthritis and spondylitis. In practice, these are difficult to apply, as many patients evolve from one subtype to another. The Moll and Wright criteria have also been considered to differentiate PsA poorly from RA, and other attempts have thus been made over the years (Helliwell and Taylor 2005, Taylor et al. 2006).

It has been stated that PsA is a more benign disease than RA (Moll and Wright 1973), but this may not always be the case. Synovitis, systemic manifestations such as fatigue and skin symptoms may cause considerable pain, functional impairment and negative impact on HRQoL. Structural damage can be prominent, e. g. in arthritis mutilans with marked shortening of digits due to resorption of phalangeal bones. Other important causes of chronically diminished hand function are enthesitis, tendonitis, and tenovaginitis. Many patients have spinal involvement with pain and limited range of spinal motion, with or without peripheral arthritis.

Spondyloarthritis

The spondyloarthritides may be regarded as a family of inflammatory joint diseases, including

ankylosing spondylitis (AS), PsA, reactive arthritis, arthritis associated with inflammatory bowel disease (IBD), undifferentiated SpA (uSpA), and the juvenile forms of AS and PsA. There is considerable overlap and potential for confusion between the various members of this family, and a number of classification criteria have been proposed (Sieper et al. 2006, Rudwaleit et al. 2009). For instance, a patient with IBD and spondylitis may be defined as having AS with concomitant IBD, or IBD-associated arthritis with spondylitis, the tradition being different among centres and countries. In this thesis, SpA is defined as AS, IBD-associated SpA or uSpA that are grouped together. The limited number of SpA patients in the SSATG register precludes meaningful analysis of subgroups other than PsA.

There is an association between the human leucocyte antigen (HLA) B27 and spondyloarthritis, particularly in the case of AS (Schlosstein et al. 1973). The male to female ratio in AS is about 2:1 to 3:1. The prevalence of AS, like other SpAs, is strongly dependent on the frequency of HLA-B27 in the population studied, and the classification criteria employed. The prevalence of AS in Scandinavia has been reported to be 0.15-1.4% in the general population, but it is higher in certain ethnic groups and much lower in Asia and Africa (Sieper et al. 2006).

Clinically, SpA is characterised by inflammatory back pain with or without asymmetrical peripheral arthritis, usually of the lower limbs, symptoms starting in the 2nd or 3rd decade of life, radiographic sacroiliitis or ankylosis of the SI joints and spine, and presence of HLA-B27 (van der Linden and van der Heijde 1998). Inflammatory back pain starts before the age of 40 years, normally has insidious onset, persists for at least 3 months, and is associated with morning stiffness and improvement on exercise but not on rest. The most frequently used classification criteria for AS are the modified 1984 New York criteria (van der Linden et al. 1984), and for SpA in general, the European Spondylarthropathy Study Group (ESSG) criteria (Dougados et al. 1991), which include the uSpA group. There is often considerable delay in diagnosis, partly due to the patients with inflammatory back pain not being recognised among the vast majority having degenerative or mechanical back problems. In AS, the most severe and frequent subgroup, axial symptoms predominate, but only a minority of patients progress to total ankylosis (van der Linden and van der Heijde 1998). In the other SpAs, peripheral arthritis may be prominent, with or without axial involvement.

Treatment of inflammatory arthritis

During the later part of the 20th century, arthritis treatment has undergone major changes through the introduction of effective disease modifying antirheumatic drugs (DMARDs), notably the biologic therapies. Anyone active in the field has had the keen feeling of the history of rheumatology evolving rapidly. Before the advent of DMARDs, rheumatology was a question rather of managing than of treating. This, to a great extent, belongs now to the past. With the emergence of new potent, sometimes expensive and potentially toxic drugs, demands on rheumatology and society have increased radically. The challenges are both clinical (giving the right treatment to the right patients and monitoring them efficiently) and economic. Even so, some major therapeutic problems remain to be solved. Still, a sizable portion of patients fail multiple treatments and their disease continues to be active; in others, treatment choices are limited by toxicity or co-morbidities. It is still hard to predict response in individual patients. Internationally, the availability of the new, expensive treatments is variable for financial reasons, and arthritis patients may face different prognoses depending on their country of residence (Sokka et al. 2009a). There thus remains much work to be done, both in basic science (development of new therapies) and clinically (learning from post-marketing observational studies).

Aims of arthritis treatment

The three main aims of arthritis treatment are:

- Abolition of symptoms;
- Prevention of functional loss;
- Preservation of longevity.

Symptoms like pain, stiffness, fatigue and general malaise prompt patients to seek medical attendance, and the elimination of these is appropriately the first priority of patient and physician alike. They are closely related to inflammation, and there are ample therapeutic options to relieve them.

Functional impairment is linked to both inflammation (reversible component) and structural damage (irreversible) (Aletaha et al. 2006). Structural damage in inflammatory arthritis is assessed as erosions on radiographs or deformity clinically. The main challenge is to prevent irreversible damage from occurring, which is the basis for the almost universally accepted mantra of early, aggressive treatment (Sokka and Makinen 2009a). Inflammation contributes to structural damage, but there is evidence of a partial disconnect between the two, demanding treatment that goes beyond reversing inflammation (Landewe et al. 2006, Saleem et al. 2009, Smolen et al. 2009a). This aspect of arthritis treatment is of long term relevance to work capacity and HRQoL.

Mortality has been found to be increased in inflammatory arthritis (Wolfe et al. 1994, Gladman 2008, Zochling and Braun 2008). The most frequent causes of death in arthritis patients are the same as in the general population, i. e. cardiovascular disease (CVD), malignancy, and infection. There is thought to be a connection between long-standing, high-degree inflammation and the risk of CVD (Turesson et al. 2008), and some types of malignancy, and also with serious extra-articular manifestations like amyloidosis and RA vasculitis. The effective reversal of inflammation thus may contribute to reduce the risk of premature death in arthritis patients (van Vollenhoven 2008). On the other hand, there has been concern regarding increased risk of malignancies, particularly lymphomas, as a consequence of treatment with TNF blockers, in addition to the already elevated lymphoma risk in RA. This seems, however, not to have been substantiated to date (Askling et al. 2009).

Traditional treatment

Spa treatment retains some of its popularity, e. g. in the form of rehabilitation abroad in warm climates. Methodologically sound trials of spa therapy in arthritis patients seem, however, largely to be lacking (Verhagen et al. 2003). From the early days of rheumatology, training and maintenance of function by the employment of physical therapy has been important, also in the form of post operative training after orthopaedic and hand surgery. Occupational therapy and aids of various kinds are needed by many patients. Team care, patient education and a holistic approach were even more emphasised before the advent of DMARDs, but they remain an essential complement to this day. There is evidence that structured patient education is beneficial in RA (Lindroth et al. 1997).

Aspirin was introduced into rheumatology soon after its invention, and the non-steroidal anti-inflammatory drugs (NSAIDs) developed from aspirin are still widely used in spite of substantial toxicity. They are a mainstay of AS and SpA treatment even today (Akkoc et al. 2006).

Disease modifying anti-rheumatic drugs and glucocorticoids

Methotrexate is the most widely used DMARD, considered to be the "anchor drug" in RA treatment (Bijlsma and Jacobs 2009), but it is also used in PsA and SpA, although the basis for this is weaker (Braun and Rau 2009). In clinical trials, new treatments are compared to methotrexate rather than placebo. The early institution of methotrexate therapy has altered the course and prognosis of RA towards fewer symptoms and less tissue damage and function loss. The traditional "therapeutic pyramid" with NSAIDs at the basis, then antimalarials and gold preparations and methotrexate at the

top, to be used in severe cases after the failure of other measures, has thus been tilted upside down (Wilske 1993).

Sulfasalazine is another traditional DMARD used for RA (Suarez-Almazor et al. 2000), PsA (Ravindran et al. 2008) and AS, mainly being effective in peripheral disease (Chen and Liu 2005). Antimalarials, gold, azathioprine, ciklosporin and leflunomide are also used to a limited extent in the patient groups focused on in this thesis.

Combination treatment with several DMARDs has been studied mainly in RA, both in early and established disease (O'Dell 2001). Most often, step-up therapy is used in cases with a suboptimal response to a single agent, usually methotrexate. There are also step-down regimens, starting with multiple agents that are tapered one by one. The problem in initial combination therapy is, that it is impossible to judge if the patient responds to one or more of the agents given, thus possibly giving rise to unnecessary toxicity.

Since their introduction in the 1950s, glucocorticoids have an important albeit somewhat controversial position in the anti-rheumatic armamentarium. The well known side effects, such as osteoporosis and increased susceptibility to infection, have caused scepticism among rheumatologists as regards long term steroid treatment. On the other hand, they have been much employed as bridging therapy for inflammatory arthritis, awaiting the onset of effect of slow-acting DMARDs (van Gestel et al. 1995). There is some evidence that low dose glucocorticoid (<10 mg Prednisolone daily) retards radiographic progression in RA in the medium term (2 years) (Kirwan 1995, Svensson et al. 2005), but the long term effects are still not clear.

Another important application for glucocorticoids is for local injection into joints or tendon sheaths (Weitoft 2005). Synthetic steroid esters with low water solubility are used to promote retention locally.

Biologic treatment

DMARDs and glucocorticoids exert a broad and unspecific immunosuppressive effect, and in most cases they have been found to be effective in rheumatic diseases empirically. Biologic agents, by contrast, are specifically engineered to target a specific mechanism or molecule (e. g., a proinflammatory cytokine), known to be relevant in the disease under consideration. This concept was first successfully pursued in oncology (Clark and Weiner 1995). Currently available biologic drugs for inflammatory arthritis include antagonists to tumour necrosis factor α (TNFα), interleukin-1 (IL-1), interleukin-6 (IL-6), co-stimulatory interaction between antigen-presenting cells and T-cells, and the B-cell antigen CD20, resulting in B-cell depletion.

The TNF antagonists studied here are the monoclonal antibodies infliximab (Maini et al. 1998) and adalimumab (Keystone et al. 2004) and the receptor fusion protein etanercept (Moreland et al. 2001). Infliximab is a chimeric mouse/human anti-TNFα antibody, administered as an intravenous infusion of 3 mg/kg (5 mg for PsA) at weeks 0, 2, 6 and every 8 weeks thereafter in conjunction with methotrexate. Adalimumab, a fully human monoclonal antibody to TNFα, is given as a subcutaneous injection of 40 mg every 2 weeks. Etanercept is given subcutaneously, 25 mg twice weekly or 50 mg once weekly. All three anti-TNF drugs are approved for RA, PsA and SpA.

The introduction of TNF blockers was a major step forward in arthritis treatment. Even though a sizable proportion of patients (25-30%) responds poorly, most patients experience rapid effect on symptoms, and there is evidence of superior effect in retarding erosiveness in RA as compared to conventional DMARDs, especially in combination with methotrexate (van der Heijde et al. 2006).

The problem of treatment failure can be handled in different ways. In both primary and secondary failure, it has turned out that switching between anti-TNF drugs is useful in many patients (Karlsson et al. 2008, Smolen et al. 2009b). In the case of infliximab, the dose interval is frequently diminished or the dose increased in cases of insufficient effect, although there seems to be little justification for the latter approach (Pavelka et al. 2009). It may seem more appropriate to try biologics with other

mechanisms of action in the case of primary anti-TNF failure, but the options for this have been limited until recently.

Safety data on the TNF blockers are now beginning to accumulate. Due to the anti-tumour effect of TNFα (at least in vitro), there was some concern as to the risk of malignancy, particularly lymphoma, but available data does not suggest obvious risk increase (Askling and Bongartz 2008). Also, there seems not to be any enhanced risk of cardiovascular disease, an issue brought up by the harmful effect of TNF blockade in severe heart failure; rather the opposite seems to be the case (van Vollenhoven 2008). Allergic reactions may limit the use of these remedies, particularly infliximab without concomitant methotrexate (Kapetanovic et al. 2006). The role of TNFα in defence against bacterial infection, not least tuberculosis, has prompted special vigilance concerning infectious complications during TNF blockade. Screening for occult infection prior to institution of treatment now is mandatory. The role of TNF blockers in surgical complications is not clear. Even though TNF blockade may carry some increased risk for infection and certain malignancies, so far the overall impression is that this risk is outweighed by its therapeutic effect. The risk of cardiovascular disease seems to be lowered (Zink et al. 2009).

Even though the first TNF blockers were marketed 10 years ago, it must be kept in mind that they are used for chronic conditions with the potential need of treatment lasting a lifetime. New safety issues may thus arise in the future.

Alternative biologics, approved to date in Sweden for RA only, are the co-stimulation blocker abatacept, the anti-CD20 antibody rituximab and the anti-IL-6-antibody tocilizumab, which were not studied in the current investigation and thus are outside the scope of this overview.

Treatment strategies and recommendations

Given the recent availability of many new, effective treatment modalities, therapeutic thinking in arthritis care has changed fundamentally. The traditional therapeutic pyramid referred to above is now obsolete. There are many possible sequences and combinations of options, and optimal strategies are largely unknown, with few head-to-head comparisons of drugs and regimens. Thus, there are today many alternatives to the conventional "saw tooth" method of trying one DMARD after another until, hopefully, achieving response, which may take many months during which irreversible damage might occur.

"Tight control" of RA (TICORA) by means of frequent evaluations, activity index (DAS28) driven modifications of conventional DMARD treatment and local and systemic steroids was inspired by strategies employed in diabetes (Grigor et al. 2004). The TICORA regimen resulted in better disease activity control than usual care, perhaps in part explained by generous steroid usage; however, the concept is interesting and worth further studies.

The concept of a "therapeutic window" in RA has prompted studies of very early, intensive treatment regimens such as the COBRA (Boers et al. 1997) and BeST studies(Goekoop-Ruiterman et al. 2005). In these, initial high dose glucocorticoid and infliximab, respectively, were given in early RA with the hope of turning off the disease process and maintaining remission in the long term by modest, conventional treatment. In a proportion of patients, this was possible, at least in the medium term perspective. These strategies remain controversial and have not been widely accepted in routine care.

In order to support rheumatologists in making the important treatment decisions in everyday practice, a number of guidelines, algorithms and recommendations have emerged (www.NICE.org.uk/CG79, Maksymowych et al. 2007, Pham et al. 2007, Gottlieb et al. 2008, Saag et al. 2008, Luqmani et al. 2009). The current Swedish recommendations for treatment of RA are given in Figure 1 (www.svenskreumatologi.se). Naturally, guidelines need frequent revision as new drugs and new data on existing drugs emerge. They are generally issued by rheumatology organisations or health authorities pertaining to the various conditions in each country or region. They are based on a

combination of literature evidence and expert opinion, i. e. on group level data. Consequently, they provide no more than a framework of reasoning about the therapeutic choices in individual patients, as they cannot take all circumstances such as co-morbidities, economic and logistic factors, local treatment tradition or experience and preference by patient or rheumatologist into account. In the end, the treatment choice rests with the treating rheumatologist in dialogue with the well informed patient,and guidelines will not for long eliminate the need for clinical judgement (Kavanaugh 2009).

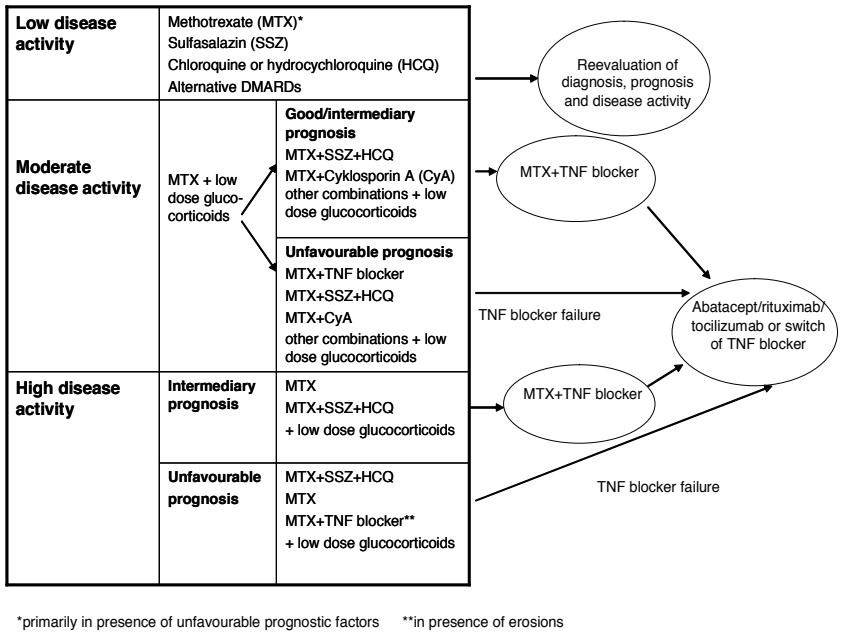


Figure 1. Treatment guidelines for RA, issued by the Swedish Society of Rheumatology. Translated from(www.svenskreumatologi.se).

Assessment of treatment outcome

Disease activity

One important and desirable outcome of arthritis treatment is the reduction of disease activity, if possible to nil. The concept of disease activity is closely linked with inflammation, but there is no single clinical or laboratory measure encompassing this, nor is there a gold standard. A variety of composite measures exist. These may comprise patient, physician and laboratory derived measures, variously chosen and weighted (Table 1).

Some indices include all of these, e. g. the DAS28 (www.DAS-score.nl) and the SDAI, others excludes the laboratory measure, e. g. the CDAI (Aletaha 2005a), and there are indices with only patient derived components, such as the PAS (Wolfe et al. 2005), to choose a few examples from RA. There are activity indices and response criteria for PsA and SpA also (Garrett et al. 1994, Fransen et al. 2006, Lukas et al. 2009). There is an ongoing debate concerning the optimal activity index from aspects of feasibility, sensitivity to change and ability to cover important aspects of the disease. Generally, no index can be said to summarize all important aspects in a single numeric measure, and the uncritical acceptance of a score has its pitfalls, as some of the components may be heavily influenced by factors unrelated to

inflammatory arthritis. One example of this is general health and pain measures in the context of intercurrent conditions such as fibromyalgia (Ranzolin et al. 2009).

There are two different ways of looking upon disease activity outcomes in inflammatory arthritis treatment: (i) response, i. e. the difference in activity between baseline and some later time point after the institution of treatment; and (ii) achievement of a certain level of activity, e. g. remission or low disease activity. Response criteria, such as the ACR20-50-70% (Felson et al. 1995), were developed for clinical trials, where the efficacy of a treatment was given as the proportion of subjects fulfilling a given response criterion. Later, disease activity states were included among the outcomes in RCTs, and indeed the EULAR response criteria for RA (van Gestel et al. 1996) incorporate both improvement and the absolute level of disease activity at the endpoint. Even though designed for RA, the EULAR response criteria have been shown to perform well also in PsA (Fransen et al. 2006).

For the individual patient, as well as from a tissue damage and functional perspective, it seems to be more relevant to be in remission or low disease activity, than to have made a leap downward on the activity scale of a certain magnitude. Thus, the various disease activity index scales have been divided by cut-points into high, medium, low disease activity and remission (Table 1). This simplifies the interpretation of index measurement values, but also reduces the information content of

Table 1

A. Cut-points for disease states according to selected activity indices.

	High	Moderate	Low	Remission
DAS28 original#	>5,1	<5,1	<3,2	<2,6
DAS28 modified cut off #	>5,5	<5,5	<3,6	<2,4
SDAI##	>26	<26	<11	<3,3
CDAI###	>22	<22	<10	<2,8

B. Components of selected response criteria and activity indices, and algorithms for their calculation.

	TJC	SJC	Patient global VAS	Patientpain VAS	Evaluatorglobal	HAQ	ESR	CRP
ACR	yes	yes	yes/no	yes/no	yes/no	yes/no	yes/no	yes/no
EULAR	yes	yes	yes	no	no	no	yes	no
PsARC£	Yes	Yes	Yes(Likert)	No	Yes(Likert)	No	No	no
SDAI	yes	yes	yes	no	yes	no	no	yes
CDAI	yes	yes	yes	no	yes	no	no	no
PAS££	No	No	Yes	Yes	No	Yes	No	no
RAID£££	no	no	n/a	yes	no	yes	No	no
BASDAI$	No	No	No	Yes	No	No	No	no
ASDAS B$$	No	No	Yes	Yes	No	No	Yes	no

DAS28=0,56x√TJC28+0,28x√SJC28x0,7xlnESR+0,014xPat global VAS(in mm)

SDAI=SJC+TJC+Pat global VAS(in cm)+Eval global VAS(in cm)+CRP(in mg/dL)

CDAI=SJC+TJC+Pat global VAS(in cm)+Eval global VAS(in cm)

£ PsARC response=improvement in at least 2 of Pat global Likert 1-5, Eval global, TJC improved 30% or more, SJC improved 30% or more (1 of TJC or SJC mandatory)

££ PAS=(3,33xHAQ+VASpain+VASglobal)/3

£££ RAID=weighted sum of pain(21%), functional disability (16%), fatigue (15%), emotional well-being (12%), sleep (12%), coping (12%), physical well-being (12%)

$ BASDAI=(VASfatigue(in cm)+VASneck, back and hip pain+VASpain and swelling, other+VASdiscomfort/tenderness+[VASmorning stiffness, level+VAS morning stiffness, duration]/2)/5

$$ ASDAS B=0,079xVASback pain(in cm)+0,069xVASmorning stiffness duration+0,113xPat global VAS+0,293x√ESR

continuous variables into a few categories. It has been stated that "it's good to feel better, but it's better to feel good" (Dougados 2005). The two aspects of treatment outcome have been codified into the concepts of minimal clinically important improvement or difference, MCII/MCID (feeling better), and patient acceptable symptom state, PASS (feeling good). It has been demonstrated that MCII is dependent on the baseline level of the outcome studied (Aletaha et al. 2009), whereas the PASS is fairly constant. In the case of knee osteoarthritis and rotator cuff syndrome, it seems that MCII by patients is perceived as the improvement necessary to reach PASS (Tubach et al. 2006).

Function

The ability to perform without distress the various activities of daily living, including work, leisure activities, family and social life, is summarized in the comprehensive concept of function. Functional impairment in arthritis is dependent on both disease activity (reversible) and structural damage (irreversible component) (Aletaha et al. 2006), but also on social and mental factors not easily grasped in arthritis assessment. The irreversible component translates, from the physician perspective, into structural damage (radiographic erosiveness and deformity, irreversible joint malfunction). Functional impairment is also considered a measure of disease severity, and cohorts of patients are frequently stratified according to function. Traditionally, the four Steinbrocker functional classes for RA (Steinbrocker et al. 1949) were used, but the classes were felt to be too broad with too many patients falling into the middle categories. For many years, the Stanford Health Assessment Questionnaire disability index (HAQ) has been the standard tool to measure function in RA (Fries et al. 1980), but it has been validated also in PsA (Blackmore et al. 1995). There are many modifications to the HAQ; in the current investigation, a validated Swedish translation has been employed (Ekdahl et al. 1988) (Appendix). Unlike many other HAQ versions, this takes the use of aids into account. In SpA, there are specially designed functional measures, e. g. the Bath Ankylosing Spondylitis Functional Index (BASFI) and the Dougados functional index (DFI) (Haywood et al. 2005).

Health related quality of life

Inflammatory arthritis tends to affect every aspect of patients' lives. With the increased recognition of patient derived outcomes, interest in the broader and "softer" concept of health related quality of life (HRQoL) (WHO 1995, Sajid et al. 2008) is spreading. The WHO definition of QoL is "individuals' perception of their position in life in the context of culture and value systems in which they live and in their goals, expectations, standards and concerns". Many RCTs and observational studies of arthritis treatment now include a HRQoL outcome measure. In part, this is due to the inability of the usual measures of disease activity and function to include all the different aspects of the impact that arthritis imposes on the life of patients. Furthermore, health economics is important in new, efficient but costly treatment modalities, and assigning an economic value to treatment gain entails HRQoL measurement. This has led to the development of instruments, some generic and some disease specific, to measure HRQoL and to translate patients' perceptions of it into quantitative data.

In principle, HRQoL can be measured (i) directly by relating the individual's own perception of his or her health status to an ideal state of perfect health, or (ii) indirectly by relating the individual's preferences assessed by a questionnaire, calibrated by a direct method in a reference population. Direct methods include standard gambling (SG), time trade-off (TTO) and visual analogue scale (VAS) rating. SG and TTO are time consuming and do not perform very well in chronic diseases. Thus, questionnaire based, indirect methods, being more feasible, are widely used, such as the EQ-5D (The Euro-QoL Group 1990), or SF-36 (Ware and Sherbourne 1992), which are both generic scales. The EQ-5D questionnaire comprises the dimensions mobility, self care, usual activities, pain/discomfort and anxiety/depression, all of which are relevant in rheumatic diseases. SF-6D is 6 items from SF-36

employed to estimate "utility" (see below) (Brazier et al. 2002). The 6 dimensions of SF6-D are physical functioning, role limitations, social functioning, pain, mental health and vitality.

Questionnaire data can be translated into "utility", a quantitative measure between 0 (death) and 1 (perfect health) with negative values being possible. This is brought about by valuation of the various health states (all combinations of responses to the items), described by the indirect instrument, in a reference population, e.g. the general public, using one of the direct methods. The way from the questionnaire raw scores to a utility value may seem complicated and perilous, at least to the non-expert, as it involves an algorithm that can be variously designed, and weighting of the items. Indeed, EQ-5D utilities depend strongly on the tariff used (Nan et al. 2007). Thus, caution must be exercised in attributing absolute levels of HRQoL to utility measurements. This is even more important, if utilities in various studies are derived by means of different tariffs.

Utilities can be used to calculate quality adjusted life years, QALYs (Rasanen et al. 2006). One QALY is equal to 1 year spent in perfect health (utility=1); thus 1 year spent in a health state with utility 0.5 yields 0.5 QALY, etc, provided utility has been stable for 1 year. More generally, QALY is the area under the utility curve (AUC) for 1 year. QALYs can be assigned a price that founding sources are considered willing to pay, and this forms the basis for health economic modelling. It must again be emphasized, that all QALYs are not alike, but depend on the methodology behind utility measurements. By using different EQ-5D tariffs on the same data, contradictory judgements of cost-efficacy may be arrived at (Noyes et al. 2007).

Aims of the present investigation

The main aims of the present thesis were to

- Examine and compare the performance of and agreement between various activity indices and response criteria in patients with established RA, subjected to TNF blockade in clinical practice;
- Investigate the ability of the same criteria sets to predict continuation of anti-TNF treatment in RA;
- Describe the change in utility, measured longitudinally by the EQ-5D generic instrument, during TNF blockade in established RA, PsA and SpA in routine care, and to investigate secular trends in baseline utility values;
- Evaluate the feasibility of EQ-5D in routine arthritis care and its performance across 3 diagnostic entities;
- Explore whether utility baseline values and improvement are different in 1st, 2nd and 3rd or more anti-TNF treatment courses;
- Develop a simple, utility based outcome measure, the number needed (to treat) per QALY gained (NNQ), and apply it in TNF blockade of established RA, PsA and SpA in day-to-day arthritis care.

Patients and methods

The SSATG register

The South Swedish Arthritis Treatment Group was established in 1999 as an informal network of rheumatology departments and clinics in southern Sweden (Figure 2). The chief objective was to gather information on biologic treatment for arthritis, which was introduced on the Swedish market that year, and manage these in a database (Geborek and Saxne 2000). Primarily, this was undertaken to survey quality and safety, and legally, the database is considered to form part of the routine care of the patients. This was why the Research Ethics Committee of the Faculty of Medicine, Lund University, has determined that ethical approval of the various research projects, based on data from the register, was not necessary. A comparison with the pharmaceutical sales of TNF blockers in the catchment area showed a 95% coverage by the register (Geborek et al. 2005).

When starting a new course of biologic treatment, diagnosis, baseline demographic data, previous and current treatment (DMARDs, glucocorticoids, NSAID and analgesic use) are entered by the treating rheumatologist. Diagnosis is as judged by the treating physician; however, check-lists with the 1987 ACR criteria for RA and the 1984 revised New York criteria for AS are provided. A systematic review of the case records of a sample of RA patients (Geborek et al. 2002) demonstrated that 98% fulfilled the 1987 ACR classification criteria.

Patients are evaluated at baseline, 6 weeks, 3, 9 and 12 months and every 6 months thereafter. During the first years, patients were also evaluated at 2 weeks. At each visit, data on treatment, complications, side effects, patient global and pain visual analogue scale (VAS), evaluator global assessment on a 5-point Likert scale, HAQ, EQ-5D along with swollen and tender joint counts, ESR and CRP are gathered. In SpA, data on enthesitis, spondylitis, dactylitis, eye involvement and in PsA nail involvement are also gathered, and BASFI and BASDAI are included in cases of axial disease. Early experiences with the register have been published (Geborek et al. 2002).

On termination of therapy, patients are no longer monitored in the register. The treating physician is to provide data on time and cause of stopping treatment (failure, toxicity or other; only one reason can be given). There is some weakness in the physician-provided data on cause of treatment termination. Minor toxicity could, for instance, be taken as a reason to stop therapy with suboptimal response; the most relevant data is the overall termination rate (Zink et al. 2005, Kristensen et al. 2006).

Switchers between anti-TNF therapies are monitored in the same way for their 2nd, 3rd or more courses. In this thesis, data generally refers to treatment courses rather than to individual patients, which must be taken into consideration before making conclusions. The same individual may thus contribute data from several treatment courses. In papers I and II, however, only anti-TNF naïve patients were included.

Study populations

Paper I: One hundred eighty four RA patients from Lund, entered in the SSATG register between January 1, 1999 and December 31, 2001 due to commencement of a 1st course of TNF blockade, were investigated regarding agreement of response criteria at 3 and 6 months.

Paper II: From the SSATG catchment area, 1789 RA patients starting their 1st course of TNF blockade were included, and the fulfilment of and agreement between disease activity states, defined by various indices, and response criteria (in the 1258 patients with complete data at 3 months) were investigated. Recruitment was between March 1999 and December 2006.

Paper III: The same cohort as in paper II was examined for the predictive ability of fulfilling various activity states and response criteria sets as regards continuation of therapy.

Paper IV: In the SSATG registry, anti-TNF treatment courses were identified from March 2002 through December 2008 and EQ-5D utility data were retrieved from patients with RA, PsA or SpA in order to study the development of utility over time. There were 1584, 742 and 228 eligible 1st, 2nd and 3rd or more RA treatment courses, respectively; corresponding numbers for PsA were 401, 135 and 38; and for SpA 430, 117 and 39, respectively. Patients were followed for up to 7 years.

Paper V: Mean EQ-5D utility gain was calculated for 1st, 2nd and 3rd or more anti-TNF courses of 1001 RA, 255 SpA and 241 PsA patients. This was utilized to calculate the NNQ. The treatment courses were started January 2002 through December 2007, and data extraction was closed December 1, 2008.

Instruments of outcome measurement

The background and purpose of the various instruments are briefly outlined under "Assessment of treatment outcome". The actual questionnaires employed in this study (Appendix) were:

- SSATG forms for the physician (baseline data at treatment start; joint counts, evaluator's global assessment, medications and safety data at each visit including time and cause of treatment cessation);
- SSATG forms for the patient (comorbidities at start of treatment; general health and pain VAS, comorbidities arisen since last visit, side effects at each visit);
- The Swedish version of the Health Assessment Questionnaire (HAQ);
- The Swedish translation of the descriptive part of Euro-QoL-5-dimensions (EQ-5D).

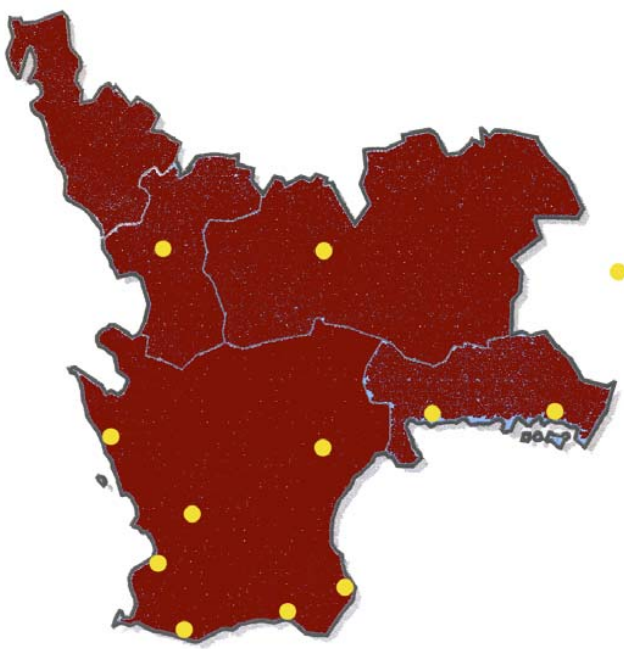


Figure 2. Catchment area for the South Swedish Arthritis Treatment Group (SSATG) register.

Statistics

Descriptive statistics was used. Due to the observational nature of data, completer analysis was employed. Missing data were not imputed; however, the characteristics of drop-outs have been analysed when possible. Furthermore, the number of valid observation at each time point has been given when relevant. All calculations were made in SPSS v 14.0 and p values less than 5% were considered significant.

To estimate agreement between various criteria sets, different methods were used in paper I and II, respectively. In paper I, "agreement" is defined as the fraction of subjects fulfilling 2 criteria simultaneously. In paper II, this is termed "positive agreement", whereas "negative agreement" denotes subjects *not* fulfilling any of the 2 criteria compared. The sum of these fractions constitutes "agreement" in paper II (Table 2). Furthermore, agreement is given as percentages in paper I and II, but in paper II also as κ values. The κ statistic is a measure of the degree to which agreement is not explained by chance alone. It is a value between -1 and 1, where 1 denotes perfect agreement, 0 chance agreement and -1 perfect non-agreement. For interpretation of κ values, please refer to paper II.

The proportion of patients remaining on therapy was calculated as Kaplan-Meier estimates, and drug adherence and discontinuation was estimated with life-table analysis. Predictors of drug continuation were determined by Cox regression models corrected for various confounders (Paper III).

In paper V, two methods for utility AUC calculation, mean utility for 1 year and the trapezium rule, were compared using Spearman's ρ.

Table 2 Positive and negative agreement, explained by comparing 2 response criteria (paper II).

	EULAR overall yes	EULAR overall no
ACR20 yes	Positive agreement	No agreement
ACR20 no	No agreement	Negative agreement

Results and discussion

Agreement of various disease activity indices and response criteria (Papers I and II)

The various activity indices, their cut-points for remission, low, moderate and high disease activity, as well as response criteria sets, have been compared previously using RA RCT data. In order to investigate their performance and agreement in routine care at the individual level, 2 studies of established RA patients, receiving their first anti-TNF treatment were conducted (paper I and II). Among activity indices and their disease state cut-points, DAS28 (original and modified cut-points), SDAI and CDAI were investigated along with their respective response criteria sets and the ACR20-50-70% response.

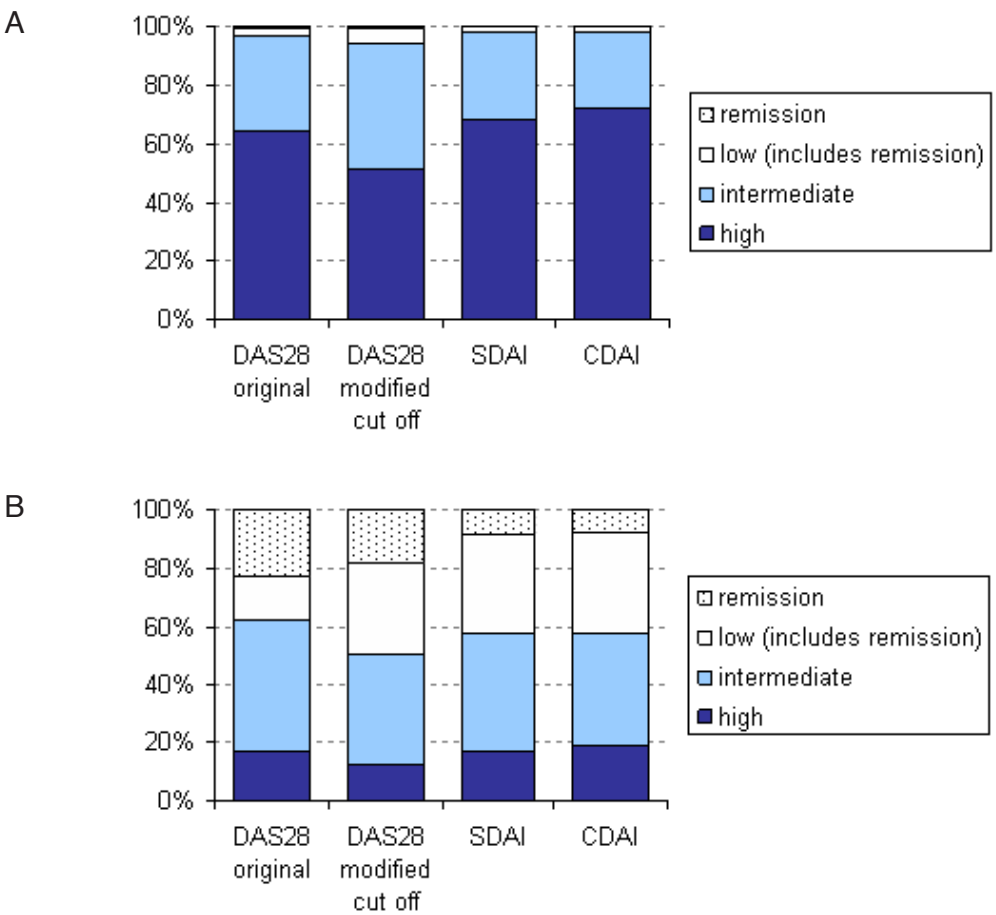


Figure 3. Distribution of activity states at baseline (N=1789, Panel A) and 3 months (N=1258, Panel B) in the RA cohort in paper II.

Figure 3 displays the distribution of activity states among the 1789 RA patients at baseline (Panel A) and the 1258 patients with 3 month data (Panel B). As expected, the great majority of cases are in high or moderate activity at baseline; however, a small number are in low activity. This may be explained by the use of high glucocorticoid doses, an indication for TNF blockade according to the Swedish treatment guidelines (www.svenskreumatologi.se). At 3 months, just fewer than 20% of patients according to all criteria sets are in high disease activity, and 40-50% are in low disease activity (including remission). This may be perceived as a modest degree of treatment success; it is not, though, altogether unexpected in an RA population with a mean disease duration of 12 years who have failed at least two DMARDs. In established RA, there will often be symptoms due to joint damage/erosions not amenable to treatment of inflammation, influencing pain and general health scores. Furthermore, our findings are in line with other observational data (Wolfe et al. 2007).

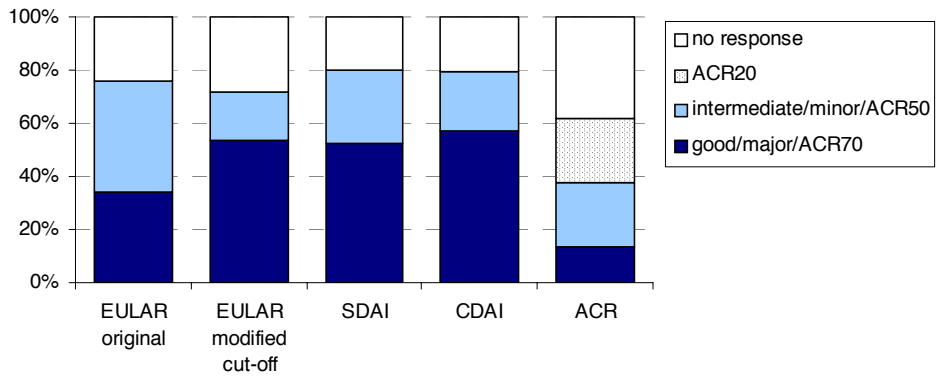


Figure 4. Response rates at 3 months (N=1258) of the patients in the RA cohort in paper II according to various criteria sets.

Response rates of the cohort in paper II according to the various criteria sets are shown in Figure 4. The rates are somewhat higher than in the paper I cohort of 184 RA patients from Lund. Among these, 51, 74 and 67% responded according to ACR20, EULAR overall and SDAI overall, respectively; the corresponding rates for ACR50, EULAR good and SDAI major were 29, 30 and 30%. In part, this may be due to the patients in paper I being recruited earlier (1999-2001), when TNF blockers just had been introduced, and they might represent more severe cases than the paper II patients, the latest of whom were recruited in 2006. We have been able to demonstrate falling trends of baseline disease activity and HAQ disability over the period 1999-2006, supporting this (Soderlin and Geborek 2008).

In general, agreement among disease states was moderate or substantial, with κ=0.5 or better, excepting remission agreement between the DAS28-based states and CDAI/SDAI. This is possibly due to low number of patients in these categories. Agreement between SDAI and CDAI was excellent; the only difference being that CRP is included in SDAI but not in CDAI. Accordingly, this supports that acute phase reactants contribute little to activity indices (Aletaha et al. 2005b). At 3 months, there is a tendency towards lower agreement between DAS28-based states and SDAI/CDAI, but κ is still 0.5-0.6. For details, please refer to table 2 of paper II.

As the various activity indices are comprised of much the same components, albeit weighted and handled differently, it is not surprising that they agree rather well. In fact, the concept of disease activity states seems consistent and uniform irrespective of the details of calculation. This may be encouraging to the clinical rheumatologist, trying to orient him- or herself among ever increasing numbers of outcome measures. As the indices investigated seem to perform equally well, the choice of which one to prefer in daily practice must largely remain a matter of individual preference or local

tradition.

Response criteria sets, by contrast, seem to agree less well, at least at the higher levels of response. The positive agreement among ACR20, EULAR overall (original cut-point) and SDAI overall in paper I was around 50%, whereas, at the ACR50/EULAR good/SDAI major level, it was about or below 20% (for examples, see Figure 5). This is in accordance with the findings in paper II, with positive agreement 60-70% and κ around 0.6 at the ACR20/overall level, but much more variable agreement at the ACR70/good/major level and very low agreement at the intermediate level, often with κ about 0. (For details, please refer to table 4 of paper II.) As for the intermediate level, the poor agreement may, at least partly, be explained by the construction of the criteria sets. Thus ACR50 includes ACR20 and ACR70 includes ACR50, but EULAR good does not include EULAR moderate.

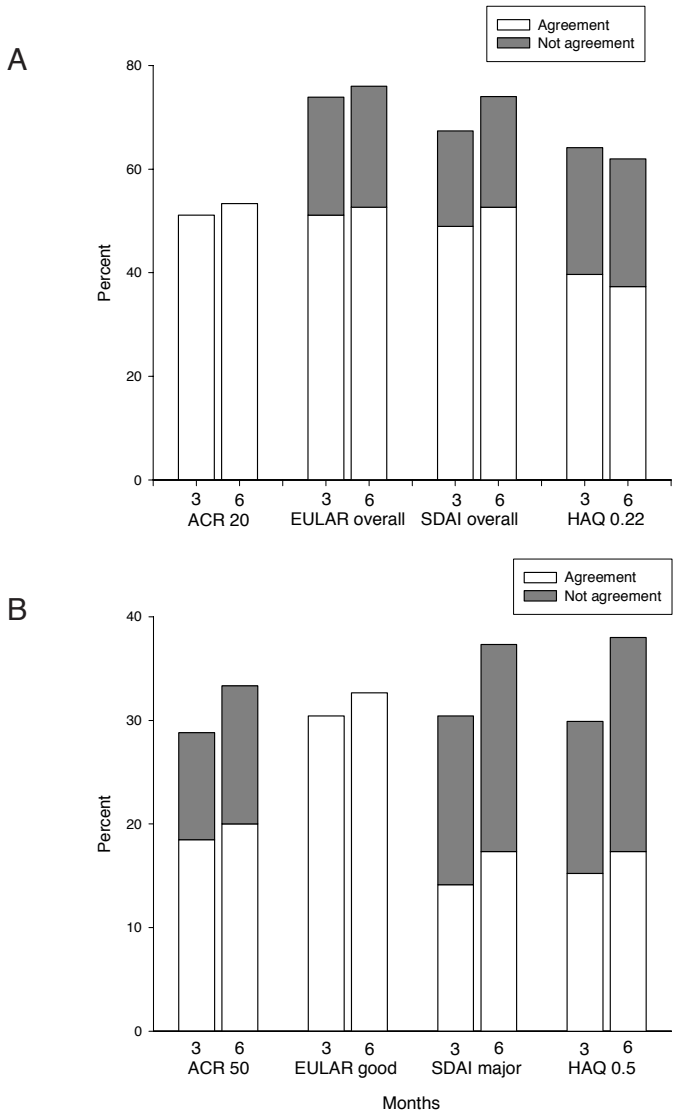


Figure 5. A: Proportion of RA patients fulfilling less strict response criteria sets at 3 and 6 months. Agreement using ACR20 as reference. B: Proportion of patients fulfilling more strict response criteria sets at 3 and 6 months. Agreement using EULAR good as reference. N=184.

At the higher level of response, the different criteria are fulfilled by approximately the same proportion of patients, but not by the same individuals. Thus, response criteria perform well at the group level, which is the point of interest in RCTs, where groups of patients receiving different treatments are compared. By contrast, when applied in clinical practice to individual RA patients, the response, at least at the higher levels, depends heavily on what criteria set is chosen. This will be increasingly relevant, as treatment in day-to-day care more often aims at remission or low disease activity. Caution must thus be exercised in the application of response criteria to individual RA patients in routine care. At the group level, however, they may give some information on the performance of various treatments in the clinical setting.

In summary, disease activity indices and their corresponding disease states seem to be more useful in the daily care of individual RA patients than response criteria. This is in line with the concept of "feeling good rather than feeling better" referred to in the section on outcome assessment.

Prediction of treatment continuation (paper III)

A major challenge in arthritis treatment is prediction of treatment response. To this day, many patients fail to improve convincingly in spite of the powerful remedies now available to the rheumatologist, notably the biologic treatments. In RA treatment with TNF blockers, baseline variables have been shown to predict response to treatment only to a limited extent (Hyrich et al. 2006, Kristensen et al. 2008). To address the problem, if treatment response or achievement of a particular disease state at 6 weeks or 3 months could predict treatment continuation, 1789 RA patients (the same cohort as in paper II) were assessed. DAS28 original and modified cut-points, SDAI and CDAI disease states and the corresponding response criteria along with ACR20-50-70 were investigated. For details on baseline characteristics, please refer to paper II. The fulfilment of the various response criteria and disease states are displayed in Table 3.

Predictors for drug discontinuation were determined by Cox regression models corrected for age, disease duration, baseline HAQ and CRP as well as concomitant use of methotrexate, as these variables all have been demonstrated to influence response at 3 months (Kristensen et al. 2008).

Response at the ACR20/overall level as well as the good/major level (except ACR70) at 6 weeks significantly predicted drug continuation, and so did response by all criteria sets at the same levels at 3 months. Moderate/minor response was not a predictor at any time point for any criteria set (Table 4).

Remaining in high disease activity strongly predicted stopping of therapy at both time points by all criteria sets (Table 4), which is not surprising, as this is in line with the current treatment recommendations in Sweden (www.svenskreumatologi.se). Conversely, achieving remission or low disease activity predicts continuation of treatment, especially after 3 months. At 6 weeks, neither achieving remission nor low disease activity (except DAS28 low) is a predictor, and neither is moderate disease activity.

Thus, at the group level, response already at 6 weeks and even more at 3 months is a significant predictor of drug continuation. ACR70 is not, however, possibly due to low number of responders. Achieving remission or low disease activity also predicts drug continuation at 3 months, but less clearly so at 6 weeks, perhaps due to more patients gaining treatment effect as time goes by. Moderate/minor response or achieving moderate activity is not a predictor.

Reason for stopping treatment may influence predictability. To examine this, hazard ratios (HRs) for drug discontinuation by the strongest predictor (EULAR original overall response) was computed separately for failure and adverse events. At 6 weeks, the HR for stopping treatment due to adverse events was 1.17 (95% CI: 0.61-2.26), P=0.64, and for failure 0.43 (0.24-0.76), P=0.004. The corresponding values at 3 months were 0.44 (0.31-0.63), P<0.0001 and 0.28 (0.19-0.41), P<0.0001, respectively. It should

be kept in mind at this point, that the reason for stopping treatment was determined by the treating rheumatologist, and that overall stopping rate rather than stopping reason should be emphasized (Cf. "Patients and methods", section on the SSATG register).

By unadjusted Cox regression, higher age, higher HAQ score and no concomitant methotrexate was found to significantly predict drug discontinuation, independent of disease activity or response. However, the HRs were so low, that these predictors were considered to be of minor importance (for details, see paper III).

Table 3. Fulfilment of response criteria and disease states of the RA cohort in paper III. Total N=1789.

	6 weeks		3 months	
	Percent	Valid N	Percent	Valid N
Response criteria				
ACR				
ACR20	56	536	61	1234
ACR50	29	536	38	1234
ACR70	8.5	542	14	1234
EULAR original				
overall	73	499	76	1163
moderate	45	505	42	1163
good	29	499	34	1163
EULAR modified				
overall	69	499	72	1163
moderate	25	505	18	1163
good	44	499	54	1163
SDAI				
overall	75	512	80	1184
minor	37	518	28	1184
major	39	512	52	1184
CDAI				
overall	74	523	79	1195
minor	30	529	22	1195
major	43	523	57	1195
Disease stages				
EULAR original				
remission	18	524	23	1214
low	35	518	38	1214
moderate	47	518	46	1214
high	18	518	17	1214
EULAR modified				
remission	12	524	18	1214
low	46	518	49	1214
moderate	41	518	38	1214
high	13	518	12	1214
SDAI				
remission	4.1	534	8.3	1224
low	35	528	42	1224
moderate	20	528	41	1224
high	18	528	17	1224
CDAI				
remission	4.5	539	8.0	1232
low	36	533	45	1232
moderate	44	533	39	1232
high	20	533	19	1232

Table 4. Hazard ratios (HR) for discontinuation of TNF blockade in RA according to various response criteria and disease states (paper III). Total N=1789.

	6 week				3 month			
	Hazard Ratio (95 % CI)		p-value	Valid N	Hazard Ratio (95 % CI)		p-value	Valid N
Response criteria								
ACR								
ACR20	0.55	(0.43—0.71)	<0.001	526	0.56	(0.47—0.66)	<0.001	1208
ACR50	0.64	(0.48—0.85)	0.002	526	0.55	(0.46—0.66)	<0.001	1208
ACR70	1.06	(0.59—1.90)	0.850	526	0.56	(0.42—0.75)	<0.001	1208
EULAR original								
overall	0.59	(0.44—0.77)	<0.001	493	0.47	(0.39—0.56)	<0.001	1151
moderate	0.95	(0.66—1.34)	0.763	493	0.87	(0.73—1.04)	0.123	1151
good	0.72	(0.53—0.97)	0.032	493	0.58	(0.47—0.71)	<0.001	1151
EULAR modified								
overall	0.63	(0.48—0.83)	0.001	493	0.48	(0.40—0.57)	<0.001	1151
moderate	1.03	(0.69—1.53)	0.895	493	0.96	(0.77—1.19)	0.695	1151
good	0.61	(0.49—0.79)	<0.001	493	0.56	(0.47—0.67)	<0.001	1151
SDAI								
overall	0.62	(0.45—0.81)	0.001	512	0.47	(0.39—0.57)	<0.001	1179
minor	0.80	(0.56—1.15)	0.229	512	1.06	(0.87—1.28)	0.557	1179
major	0.65	(0.50—0.86)	0.002	512	0.58	(0.48—0.69)	<0.001	1179
CDAI								
overall	0.75	(0.57—0.99)	0.044	517	0.53	(0.44—0.65)	<0.001	1181
minor	0.94	(0.64—1.37)	0.742	517	1.20	(0.99—1.47)	0.067	1181
major	0.69	(0.54—0.90)	0.005	517	0.57	(0.48—0.68)	<0.001	1181
Disease states								
EULAR original								
remission	1.03	(0.65—1.62)	0.911	502	0.65	(0.51—0.82)	<0.001	1173
low	0.69	(0.52—0.93)	0.015	502	0.60	(0.49—0.72)	<0.001	1173
moderate	1.04	(0.81—1.34)	0.769	502	0.98	(0.83—1.16)	0.803	1173
high	1.60	(1.15—2.23)	0.005	502	2.18	(1.78—2.69)	<0.001	1173
EULAR modified								
remission	1.33	(0.82—2.16)	0.242	502	0.65	(0.50—0.85)	0.001	1173
low	0.60	(0.45—0.79)	<0.001	502	0.61	(0.50—0.73)	<0.001	1173
moderate	1.24	(0.96—1.60)	0.108	502	1.08	(0.91—1.29)	0.386	1173
high	1.86	(1.28—2.69)	0.001	502	2.15	(1.72—2.69)	<.001	1173
SDAI								
remission	1.85	(0.86—4.02)	0.118	512	0.60	(0.41—0.87)	0.006	1182
low	0.81	(0.62—1.06)	0.125	512	0.67	(0.56—0.81)	<0.001	1182
moderate	0.90	(0.70—1.15)	0.389	512	0.90	(0.76—1.07)	0.224	1182
high	1.73	(1.26—2.39)	0.001	512	2.19	(1.79—2.68)	<0.001	1182
CDAI								
remission	1.82	(0.84—3.945)	0.129	517	0.64	(0.441—0.918)	0.016	1189
low	0.84	(0.64—1.10)	0.192	517	0.70	(0.58—0.84)	<0.001	1189
moderate	0.87	(0.68—1.12)	0.237	517	0.84	(0.70—1.00)	0.048	1189
high	1.62	(1.20—2.19)	0.002	517	2.24	(1.84—2.73)	<0.001	1189

Interestingly, response criteria perform poorly at the individual level in our cohorts as demonstrated in papers I and II, but they seem useful at the group level in predicting response to TNF blockade already at 6 weeks. It could, however, not be generally advised to make decisions on treatment continuation at this early stage in individuals with RA, as many patients will respond later and achieve low activity or remission at 3 months. However, the data seem to indicate, that cessation of treatment in selected non-responders to TNF blockade might be considered earlier than after 3 months. This may become an issue when future guideline revisions are discussed.

Health utility development in anti-TNF treatment of RA, PsA and SpA (paper IV)

This longitudinal study was undertaken in order to explore the EQ-5D utility development during up to 7 years of anti-TNF treatment of RA, PsA and SpA in clinical practice. Other objectives were to investigate trends over time in baseline utilities, the impact of switching between TNF blockers and the feasibility of the EQ-5D instrument across three arthritis diagnoses in our routine care setting.

A total of 3714 anti-TNF courses were followed with EQ-5D utilities at baseline, 2, 6, 12 weeks and every 3-6 months thereafter. First, 2nd and 3rd or more treatment courses were examined separately; as most of the switchers also were included as 1st and sometimes 2nd courses, data should be interpreted cautiously.

Secular trends for baseline utilities at 1st treatment course were weak and not significant for RA, PsA and SpA. By contrast, there are trends towards lower disease activity and disability during the study period (Soderlin and Geborek 2008), but this does not seem to be reflected in utilities.

Utility improvement occurred rapidly, already after 2 weeks in RA and PsA and somewhat later in SpA (Figure 6), and utility levels then remained relatively constant over 5 years in patients remaining on therapy. In the situation under consideration, utility gain thus may be regarded as momentary. This has important implications for AUC and QALY calculations. Of course, the situation may be entirely different with other interventions and diseases; however, the findings illustrate the importance of utility (and other outcomes) measurement early during treatment.

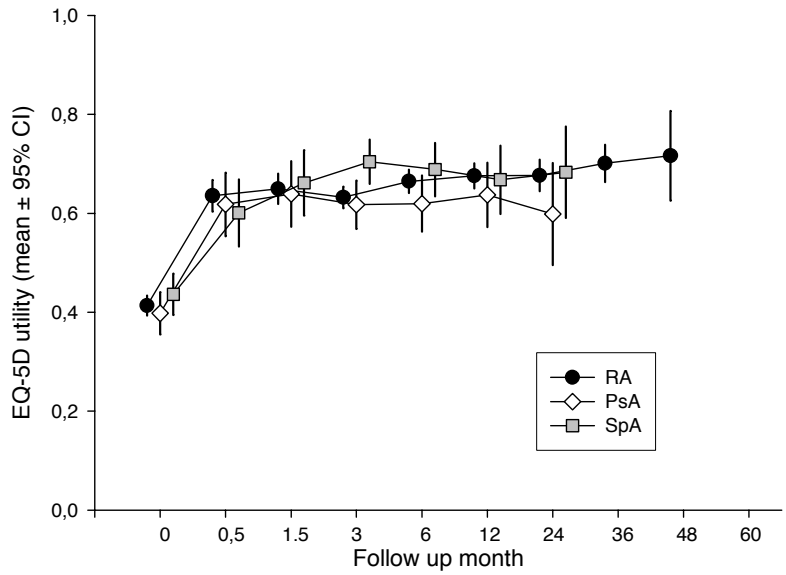
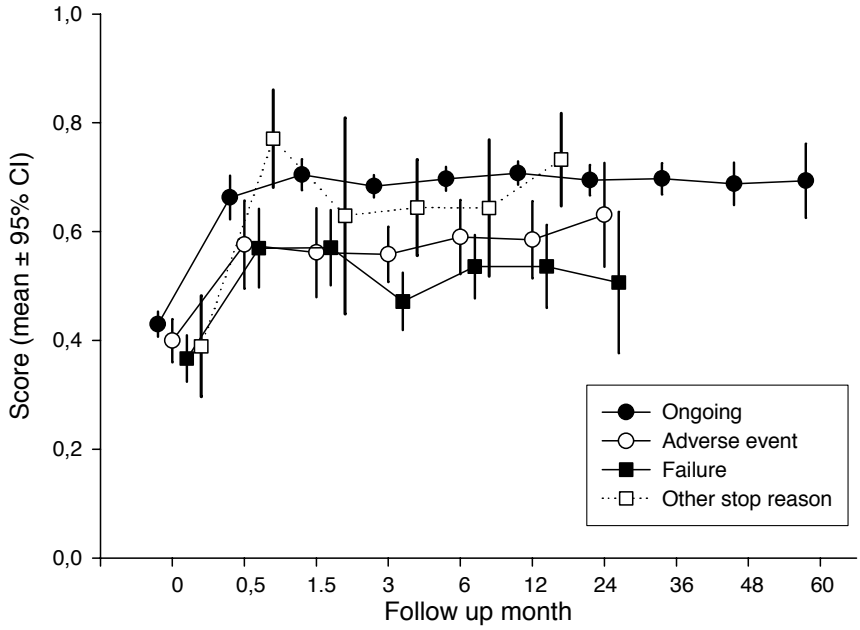


Figure 6. EQ-5D utility development during 1 st course of TNF blockade of RA, PsA and SpA.

Interestingly, treatment order seemed to have little influence on utility gain. First and 2nd treatment courses performed similarly as regards utility development. In RA, where there were fair numbers of switchers, 3rd courses started at a somewhat lower baseline utility level, perhaps due to selection of severe cases, but the absolute gain was of the same magnitude as for 1st and 2nd courses. Patients stopping therapy had lower utility gains irrespective of cause (Figure 7).



Valid N (number of EQ-5D values)										
Ongoing	772	138	204	515	401	382	288	234	140	57
Adverse event	255	41	55	148	79	66	24	13	3	1
Failure	231	47	64	157	97	64	23	10	2	0
Other	63	11	12	39	30	20	11	7	5	1

Figure 7. Utility gain for RA patients with ongoing therapy and stopping because of adverse event, failure or other reason.

Missing data is always a concern in the observational setting. In the current, descriptive study, it was felt to be crucial that data be derived from all available observations in order to make the conclusions more generally applicable to routine care. The number of observations contributing to each point in Figure 7 is thus given. To evaluate possible bias due to patients with missing measurements, we compared utility development in RA patients with complete data at time points 0, 3, 6, 12 and 24 months with all available data (Figure 8). Although utilities at some time points were higher in 1st course patients with complete data, the pattern is similar. In the observational setting, there may be missing data for many reasons. Withdrawal for whatever cause, however, may be one major reason, and we have demonstrated, that courses terminated for failure or adverse events resulted in less utility gain prior to stopping of treatment.

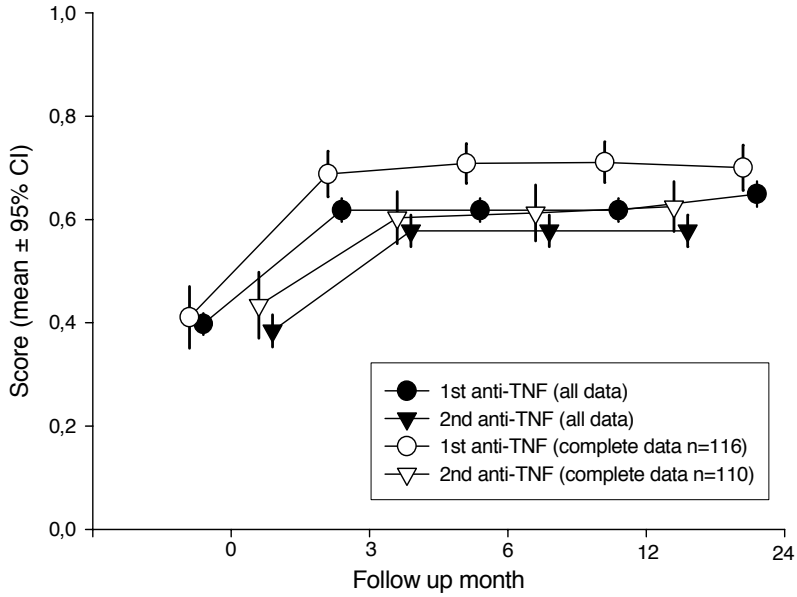


Figure 8. Utility development for 1st and 2nd RA treatment courses. Treatment courses with complete data at time points 0, 3, 6, 12 and 24 months and all courses.

To assess feasibility of the EQ-5D instrument, the proportion of complete HAQ and EQ-5D questionnaires was determined for the 17768 follow-up visits included. Ninety seven to 98% of HAQ and 93-94% of EQ-5D forms were found to be complete irrespective of diagnosis. The marginally lower value for the EQ-5D was probably due to the fact, that all five items must be answered in order to make calculation of a utility value possible, whereas the HAQ has more than one item for some of the domains, making scoring of some incomplete forms possible. Patients found the EQ-5D forms easy and rapid to fill out.

The number needed to treat per QALY gained, NNQ (Paper V)

The NNQ concept was launched in order to provide a simple, easy to understand, utility based, group level outcome measure that would help the non-economist to appreciate, to what extent an intervention is worth while from a HRQoL perspective. The NNQ is defined as the number of patients one has to subject to an intervention in order to gain 1 QALY. A QALY is the time elapsed multiplied by the utility, i.e. the AUC of the utility curve. NNQ is calculated as the inverted value of the AUC of the utility gain plotted against time, usually 1 year.

In the current investigation, EQ-5D utility was measured at baseline, 2, 6, 12 weeks and every 3-6 months thereafter. The QALY (AUC) gain might be estimated in several ways, the simplest being subtracting baseline utility values from mean utility during one year. Alternatively, the trapezium rule or more refined mathematics may be employed. We have been able to demonstrate that utility gain in TNF blockade of inflammatory arthritis occurs almost instantly (paper IV), but in situations with a more gradual onset of improvement, the method of AUC calculation is more critical. When applied to anti-TNF treatment courses, however, the mean utility method and the trapezium rule yielded almost

identical results. In a subgroup of 1st course RA treatments (N=696), the correlation coefficient was found to be 0.99.

First, 2nd and 3rd treatment courses were studied separately. For RA, NNQ was 4.5, 6.4 and 5.2 for 1st, 2nd and 3rd courses, respectively. For PsA and SpA, NNQ was 4.2-4.5 irrespective of treatment order, i.e. one has to treat 4-5 patients in order to gain 1 QALY. Treatment groups with N<50 were not analysed. Thus, NNQ varied relatively little across diagnoses and treatment course orders, with the possible exception of 2nd course RA treatment. The latter finding may be due to selection of therapy resistant cases; however, no such NNQ increase was found in 3rd RA courses or in PsA and SpA, and no firm conclusions could be drawn from this observation. These figures refer to mean utility gain for 1 year for all treatment courses, irrespective of real time spent on treatment (last observation carried forward, LOCF). NNQ was also calculated by multiplying utility gain with actual time on drug for each treatment course (time corrected values), taking shorter courses into account. As expected, this yielded slightly higher NNQ values: 4.7, 6.7 and 5.7 for 1st, 2nd and 3rd RA courses, respectively. Uncorrected NNQ for RA courses <1 year of duration were found to be 10.4, 12.6 and 9.6, respectively, whereas the corresponding corrected values were 16.5, 20.6 and 17.0. This reflects the fewer QALYs gained by patients withdrawing from therapy, conceivably often due to failure or toxicity. On the other hand, NNQ for patients remaining on therapy >1 year was 3.8, 5.4 and 4.4 (time correction not needed). This represents a group of responders.

During the study period 2002-2007, there were no secular trends of utility gains.

Conclusions

- Disease activity states as defined by DAS28 (original and modified cut-points), SDAI and CDAI exhibited moderate or good agreement at the individual level when applied to TNF blockade in day-to-day care of long-standing RA patients. The various indices performed similarly.
- Response criteria (ACR20-50-70, EULAR original and modified, SDAI and CDAI) had poor agreement at the individual level, except at the lowest level of response. Caution should be exercised when applying response criteria as an aid in treatment decisions in routine RA care.
- Response at the ACR20/overall and at the ACR50/good/major level predicted continuation of TNF blockade in RA, for most criteria sets already at 6 weeks. Achieving low disease activity predicted treatment continuation at 3 months. Remaining in high disease activity strongly predicted cessation of treatment at both time points.
- Gain in EQ-5D utilities during anti-TNF treatment of RA, PsA and SpA was demonstrated to occur rapidly, in most cases within 2 weeks, and utility remained relatively stable for 5 years in patients still on therapy. First, 2nd and 3rd treatment courses performed similarly. Patients withdrawing from therapy made less utility gain irrespective of cause. There were no secular trends in baseline utilities during the study period 2002-2008. The EQ-5D was found to be feasible and perform well across 3 arthritis diagnoses.
- The NNQ concept (number needed [to treat] per QALY gained) was developed in order to provide a simple, utility based outcome measure, easy to understand for the non-expert in health economy. NNQ is the number of patients one has to treat in order to gain 1 QALY, and it is calculated as the inverted value of the utility gain AUC for 1 year. In our setting, NNQ was found to be 4-6 for TNF blockade irrespective of diagnostic entity and treatment order. As an outcome measure, NNQ was easy to calculate and understand, and it performed well across 3 different arthritis diagnoses.

Perspectives for the future

Observational studies will continue to provide valuable information on the safety and efficacy of new treatment modalities. This will be especially important in the long term perspective, which is not covered by RCTs. Future research in the field of arthritis treatment may include:

- Long term disease activity, utility and safety data on new biologic treatments, such as B-cell depletion, co-stimulation and IL-6 blockade;
- Further elucidation of the relationship between activity, function and utility in various arthritides;
- Application of the NNQ concept in other diagnostic entities and using other treatment principles;
- Exploration of utility response and utility states;
- Determination of utility MCII and PASS in clinical practice.

Summary in Swedish/Populärvetenskaplig sammanfattning på svenska

Reumatoid artrit (ledgångsreumatism, RA), psoriasisartrit (PsA) och andra s k spondartriter (SpA), t ex ankyloserande spondylit (AS, Bechterews sjukdom), medför betydande bördor för både den som drabbas och för samhället. Ekonomiska förluster i form av sjukfrånvaro och stigande vårdkostnader är ett växande problem. Nya, effektiva behandlingsmetoder, t ex biologiska preparat som TNF-blockerarna adalimumab (Humira), etanercept (Enbrel) och infliximab (Remicade), inger både hopp om förbättrade behandlingsresultat och farhågor om hastigt ökande läkemedelskostnader. Det har därför blivit allt viktigare att på ett tillförlitligt sätt kunna mäta behandlingseffekt. Inget enskilt laboratorie- eller kliniskt mått kan fånga ens de viktigaste aspekterna av sjukdomsaktivitet vid artritsjukdom, och därför har olika sammansatta index som t ex DAS28, SDAI och CDAI utvecklats. Dessa innehåller patientskattade mått på allmän hälsa och smärta, undersökarens fynd av antal svullna och ömma leder och i flertalet fall något laboratoriemått på inflammation. Dessa vägs sedan ihop på olika sätt. Med respons (behandlingssvar) menas skillnaden i sjukdomsaktivitet före behandling och efter viss tid enligt bestämda kriterier, t ex ACR eller EULAR. För att mäta funktion har man utvecklat patientenkäter som HAQ och BASFI. På senare år har intresset ökat för hälsorelaterad livskvalitet, som också kan mätas med olika enkäter som EQ-5D och SF-6D. Förbättring av livskvalitet är ett bredare och ”mjukare” mått på behandlingseffekt, som även täcker andra aspekter av sjukdomsförloppet än inflammatorisk aktivitet och funktion. Dessutom möjliggör mätning av livskvalitet beräkning av ”utility”, ett numeriskt mått mellan 0 och 1, där 0 betecknar död och 1 perfekt hälsa. Sådana kvantitativa data kan sedan användas för beräkning av livskvalitetsjusterade levnadsår, QALY, som i sin tur kan åsättas ett pris som myndigheter eller försäkringssystem är beredda att betala. Detta är grundvalen för hälsoekonomiska beräkningar.

Avhandlingen består av fem delarbeten. Den grundar sig på data från SSATG-registret (South Sweden Arthritis Treatment Group), ett samarbete mellan reumatologiska vårdgivare i södra Sverige, som samlar effekt- och säkerhetsdata om biologiska behandlingar i ett kvalitets- och säkerhetsregister sedan 1999.

I delarbete I och II undersöks överensstämmelsen på individnivå mellan olika aktivitets- och responsmått vid behandling av långvarig RA med TNF-hämmare i rutinsjukvård. Aktivitetsindex kan med hjälp av brytpunkter delas in i aktivitetsstadier, d.v.s. hög, måttlig, låg och ingen sjukdomsaktivitet (remission). Det visade sig, att de undersökta index stämde överens ganska väl, när det gäller att karaktärisera enskilda patienter. Å andra sidan var överensstämmelsen mellan olika responskriterier dålig, utom för den lägsta graden av respons. Man bör alltså vara försiktig med att använda responskriterier som hjälpmedel vid behandlingsbeslut för enskilda patienter. De olika index och deras aktivitetsstadier fungerar bra och är ungefär likvärdiga på individnivå. De är alltså lämpligare i rutinsjukvård, vilket stämmer med att de flesta patienter upplever det vara viktigare att uppnå låg sjukdomsaktivitet (må bra), än att genomgå en viss absolut grad av förbättring (må bättre).

Det är svårt att förutse vilka patienter som kommer att svara på en viss behandling, och c:a 25% av patienter med artritsjukdom uppnår fortfarande inget bra behandlingssvar, trots tillgång på moderna, högeffektiva läkemedel. Egenskaper vid behandlingsstart som ålder, kön, funktion eller olika laboratorieprov är dåliga på att förutsäga behandlingseffekt. I delarbete III har vi därför undersökt, ifall respons eller uppnåendet av ett visst aktivitetsstadium efter 6 veckors eller 3 månaders TNF-blockad kan förutsäga fortsatt behandling vid RA. Statistiska modeller, korrigerade för variabler som man vet påverkar behandlingsutfallet (t ex ålder och samtidig methotrexatebehandling), konstruerades därför för varje aktivitetsstadium och responsnivå enligt olika kriteriesystem. Respons på låg och hög (men

inte måttlig) nivå kunde för de flesta responskriterier förutse fortsatt behandling redan efter 6 veckor. Aktivitetsstadierna gav mindre tydliga resultat vid denna tidpunkt, men att uppnå låg sjukdomsaktivitet efter 3 månader förutsade fortsatt behandling för alla kriterier. Att fortsätta i hög sjukdomsaktivitet förutsade, som väntat, behandlingsavbrott. Trots att responskriterier fungerar dåligt på individnivå i klinisk verksamhet, utgör de således de tydligaste måtten för att förutse fortsatt terapi (på gruppnivå), i flera fall redan efter 6 veckor.

Hälsorelaterad livskvalitet, mätt med EQ-5D, undersöktes under anti-TNF-behandling av patienter med RA, PsA och SpA, som följdes i upp till 7 år i delarbete IV. Man fann att ”utility”, som när det gäller EQ-5D kan variera mellan –0.56 och 1, steg snabbt efter behandlingsstart, i regel redan efter 2 veckor, något långsammare för SpA. För de patienter som fortsatte med behandling, förblev värdet sedan stabilt i upp till 5 år. Vid behandlingsavbrott, oavsett orsak, steg utility mindre. Livskvalitetsförbättring undersöktes separat för patienter som fick sitt första, andra eller tredje anti-TNF-medel. Ordningsföljden för behandlingsomgångarna visade sig ha ringa betydelse för utilityutvecklingen. Andra och tredje behandlingsomgången resulterade alltså i ungefär lika snabb och lika stor livskvalitetsförbättring som första. Under 2002-2008 förelåg inga trender i utility vid behandlingsstart, ett något överraskande fynd, eftersom vi tidigare visat att funktion och sjukdomsaktivitet vid start förbättrats över tid. EQ-5D upplevdes lätthanterlig av patienterna och fungerade väl med tre olika diagnoser.

Hälsoekonomiska beräkningar och resonemang är sällan transparenta och lättförståeliga. Det föreligger därför ett behov av ett livskvalitetsbaserat effektmått, som på ett enkelt sätt bidrar till att tydliggöra, i vilken utsträckning det är meningsfullt att genomföra en viss behandling. För detta ändamål har vi utvecklat NNQ (number needed to treat per QALY gained), definierat som det antal patienter man måste behandla för att vinna ett livskvalitetsjusterat levnadsår (QALY)(delarbete V). NNQ beräknas som det inverterade värdet av antalet vunna QALY under ett år (ytan under kurvan för utility under ett år minus utility vid behandlingsstart). I vårt material av anti-TNF-behandling av RA, PsA och SpA var NNQ=4-6 oavsett diagnos och behandlingens ordningsföljd, d.v.s. det krävdes behandling av 4-6 patienter för att vinna 1 QALY. Patienter med behandlingstider under 1 år hade högre NNQ, en konsekvens av lägre QALY-vinst på grund av behandlingsavbrott till följd av t ex bristande effekt eller biverkan. I princip kan NNQ tillämpas på andra sjukdomsgrupper och behandlingar; i vår studie fungerade det väl på de 3 olika artritdiagnoserna. Man kan betrakta NNQ som ett mycket starkt förenklat hälsoekonomiskt mått, där vinsten består av QALY-ökningen till följd av behandlingen, och kostnaden motsvaras av läkemedelskostnaden.

Standardiserade aktivitetsindex, funktions- och livskvalitetsmått har underlättat tolkningen och jämförelsen av kliniska prövningar och observationsstudier på artritbehandlingens område. Statistiska metoder för att göra resultatförutsägelser och jämförelser har också starkt bidragit till den kliniska reumatologins utveckling. Man måste dock komma ihåg, att de resultat som föreligger gäller för grupper av patienter. Inget index eller resultatmått kan i sig omfatta alla viktiga aspekter av en viss sjukdom hos en enskild patient. För överskådlig tid kommer behandlingsbeslut i det enskilda fallet att ytterst åvila den behandlande läkaren i dialog med sin välinformerade patient. Ingen kan frånhända sig den kunskap som vunnits på gruppnivå, men kliniskt omdöme kommer att förbli en omistlig del i det dagliga omhändertagandet av patienter med artritsjukdom.

Acknowledgements

This thesis is the result of the combined labour of many people; first and foremost *the patients and colleagues contributing data to the SSATG register*. The fact, that it is possible to mention only a few of those who have helped and supported my work, does not mean that I am less grateful to all the rest. Particularly, however, I would like to mention:

Pierre Geborek, my principal supervisor, for his commitment, many brilliant ideas and for many profound discussions, in moments of success and despair alike, and also for being the initiator and manager of the SSATG register;

Tore Saxne, my co-supervisor, for never-failing support in matters big and small, valuable suggestions and total reliability;

Daniel Aletaha, Lennart Jacobsson, Lars Erik Kristensen and *Ingemar Petersson,* my co-authors, for many rewarding exchanges and helpful criticisms;

Elisabeth Lindqvist, my chief of department, and *all my colleagues* at the department of rheumatology in Lund for creating a research friendly working environment and helping and supporting me in many ways;

Jan-Åke Nilsson, for gently but efficiently guiding me in the mysteries of statistics;

Eleonor Danielsson, Elna Haglund, Lena Jonsson, Eva-Karin Kristoffersson, and *all of the nursing staff* of the department of rheumatology in Lund, for their devotion both to the care of patients and the gathering of questionnaires and blood samples;

Lotta Larsson, for hard work with the SSATG database and keeping a vigilant eye on the providers of data, including myself, in order to fill the gaps and keep high standards of quality;

Ingrid Mattsson Geborek, for expert secretarial help with manuscripts and professional lay-out of posters and this thesis;

Ingrid Jönsson, for keeping me clear of the perils of university administrative matters;

Lena Billquist Santacruz, for keeping track of my being on and off clinical work and many other practical matters;

Roger Hesselstrand , Meliha Kapetanovic and Christina Ståhl Hallengren, for encouragement and a generally positive attitude;

Bengt Månsson, for many funny and serious chats on medicine, research and life in general;

Gunnar Sturfelt, for (among other things) helping with my very first publication in rheumatology;

Frank Wollheim, for unrelenting enthusiasm and many nice exchanges on the topic of music;

Lars Rydgren, for general support and for sharing my interest in organ music;

My deceased colleagues *Margareta Simonsson* and *Eva Fex,* for encouraging me to commence with this work and support in the early phase;

Ido Leden, for first introducing me to clinical rheumatology;

My former *colleagues at the department of internal medicine, Simrishamn hospital*, for training me to work clinically, and teaching the art and craft of internal medicine;

My family.

References

Akkoc N, van der Linden S, Khan A. Ankylosing spondylitis and symptom-modifying vs disease-modifying therapy. Best Pract Res Clin Rheumatol **2006**; 20(3): 539-57.

Aletaha D, Smolen JS. The Simplified Disease Activity Index (SDAI) and the Clinical Disease Activity Index (CDAI): A review of their usefulness and validity in rheumatoid arthritis. Clin Exp Rheumatol **2005a**; 23(Suppl 39): S100-S108.

Aletaha D, Nell V, Stamm T, Uffmann M, Pflugbeil S, Machold K, Smolen JS. Acute phase reactants add little to composite disease activity indices for rheumatoid arthritis: validation of a clinical activity score. Arthritis Res Ther **2005b**; 7: R796-R806.

Aletaha D, Smolen JS, Ward MM. Measuring function in rheumatoid arthritis: Identifying reversible and irreversible components. Arthritis Rheum **2006**; 54(9): 2784-92.

Aletaha D, Funovits J, Ward MM, Smolen JS, Kvien TK. Perception of improvement in patients with rheumatoid arthritis varies with disease activity levels at baseline. Arthritis Rheum **2009**; 61(3): 313-20.

Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, Healey LA, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum **1988**; 31(3): 315-24.

Askling J and Bongartz T. Malignancy and biologic therapy in rheumatoid arthritis. Curr Opin Rheumatol **2008**; 20(3): 334-9.

Askling J, Baecklund E, Granath F, Geborek P, Fored M, Backlin C, Bertilsson L, et. al. Anti-tumour necrosis factor therapy in rheumatoid arthritis and risk of malignant lymphomas: relative risks and time trends in the Swedish Biologics Register. Ann Rheum Dis **2009**; 68(5): 648-53.

Banal F, Dougados M, Combescure C, Gossec L. Sensitivity and specificity of the American College of Rheumatology 1987 criteria for the diagnosis of rheumatoid arthritis according to disease duration: a systematic literature review and meta-analysis. Ann Rheum Dis **2009**; 68(7): 1184-91.

Bijlsma JW and Jacobs JW. Methotrexate: Still the anchor drug in RA treatment. Joint Bone Spine **2009**; 76(5): 452-4.

Blackmore MG, Gladman DD, Husted J, Long JA, Farewell VT. Measuring health status in psoriatic arthritis: the Health Assessment Questionnaire and its modification. J Rheumatol **1995**; 22(5): 886-93.

Boers M, Verhoeven AC, Markusse HM, van de Laar MA, Westhovens, van Denderen RJC, van Zeben D, et al. Randomised comparison of combined step-down prednisolone, methotrexate and sulphasalazine with sulphasalazine alone in early rheumatoid arthritis. Lancet **1997**; 350(9074): 309-18.

Braun J and Rau R. An update on methotrexate. Curr Opin Rheumatol **2009**; 21(3): 216-23.

Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J Health Econ **2002**; 21(2): 271-92.

Chen J and Liu C. Sulfasalazine for ankylosing spondylitis. Cochrane Database Syst Rev (2) **2005**: CD004800.

Clark JI and Weiner LM. Biologic treatment of human cancer. Curr Probl Cancer **1995**; 19(4): 185-262.

Dougados M, van der Linden S, Juhlin R, Huitfeldt B, Amor B, Calin A, Cats A, et al. The European Spondylarthropathy Study Group preliminary criteria for the classification of spondylarthropathy. Arthritis Rheum **1991**; 34(10): 1218-27.

Dougados M. It's good to feel better but it's better to feel good. J Rheumatol **2005**; 32(1): 1-2.

Ekdahl C, Eberhardt K, Andersson SI, Svensson B. Assessing disability in patients with rheumatoid arthritis. Use of a Swedish version of the Stanford Health Assessment Questionnaire. Scand J Rheumatol **1988**; 17(4): 263-71.

Felson DT, Anderson JJ, Boers M, Bombardier C, Furst D, Goldsmith C, Katz LM, Lightfoot Jr R, et al. American College of Rheumatology. Preliminary definition of improvement in rheumatoid arthritis. Arthritis Rheum **1995**: 38(6): 727-35.

Fitzgerald O and Dougados M. Psoriatic arthritis: one or more diseases? Best Pract Res Clin Rheumatol **2006**; 20(3): 435-50.

Fransen J, Antoni C, Mease PJ, Uter W, Kavanaugh A, Kalden JR, Van Riel PL, et al. Performance of response criteria for assessing peripheral arthritis in patients with psoriatic arthritis: analysis of data from randomised controlled trials of two tumour necrosis factor inhibitors. Ann Rheum Dis **2006**; 65(10): 1373-8.

Fries JF, Spitz P, Kraines RG, Holman HR, et al. Measurement of patient outcome in arthritis. Arthritis Rheum **1980**; 23(2): 137-45.

Gabriel SE and Michaud K. Epidemiological studies in incidence, prevalence, mortality, and comorbidity of the rheumatic diseases. Arthritis Res Ther **2009**; 11(3): 229.

Garrett S, Jenkinson T, Kennedy LG, Whitelock H, Gaisford P, Calin A. A new approach to defining disease status in ankylosing spondylitis: the Bath Ankylosing Spondylitis Disease Activity Index. J Rheumatol **1994**; 21(12): 2286-91.

Geborek P and Saxne T. Clinical protocol for monitoring of targeted therapies in rheumatoid arthritis. Rheumatology (Oxford) **2000**; 39(10): 1159-61.

Geborek P, Crnkic M, Petersson IF, Saxne T. Etanercept, infliximab, and leflunomide in established rheumatoid arthritis: clinical experience using a structured follow up programme in southern Sweden. Ann Rheum Dis **2002**; 61(9): 793-8.

Geborek P, Nitelius E, Noltorp S, Petri H, Jacobsson L, Larsson L, Saxne T, et al. Population based studies of biological antirheumatic drug use in southern Sweden: comparison with pharmaceutical sales. Ann Rheum Dis **2005**; 64(12): 1805-7.

Gladman DD. Psoriatic arthritis. Rheum Dis Clin North Am **1998**; 24(4): 829-44, x.

Gladman DD. Mortality in psoriatic arthritis. Clin Exp Rheumatol **2008**; 26(5 Suppl 51): S62-5.

Goekoop-Ruiterman YP, de Vries-Bouwstra JK, Allaart CF, van Zeben D, Kerstens PJ, Hazes JM, Zwinderman AH, 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(11): 3381-90.

Gottlieb A, Korman NJ, Gordon KB, Feldman SR, Lebwohl M, Koo JY, Van Voorhees AS, et al. Guidelines of care for the management of psoriasis and psoriatic arthritis: Section 2. Psoriatic arthritis: overview and guidelines of care for treatment with an emphasis on the biologics. J Am Acad Dermatol **2008**; 58(5): 851-64.

Grigor C, Capell H, Stirling A, McMahon AD, Lock P, Vallance R, Kincaid W, et al. Effect of a treatment strategy of tight control for rheumatoid arthritis (the TICORA study): a single-blind randomised controlled trial. Lancet **2004**; 364 (9430): 263-9.

Haywood KL, Garratt AM, Dawes PT. Patient-assessed health in ankylosing spondylitis: a structured review. Rheumatology (Oxford) **2005**; 44(5): 577-86.

Helliwell PS and Taylor WJ. Classification and diagnostic criteria for psoriatic arthritis. Ann Rheum Dis **2005**; 64 Suppl 2: ii3-8.

Hyrich KL, Watson KD, Silman AJ, Symmons DP, et al. Predictors of response to anti-TNF-alpha therapy among patients with rheumatoid arthritis: results from the British Society for Rheumatology Biologics Register. Rheumatology (Oxford) **2006**; 45(12): 1558-65.

Kapetanovic MC, Larsson L, Truedsson L, Sturfelt G, Saxne T, Geborek P. Predictors of infusion reactions during infliximab treatment in patients with arthritis. Arthritis Res Ther **2006**; 8(4): R131.

Karlsson JA, Kristensen LE, Kapetanovic MC, Gulfe A, Saxne T, Geborek P. Treatment response to a second or third TNF-inhibitor in RA: results from the South Swedish Arthritis Treatment Group Register. Rheumatology (Oxford) **2008**; 47(4): 507-13.

Kavanaugh A. Therapy: Guidelines in rheumatology: quo vadis? Nat Rev Rheumatol **2009**; 5(8): 423-4.

Keystone EC, Kavanaugh AF, Sharp JT, Tannenbaum H, Hua Y, Teoh LS, Fischkoff SA, et al. Radiographic, clinical, and functional outcomes of treatment with adalimumab (a human anti-tumor necrosis factor monoclonal antibody) in patients with active rheumatoid arthritis receiving concomitant methotrexate therapy: a randomized, placebo-controlled, 52-week trial. Arthritis Rheum **2004**; 50(5): 1400-11.

Kirwan JR. The effect of glucocorticoids on joint destruction in rheumatoid arthritis. The Arthritis and Rheumatism Council Low-Dose Glucocorticoid Study Group. N Engl J Med **1995**; 333(3): 142-6.

Klareskog L, Stolt P, Lundberg K, Kallberg H, Bengtsson C, Grunewald J, Ronnelid J, et al. A new model for an etiology of rheumatoid arthritis: smoking may trigger HLA-DR (shared epitope)-restricted immune reactions to autoantigens modified by citrullination. Arthritis Rheum **2006**; 54(1): 38-46.

Klareskog L, Catrina AI, Paget S. Rheumatoid arthritis. Lancet **2009**; 373(9664): 659-72.

Kristensen LE, Saxne T, Nilsson JA, Geborek P. Impact of concomitant DMARD therapy on adherence to treatment with etanercept and infliximab in rheumatoid arthritis. Results from a six-year observational study in southern Sweden. Arthritis Res Ther **2006**; 8(6): R174.

Kristensen LE, Kapetanovic MC, Gulfe A, Soderlin M, Saxne T, Geborek P. Predictors of response to anti-TNF therapy according to ACR and EULAR criteria in patients with established RA: results from the South Swedish Arthritis Treatment Group Register. Rheumatology (Oxford) **2008**; 47(4): 495-9.

Landewe R, van der Heijde D, Klareskog L, van Vollenhoven R, Fatenejad S. Disconnect between inflammation and joint destruction after treatment with etanercept plus methotrexate: results from the trial of etanercept and methotrexate with radiographic and patient outcomes. Arthritis Rheum **2006**; 54(10): 3119-25.

Lawrence RC, Hochberg MC, Kelsey JL, McDuffie FC, Medsger Jr TA, Felts WR, Shulman LE. Estimates of the prevalence of selected arthritic and musculoskeletal diseases in the United States. J Rheumatol **1989**; 16(4): 427-41.

Lawrence RC, Helmick CG, Arnett FC, Deyo RA, Felson DT, Giannini EH, Heyse SP, et al. Estimates of the prevalence of arthritis and selected musculoskeletal disorders in the United States. Arthritis Rheum **1998**; 41(5): 778-99.

Lindroth Y, Brattstrom M, Bellman I, Ekestaf G, Olofsson Y, Strombeck B, Stenshed B, et al. A problem-based education program for patients with rheumatoid arthritis: evaluation after three and twelve months. Arthritis Care Res **1997**; 10(5): 325-32.

Lukas C, Landewe R, Sieper J, Dougados M, Davis J, Braun J, van der Linden S, et al. Development of an ASAS-endorsed disease activity score (ASDAS) in patients with ankylosing spondylitis. Ann Rheum Dis **2009**; 68(1): 18-24.

Luqmani R, Hennell S, Estrach C, Basher D, Birrell F, Bosworth A, Burke F, et al. British Society for Rheumatology and British Health Professionals in Rheumatology guideline for the management of rheumatoid arthritis (after the first 2 years). Rheumatology (Oxford) **2009**; 48(4): 436-9.

Maini RN, Breedveld FC, Kalden JR, Smolen JS, Davis D, Macfarlane JD, Antoni C, et al. Therapeutic efficacy of multiple intravenous infusions of anti-tumor necrosis factor alpha monoclonal antibody combined with low-dose weekly methotrexate in rheumatoid arthritis. Arthritis Rheum **1998**; 41(9): 1552-63.

Maksymowych WP, Gladman DD, Rahman P, Boonen A, Bykerk V, Choquette D, Dimond S, et al. The Canadian Rheumatology Association/ Spondyloarthritis Research Consortium of Canada treatment recommendations for the management of spondyloarthritis: a national multidisciplinary stakeholder project. J Rheumatol **2007**; 34(11): 2273-84.

Moll JM and Wright V. Psoriatic arthritis. Semin Arthritis Rheum **1973**; 3(1): 55-78.

Moreland LW, Cohen SB, Baumgartner SW, Tindall EA, Bulpitt K, Martin R, Weinblatt M, et al. Long-term safety and efficacy of etanercept in patients with rheumatoid arthritis. J Rheumatol **2001**; 28(6): 1238-44.

Nan L, Johnson JA, Shaw JW, Coons SJ. A comparison of EQ-5D index scores derived from the US and UK population-based scoring functions. Med Decis Making 2007;27(3):321-6.

Noyes K, Dick AW, Holloway RG. The implications of using US-specific EQ-5D preference weights for cost-effectiveness evaluation. Med Decis Making **2007**; 27(3): 327-34.

O'Dell JR. Combinations of conventional disease-modifying antirheumatic drugs. Rheum Dis Clin North Am **2001**; 27(2): 415-26, x.

Pavelka K, Jarosova K, Suchy D, Senolt L, Chroust K, Dusek L, Vencovsky J. Increasing the infliximab dose in rheumatoid arthritis patients: a randomised, double blind study failed to confirm its efficacy. Ann Rheum Dis **2009**; 68(8): 1285-9.

Pham T, Fautrel B, Dernis E, Goupille P, Guillemin F, Le Loet X, Ravaud P, et al. Recommendations of the French Society for Rheumatology regarding TNFalpha antagonist therapy in patients with ankylosing spondylitis or psoriatic arthritis: 2007 update. Joint Bone Spine **2007**; 74(6): 638-46.

Ranzolin A, Brenol JC, Bredemeier M, Guarienti J, Rizzatti M, Feldman D, Xavier RM. Association of concomitant fibromyalgia with worse disease activity score in 28 joints, health assessment questionnaire, and short form 36 scores in patients with rheumatoid arthritis. Arthritis Rheum **2009**; 61(6): 794-800.

Rasanen P, Roine E, Sintonen H, Semberg-Konttinen V, Ryynanen OP, Roine R. Use of quality-adjusted life years for the estimation of effectiveness of health care: A systematic literature review. Int J Technol Assess Health Care **2006**; 22(2): 235-41.

Ravindran V, Scott DL, Choy EH. A systematic review and meta-analysis of efficacy and toxicity of disease modifying anti-rheumatic drugs and biological agents for psoriatic arthritis. Ann Rheum Dis **2008**; 67(6): 855-9.

Rudwaleit M, Landewe R, van der Heijde D, Listing J, Brandt J, Braun J, Burgos-Vargas R, et al. The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part I): classification of paper patients by expert opinion including uncertainty appraisal. Ann Rheum Dis **2009**; 68(6): 770-6.

Saag KG, Teng GG, Patkar NM, Anuntiyo J, Finney C, Curtis JR, Paulus HE, et al. American College of Rheumatology 2008 recommendations for the use of nonbiologic and biologic disease-modifying antirheumatic drugs in rheumatoid arthritis. Arthritis Rheum **2008**; 59(6): 762-84.

Sajid MS, Tonsi A, Baig MK. Health-related quality of life measurement. Int J Health Care Qual Assur **2008**; 21(4): 365-73.

Saleem B, Brown AK, Keen H, Nizam S, Freeston J, Karim Z, Quinn M, et al. Disease remission state in patients treated with the combination of tumor necrosis factor blockade and methotrexate or with disease-modifying antirheumatic drugs: A clinical and imaging comparative study. Arthritis Rheum **2009**; 60(7): 1915-22.

Schellekens GA, Visser H, de Jong BA, van den Hoogen FH, Hazes JM, Breedveld FC, van Venrooij WJ. The diagnostic properties of rheumatoid arthritis antibodies recognizing a cyclic citrullinated peptide. Arthritis Rheum **2000**; 43(1): 155-63.

Schlosstein L, Terasaki PI, Bluestone R, Pearson CM. High association of an HL-A antigen, W27, with ankylosing spondylitis. N Engl J Med **1973**; 288(14): 704-6.

Sieper J, Rudwaleit M, Khan MA, Braun J. Concepts and epidemiology of spondyloarthritis. Best Pract Res Clin Rheumatol **2006**; 20(3): 401-17.

Simonsson M, Bergman S, Jacobsson LT, Petersson IF, Svensson B. The prevalence of rheumatoid arthritis in Sweden. Scand J Rheumatol **1999**; 28(6): 340-3.

Smolen JS, Han C, van der Heijde DM, Emery P, Bathon JM, Keystone E, Maini RN, et al. Radiographic changes in rheumatoid arthritis patients attaining different disease activity states with methotrexate monotherapy and infliximab plus methotrexate: the impacts of remission and tumour necrosis factor blockade. Ann Rheum Dis **2009a**; 68(6): 823-7.

Smolen JS, Kay J, Doyle MK, Landewe R, Matteson EL, Wollenhaupt J, Gaylis N, et al. Golimumab in patients with active rheumatoid arthritis after treatment with tumour necrosis factor alpha inhibitors (GO-AFTER study): a multicentre, randomised, double-blind, placebo-controlled, phase III trial. Lancet **2009b**; 374(9685): 210-21.

Soderlin MK and Geborek P. Changing pattern in the prescription of biological treatment in rheumatoid arthritis. A 7-year follow-up of 1839 patients in southern Sweden. Ann Rheum Dis **2008**; 67(1): 37-42.

Sokka T, Kautiainen H, Pincus T, Toloza S, Castelar Pinheiro GD, Lazovskis J, Hetland ML et al. Disparities in rheumatoid arthritis disease activity according to gross domestic product in 25 countries in the QUEST-RA database. Ann Rheum Dis **2009a**; 68(11): 1666-72.

Sokka T and Makinen H. Drug management of early rheumatoid arthritis - 2008. Best Pract Res Clin Rheumatol **2009b**; 23(1): 93-102.

Steinbrocker O, Traeger CH, Batterman RC. Therapeutic criteria in rheumatoid arthritis. J Am Med Assoc **1949**; 140(8): 659-62.

Suarez-Almazor ME, Belseck E, Shea B, Wells G, Tugwell P. Sulfasalazine for rheumatoid arthritis. Cochrane Database Syst Rev (2) **2000**: CD000958.

Svensson B, Boonen A, Albertsson K, van der Heijde D, Keller C, Hafstrom I. Low-dose prednisolone in addition to the initial disease-modifying antirheumatic drug in patients with early active rheumatoid arthritis reduces joint destruction and increases the remission rate: a two-year randomized trial. Arthritis Rheum **2005**; 52(11): 3360-70.

Taylor W, Gladman D, Helliwell P, Marchesoni A, Mease P, Mielants H. Classification criteria for psoriatic arthritis: development of new criteria from a large international study. Arthritis Rheum **2006**; 54(8): 2665-73.

Tubach F, Dougados M, Falissard B, Baron G, Logeart I, Ravaud P. Feeling good rather than feeling better matters more to patients. Arthritis Rheum **2006**; 55(4): 526-30.

Turesson C, Jacobsson LT, Matteson EL. Cardiovascular co-morbidity in rheumatic diseases. Vasc Health Risk Manag **2008**; 4(3): 605-14.

Waaler E. [Rheumatoid factor in the 1930s and today]. Nord Med **1970**; 83(44): 1385-9.

van der Heijde D, Klareskog L, Rodriguez-Valverde V, Codreanu C, Bolosiu H, Melo-Gomes J, Tornero-Molina J, et al. Comparison of etanercept and methotrexate, alone and combined, in the treatment of rheumatoid arthritis: two-year clinical and radiographic results from the TEMPO study, a double-blind, randomized trial. Arthritis Rheum **2006**; 54(4): 1063-74.

van der Linden S, Valkenburg HA, Cats A. Evaluation of diagnostic criteria for ankylosing spondylitis. A proposal for modification of the New York criteria. Arthritis Rheum **1984**; 27(4): 361-8.

van der Linden S and van der Heijde D. Ankylosing spondylitis. Clinical features. Rheum Dis Clin North Am **1998**; 24(4): 663-76, vii.

van Gestel AM, Laan RF, Haagsma CJ, van de Putte LB, van Riel PL. Oral steroids as bridge therapy in rheumatoid arthritis patients starting with parenteral gold. A randomized double-blind placebo-controlled trial. Br J Rheumatol **1995**; 34(4): 347-51.

van Gestel AM, Prevoo ML, van 't Hof MA, van Rijswijk MH, van de Putte LB, van Riel PL. Development and validation of the European League Against Rheumatism response criteria for rheumatoid arthritis. Comparison with the preliminary American College of Rheumatology and the World Health Organization/International League Against Rheumatism Criteria. Arthritis Rheum **1996**; 39(1): 34-40.

van Vollenhoven RF. Rheumatologists, take heart! We may be doing something right. Arthritis Res Ther **2008**; 10(2): 105.

Ware JE, Jr and Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care **1992**; 30(6): 473-83.

Weitoft T. Intra-articular glucocorticoid treatment. Thesis. **2005**. Dept of medical sciences. Uppsala

Verhagen AP, Bierma-Zeinstra SM, Cardoso JR, de Bie RA, Boers M, de Vet HC. Balneotherapy for rheumatoid arthritis. Cochrane Database Syst Rev (4) **2003**: CD000518.

WHO. The World Health Organization Quality of Life assessment (WHOQOL): position paper from the World Health Organization. Soc Sci Med **1995**; 41(10): 1403-9.

Wilske KR. Inverting the therapeutic pyramid: observations and recommendations on new directions in rheumatoid arthritis therapy based on the author's experience. Semin Arthritis Rheum **1993**; 23(2 Suppl 1): 11-8.

Wolfe F, Mitchell DM, Sibley JT, Fries JF, Bloch DA, Williams CA, Spitz PW, et al. The mortality of rheumatoid arthritis. Arthritis Rheum **1994**; 37(4): 481-94.

Wolfe F, Michaud K, Pincus T. A composite disease activity scale for clinical practice, observational studies, and clinical trials: the patient activity scale (PAS/PAS-II). J Rheumatol **2005**; 32(12): 2410-5.

Wolfe F, Rasker JJ, et al. Minimal disease activity, remission, and the long-term outcomes of rheumatoid arthritis. Arthritis Rheum **2007**; 57(6): 935-42.

www.DAS-score.nl. Retrieved Dec 25, **2008**.

www.NICE.org.uk/CG79. Retrieved Oct 6, **2009**.

www.svenskreumatologi.se. Retrieved Oct 6, **2009**.

Zink A, Listing J, Kary S, Ramlau P, Stoyanova-Scholz M, Babinsky K, von Hinueber U, et al. Treatment continuation in patients receiving biological agents or conventional DMARD therapy. Ann Rheum Dis **2005**; 64(9): 1274-9.

Zink A, Askling J, Dixon WG, Klareskog L, Silman AJ, Symmons DP. European biologicals registers: methodology, selected results and perspectives. Ann Rheum Dis **2009**; 68(8): 1240-6.

Zochling J and Braun J. Mortality in ankylosing spondylitis. Clin Exp Rheumatol **2008**; 26(5 Suppl 51): S80-4.

Appendix 1

Basline data, to be filled out by the treating physician. Includes drug to be started, concomitant and earlier treatment, diagnosis (checklists on reverse), and indication for biologic treatment.

 SSATG

Basdata för DMARD- biologisk behandling

(Instruktioner se baksidan)

Personnr:

Namn:

- Inklusionsdatum:.....
- Pat ansvarig läkare.....
- ***Behandlings regim:***

  Preparat:.....

  Kombination (med annat DMARD) ❑ ⟶ Tillägg till tidigare DMARD Ja ❑ Nej ❑

  Monoterapi (ej annat DMARD) ❑
- ***Mål / Subjektiv destruktionsgradering:***

  Mindre destruktiv sjukdom ❑ (HAQ reducerat med 50%)

  Kraftigt destruktiv sjd ❑ (HAQ reducerat med 20%)
- ***Behandlingsdiagnos och debutår:***

  RA ❑ År.....ICD10nr.....

  Annan ❑ Vilken?.....År.....ICD10nr.....
- ***Indikation:***
  (endast 1 alternativ)

  Svikt / Intolerans på DMARD
  (avser någonsin, inklusive biologiskbeh)
  - på DMARD i komb.beh oavsett antal ❑
  - på >3 DMARD prep ❑
  - på 3 DMARD prep ❑
  - på 2 DMARD prep ❑
  - på 1 DMARD prep ❑

  Inflam. aktiv *ej tidigare* DMARD beh ❑

  Pats önskemål (ovanstående stämmer ej) ❑
- ***Patientens egna uppgifter om längd .....cm och vikt.....kg***
- ***Tidigare/pågående systemiskt kortison:*** ❑ **Nej** ❑ **Ja**
- ***Tidigare/pågående DMARD beh :*** ❑ **Nej** ❑ **Ja,** kryssa för vilka nedan

❑ Antimalaria	❑ Leukeran
❑ Arava	❑ Methotrexate
❑ Azatioprin	❑ Penicillamin
❑ Ciklosporin	❑ Podofyllotoxin prep
❑ Cyklofosfamid	❑ Remicade
❑ Enbrel	❑ Salazopyrin
❑ Entocort	Övrigt
❑ Guld i.m	❑ .....
❑ Guld p.o	❑ .....
❑ Humira	❑ .....
❑ Kineret	❑ .....

OBS!!! Fortsättning på baksidan

• *För nydebuterad RA*	*Nej*	*Ja*	
Morgonstelhet	❑	❑	(≥1 timme)
Artrit i ≥ 3 ledområden	❑	❑	(PIP, MCP, handled, armbåge, knä, fotled, MTP – hö eller vä)
Artrit i hand	❑	❑	(PIP, MCP, handled)
Symmetrisk artrit	❑	❑	(symmetri mellan ledområden – hö och vä sida)
Reumatiska noduli	❑	❑	
Reumatoid faktor	❑	❑	
Röntgenförändringar	❑	❑	(usurer el. otvetydig periartikulär osteopeni i handskelett)

• *För spondylartriter**	*Nej*	*Ja*	*Vet ej - irrelevant*
Säker ankyloserande spondylit**	❑	❑	
Tidigare/aktuell klinisk spondylit	❑	❑	
Perifer artritsjukdom (dist. om axlar och höfter)	❑	❑	
Radiologisk sacroiliit	❑	❑	❑
Radiologisk spondylit	❑	❑	❑
Tidigare/aktuell klinisk daktylit	❑	❑	❑
Tidigare/aktuell klinisk entesit	❑	❑	❑
Tidigare/aktuell nagel psoriasis	❑	❑	❑
Tidigare/aktuell PPP (palmoplantar pustulos)	❑	❑	❑
IBD (inflammatory bowel disease)	❑	❑	

→ Debut år..... Mb Crohn ❑ Ulcerös Colit ❑
Annan .....

***Kriterier för spondylartropati enligt ESSG (European Spondylarthropathy Study Group)**
Inflammatorisk ryggsmärta eller perifer synovit (asymmetrisk, huvudsakligen i nedre extremiteterna) och en av följande:

- Familjäritet
- Psoriasis
- Inflammatorisk tarmsjukdom
- Uretrit, cervicit eller akut diarré inom en månad för artrit debut
- Glutealsmärta alternerande på vänster och höger sida
- Entesopati
- Sacroiliit (röntgenologisk)
-

****Modifierade NewYork-kriterier (1984) för ankyloserande spondylit (AS)**
1. Ländryggsvärk under minst tre månader som förbättras av rörelse men ej av vila.
2. Begränsad rörlighet i ländryggen sagitalt (i sidled) och frontalt (framåt och bakåt).
3. Minskad bröstkorgsexpansion (ålders- och könsjusterat).
4. Bilateral sakroiliit grad II-IV eller unilateral sakroiliit grad III eller IV.

Fotnot: Definitivt AS föreligger vid kriterium 4 och minst en av punkt 1-3.

==========

Information / Instruktion för formuläret.

Behandlingsregim skall alltid anges vid preparat. Exempelvis Enbrel sc, Remicade iv, Arava po, Mtx po, Mtx sc m fl. Man skall alltid kryssa i rutan monoterapi eller kombinationsterapi. Om den nya huvudterapin ges som tillägg till tidigare DMARD-terapi kryssas detta för i rutan efter kombinationsterapi.
I framtiden kan exempelvis en behandlingsregim vara att man samtidigt startar en kombinationsterapi med flera DMARD/biologiska behandlingar. Denna gäller då som behandlingsregim och innebär att denna regim kan vara monoterapi om annan DMARD-terapi inte fortsättes. Det är behandlingsregimen som utvärderas, inte de enskilda preparaten i detta sammanhang.
Mål/destruktionsgradering skall mest ses som en stratifiering i gruppen kraftigt destruerade (HAQ-reduktion 20%) och mindre kraftigt destruerade (HAQ-reduktion 50%). Detta är naturligtvis en subjektiv värdering men ger i alla fall en uppfattning om hur destruerad patienten bedöms vara av läkaren. *Uppgiften är obligatorisk och ingen data- inmatning är möjlig utan denna.*
Behandlingsdiagnos och debutår här anges diagnos debut år samt ICD10nr.
Indikation endast ett alternativ skall väljas. Om svikt på kombinationer av DMARD inklusive biologiskbeh skall alltid detta väljas.
Tidigare Kortison & DMARD-behandling längst ner på sidan skall ses som en strukturerad genomgång av patientens tidigare DMARD-behandlingar. OBS att Entocort i detta sammanhang pga sin speciella profil räknas som DMARD och således inte skall anges som steroidpreparat.
Det kan med fördel fyllas i innan indikationsrutan kryssas i. Obs att även kortvariga försök med DMARD skall anges även om den avbrutits av mer irrelevant orsak. Alla testade DMARD-preparat skall anges.
För nydebuterad RA frågorna besvaras med kryss i tillämplig ruta
För spondylartriter frågorna besvaras med kryss i tillämplig ruta

Appendix 2

Visit form, to be filled out by physician. Includes treatment taken up to and after the visit, joint counts, evaluator's global assessment, and information on axial symptoms, dactylitis, enthesitis, psoriatic nail involvement and eye involvement in SpA.

SSATG

Besöksblankett – DMARD- biologisk behandling

Evalueringsdatum:.....

Läkare vid besöket:.....

Personnr:

Namn:

Licensnr:

• *Aktuell sjukdomsaktivitet alla diagnoser*

Allmän hälsa	Smärta	HAQ	SR	CRP	Leder svullna 28-leder	Leder ömma 28-leder

Läkar-/eval bedömning → Ingen ❑ Låg ❑ Måttlig ❑ Hög ❑ Maximal ❑

• *Medicinering*

Huvudterapi →	Genomsnittlig intagen dos (per dag eller vecka) under de senaste 2 veckorna, för längre dosintervall ex. *Remicade, MabThera mfl, anges senste dos och datum före aktuell evaluering*	Ordinerad dos mg och frekvens ex. 200mg var 8:e vecka, 40mg v a v, 25mg 2ggr/v, 50mg 1ggr/v, 100mg/d
.....	.....mg/.....	.....mg/.....
DMARD Nej ❑ Ja ❑ →		
.....	.....mg/.....	.....mg/.....
.....	.....mg/.....	.....mg/.....
.....	.....mg/.....	.....mg/.....
System steroider	Nej ❑ Vb ❑ Ja ❑ .....mg	Nej ❑ Vb ❑ Ja ❑ .....mg
NSAID	Nej ❑ Vb ❑ Ja ❑	Nej ❑ Vb ❑ Ja ❑
Analgetika	Nej ❑ Vb ❑ Ja ❑	Nej ❑ Vb ❑ Ja ❑

• *För Spondartriter inklusive Psoriasisartrit dessutom*

Leder svullna 66-leder	Leder ömma 68-leder	PASI-Score (ev.)

Om pågående/tidigare klinisk spondylit skall pat fylla i föjande formulär

	Nej	Ja
BASDAI	❑	❑
BASFI	❑	❑
BASG1+G2	❑	❑

Klinisk pågående:	Nej	Ja	Ej relevant	
Spondylit	❑	❑	❑	
Daktylit	❑	❑	❑	
Entesit	❑	❑	❑	
Nagel psoriasis	❑	❑	❑	
PPP (palmo-plantar-pustulos)	❑	❑	❑	
			Hö	Vä
Infl. reumatisk ögonsjd (uveit, sclerit, keratit)	❑	❑ →	❑	❑

Antal akuta uveiter sedan föregående evaluering .....

Registreringsanvisningar vid DMARD- biologisk behandling

Denna Besöksblankett skall fyllas i vid insättningen och sedan vid 3, 6 och 12 månader, samt därefter var 6:e månad, eller så snart behandlingen avbryts oavsett orsak. Vissa preparat har tätare evalueringsintervall.

Patientblanketten fylls i av patienten på båda sidorna inför läkarbesöket, frågorna är självinstruerande. Läkaren räknar ut värdet för HAQ på sidan två, genom att kryss i den första kolumnen ger poäng =0, den andra poäng = 1, den tredje till femte kolumnen poäng = 2 och den sista kolumnen = 3 poäng. Varje grupp med frågor poängsätts på strecket till höger efter den högsta poängen på någon fråga i gruppen. Om patienten t.ex. behöver hjälp med att tvätta håret (=2) räknas detta ut till höger, även om patienten inte har några svårigheter att klä sig (=0). Poängen från varje frågegrupp till höger (0-3) summeras, och summan blir då mellan 0 och 24, varefter ett genomsnitt räknas ut enligt lathunden längst ned, t.ex. ger summan 11 ett HAQ på 1.38.

Effektregistreringen på denna besöksblanketts framsida ***för alla diagnoser*** fylls i enligt:

LVnr/Personnummer = en unik patientkod i form av diarienummer på licensen/patientens personnr.
Patientnamn = Efternamn, förnamn
Datum = Undersökningsdatum
Läk Signatur = Vid evalueringstillfället behandlande läkare
Allmän hälsa = VAS värdet mätt i mm från vänster på skalan på patientblanketten, dvs "Helt bra blir 0 och "så dålig som tänkas kan" blir 100
Smärta = VAS värdet mätt på samma sätt från vänster på skalan
HAQ = värdet uträknat enligt ovanstående instruktion
SR = värdet i mm från laboratoriet
CRP = värdet från lokalt laboratorium.
Leder svullna = antalet svullna leder enligt 28-ledsindex, i vilket inkluderas följande leder:
10 PIP-leder, 10 MCP-leder, 2 handleder, 2 armbågsleder, 2 axelleder och 2 knäleder.
Leder ömma = antalet ömma leder enligt 28-ledsindex
Läkarbedömning med övriga variabler kända (de föregående rutorna på raden) gör läkaren en global bedömning av sjukdomsaktiviteten - kryssa för **ett** alternativ.

Medicinering – intagen och ordinerad vid besöket

Biologiskt läkemedel/DMARD anges med preparatnamn. Genomsnittlig veckodos senaste 2 veckorna anges i mg/dag eller vecka, för längre dosintervall än 2 veckor, exempelvis vid infusionsbehandling anges senaste dos och datum ex. 200mg 020405. Ordinerad dos anges i mg och frekvens ex. 200 mg var 8:e vecka, 40mg v a v, 25 mg 2ggr/vecka, 50mg1 ggr/vecka, 100mg/d.
Övrig medicinering besvaras med kryssrutor för intagen resp ordinerad varvid kortisondosen anges vid regelbunden medicinering, som dygnsdos i mg ekvivalent med Prednisolon.

För Spondartrit inklusive psoriasis

Bör även 66/68-ledsindex göras (OBS! Både 28-ledsindex och 66/68 ledsindex skall anges).
Eventuellt PASI-Score (http://www.medscape.com/viewarticle/507681_6)
Frågorna vid kryssrutorna besvaras med Ja, Nej eller Ej relevant
Om pågående/tidigare klinisk spondylit bör pat fylla i BASDAI, BASFI och eventuellt BASG1+G2

Appendix 3

Security form, to be filled out by patient at each visit. Includes data on hospitalisation, surgery, side effects, intercurrent infection and comments by patient. Physician's comments on reverse, including classification of adverse event, judgement of causal relation to biologic drug, stopping of therapy (date and cause: side effect, failure or other).

Lund 060313/LL

Säkerhetsformulär för patienter med antireumatisk behandling.

Personnr:

Namn:

Datum .....

Var vänlig och besvara nedanstående frågor *före* läkarbesöket genom att kryssa i Ja eller Nej, vid kryss i Ja rutan besvaras även följdfrågan. Lämnas sedan till läkaren/sjuksköterskan som Du träffar idag.

Har Du sedan föregående besök här på Reumatologen :

- Varit inlagd på sjukhus? Nej ❑ Ja ❑ Varför?.....
  Vilket sjukhus och vilken avdelning?.....
  .....
- Blivit opererad? Nej ❑ Ja ❑ Vilken operation? .....
  .....
- Upplevt någon biverkning av Din behandling? Nej ❑ Ja ❑ Vad?.....
  .....
  .....
- Haft någon infektion såsom:

Halsont / snuva	Nej ❑	Ja ❑
Tandinfektion	Nej ❑	Ja ❑
Bältros	Nej ❑	Ja ❑
Hudinfektion	Nej ❑	Ja ❑
Lunginflammation	Nej ❑	Ja ❑
Maginfluenza	Nej ❑	Ja ❑
Urinvägsinfektion	Nej ❑	Ja ❑
Övrig infektion	Nej ❑	Ja ❑ Vilken?.....

.....

Skriv ev kommentarer på baksidan under rubriken patient kommentar

Patient kommentar

Läkarens kommentar

Datum	Händelse	Beskrivning av symptom, organengagemang, tidsrelation till huvudterapi, utredning samt ev slutenvård.

- **Allvarlighetsgrad (*= rapportskyldighet)**

❑ Mild
❑ Måttlig
❑ Allvarlig *
❑ Livshotande *
❑ Dödlig *
❑ Övrig rapportskyldighet *

- **Förlopp**

❑ Tillfrisknat utan men
❑ Tillfrisknat med men
❑ Okänt
❑ Ännu ej tillfrisknat
❑ Avliden

- **Relation till huvudterapi**

❑ Sannolik
❑ Möjlig
❑ Osannolik
❑ Ej bedömbart

- **Åtgärd huvudterapi**

❑ Ingen
❑ Tillfälligt utsatt från-till eller antal veckor....
.....
❑ Definitivt utsatt

Skall skickas till Läkemedelsverket: Ja ❑ Avvakta ❑ Ej till LMV ❑

- **Avslutad terapi övrig orsak**

Huvudterapi: .....Datum.....

❑ Uppnådd effekt avtagit ❑ Terapisvikt ❑ Övrigt .....

.....

Appendix 4

Patient form to be filled out at each visit. Includes pain and general health VAS, descriptive part of EQ-5D and PASS question (satisfaction with present state, yes/no); HAQ on reverse.

Reumatologiska kliniken, Lund, 080430/LL

SSATG

PATIENTBLANKETT

Datum:.....

Ange din vikt: .....Kg

Personnr:

Namn:

Svara på frågorna här nedan och på baksidan.
Var noga med att läsa varje fråga och instruktionen.
Tack för din hjälp.

Sätt ett streck tvärs över linjen under frågan hur du har upplevt den senaste veckan:

Hur mycket smärta har du haft på grund av din ledsjukdom?

0 1 2 3 4 5 6 7 8 9 10

Ingen smärta alls _____ Värsta tänkbara smärta

Hur har du känt dig, allmänt sett, med tanke på din ledsjukdom?

0 1 2 3 4 5 6 7 8 9 10

Helt bra _____ Så dålig som tänkas kan

Markera genom att kryssa i en ruta, vilket påstående som bäst beskriver ditt hälsotillstånd idag.

Rörlighet

Jag går utan svårigheter	❑	Kryssa
Jag kan gå men med viss svårighet	❑	endast
Jag är sängliggande	❑	i en ruta

Hygien

Jag behöver ingen hjälp med min dagliga hygien, mat eller påklädning	❑	Kryssa
Jag har vissa problem att tvätta eller klä mig själv	❑	endast
Jag kan inte tvätta eller klä mig själv	❑	i en ruta

Huvudsakliga aktiviteter (tex arbete, studier, hushållssysslor, familje- och fritidsaktiviteter)

Jag klarar av min huvudsakliga sysselsättning	❑	Kryssa
Jag har vissa problem med att klara av min huvudsakliga sysselsättning	❑	endast
Jag klarar inte min huvudsakliga sysselsättning	❑	i en ruta

Smärtor/besvär

Jag har varken smärtor eller besvär	❑	Kryssa
Jag har måttliga smärtor eller besvär	❑	endast
Jag har svåra smärtor eller besvär	❑	i en ruta

Rädsla/nedstämdhet

Jag är inte orolig eller nedstämd	❑	Kryssa
Jag är orolig och nedstämd i viss utsträckning	❑	endast
Jag är i högsta grad orolig eller nedstämd	❑	i en ruta

Jämfört med mitt allmänna hälsotillstånd de senaste tolv månaderna är mitt hälsotillstånd i dag:

Bättre	❑	Kryssa
Oförändrat	❑	endast
Sämre	❑	i en ruta

Är ditt nuvarande tillstånd tillfredsställande vad gäller din allmänna funktionsnivå och de smärtor du har nu?

Ja ❑ Nej ❑

Funktionsfrågeformulär
SRF 1996:1

Datum _____

Sätt ett kryss i den ruta som bäst beskriver Din situation under den senaste veckan	Utan svårighet	Med viss svårighet	Mycket svårt	Använder hjälpmedel	Hjälp av annan person	Kan inte alls	
Kan Du tvätta håret?	❑	❑	❑	❑	❑	❑	
Kan Du klä på Dig, inklusive knyta skoband och knäppa knappar?.....	❑	❑	❑	❑	❑	❑	___
Kan Du resa Dig från en stol som saknar armstöd?.....	❑	❑	❑	❑	❑	❑	
Kan Du komma i och ur sängen?.....	❑	❑	❑	❑	❑	❑	___
Kan Du skära kött?	❑	❑	❑	❑	❑	❑	
Kan Du laga Din egen mat?	❑	❑	❑	❑	❑	❑	
Kan Du lyfta ett fullt glas till munnen?.	❑	❑	❑	❑	❑	❑	___
Kan Du gå ned för fem trappsteg?.....	❑	❑	❑	❑	❑	❑	
Kan Du gå utomhus på plan mark?.....	❑	❑	❑	❑	❑	❑	___
Kan Du bada i badkar?	❑	❑	❑	❑	❑	❑	
Kan Du sätta Dig på och resa Dig från en toalettstol?	❑	❑	❑	❑	❑	❑	
Kan Du tvätta och torka Dig överallt?	❑	❑	❑	❑	❑	❑	___
Kan Du ta ned ett 2kg paket med t.ex. socker från en hylla i huvudhöjd?.....	❑	❑	❑	❑	❑	❑	
Kan Du böja Dig ned och ta upp kläder från golvet?	❑	❑	❑	❑	❑	❑	___
Kan Du öppna bildörrar?	❑	❑	❑	❑	❑	❑	
Kan Du öppna burkar med skruvlock, som varit öppnade förut?.....	❑	❑	❑	❑	❑	❑	
Kan Du vrida på en vanlig vattenkran?	❑	❑	❑	❑	❑	❑	___
Kan Du dammsuga?	❑	❑	❑	❑	❑	❑	
Kan Du klara Dina inköp till hushållet?	❑	❑	❑	❑	❑	❑	
Kan Du komma i och ur en bil?.....	❑	❑	❑	❑	❑	❑	___

1/0,13 **2**/0,25 **3**/0,38 **4**/0,5 **5**/0,63 **6**/0,75 **7**/0,88 **8**/1,0 **9**/1,13 **10**/1,25 **11**/1,38 **12**/1,5 **13**/1,63
14/1,75 **15**/1,88 **16**/2,0 **17**/2,13 **18**/2,25 **19**/2,38 **20**/2,5 **21**/2,63 **22**/2,75 **23**/2,88 **24**/3,0

I

EXTENDED REPORT

Response criteria for rheumatoid arthritis in clinical practice: how useful are they?

A Gülfe, P Geborek, T Saxne

Ann Rheum Dis 2005;**64**:1186–1189. doi: 10.1136/ard.2004.027649

See end of article for authors' affiliations

Correspondence to: Dr Anders Gülfe, Department of Rheumatology, Lund University Hospital, SE-221 85 Lund, Sweden; andersgulfe@hotmail.com

Accepted 30 January 2005
Published Online first 10 March 2005

Objective: To compare the performance of the American College of Rheumatology (ACR), European League Against Rheumatism (EULAR), and simple disease activity index (SDAI) response criteria for rheumatoid arthritis at the individual level in an observational cohort.
Methods: 184 outpatients were followed using a structured protocol. For each patient, the responses according to ACR 20% and 50%, EULAR moderate and good, and SDAI minor and major responses were calculated. For comparison, improvements in health assessment questionnaire (HAQ) score of 0.22 and 0.5 were calculated. The numbers of individuals fulfilling the criteria at each level were compared, and the numbers fulfilling any two sets of response criteria calculated. The EULAR ''moderate'' and ''good'' responses were grouped together as ''overall,'' and SDAI ''minor'' and ''major'' were merged into SDAI ''overall''.
Results: All 94 ACR 20 responders were found in the EULAR and SDAI ''overall'' response groups, and 118 of 124 SDAI ''overall'' responders were found in the EULAR ''overall'' group. In contrast, of 53 ACR 50 responders, only 34 were found in the EULAR ''good'' or SDAI ''major'' group. Among the 56 patients in the EULAR ''good'' response group, only 26 met the SDAI ''major'' response. Improvement in HAQ score performed similarly to the other response criteria sets at the group levels.
Conclusions: For individual patients, agreement is good at the level of ACR 20 response, when EULAR overall, SDAI overall, or HAQ 0.22 criteria are applied. Agreement between ACR 50, EULAR good, SDAI major, and HAQ 0.5 response is poor. This should be considered when response criteria are used for clinical decisions.

Rheumatoid arthritis is a chronic, disabling disease affecting about 0.5% of the population.[1] Pharmacological treatment tends to be of long duration and may be complex. Response is often suboptimal, and toxic side effects are not uncommon. No single measure of disease activity or changes in activity (that is, the difference in disease activity between two observations) has proven sufficient, and a variety of composite indices have thus been developed. The utility of such standardised response criteria, for example the ACR (American College of Rheumatology) 20–50–70% response[2] and the European League Against Rheumatism (EULAR) response criteria[3] are well established for use in clinical trials, where the proportion of patients responding constitutes a measure of efficacy compared to placebo or a standard treatment such as methotrexate. This practice has greatly facilitated the evaluation of novel treatments. The disease activity score (DAS) and its variants,[4 5] and the simple disease activity index (SDAI),[6] are intended for routine clinical use. The components of the various response criteria sets are shown in table 1.

Various response criteria sets have been validated against each other in randomised, controlled trials (RCTs) of antirheumatic treatment regimens.[7] In general, the degree of agreement between different response criteria sets is fair in RCTs. The problem of different responsiveness at the individual patient level is, however, seldom addressed.

In a previous communication we described a clinical protocol for monitoring treatment in patients with rheumatoid arthritis.[8] The protocol is suitable for monitoring patients seen in routine clinical practice and can be used to estimate the efficacy and tolerability of different treatment regimens in spite of possible confounding by indication. The individual patient's reaction to treatment according to sets of response criteria is easily determined using this protocol. We have recently reported the ACR responses for etanercept, infliximab, and leflunomide in patients completing the first year of treatment.[9]

Our aim in the present study was to compare the performance of the ACR, EULAR, and SDAI response criteria sets in individual patients, in an observational study of patients with long standing established rheumatoid arthritis treated with tumour necrosis factor α (TNFα) blockers. Improvement in a patient administered instrument, the health assessment questionnaire (HAQ), was used for comparison.

METHODS

Patients attending the Department of Rheumatology, Lund University Hospital, and who had started treatment with TNFα blockers were entered consecutively into a database.

Requirements for inclusion in the study were a diagnosis of rheumatoid arthritis according to the ACR 1987 diagnostic criteria,[10] and treatment failure on at least two disease modifying antirheumatic drugs (DMARDs) including methotrexate. The patients had to be included in the database between 1 January 1999 and 31 December 2001, and a complete dataset at baseline and at three months had to be available. To investigate whether the response pattern changed with longer treatment time, similar analyses were also carried out for patients with a complete dataset at six months.

The protocol comprises the following variables: the 28 joint swollen and tender joint count, patient's global visual

Abbreviations: ACR, American College of Rheumatology; DAS, disease activity score; EULAR, European League Against Rheumatism; HAQ, health assessment questionnaire; SDAI, simple disease activity index

Table 1 Components of the various response criteria sets

Criteria set	Tender joint count	Swollen joint count	Patient global VAS	Patient pain VAS	Evaluator's global	HAQ	ESR	CRP
ACR	+	+	+/−	+/−	+/−	+/−	+/−	+/−
EULAR (DAS)	+	+	+	−	−	−	+	−
SDAI	+	+	+	−	+	−	−	+

+, required; −, not required; in the ACR response criteria, any three of the variables marked "+/−" are required. For details about the response criteria, see references 1–5.
ACR, American College of Rheumatology; CRP, C reactive protein; DAS, disease activity score; ESR, erythrocyte sedimentation rate; EULAR, European League Against Rheumatism; SDAI, simple disease activity index; VAS, visual analogue scale.

analogue scale (VAS) (a 10 cm non-anchored horizontal line[11]), patient's pain VAS, the health assessment questionnaire (HAQ),[12 13] and the evaluator's global assessment of disease activity (five degrees: inactive, low, moderate, high, or maximal), erythrocyte sedimentation rate (ESR) according to Westergren, and C reactive protein.

These variables were used to calculate fulfilment of the following response criteria: ACR 20% and 50%; EULAR non-response, moderate response, and good response; SDAI minor and major responses; and improvement in HAQ score of 0.22 (HAQ 0.22) and 0.5 (HAQ 0.5). The reason for using these HAQ levels of improvement are that 0.22 has been shown to be a level of improvement perceived beneficial by the patient,[14] and 0.5 has been used in health economics models.[15] For the purpose of the present study, EULAR moderate and good responders and SDAI minor and major were grouped together as "overall" responders. The numbers of individuals fulfilling the respective response criteria at each level were compared and the agreement between individual patients fulfilling two sets of response criteria was calculated for each possible pair of response criteria at the actual level.

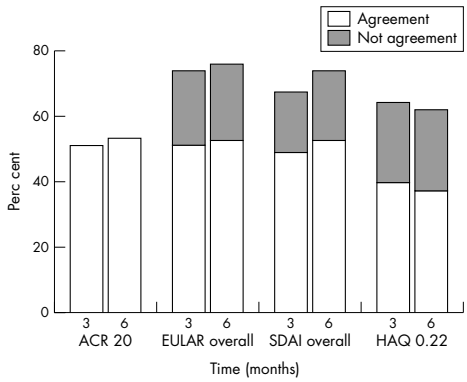


Figure 1 Proportion of patients fulfilling less strict response criteria sets at three and six months. The agreement between the different criteria sets using ACR 20% as reference is demonstrated.

RESULTS

During the period 184 rheumatoid patients fulfilled the requirements for evaluation in the study. The characteristics of the patients at baseline and for some of the variables at three and six months are shown in table 2.

For ACR 20, EULAR overall, SDAI overall, and HAQ 0.22 response, the proportion of responders was 51%, 74%, 67%, and 64%, respectively. All 94 ACR 20 patients were found in the EULAR overall and the SDAI overall groups, while the HAQ 0.22 showed agreement in 73 of these patients. The absolute majority of the SDAI overall (118/124) are found in the EULAR overall group, which comprises 136 patients. For HAQ improvement 0.22 the agreement with EULAR overall is 94/118 and for SDAI overall 90/118 (table 3, fig 1).

For ACR 50, EULAR good, SDAI major, and HAQ 0.5 response, the response rates were 29%, 30%, 30%, and 29%, respectively. However, at the individual level only 34 of the 53

Table 2 Characteristics of the cohort (n = 184) at inclusion and at three and six months

	Baseline	3 Months	6 Months
Numbers	184	184	150
Male/female	46/138		
Disease duration (years)	12 (0 to 55)		
Age at inclusion (years)	56 (20 to 84)		
Previous number of DMARDs	3 (2 to 5)		
Steroid dose (mg/week)	35 (17 to 57)	35 (0 to 53)*	18 (0 to 35)*
ESR (mm/h)	31.5 (20 to 54)	20 (12 to 34)*	20 (10 to 38)*
C reactive protein (mg/litre)	21 (10 to 42)	9 (1 to 22)*	9 (1 to 23)*
VAS pain (mm)	64 (47 to 78)	34 (16 to 55)*	29 (15 to 55)*
VAS global (mm)	66.5 (49 to 81)	34 (15 to 55)*	29 (15 to 57)*
Physician's global assessment	2 (2 to 3)	1 (1 to 2)*	1 (1 to 2)*
28 tender joint count	7.5 (3 to 13)	2 (0 to 5.5)*	2 (0 to 6)*
28 swollen joint count	9 (5 to 12)	3 (1 to 6)*	3 (0 to 6)*
HAQ	1.4 (1 to 1.9)	1.1 (0.6 to 1.6)*	1.1 (0.6 to 1.6)*
DAS 28	5.6 (4.7 to 6.4)	3.8 (2.8 to 4.8)*	3.7 (2.7 to 4.8)*
SDAI	31 (24 to 41)	14 (9 to 24)*	13 (7 to 22)*

Values for median and (25th to 75th centile). Physician's global assessment is recorded at a 5 point Likert scale.
*$p<0.001$ *v* baseline.
DAS 28; 28 joint disease activity score; DMARD, disease modifying antirheumatic drug; ESR, erythrocyte sedimentation rate; HAQ, health assessment questionnaire; RF, rheumatoid factor; SDAI; simple disease activity index; VAS, visual analogue scale.

Table 3 Agreement of response criteria fulfilment in individual patients

		Agreement (n)						
	Total responders	ACR 20	EULAR overall	SDAI overall	HAQ 0.22	ACR 50	EULAR good	SDAI major
ACR 20	94							
EULAR overall	136	94						
SDAI overall	124	90	118					
HAQ 0.22	118	73	94	87				
ACR 50	53	53	53	51	44			
EULAR good	56	46	56	50	42	34		
SDAI major	56	49	56	56	38	34	26	
HAQ 0.5	55	30	48	47	55	31	28	31

For details, see text.
ACR 20, American College of Rheumatology criteria, 20% response; EULAR, European League Against Rheumatism; HAQ 0.5, improvement by 0.5 in health assessment questionnaire; SDAI; simple disease activity index.

ACR 50 responders fulfilled EULAR good or SDAI major responses or both. The EULAR good response agreement with SDAI major response was only 26/56. HAQ 0.5 agreement was of the same magnitude (28/53 *v* EULAR good response, 34/53 *v* SDAI major response (table 3, fig 2)).

At six months, data from 150 patients were available (table 2). Agreement was similar (examples shown in figs 1 and 2).

DISCUSSION

Evaluation of treatment effects in rheumatoid arthritis has received more attention with the introduction of novel therapies, notably the TNFα blockers, which are very expensive and for which the long time effects are unknown. Standardised measures of efficacy should be reliable and simple to use in everyday practice. The criteria sets studied contain much the same variables, but they are weighted or handled somewhat differently. The history and philosophy behind the criteria sets are also different. It is therefore interesting that the agreement between the three criteria sets at the individual patient level was fair at the lower levels of response. More patients respond on the EULAR and SDAI scales than on ACR 20, but the agreement was very good, indicating that, in this observational cohort, patients responded similarly on all three criteria sets.

HAQ, a measure of function, and HAQ improvement have been employed for testing construct validity of the EULAR and SDAI criteria, as it is not included in these.[3 6] It contributes only little to the ACR response criteria set. The fixed levels of 0.22 and 0.5 were in some aspects arbitrarily chosen, but they have been used in previous studies. One reason for using HAQ improvement is the concept that patient self report questionnaires are sufficient to evaluate the efficacy of rheumatoid arthritis treatment.[16] Indeed, at the levels chosen, HAQ improvement did not perform very differently from the other criteria sets tested at the group level in our study.

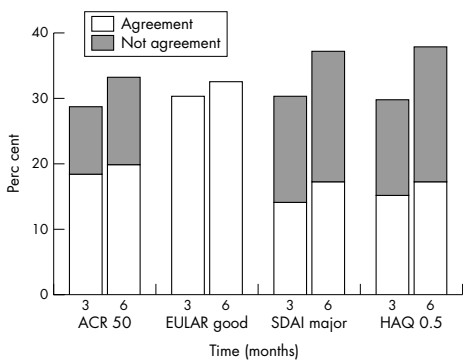


Figure 2 The proportion of patients fulfilling more strict response criteria sets at three and six months. Agreement using EULAR good response is given as reference.

Results were different when comparing the ACR 50 with the EULAR good and the SDAI major responses, respectively. At the group level, they performed similarly—that is, the same numbers of patients tended to respond, irrespective of the criteria sets applied. When individual responses were analysed, however, agreement was poor. At the higher level of response, EULAR good and SDAI major showed agreement in only 34 of 56 patients. If used as a basis for treatment decisions in the individual patient, the choice of criteria would have a major impact. The clinical importance of this may be limited, given the fact that many patients do not respond at this level in routine care, and that the lower degree of response will often be considered sufficient to continue with a particular treatment. However, as treatment modes become more effective, goals of therapy will change, and aiming at remission or near remission will be increasingly realistic. Thus the verification of response at higher levels will become more important.

Standardised response criteria and activity scores in rheumatoid arthritis have proven their utility in clinical trials, when groups of patients are analysed statistically and biological variation tends to be levelled out. They also perform well in observational studies to estimate the response at group level, and in this context they can be indicative of the efficacy of various treatments in clinical practice. In this non-randomised, observational cohort of long standing rheumatoid patients treated with TNFα blockers, the various criteria sets appeared to perform differently at the individual level at the higher degree of response. It may thus be wise to consider response criteria fulfilment in individuals with some scepticism and not to rely heavily on them in clinical practice, but to look upon them as one of several aids for treatment decisions. Absolute measures of disease activity, such as DAS or SDAI levels, are probably better for treatment decisions in the daily care of individual patients. Clinical judgement remains crucial in the management of rheumatoid patients, but currently available response criteria, although not perfect, may be included in the evaluation of treatment with antirheumatic drugs to facilitate monitoring of treatment response.

ACKNOWLEDGEMENTS

This study was supported by grants from Österlund and Kock Foundations, King Gustav V 80 year fund, and Reumatikerförbundet.

Authors' affiliations
A Gülfe, P Geborek, T Saxne, Department of Rheumatology, Lund University Hospital, Lund, Sweden

REFERENCES

1 **Simonsson M**, Bergman S, Jacobsson LT, Petersson IF, Svensson B. The prevalence of rheumatoid arthritis in Sweden. *Scand J Rheumatol* 1999;**28**:340–3.
2 **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.
3 **van Gestel AM**, Prevoo ML, van't Hof MA, van Rijswijk MH, van de Putte LB, van Riel PL. Development and validation of the European League Against Rheumatism response criteria for rheumatoid arthritis. Comparison with the preliminary American College of Rheumatology and the World Health Organization/International League Against Rheumatism Criteria. *Arthritis Rheum* 1996;**39**:34–40.
4 **van der Heijde DM**, van't Hof MA, van Riel PL, Theunisse LA, Lubberts EW, van Leeuwen MA, *et al*. Judging disease activity in clinical practice in rheumatoid arthritis: first step in the development of a disease activity score. *Ann Rheum Dis* 1990;**49**:916–20.
5 **Prevoo ML**, van't Hof MA, Kuper HH, van Leeuwen MA, van de Putte LB, van Riel PL. Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. *Arthritis Rheum* 1995;**38**:44–8.
6 **Smolen JS**, Breedveld FC, Schiff MH, Kalden JR, Emery P, Eberl G, *et al*. A simplified disease activity index for rheumatoid arthritis for use in clinical practice. *Rheumatology (Oxford)* 2003;**42**:244–57.
7 **Verhoeven AC**, Boers M, van Der LS. Responsiveness of the core set, response criteria, and utilities in early rheumatoid arthritis. *Ann Rheum Dis* 2000;**59**:966–74.
8 **Geborek P**, Saxne T. Clinical protocol for monitoring of targeted therapies in rheumatoid arthritis. *Rheumatology (Oxford)* 2000;**39**:1159–61.
9 **Geborek P**, Crnkic M, Petersson IF, Saxne T, South Swedish Arthritis Treatment Group. Etanercept, infliximab, and leflunomide in established rheumatoid arthritis: clinical experience using a structured follow up programme in southern Sweden. *Ann Rheum Dis* 2002;**61**:793–8.
10 **Arnett FC**, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, *et al*. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 1988;**31**:315–24.
11 **Ferraz MB**, Quaresma MR, Aquino LR, Atra E, Tugwell P, Goldsmith CH. Reliability of pain scales in the assessment of literate and illiterate patients with rheumatoid arthritis. *J Rheumatol* 1990;**17**:1022–4.
12 **Ekdahl C**, Eberhardt K, Andersson SI, Svensson B. Assessing disability in patients with rheumatoid arthritis. Use of a Swedish version of the Stanford Health Assessment Questionnaire. *Scand J Rheumatol* 1988;**17**:263–71.
13 **Fries JF**, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. *Arthritis Rheum* 1980;**23**:137–45.
14 **Wells GA**, Tugwell P, Kraag GR, Baker PR, Groh J, Redelmeier DA. Minimum important difference between patients with rheumatoid arthritis: the patient's perspective. *J Rheumatol* 1993;**20**:557–60.
15 **Kobelt G**, Jonsson L, Young A, Eberhardt K. The cost-effectiveness of infliximab (Remicade) in the treatment of rheumatoid arthritis in Sweden and the United Kingdom based on the ATTRACT study. *Rheumatology (Oxford)* 2003;**42**:326–35.
16 **Pincus T**, Strand V, Koch G, Amara I, Crawford B, Wolfe F, *et al*. An index of the three core data set patient questionnaire measures distinguishes efficacy of active treatment from that of placebo as effectively as the American College of Rheumatology 20% response criteria (ACR20) or the Disease Activity Score (DAS) in a rheumatoid arthritis clinical trial. *Arthritis Rheum* 2003;**48**:625–30.

II

BMC Musculoskeletal Disorders

Research article

Disease activity level, remission and response in established rheumatoid arthritis: Performance of various criteria sets in an observational cohort, treated with anti-TNF agents

Anders Gülfe*[1], Daniel Aletaha[2], Tore Saxne[1] and Pierre Geborek[1]

Address: [1]Dept of Rheumatology, Lund University Hospital, Lund, Sweden and [2]Medical University of Vienna, Vienna, Austria

Email: Anders Gülfe* - anders.gulfe@med.lu.se; Daniel Aletaha - daniel.aletaha@meduniwien.ac.at; Tore Saxne - tore.saxne@med.lu.se; Pierre Geborek - pierre.geborek@med.lu.se

* Corresponding author

Published: 23 April 2009

BMC Musculoskeletal Disorders 2009, **10**:41 doi:10.1186/1471-2474-10-41

Received: 16 October 2008
Accepted: 23 April 2009

This article is available from: http://www.biomedcentral.com/1471-2474/10/41

Abstract

Background: Most composite indices of disease activity and response criteria in RA have been validated and compared in clinical trials rather than routine care. We therefore wanted to compare the performance of the DAS28, SDAI and CDAI activity indices, their activity states, their response criteria, and also compare with the ACR response criteria in an observational clinical setting.

Methods: Agreement between the criteria sets was investigated using κ statistics in a non-randomized cohort of 1789 RA patients from southern Sweden, starting their first course of anti-TNF-treatment. Mean disease duration was 12 years. Completer analysis was used.

Results: Agreement between high, moderate and low activity states was moderate or substantial, with κ = 0.5 or better for all criteria. Agreement between SDAI and CDAI disease states was > 90% in these categories with κ > 0.8. DAS28 original and modified cut point remission had good agreement (κ = 0.91). Agreement between responses was substantial at the overall/ACR20 level (about 95%, κ = 0.7 or better) for all criteria. By contrast, agreement was poor between moderate and high level responses.

Conclusion: Disease activity states according to the various indices perform similarly and show substantial agreement at all levels except remission. Agreement between SDAI and CDAI states is excellent. Response criteria, applied at the individual patient level, are hard to interpret and show poor agreement, except at the lowest level of response. Thus, they should not be applied uncritically in clinical practice.

Background

Indices of disease activity in RA, such as the Disease Activity Score in 28 Joints (DAS28) [1], the Simple Disease Activity Index (SDAI) [2] and the Clinical Disease Activity Index (CDAI) [3] and their respective cut-off levels for low disease activity (LDA) and remission (no activity) are tools that can be used in routine care. However, they have been validated mainly in clinical trials, where patients are meeting rigorous inclusion criteria and not always reflect the "real world" situation [4].

Response to treatment in rheumatoid arthritis (RA), as opposed to disease activity, denotes the improvement between two time points due to some intervention. In the

trial setting, where the aim is to compare one treatment to another (standard) or none at all (placebo), response is the outcome measure of choice. The efficacy of a treatment is expressed as the proportion of a patient group that meets a certain response criterion. Indeed, disease activity indices, their LDA and remission criteria are not recommended as primary end points in trials due to low sensitivity to change[5]. On the other hand, response criteria, such as the DAS based EULAR moderate or good [6], SDAI [2] and CDAI [3] minor or major, and American College of Rheumatology (ACR) 20, 50 or 70% response [7], may be less suitable in routine care. Unlike the group level, individual responses will depend on the criteria set chosen, at least at the stricter levels [8].

The cut-points for the various activity indexes and the components of the response criteria are summarized in Table 1.

The aim of the present study was to apply and compare the performance of various activity indices and response criteria in an observational cohort of patients from southern Sweden with established RA, treated with their first course of TNF-blockers. The findings may also serve as a reference of what levels of activity and response are to be expected in the "real life" setting without any formal or financial restrictions. The comparisons will also serve as a repository for future studies using one or the other measure as their primary end-point, relating them to results observable with other primaries.

Methods

The South Swedish Arthritis Treatment Group (SSATG), a network of hospital and office based rheumatologists in southern Sweden, maintains a database into which all courses of treatment with biologic drugs for RA and other arthritides are entered as described elsewhere [9].

Patients eligible for the study had a diagnosis of RA, as judged by the treating rheumatologist, and started their first course of treatment with infliximab, etanercept, or adalimumab from March 1999 through December 2006. Due to the quality and safety monitoring character of the register, no formal ethics committee approval was required.

Patients were evaluated at baseline and 3 months. The follow up protocol included tender and swollen 28 joint counts, visual analogue scale (VAS) for pain and patient global, evaluator global Likert scale, ESR, and CRP. Thus DAS28, SDAI and CDAI could be calculated at each time point as well as fulfilment of EULAR, ACR, SDAI and CDAI response criteria at 3 months. The number of patients falling into the different categories of disease activity (including remission) when applying the cut off levels proposed for DAS28 (original and modified), SDAI and CDAI were then calculated.

Statistical methods

Descriptive statistics was used throughout. For each level of disease activity and response, a reference criteria set was chosen. Patients fulfilling and not fulfilling this com-

Table 1: Cut points for activity states according to various indexes (Panel A) and components of response criteria (Panel B)

A. Activity index

	High	Moderate	Low	Remission
DAS28 original#	> 5,1	< 5,1	< 3,2	< 2,6
DAS28 modified cut off #	> 5,5	< 5,5	< 3,6	< 2,4
SDAI##	> 26	< 26	< 11	< 3,3
CDAI###	> 22	< 22	< 10	< 2,8

B. Response criteria

	TJC	SJC	Patient global VAS	Patient pain VAS	Evaluator global	HAQ	ESR	CRP
ACR	yes	yes	yes/no	yes/no	yes/no	yes/no	yes/no	yes/no
EULAR	yes	yes	yes	no	no	no	yes	no
SDAI	yes	yes	yes	no	yes	no	no	yes
CDAI	yes	yes	yes	no	yes	no	no	no

DAS28 = 0,56x$\sqrt{}$TJC28+0,28x$\sqrt{}$SJC28x0,7xlnESR+0,014xPat global VAS(in mm)
SDAI=SJC+TJC+Pat global VAS(in cm)+Eval global VAS(in cm)+CRP(in mg/dL)
CDAI = SJC+TJC+Pat global VAS(in cm)+Eval global VAS(in cm)
Cut points for activity states according to various indexes (Panel A) and components of response criteria (Panel B). Yes, required; no, not required; yes/no, in the ACR criteria, 3 of the variables marked "yes/no" are required (0 or 1 laboratory measure). TJC, tender joint count; SJC, swollen joint count; VAS, visual analogue scale; HAQ, health assessment questionnaire.

prised the basis for comparison, as another criteria set was applied to the same patients. The agreement, i e patients fulfilling both sets (positive agreement) or neither (negative agreement) were then calculated as percentages and assessed with κ statistics[10]. The procedure was then repeated with a new reference criteria set. Thus, all comparisons were made between pairs of disease states and according to different indices, but each comparison was on the same level of disease activity. Likewise, pair wise comparisons of response criteria fulfilment at each level of response were performed. For example, DAS28 original cut point low was compared to SDAI low, and ACR20 was compared to EULAR overall response, etc. The κ values indicate the level of agreement beyond chance between two dichotomous variables. A frequently cited rule of thumb [11] is that κ values > 0.8 correspond to almost perfect agreement, 0.61–0.8 to substantial, 0.41–0.6 to moderate and 0.2–0.4 to fair agreement. κ = 0 denotes chance agreement. Completer analysis has been applied due to the observational nature of the study.

Results

The eligibility criteria were met by 1789 patients. The baseline characteristics for all patients, the 1258 with 3 month data and the 531 patients lacking data at 3 months are given in Table 2. At baseline the majority (> 95%) of patients were in high/intermediate disease activity irrespective of criteria sets used (Figure 1A). At 3 months, 12–19% of patients had high, 39–46% had moderate, and 38–49% had low disease activity depending on criteria set. According to DAS28 original and modified cut points, 23 and 19%, respectively, were in remission, whereas about 8% were in remission according to SDAI and CDAI (Figure 1B).

The agreement between each defined disease state at baseline and 3 months, e.g. DAS28 low original cut point, DAS28 low modified cut point, SDAI low and CDAI low, are given in Table 3, along with their respective κ values. In general, agreement between the activity states determined by the cut-points of the various activity indices was moderate or substantial with κ values of about 0.5 or higher. At 3 months, remission according to the 2 DAS28-based categories had excellent agreement, κ = 0.84, and so did SDAI and CDAI, κ = 0.91, whereas agreement between DAS28-based agreement criteria and SDAI/CDAI was moderate, κ = 0.40–0.47. SDAI and CDAI had excellent agreement at all levels of disease activity (Table 3).

For the calculation of response at 3 months, 1258 patients had complete data, with baseline characteristics similar to the total cohort, but also to those with missing 3 month data (table 2). Response rates were 60–70% at the ACR20/overall level, 42% for EULAR original moderate, 19% for EULAR modified moderate, 25% for SDAI and CDAI minor response. At the major/good level, 35% responded according to EULAR original and ACR50 and 50–55% according to EULAR modified, SDAI and CDAI. Fourteen per cent were ACR70 responders (Figure 2).

The responder agreements at 3 months are summarized in Table 4. Substantial agreement (κ = 0.54–0.91) was found at the modest response level of ACR20, EULAR overall, SDAI overall and CDAI overall. Agreement at good/major level between EULAR original on one hand and, ACR70, SDAI, or CDAI on the other was poor (κ = 0.17–0.27), whereas EULAR modified was in much better agreement with SDAI and CDAI, κ = 0.69. SDAI and CDAI showed good agreement with κ values between 0.68 and 0.91 at the different response levels, while the agreement between other criteria sets at different response levels was more variable.

Discussion

The major findings in this observational study of a non-randomized cohort of established RA patients, receiving their first course of anti-TNF treatment and followed for 3 months, were that the disease activity states according to the various indices performed similarly and showed moderate or substantial agreement at all levels except remission. SDAI and CDAI stages show excellent agreement. Agreement of response criteria is substantial at low

Table 2: Baseline characteristics of all included and those with and without 3 month data

	All included N = 1789		With 3 month data N = 1258		Missing at 3 months N = 531	
	Mean	Std dev	Mean	Std dev	Mean	Std dev
Percent female	77.2		77.8		77.8	
Age	55.9	13.4	55.6	13.2	56.3	13.7
Disease duration, years	12.1	10.2	12.0	10.0	12.2	10.6
Ongoing DMARDs	0.85	0.57	0.85	0.57	0.84	0.57
DAS28 (0–10)	5.54	1.18	5.58	1.16	5.43	1.21
CDAI	30.6	12.3	31.0	12.1	32.2	13.9
HAQ (0–3)	1.34	0.64	1.36	0.64	1.29	0.62
CRP (mg/L)	30.8	33.0	31.4	33.6	28.9	31.3

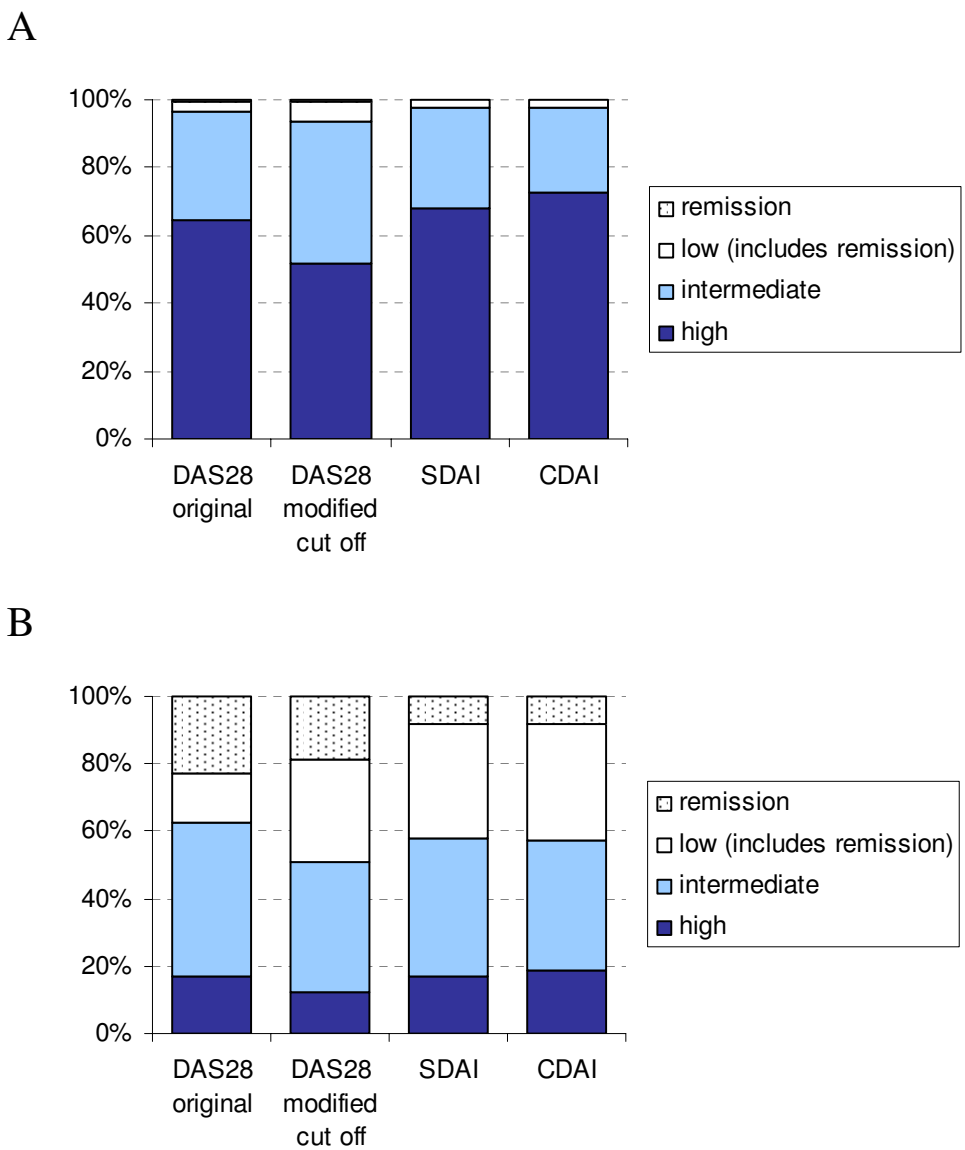


Figure 1
Disease stages at A) baseline and B) at 3 month follow up according to the different criteria sets. Notice that patients in remission are included in the low disease activity category, which thus comprises white and dotted areas together.

Table 3: Agreement between disease activity states at baseline and three month according to various indices, expressed as percentages and κ values.

Baseline							
Valid N		**DAS28 modified cut off *1657***		**SDAI *1657***		**CDAI *1681***	
Reference criterion	*Valid N*	**%**	**κ**	**%**	**κ**	**%**	**κ**
High disease activity							
DAS28 original	*1657*	52/35	**0,73**	60/27	**0,69**	61/24	**0,65**
DAS28 modified cut off	*1657*			50/31	**0,62**	51/27	**0,54**
SDAI	*1672*					67/26	**0,83**
Moderate disease activity							
DAS28 original	*1657*	29/55	**0,66**	23/61	**0,63**	20/62	**0,58**
DAS28 modified cut off	*1657*			25/52	**0,52**	21/53	**0,42**
SDAI	*1672*					24/69	**0,80**
Low disease activity							
DAS28 original	*1657*	2.7/94	**0,64**	1.3/96	**0,55**	1.3/96	**0,54**
DAS28 modified cut off	*1657*			1.8/94	**0,46**	1.6/94	**0,41**
SDAI	*1672*					1.7/97	**0,81**
Remission							
DAS28 original	*1657*	0.6/99	**0,91**	0.1/99	**0,15**	0.1/99	**0,15**
DAS28 modified cut off	*1657*			0.1/99	**0,18**	0.1/99	**0,18**
SDAI	*1672*					0/100	**1,00**
3 month							
Valid N		***1214***		***1224***		***1232***	
High disease activity							
DAS28 original	*1206*	12/83	**0,82**	13/79	**0,72**	13/78	**0.69**
DAS28 modified cut off	*1214*			11/82	**0,72**	11/80	**0.67**
SDAI	*1224*					15/80	**0,83**
Moderate disease activity							
DAS28 original	*1206*	34/50	**0,67**	32/46	**0,55**	30/45	**0,49**
DAS28 modified cut off	*1214*			29/50	**0,56**	26/49	**0,47**
SDAI	1224					35/55	**0,79**
Low disease activity							
DAS28 original	1206	38/51	**0,77**	33/53	**0,71**	32/52	**0,67**
DAS28 modified cut off	1214			39/47	**0,71**	37/46	**0,66**
SDAI	1224					40/55	**0,88**
Remission							
DAS28 original	1206	18/77	**0,87**	7.7/77	**0,42**	7.3/76	**0,40**
DAS28 modified cut off	1214			7.1/80	**0,47**	6.7/80	**0,44**
SDAI	1224					7.5/91	**0,91**

Percentages denote positive agreement/negative agreement, i.e. per cent patients achieving both compared activity states/per cent patients achieving neither.

response levels such as ACR20 and EULAR, SDAI and CDAI overall. At the moderate/minor level, when considered separately, agreement is poor, and this also holds true for the good/major level.

The κ value tells to which extent agreement is not explained by chance. The positive/negative agreements give, however, an idea of the degree of agreement; the closer total agreement (positive + negative) is to 100%, the higher degree of agreement. Furthermore, κ depends on sample size, and it is hard to interpret, if the compared groups are small. This is illustated by the limited κ agreement regarding both ACR70 responders and those reaching remission.

Completer analysis may lead to bias, if the drop-outs at 3 months substantially differ from the completers. The baseline data of the drop-outs and patients with 3 month data was therefore examined separately and not found to differ from the whole cohort in any clinically relevant way (Table 2). Also, in the observational setting, only patients remaining on therapy are contributing their data as time

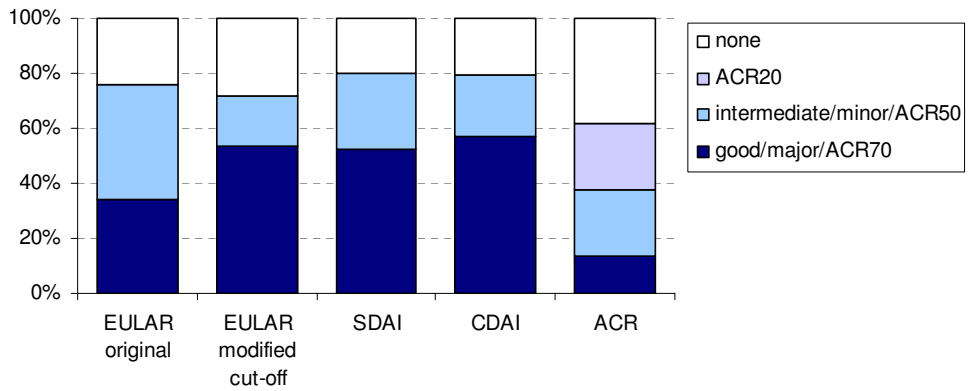


Figure 2
Response at 3 months according to different criteria sets. Notice that all ACR70 responders are include in the ACR50 responders, and all ACR50 responders are included in the ACR20 responders. Similarly EULARoverall responders include both good and intermediate responders, while SDAIoverall and CDAIoverall responders includes both major and minor responders

Table 4: Agreement of response according to various criteria sets at 3 months, expressed as percentages and κ values

Reference criterion		**EULAR**		**EULAR modified cut off**		**SDAI**		**CDAI**	
Valid N		**1163**		**1163**		**1184**		**1195**	
	Valid N	%	κ	%	κ	%	κ	%	κ
Overall response		***good+moderate***		***good+moderate***		***major+minor***		***major+minor***	
ACR20	1234	61/22	**0,61**	59/24	**0,62**	62/19	**0,54**	61/19	**0,54**
EULAR	1163			71/23	**0,83**	73/17	**0,71**	72/17	**0,68**
EULAR modified cut off	1163					70/18	**0,70**	69/18	**0,67**
SDAI	1184							78/19	**0,91**
Moderate/minor response		***moderate***		***moderate***		***minor***		***minor***	
ACR50	1234	13/33	**0,00**	3.5/47	**0,00**	7.3/41	**0,00**	5.6/45	**0,00**
EULAR	1163			14/54	**0,27**	13/43	**0,04**	10/46	**0,03**
EULAR modified cut off	1163					11/65	**0,31**	8.7/68	**0,28**
SDAI	1184							20/70	**0,74**
Good/major response		***good***		***good***		***major***		***major***	
ACR70	1234	12/64	**0,40**	13/46	**0,22**	12/46	**0,19**	12/42	**0,17**
EULAR	1163			31/43	**0,49**	25/38	**0,27**	26/36	**0,27**
EULAR modified cut off	1163					45/39	**0,69**	47/37	**0,69**
SDAI	1184							51/42	**0,86**

Percentages denote positive agreement/negative agreement, i.e. per cent patients fulfilling both compared response criteria/per cent patients fulfilling neither.

goes by, which tends to inflate the results. Missing information at 3 month follow up in our setting is higher than in the British Biologics register [12]. There may be several explanations for this, but the mandatory response demanded for drug continuation in the British setting may be one important factor. Also, the patients in the British register differ on numerous baseline characteristics such as higher disease activity and very high disability measured by the HAQ, all factors that may influence response rates and future disease states [12,13]. Thus each setting has its features and face to face comparison should always be done with great care.

The distribution between disease state categories at baseline and after 3 months is what may be expected in RA patients with mean disease duration of 12 years and who failed at least 2 different DMARDs. Achieving remission and even low disease activity thus appears not to be a too common event in established RA, treated with TNF-blockers. This is in agreement with other observational studies [16]. In our setting with many long-standing cases, erosive disease may preclude reaching remission in many patients although inflammation per se is suppressed by treatment. The small number of patients in low disease activity (LDA) at baseline is explained by patients taking high doses of prednisolone, also an accepted indication for commencing TNF-blocker treatment according to the Swedish guidelines.

The agreement between disease states, as defined by the DAS28 original and modified cut points, SDAI and CDAI proposed cut points, seems to be substantial. This is in agreement with comparisons based on trial data [14]. The excellent agreement of the disease state groups based on SDAI and CDAI irrespective of activity level may be accounted for by their great similarity: they contain the same components, added together, except for the CRP that is excluded from the CDAI. Our data thus support the notion of acute phase reactant values contributing little to the overall disease activity estimate [3]. The SDAI and CDAI remission is achieved in fewer patients than the DAS28 remission that appears less strict in this cohort, in agreement with previous findings [15]. At the remission level, agreement is almost perfect between EULAR original and modified cut off, and also between SDAI and CDAI at 3 months, as may be expected (Table 3). The moderate agreement between the DAS28-based and SDAI/CDAI remission, may partly be accounted for by the greater strictness of the latter (Figure 1B). Due to low patient numbers reaching remission, κ must be interpreted with caution in this category.

Response criteria are intended for clinical trials, and it is thus not surprising that they perform poorly when applied to individual RA patients in clinical practice. Agreement (Table 3) is variable across criteria and response levels with a tendency to be better at the less strict overall level, where EULAR moderate and good are merged to one group, as are SDAI and CDAI minor and major responders. In this manner, substantial agreement is achieved with ACR20, which includes all ACR50 and 70 responders, in accordance with previous findings [5,8].

The very poor κ values (often close to 0) for moderate response comparisons seem to indicate random agreement at the individual level. This may in part be due to the construction of the criteria. Thus, the ACR50 responders include all the ACR70 responders, whereas the EULAR moderate category does not include the EULAR good. The same mechanism may be operative concerning SDAI and CDAI minor and major responders. EULAR response according to original and modified definitions also exhibit poor agreement, especially at the moderate level. This is an expected finding, given the different cut-points of DAS28 used in each case. The value of response criteria in monitoring patients in routine care thus seems to be limited. In contrast, group level responses in a clinical setting can give some indication of the value of a particular treatment in routine care. As far as the management of individual RA patients with established disease is concerned, at least in our setting, the achievement of an absolute degree of disease activity seems to be a more relevant treatment goal than fulfilling a response criterion, i.e. achieving a given degree of improvement[8]. Our data thus do not support the use of response criteria as aid in the monitoring of RA-patients treated routinely with TNF-blockers, but this should be verified in other clinical cohorts.

In the development and evaluation of new treatment modalities, as well as in routine care, a unified concept of disease activity measurement and treatment aims will be beneficial. The widespread use of reproducible and simple composite measures of RA activity will facilitate this development. The present study provides support for this, but further validation of the indices in other cohorts is desirable.

Conclusion

Disease activity states according to the various indices perform similarly and show substantial agreement at all levels except remission. Agreement between SDAI and CDAI states is excellent. Response criteria, applied at the individual patient level, are hard to interpret and show poor agreement, except at the lowest level of response. Thus, they should not be applied uncritically in clinical practice.

Abbreviations

RA: rheumatoid arthritis; TNF: tumour necrosis factor; SSATG: South Swedish Arthritis Treatment Group; DAS:

Disease activity score; ACR: American College of Rheumatology; SDAI: Simple disease activity index; CDAI: Clinical disease activity index; LDA: low disease activity; CRP: C-reactive protein; HAQ: Health Assessment Questionnaire.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
AG participated in the design of the study, performed the statistical calculations and wrote the manuscript. DA participated in the design of the study and suggested the statistical methodology. TS participated in the design of the study and helped drafting the manuscript. PG conceived the study and helped drafting the manuscript. All authors read and approved the final manuscript.

Acknowledgements
We are indebted to all colleagues and staff in the South Swedish Arthritis Treatment Group for cooperation and data supply and to Jan-Åke Nilsson for help with statistical calculations. This study was supported by grants from Österlund and Kock Foundations, King Gustav V 80 year fund, Lund University Hospital, Södra sjukvårdsregionen, Faculty of Medicine, Lund University and Reumatikerförbundet.

References
1. **Afdelning Reumatologie UMC Sint Radboud Nijmegen** [http://www.das-score.nl]
2. Smolen PJS, *et al.*: **A simplified disease activity index for rheumatoid arthritis for use in clinical practice.** *Rheumatology (Oxford)* 2003, **42(2):**244-57.
3. Aletaha DNV, Stamm T, Uffmann M, Pflugbeil S, Machold K, Smolen JS: **Acute phase reactants add little to composite disease activity indices for rheumatoid arthritis: validation of a clinical activity score.** *Arthritis Research & Therapy* 2005, **7:**R796-R806.
4. Zink A, *et al.*: **Effectiveness of tumor necrosis factor inhibitors in rheumatoid arthritis in an observational cohort study: comparison of patients according to their eligibility for major randomized clinical trials.** *Arthritis Rheum* 2006, **54(11):**3399-407.
5. **A proposed revision to the ACR20: the hybrid measure of American College of Rheumatology response.** *Arthritis Rheum* 2007, **57(2):**193-202.
6. van Gestel AM, *et al.*: **Development and validation of the European League Against Rheumatism response criteria for rheumatoid arthritis. Comparison with the preliminary American College of Rheumatology and the World Health Organization/International League Against Rheumatism Criteria.** *Arthritis Rheum* 1996, **39(1):**34-40.
7. Felson DT, *et al.*: **American College of Rheumatology. Preliminary definition of improvement in rheumatoid arthritis.** *Arthritis Rheum* 1995, **38(6):**727-35.
8. Gulfe A, Geborek P, Saxne T: **Response criteria for rheumatoid arthritis in clinical practice: how useful are they?** *Ann Rheum Dis* 2005, **64(8):**1186-9.
9. Geborek P, Saxne T: **Clinical protocol for monitoring of targeted therapies in rheumatoid arthritis.** *Rheumatology (Oxford)* 2000, **39(10):**1159-61.
10. Viera AJ, Garrett JM: **Understanding interobserver agreement: the kappa statistic.** *Fam Med* 2005, **37(5):**360-3.
11. Landis JRKG: **The measurement of observer agreement for categorical data.** *Biometrics* 1977, **33:**159-174.
12. Hyrich KL, *et al.*: **Predictors of response to anti-TNF-alpha therapy among patients with rheumatoid arthritis: results from the British Society for Rheumatology Biologics Register.** *Rheumatology (Oxford)* 2006, **45(12):**1558-65.
13. Kristensen LE, Kapetanovic MC, Gülfe A, Söderlin M, Saxne T, Geborek P: **Predictors of response to anti-TNF therapy according to ACR and EULAR criteria in patients with established RA: results from the South Swedish Arthritis Treatment Group Register.** *Rheumatology (Oxford)* 2008, **47(4):**495-9.
14. Aletha D, S J: **The Simplified Disease Activity Index (SDAI) and the Clinical Disease Activity Index (CDAI): A review of their usefulness and validity in rheumatoid arthritis.** *Clin Exp Rheumatol* 2005, **23(5 Suppl 39):**S100-S108.
15. Aletaha D, *et al.*: **Remission and active disease in rheumatoid arthritis: defining criteria for disease activity states.** *Arthritis Rheum* 2005, **52(9):**2625-36.
16. Wolfe F, *et al.*: **Minimal disease activity, remission, and the long-term outcomes of rheumatoid arthritis.** *Arthritis Rheum* 2007, **57(6):**935-42.

Pre-publication history
The pre-publication history for this paper can be accessed here:

http://www.biomedcentral.com/1471-2474/10/41/prepub

III

Six and 12 Weeks Treatment Response Predicts Continuation of Tumor Necrosis Factor Blockade in Rheumatoid Arthritis: An Observational Cohort Study from Southern Sweden

ANDERS GÜLFE, LARS ERIK KRISTENSEN, and PIERRE GEBOREK

ABSTRACT. ***Objective.*** To investigate if treatment response predicts continuation of anti-tumor necrosis factor (TNF) treatment in patients with rheumatoid arthritis (RA).

Methods. We investigated if treatment response and/or achieving a certain activity state at 6 weeks or 3 months predicts continuation of treatment in an observational cohort of 1789 anti-TNF-naive patients with established RA disease from southern Sweden.

Results. Response to treatment at 6 weeks at overall/American College of Rheumatology (ACR20) or good/major level (except ACR70) significantly predicted drug continuation. Response according to all criteria sets at overall/ACR20 and at good/major/ACR70 level predicted drug continuation at 3 months, as did achieving low disease activity at 3 months irrespective of activity index applied. Remaining in a high disease activity state predicted drug discontinuation at both timepoints and according to all criteria sets.

Conclusion. Response criteria may be useful aids in deciding on continuation of TNF blockade in RA as early as after 6 weeks of treatment. The various criteria sets perform similarly. (First Release Jan 15 2009; J Rheumatol 2009;36:517-21; doi:10.3899/jrheum.080509)

Key Indexing Terms:

OBSERVATIONAL STUDY | RHEUMATOID ARTHRITIS
ANTI-TUMOR NECROSIS FACTOR TREATMENT | TREATMENT CONTINUATION
PREDICTORS | RESPONSE | DISEASE ACTIVITY STATES

Previously, baseline characteristics have been found to only weakly predict continuation of anti-tumor necrosis factor (TNF) treatment in patients with rheumatoid arthritis (RA)[1,2]. Even though amelioration of symptoms appears to occur early (within weeks) after starting TNF blockade in responding patients with RA, there is a paucity of data regarding the timepoint best suited for deciding on the continuation or stopping of treatment. On the individual level, no rigid guidelines can be given in this regard, but predictors of drug continuation at the group level will give a notion of the (minimal) time needed to judge the meaningfulness of going on with therapy. To address this, we conducted a study of an observational cohort of patients with established RA starting their first course of adalimumab, etanercept, or infliximab. The aim was to investigate whether treatment response or achieving a certain disease activity state at 6 weeks or 3 months predicted continuation of anti-TNF therapy. We also wanted to study whether any specific criteria set was superior in this aim.

From the Department of Rheumatology, Lund University Hospital, Lund, Sweden.

Supported by grants from the Österlund and Kock Foundations, King Gustav V 80 year Fund, Region Skåne, Lund University Hospital, and Reumatikerförbundet.

A. Gülfe, MD; L.E. Kristensen, MD, PhD; P. Geborek, MD, PhD, Lund University Hospital.

Address reprint requests to Dr. A. Gülfe, Department of Rheumatology, Lund University Hospital, SE-221 85 Lund, Sweden.
E-mail: anders.gulfe@med.lu.se

Accepted for publication October 6, 2008.

MATERIALS AND METHODS

Patients with RA and other arthritides who start treatment with TNF blockers and other biologics in southern Sweden are entered into a database maintained by the South Swedish Arthritis Treatment Group (SSATG) as described[3]. The catchment area has a population of about 1.3 million, and the coverage as compared to the sales of the relevant drugs through the pharmacies is 95%[4]. Because of the safety surveillance character of the registry, no ethics committee approval was needed.

Patients eligible for the study had a diagnosis of RA, as judged by the treating rheumatologist, and started their first course of treatment with infliximab, etanercept, or adalimumab from March 1999 through December 2006.

Patients were enrolled continuously. The 3 TNF blockers were studied together, since they have been shown to perform similarly in our cohort[2].

Fulfilment of American College of Rheumatology (ACR) 20%, 50% and 70% response[5]; European League Against Rheumatism (EULAR) overall (moderate + good), moderate, and good response according to original and modified cutoff values[6]; and overall (minor + major), minor, and major response according to the Simple Disease Activity Index (SDAI)[7] and the Clinical Disease Activity Index (CDAI)[8] were calculated at 6 weeks and 3 months. Achievement of low, moderate, or high disease activity

according to the Disease Activity Score in 28 joints (DAS28) original and modified cutoff values[9,10], SDAI[10], and CDAI[11] were also calculated at the same timepoints. Hazard ratios (HR) for stopping treatment were then calculated together with their 95% confidence intervals (CI) and p values for each of the criteria.

Completer analysis was used for response rates at 6 weeks and 3 months due to the limited followup time and observational design of the study. The proportion of patients remaining on therapy at 6 and 12 weeks was estimated using Kaplan-Meier statistics. Drug adherence and discontinuation (due to failure or adverse event) was estimated by life-table analysis with stop date December 2006. Missing data were requested from treating physicians yearly. Cox regression analysis was employed to study predictors of treatment continuation. The regression models included correction for age, disease duration, baseline Health Assessment Questionnaire (HAQ) score[12], baseline C-reactive protein (CRP) level, and concomitant methotrexate use, since all these variables have previously been shown to influence 3-month treatment response[2]. Due to colinearity, a regression model was computed for each of the response criteria and disease activity states studied. P values less than 0.05 were considered significant.

RESULTS

There were 1789 patients with established RA meeting the eligibility criteria, 77% of whom were female. As reporting at 6 weeks in the SSATG system is optional, there were considerably fewer patients with data at 6 weeks than at 3 months. At baseline, mean (standard deviation) age was 56.9 (13.4) years and disease duration 12.1 (10.2) years, number of ongoing disease modifying antirheumatic drugs (DMARD) 0.9 (0.6), HAQ score 1.3 (0.6), DAS28 score 5.5 (1.2), erythrocyte sedimentation rate 35.6 (24.7) mm/h, and CRP 30.8 (33.0) g/l. There were few early withdrawals, with 96% and 94% of patients remaining on therapy at 6 weeks and 3 months, respectively.

The proportions of patients meeting the various response criteria and disease activity stages are given in Table 1. At the lowest level (ACR20, EULAR overall original and modified cutpoints, SDAI overall, and CDAI overall) response rates at 6 weeks were 56%–75%. At 3 months the corresponding rates were 61%–80%. The higher response level (ACR50, EULAR good, SDAI and CDAI major) was achieved by 29%–44% at 6 weeks and 34%–57% at 3 months.

Drug continuation was predicted significantly by achieving ACR20 and ACR50 responses at 6 weeks and all levels of ACR response at 3 months. Similarly, achieving EULAR overall or good responses, both original and modified, predicted continuation of treatment at both timepoints, and this was also true for SDAI and CDAI overall and major. Isolated moderate/minor response did not predict drug continuation for any of the criteria sets at any timepoint (Table 2).

To assess the effect of reason for discontinuation of therapy, the HR for stopping treatment for the strongest predicting variable, the EULAR original overall response, were calculated separately for adverse events and failure. At 6 weeks, the HR for stopping treatment due to adverse events was 1.17 (95% CI 0.61–2.26, p = 0.64), and for failure it was 0.43 (95% CI 0.24–0.76, p = 0.004). The corresponding values at 3 months were 0.44 (95% CI 0.31–0.63, p < 0.0001) and 0.28 (95% CI 0.19–0.41, p < 0.0001), respectively.

Table 1. Response and achievement of disease states according to various criteria. Total n = 1789.

	6 Weeks		3 Months	
	%	Valid n	%	Valid n
Response criteria				
ACR				
ACR20	55	536	61	1234
ACR50	29	536	38	1234
ACR70	8.5	542	14	1234
EULAR original				
Overall	73	499	76	1163
Moderate	45	505	42	1163
Good	29	499	34	1163
EULAR modified				
Overall	69	499	72	1163
Moderate	25	505	18	1163
Good	44	499	54	1163
SDAI				
Overall	75	512	80	1184
Minor	37	518	28	1184
Major	39	512	52	1184
CDAI				
Overall	74	523	79	1195
Minor	30	529	22	1195
Major	43	523	57	1195
Disease stages				
EULAR original				
Remission	18	524	23	1214
Low	35	518	38	1214
Moderate	47	518	46	1214
High	18	518	17	1214
EULAR modified				
Remission	12	524	18	1214
Low	46	518	49	1214
Moderate	41	518	38	1214
High	13	518	12	1214
SDAI				
Remission	4.1	534	8.3	1224
Low	35	528	42	1224
Moderate	20	528	41	1224
High	18	528	17	1224
CDAI				
Remission	4.5	539	8.0	1232
Low	36	533	45	1232
Moderate	44	533	39	1232
High	20	533	19	1232

ACR: American College of Rheumatology; EULAR: European League Against Rheumatism; SDAI: Simple Disease Activity Index; CDAI: Clinical Disease Activity Index.

There were 396 patients with complete data at 6 weeks and 3 months. For these patients, the mean HR for the EULAR original overall response to predict discontinuation of therapy was 0.80 (95% CI 0.51–1.25, p = 0.32) at 6 weeks and 0.44 (95% CI 0.29–0.66, p < 0.0001) at 3 months.

Not surprisingly, achieving remission or low disease activity state generally predicts continuation of treatment, especially at 3 months. At 6 weeks, achieving disease remis-

Table 2. Hazard ratios (HR) for stopping tumor-necrosis factor blockade for patients achieving defined response or disease states according to various criteria. Total n = 1789.

	6 Weeks			3 Months		
	HR (95% CI)	p	Valid n	HR (95% CI)	p	Valid n
Response criteria						
ACR						
ACR20	0.55 (0.43–0.71)	< 0.001	526	0.56 (0.47–0.66)	< 0.001	1208
ACR50	0.64 (0.48–0.85)	0.002	526	0.55 (0.46–0.66)	< 0.001	1208
ACR70	1.06 (0.59–1.90)	0.850	526	0.56 (0.42–0.75)	< 0.001	1208
EULAR original						
Overall	0.59 (0.44–0.77)	< 0.001	493	0.47 (0.39–0.56)	< 0.001	1151
Moderate	0.95 (0.66–1.34)	0.763	493	0.87 (0.73–1.04)	0.123	1151
Good	0.72 (0.53–0.97)	0.032	493	0.58 (0.47–0.71)	< 0.001	1151
EULAR modified						
Overall	0.63 (0.48–0.83)	0.001	493	0.48 (0.40–0.57)	< 0.001	1151
Moderate	1.03 (0.69–1.53)	0.895	493	0.96 (0.77–1.19)	0.695	1151
Good	0.61 (0.49–0.79)	< 0.001	493	0.56 (0.47–0.67)	< 0.001	1151
SDAI						
Overall	0.62 (0.45–0.81)	0.001	512	0.47 (0.39–0.57)	< 0.001	1179
Minor	0.80 (0.56–1.15)	0.229	512	1.06 (0.87–1.28)	0.557	1179
Major	0.65 (0.50–0.86)	0.002	512	0.58 (0.48–0.69)	< 0.001	1179
CDAI						
Overall	0.75 (0.57–0.99)	0.044	517	0.53 (0.44–0.65)	< 0.001	1181
Minor	0.94 (0.64–1.37)	0.742	517	1.20 (0.99–1.47)	0.067	1181
Major	0.69 (0.54–0.90)	0.005	517	0.57 (0.48–0.68)	< 0.001	1181
Disease states						
EULAR original						
Remission	1.03 (0.65–1.62)	0.911	502	0.65 (0.51–0.82)	< 0.001	1173
Low	0.69 (0.52–0.93)	0.015	502	0.60 (0.49–0.72)	< 0.001	1173
Moderate	1.04 (0.81–1.34)	0.769	502	0.98 (0.83–1.16)	0.803	1173
High	1.60 (1.15–2.23)	0.005	502	2.18 (1.78–2.69)	< 0.001	1173
EULAR modified						
Remission	1.33 (0.82–2.16)	0.242	502	0.65 (0.50–0.85)	0.001	1173
Low	0.60 (0.45–0.79)	< 0.001	502	0.61 (0.50–0.73)	< 0.001	1173
Moderate	1.24 (0.96–1.60)	0.108	502	1.08 (0.91–1.29)	0.386	1173
High	1.86 (1.28–2.69)	0.001	502	2.15 (1.72–2.69)	< 0.001	1173
SDAI						
Remission	1.85 (0.86–4.02)	0.118	512	0.60 (0.41–0.87)	0.006	1182
Low	0.81 (0.62–1.06)	0.125	512	0.67 (0.56–0.81)	< 0.001	1182
Moderate	0.90 (0.70–1.15)	0.389	512	0.90 (0.76–1.07)	0.224	1182
High	1.73 (1.26–2.39)	0.001	512	2.19 (1.79–2.68)	< 0.001	1182
CDAI						
Remission	1.82 (0.84–3.945)	0.129	517	0.64 (0.441–0.918)	0.016	1189
Low	0.84 (0.64–1.10)	0.192	517	0.70 (0.58–0.84)	< 0.001	1189
Moderate	0.87 (0.68–1.12)	0.237	517	0.84 (0.70–1.00)	0.048	1189
High	1.62 (1.20–2.19)	0.002	517	2.24 (1.84–2.73)	< 0.001	1189

sion according to all criteria does not significantly predict continuation, nor does moderate activity. Remaining in high disease activity state strongly predicts discontinuation of drug irrespective of activity index (Table 2).

Independent of treatment response and disease activity, significant predictors of premature treatment termination by unadjusted Cox regression analysis were: higher age (HR 1.01, 95% CI 1.00–1.01), higher HAQ score (HR 1.21, 95% CI 1.07–1.36), and no methotrexate at inclusion (HR 0.75, 95% CI 0.65–0.87). When adjusting for the strongest predicting response criterion, the EULAR overall original at 3 months, HAQ (HR 1.26, 95% CI 1.09–1.46) and no methotrexate (HR 0.80, 95% CI 0.67–0.96) at inclusion predicted termination of treatment.

Sex, year of treatment initiation, and disease duration prior to treatment initiation did not predict continuation of therapy.

DISCUSSION

The main finding of our observational study of 1789 patients from southern Sweden with longstanding RA is that the key predictors for continuation of anti-TNF treatment were response and achievement of a state of low disease activity, in most instances already after 6 weeks, and very signifi-

cantly after 3 months (Table 2). Conversely, remaining in a high disease activity state strongly predicts drug discontinuation (Table 2). On the other hand, baseline characteristics, although significant, predict treatment continuation only to a lesser degree.

It seems that relatively little change in overall treatment response takes place between the timepoints (Table 1). This is also true for the achievement of disease states (Table 1). There is, however, a clear increase in the proportion reaching a higher degree of response irrespective of criteria set used at 3 months compared to 6 weeks. The modest response rates are not surprising in a cohort of RA patients with mean disease duration of 12 years, having failed at least 2 previous DMARD.

At 6 weeks, continuation of treatment is significantly predicted by response according to ACR20 and ACR50 and EULAR overall and good original and modified, SDAI overall and major, and CDAI overall and major responses (Table 2). ACR70 fails to predict drug continuation at 6 weeks, presumably because of the small number of responders according to this strict criterion. These findings prevail and are even stronger at 3 months, when responses at all ACR levels, and according to all other criteria sets at overall and good/major levels, significantly predict continuation of treatment. By contrast, isolated moderate/minor responses fail to predict drug continuation. Conceivably, the ability of the ACR20 and overall-level responses to significantly predict drug continuation may be due to the good/major components of these merged measures.

As for disease activity states, achieving remission or low disease activity at 3 months predicts drug continuation according to all criteria sets (Table 2). The picture at 6 weeks is less clear, with the DAS28-based criteria for low activity reaching significance, unlike the stricter SDAI and CDAI. Remaining in a high disease activity state strongly predicts drug discontinuation, both at 6 weeks and 3 months, and according to all criteria sets.

Discontinuation of treatment is also independently predicted by the baseline variables higher age and HAQ and no concomitant methotrexate. However, the absolute HR of these predictors are so small that the clinical significance is minor compared to that of response to treatment.

In our study, it thus seems that major response to treatment at 6 weeks in most cases predicts treatment continuation, and this is even clearer at 3 months, applying completer analysis. The same tendency is noted for achieving low disease activity, but this is clearly evident only at 3 months. For patients with complete data at both timepoints, the strongest predicting variable, the EULAR original overall response, significantly predicted drug continuation at 3 months only, illustrating the tendency for completer analysis to inflate results and thus somewhat weakening our conclusions regarding 6-week prediction. However, the failure to predict drug continuation at 6 weeks may be due to lack of power, the number of patients being much lower than in the completer analysis.

As for the reason for treatment termination, there were significant HR for treatment failure at both timepoints and for adverse event at 3 months for the EULAR original overall. It must be remembered that the reason for termination was determined by the treating rheumatologist. Stopping due to treatment failure or adverse event should thus be interpreted with caution, and more emphasis given to the overall reason for stopping treatment[13,14].

Response criteria are intended for comparing treatments in clinical trials, and they have been shown to perform poorly at the individual level in the clinical setting[15]. Interestingly, in this observational, nonrandomized cohort, at the group level, they appear to be the most sensitive predictors of continuation with TNF blockade already after 6 weeks, rather than the achievement of low disease activity. After 3 months, more patients are in a low disease activity state or even remission, and this predicts continuation at this timepoint, but response criteria remain better predictors. Treatment failure, i.e., ongoing high disease activity, predicts stopping of anti-TNF treatment both at 6 weeks and at 3 months. This is not surprising, since it is in accord with the current Swedish guidelines[16]. In early RA trials, rapid suppression of disease activity has been shown to predict low disease activity later[17,18]. This is compatible with our findings in longstanding RA.

The various criteria sets perform similarly and seem to be useful tools aiding a decision on continuation of anti-TNF treatment. One criterion set could hardly be considered superior to another, but the simplicity of the CDAI, which requires no laboratory measure, would make this set especially suited for day to day clinical use. However, activity indexes and response criteria cannot be solely depended upon in making treatment decisions. They do not cover all the dynamic aspects of the disease, and clinical judgment will remain important in the daily care of patients with RA. Indeed, treatment continuation based on the judgment of the treating rheumatologist was the standard by which efficacy was determined in our study. This may be regarded as a surrogate measure, but the prognostication of treatment response by the use of various biomarkers has so far turned out to be difficult[19]. However, the employment of composite measures of disease activity and response to treatment seems to be useful in predicting continuation of treatment as early as 6 weeks after initiation of TNF blockade, at least at the group level. Consequently, decision-making regarding continuation of TNF blockade might be considered as early as at 6 weeks of followup, as opposed to previous guidelines. Thus our finding raises important questions regarding clinical decision-making as well as health economic issues. Even so, our results should be verified in other patient studies.

ACKNOWLEDGMENT

We are indebted to all colleagues and staff in the South Swedish Arthritis Treatment Group for cooperation and data supply, and to Jan-Åke Nilsson for help with statistical calculations.

REFERENCES

1. Hyrich KL, Watson KD, Silman AJ, Symmons DP. Predictors of response to anti-TNF-alpha therapy among patients with rheumatoid arthritis: results from the British Society for Rheumatology Biologics Register. Rheumatology Oxford 2006;45:1558-65.
2. Kristensen LE, Kapetanovic MC, Gulfe A, Soderlin M, Saxne T, Geborek P. Predictors of response to anti-TNF therapy according to ACR and EULAR criteria in patients with established RA: results from the South Swedish Arthritis Treatment Group Register. Rheumatology Oxford 2008;47:495-9.
3. Geborek P, Saxne T. Clinical protocol for monitoring of targeted therapies in rheumatoid arthritis. Rheumatology Oxford 2000;39:1159-61.
4. Geborek P, Nitelius E, Noltorp S, et al. Population based studies of biological antirheumatic drug use in southern Sweden: comparison with pharmaceutical sales. Ann Rheum Dis 2005;64:1805-7.
5. Felson DT, Anderson JJ, Boers M, et al. American College of Rheumatology. Preliminary definition of improvement in rheumatoid arthritis. Arthritis Rheum 1995;38:727-35.
6. van Gestel AM, Prevoo ML, van 't Hof MA, van Rijswijk MH, van de Putte LB, van Riel PL. Development and validation of the European League Against Rheumatism response criteria for rheumatoid arthritis. Comparison with the preliminary American College of Rheumatology and the World Health Organization/International League Against Rheumatism Criteria. Arthritis Rheum 1996;39:34-40.
7. Smolen JS, Breedveld FC, Schiff MH, et al. A simplified disease activity index for rheumatoid arthritis for use in clinical practice. Rheumatology Oxford 2003;42:244-57.
8. Aletaha D, Nell VP, Stamm T, et al. Acute phase reactants add little to composite disease activity indices for rheumatoid arthritis: validation of a clinical activity score. Arthritis Res Ther 2005;7:R796-806.
9. Prevoo ML, van 't Hof MA, Kuper HH, van Leeuwen MA, van de Putte LB, van Riel PL. Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. Arthritis Rheum 1995;38:44-8.
10. Aletaha D, Ward MM, Machold KP, Nell VP, Stamm T, Smolen JS. Remission and active disease in rheumatoid arthritis: defining criteria for disease activity states. Arthritis Rheum 2005;52:2625-36.
11. Aletaha D, Smolen J. The Simplified Disease Activity Index (SDAI) and the Clinical Disease Activity Index (CDAI): a review of their usefulness and validity in rheumatoid arthritis. Clin Exp Rheumatol 2005;23 Suppl:S100-8.
12. Ekdahl C, Eberhardt K, Andersson SI, Svensson B. Assessing disability in patients with rheumatoid arthritis. Use of a Swedish version of the Stanford Health Assessment Questionnaire. Scand J Rheumatol 1988;17:263-71.
13. Kristensen LE, Saxne T, Nilsson JA, Geborek P. Impact of concomitant DMARD therapy on adherence to treatment with etanercept and infliximab in rheumatoid arthritis. Results from a six-year observational study in southern Sweden. Arthritis Res Ther 2006;8:R174.
14. Zink A, Listing J, Kary S, et al. Treatment continuation in patients receiving biological agents or conventional DMARD therapy. Ann Rheum Dis 2005;64:1274-9.
15. Gulfe A, Geborek P, Saxne T. Response criteria for rheumatoid arthritis in clinical practice: how useful are they? Ann Rheum Dis 2005;64:1186-9.
16. Guidelines for Pharmacotherapy of Rheumatoid Arthritis. In: The Swedish Society for Rheumatology; 2004 [Swedish]. Internet. Accessed Dec 19, 2008. Available from: http://www.musconline.org/html/srf/images/documents/riktra04/riktrakort.pdf
17. van der Kooij SM, Goekoop-Ruiterman YP, de Vries-Bouwstra JK, et al. Probability of continued low disease activity in patients with recent onset rheumatoid arthritis treated according to the Disease Activity Score. Ann Rheum Dis 2008;67:266-9.
18. Aletaha D, Funovits J, Keystone EC, Smolen JS. Disease activity early in the course of treatment predicts response to therapy after one year in rheumatoid arthritis patients. Arthritis Rheum 2007;56:3226-35.
19. Smolen JS, Aletaha D, Grisar J, Redlich K, Steiner G, Wagner O. The need for prognosticators in rheumatoid arthritis. Biological and clinical markers: where are we now? Arthritis Res Ther 2008;10:208.

IV

ARD Online First, published on March 27, 2009 as 10.1136/ard.2008.103473

Rapid and sustained health utility gain in anti-TNF treated inflammatory arthritis. Observational data during seven years in southern Sweden.

Anders Gülfe, MD#, Lars Erik Kristensen, MD, PhD#, Tore Saxne, MD, PhD#, Lennart TH Jacobsson, MD, PhD##, Ingemar F Petersson, MD, PhD###, Pierre Geborek, MD, PhD#

#Dept of Rheumatology, Lund University Hospital, Lund Sweden
##Dept of Rheumatology, Malmö University Hospital, Malmö, Sweden
###South Sweden Musculoskeletal Research Centre, Dept of Orthopedics, Lund University Hospital, Lund, Sweden

Correspondence: Pierre Geborek
Dept of Rheumatology
Lund University Hospital
SE-221 85 Lund, Sweden
Telephone: +46 46171619
Fax: +46 46128468
E-mail: Pierre.geborek@med.lu.se

Abstract

Background. Rheumatoid arthritis (RA), psoriatic arthritis (PsA), and other spondylarthritides (SpA) impose great impact on the individual in addition to the costs on society, which may be reduced by effective pharmacological treatment. Industry independent health economic studies should complement studies sponsored by industry.
Objective. To study secular trends in baseline health utilities in patients commencing TNF blockade for arthritis in clinical practice over 7 years; to address utility changes during treatment; to investigate the influence of previous treatment courses; to study the feasibility of health utility measures, and to compare them across diagnostic entities.
Methods. EuroQoL 5 Dimensions (EQ-5D) utility data were collected from a structured clinical follow-up program of anti-TNF treated patients with RA (N=2554), PsA (N=574) or SpA (N=586). Time trends were calculated. Completer analysis was used.
Results. There were weak or non-significant secular trends for increasing baseline utilities over time for RA, PsA and SpA. Maximum gain in utilities occurred already after 2 weeks for all diagnoses and remained stable for patients remaining on therapy. First and second anti-TNF courses performed similarly.
Conclusions. Utilities at inclusion remained largely unchanged for RA, PsA and SpA over 7 years. Improvement occurred early during treatment and not beyond 6 weeks at the group level. Improvement during the first course was not consistently greater than the second. There were no major differences between RA, PsA and SpA. EQ-5D proved feasible and applicable across these diagnoses. These “real world” data may be useful for health economic modelling.

Key words: Anti-TNF treatment; EuroQoL-5-dimensions; health economics; health utilities; observational study; psoriatic arthritis; quality of life; rheumatoid arthritis;; spondarthritis; time trends.

Abbreviations: CRP, C-reactive protein; DMARD, disease modifying antirheumatic drug; EQ-5D, EuroQoL 5 dimensions; HAQ, Health Assessment Questionnaire; IL-1, interleukin 1; PsA, psoriatic arthritis; QoL, quality of life; RA, rheumatoid arthritis; RCT, randomized controlled trial; SD, standard deviation; SSATG, South Swedish Arthritis Treatment Group; SpA, spondarthritis; TNF, tumour necrosis factor.

Introduction

Societal costs of rheumatoid arthritis (RA), psoriatic arthritis (PsA), and other spondylarthritides (SpA) are substantial [1], and indirect costs predominate. Effective treatment for these diseases, preventing disability, should therefore be beneficial for society. On the other hand, the new effective biologic therapies (blockers of TNF, IL-1, T-cell costimulation and B-cell depletors) are costly. Therefore, the cost effectiveness of these drugs has been subject to health economic studies[2 3]. However, many studies on cost effectiveness rely upon clinical trial data [4-6] with their limited generalizability, and not on observational data from daily clinical practice [7 8]. Furthermore, several diagnoses from the same setting are rarely reported, and health utilities are derived from measures such as the health assessment questionnaire (HAQ) [9 10] in RA [11 12].

Changes in the indications for treatment with biologics can be anticipated when used in clinical practice, as physicians become more familiar with them. This could result in secular changes in baseline utilities as well as change in their improvement during treatment. Indeed, we and others have reported changes in baseline characteristics towards lower disease activity and disability at start of first treatment during the first three years after introducing anti-TNF therapy[13 14].

We have reported the costs and health economic benefits associated with early anti-TNF therapy for RA in clinical practice [7]. However, we had to use sensitivity analyses for cost effectiveness estimations, since we did not have multiple measurements during the first treatment year, and we also lacked utility values beyond this time. Another aspect not scrutinized in the study was that patients switched between different expensive biologic drugs (up to 35% in our setting)[14].

We undertook the present study on patients with RA, PsA, and other SpA treated in clinical practice in southern Sweden, with four specific goals:

- To determine if health utilities at initiation of anti-TNF treatment changed over the 7 year period between May 2002 and December 2008;
- To address changes in this measure during anti-TNF treatment both in the short and long perspective, including reasons for drug withdrawal;
- To study if previous biological therapy influenced this measure;
- To study the feasibility of the preference based health utility instrument EuroQoL-5-dimensions (EQ-5D) and apply it across different diagnostic entities.

Materials & Methods

Anti-TNF treatment courses for patients with RA, PsA and SpA according to the treating physician were retrieved from the South Swedish Arthritis Treatment Group Registry[15-17]. In 2002, collecting health utility data was introduced in routine clinical follow up. Data were collected using the five descriptive questions of the EuroQol 5 Dimension (EQ-5D)[18]. The visual analogue scale of the EQ-5D was not used. From this generic preference based instrument, utility values can be derived with a range from death (0) to full health (1), with values below 0 (-0.56) being possible[19 20]. The dimensions covered by the EQ-5D include mobility, self-care, usual activities, pain/discomfort, and anxiety/depression.

Patients eligible for this study had a diagnosis of RA, PsA, or SpA according to the opinion of the treating physician. Patients with PsA and SpA were further classified as having peripheral joint disease (arthritis distal of shoulder and hips ever) and/or clinical signs of spondylitis by the treating physician[21]. The different diagnoses and classifications have been validated in large groups of patients and found to be accurate in between 90-98 %[16 21] according to established criteria[22-25]. The patients were identified in the SSATG registry during the period May 2002 through December 2008 as starting a treatment course of infliximab, etanercept, or adalimumab. Treatment courses were classified as either first, second, or third or more anti-TNF. All EQ-5D utility values at treatment follow up time points 0, 0.5, 1.5, 3, 6, 12, 24, 30, 36, 48, and 60 months were retrieved from the database. Treatment courses lacking baseline EQ-5D were excluded. Anti-TNF treatments were grouped according to year of initiation for analyses of time trends at treatment onset. Reason for drug withdrawal was documented in the SSATG protocol as failure, adverse event, or other, but only one stop reason could be given. Distinction between primary failure (never having a response) and secondary failure (having an initial response, with deterioration later), was not always possible, and they were therefore grouped together. Other stop reasons include among others pregnancy, switching for convenience, or remission. Missing follow up data were requested from treating physicians 1-2 times per year, including possible withdrawal reason.

To assess feasibility of EQ-5D in the current observational setting, the number of follow up visits with full EQ-5D information was compared to those with data on HAQ.

Statistics: Values are given with mean and 95 % confidence interval (CI). Follow up times are given with median (range), and mean (SD) values. Generally, only observations with at least 20 valid N are presented. Patients remaining on therapy at given follow up time points were estimated from Kaplan-Meier plots. Patients with full data sets at time 0, 3, 6, 12, and 24 months were compared with all patients to see if there were relevant differences between those with complete and incomplete data.

Results

Baseline patient characteristics according to anti-TNF treatment and diagnosis are shown in table 1. Most patients receiving their second, third or more course of TNF-blocker were included in the first course group, thus making direct statistical comparisons of the patient groups hazardous. There were several differences between the diagnostic groups. RA patients were older, had tried more DMARDs, were more often treated with concomitant DMARDs, and were more often female. Overall, patients subject to more than one anti-TNF drug tended to be older, have longer disease duration and they were less often treated with concomitant methotrexate. As expected, clinical signs of spondylitis were more prevalent in the SpA group (77%); many patients belonged to the undifferentiated SpA entity. Clinical spondylitis was also present in almost 30 % of PsA patients.

The secular trends for baseline EQ-5D utility values at first anti-TNF treatment and the different diagnoses are illustrated by figure 1. There were weak, non significant trends for increasing baseline utility values for RA patients (Spearman's ρ=0.03, P=0.23), PsA (ρ=0.04, P=0.37), and SpA (ρ=0.05, P=0.29) over time.

The development of EQ-5D utilities at first, second, and third or more anti-TNF treatment course for RA, PsA and SpA patients is illustrated in figure 2A-C. For RA patients, utility

improvement during the first and second anti-TNF treatments perform in a similar way. The third or more anti-TNF treatments start from a lower utility level and groups are smaller with wider CIs, but nevertheless they perform with about the same numerical improvement as for first and second anti-TNF treatments.

Most gain in EQ-5D utilities was achieved already after 2 weeks for both first and second anti-TNF treatments for RA and PsA at the group level, while the SpA patients had a somewhat slower initial improvement.

RA patients stopping therapy demonstrated lower utility gain regardless of reason for withdrawing treatment (figure 3). Utility improvement in cases with stop reasons other than adverse event or failure seems to perform like those remaining on treatment, but numbers are limited.

To assess the feasibility of EQ-5D in clinical practice, we compared the frequency of complete EQ-5D and HAQ questionnaires. Total number of follow up visits were 12585, 2553, and 2630 with presence of HAQ values in 98%, 97%, 97%, and presence of EQ-5D values in 93%, 94%, and 94% for RA, PsA and SpA, respectively.

To investigate the possibility of bias in patients with missing values, we compared the total amount of information for RA patients at time points 0, 3, 6, 12, and 24 months with RA patients with complete data sets at all these time points, grouped according to treatment order (figure 4). The pattern of improvement is similar regardless of data set completeness, but the magnitude of improvement is somewhat higher at some time points in first anti-TNF courses with complete data sets.

To facilitate health economic modelling, we calculated median/mean follow up time and life-table estimates of drug survival, in relation to diagnoses, anti-TNF treatment sequences, and stop reasons (On-line supplemental Table 1, Figure 2A-C). Expectedly, patients with ongoing treatment had longer follow up, while those stopping because of adverse event had the shortest. Treatment courses terminated due to low response and failure had follow up times close to those due to adverse event. Other stop reasons were less common but resulted in longer follow up times. Overall follow up time decreased with increasing anti-TNF number, and follow up times were skewed towards early withdrawal as indicated by lower median compared to mean values.

Discussion

A major finding in this study was the rapid improvement in health utilities already after 2 weeks in RA patients treated with their first anti-TNF drug. The improvement was maintained for at least 5 years for patients remaining on therapy. Baseline utilities remained relatively stable during the period 2002-2008 for the first anti-TNF drug. This was somewhat unexpected, given our previously reported steady improvement of both disability (HAQ) and disease activity (DAS28) levels during 1999-2006[14]. Interestingly, there were no major differences between different chronic arthritis diagnoses, whereas there was a trend for lower baseline utilities with increasing number of anti-TNF drugs. However, limited number for third course anti-TNF in PsA and SpA patients precludes firm conclusions at present.

EQ-5D was chosen due to its simplicity, patient acceptability, and well established utilities. It is well suited for measuring diseases mainly involving locomotor organs, including dimensions such as pain, mobility, self-care, and usual activities, all of which are important in inflammatory joint diseases. We have found the visual analogue scale of the EQ-5D less suitable with low patient acceptability in clinical practice, and the core set already included two VAS scales, one of pain and one of global disease activity. Our findings of only 6-7 percent missing health utilities compared to 2-3% for HAQ scores confirms the feasibility of the EQ-5D instrument in our observational setting.

EQ-5D is a generic measure thus intended for comparing various diseases. Our findings support this. This type of comparison had not been possible using disease specific measures such as the HAQ, RA-QoL, and PsAQoL[9 27-29]. Although VAS scales have been used as surrogates in health economic models[29], it should be better to use instruments with established health utilities. However, EQ-5D entails several subjective judgements made by the patients, and therefore it has to be complemented with more objective measures before making decisions regarding start or change of biological treatment.

Observational data like ours are more generally applicable as a reference for health economic modelling than RCT data, which are derived from highly selected patients[5 30 31]. Furthermore, in Sweden there are no formal requirements for inclusion or response, few economic restrictions, and drug costs are almost entirely funded by society. This may result in more missing data, but data may be less biased towards worse utility and disease activity.

Major strengths of the present study is that the variables have been prospectively collected and the setting can be regarded as truly population based[17]. It is also, to our knowledge, the first report giving comprehensive data on the development of EQ-5D utility over 7 years for patients with RA, PsA and SpA in a clinical setting. Our findings are in line with utility gain in RCTs of TNF blockers in RA[32 33] and AS[34] and an observational study of PsA[35].

Our investigation also has limitations. Firstly, it is difficult to obtain complete sets of data in the observational setting. Using all available data increases generalizability. This, however, will yield lower improvement estimates as compared to including only subjects with complete follow up information from all visits (Figure 4). Thus, there may be a possible bias if complete data sets are required. Incompleteness could be due to either withdrawal from therapy or missing reports for other reasons. Withdrawal may be the main reason why gain in health utility is less when using all available data as compared to only subjects with complete follow-up information. Those stopping therapy, irrespective of cause (lack of effect or adverse event), had less improvement in health utility prior to the stop (Figure 3). More emphasis should be put on overall withdrawal rate than on stop reason, since insufficient effect may lower the threshold for stopping treatment due to a mild adverse event.[21 36]. Secondly, regular follow up is a prerequisite for good data provision. This can be a problem in a voluntary multi-centre observational setting such as ours, where health care is provided in organisations changing over time for economical, political, or other reasons. Even so, the professional SSATG network has remained stable over the last 10 years, and we have not been able to identify any major bias in the missing follow up data which would seriously impair our conclusions.

Previously, when we only had baseline and 1 year measures, we had to make sensitivity analyses as to when the actual improvement occurred[7]. This can now be simplified. In the present dataset, the gain can be regarded as almost instantaneous and steady over the years

after anti-TNF institution (figure 2A-C), thereby facilitating the calculation of gain in quality adjusted life years (QALYs). However, it must be kept in mind that these calculations are derived from patients remaining on therapy, and therefore selected as good responders. Drug continuation can vary substantially between different treatment remedies in our setting [37], and this must be accounted for in health economic modelling. We consider the data in Figure 2A-C and on-line supplemental Table 2 fairly robust due to the active and regular search for treatment withdrawal reason when follow-up data are missing[16 37].

Our data illustrate, that irrespective of guidelines, there are trends regarding the baseline characteristics of patients that are started on biologics[14]. These trends seem less obvious for health utility measures, but long term follow up, as in the present study, is needed in addition to analyses of RCTs and shorter observational studies to establish true utility gain in the clinical setting.

In conclusion, this study demonstrates a rapid gain in EQ-5D utility after initiation of anti-TNF-treatment in chronic arthritis, irrespective of diagnosis, and that this changes little in the subsequent five years for patients remaining on therapy. The EQ-5D utility used to describe improvement is feasible in this observational population, and the similar results obtained for various diagnoses support its generic character.

Acknowledgements
We are indebted to all colleagues and staff in the South Swedish Arthritis Treatment Group for cooperation and data supply and to Jan-Åke Nilsson for help with statistical calculations. This study was supported by grants from Österlund and Kock Foundations, King Gustav V 80 year fund, Lund University Hospital, Region Skåne, Faculty of Medicine, Lund University and Reumatikerförbundet.

Legends

Figure 1.
EQ-5D at first treatment initiation 2002-2008 for RA, PsA and SpA patients.

Figure 2A.
EQ-5D during follow up for RA patients, starting anti-TNF 2002-2008 and with baseline EQ-5D values.

Figure 2B.
EQ-5D during follow up for PsA patients, starting anti-TNF 2002-2008 and with baseline EQ-5D values.

Figure 2C.

EQ-5D during follow up for SpA patients, starting anti-TNF 2002-2008 and with baseline EQ-5D values.

Figure 3.
EQ-5D during follow up – stop reason. RA patients starting first anti-TNF 2002-2008 and with baseline EQ-5D values.

Figure 4.
EQ-5D during follow up for RA patients. All patients and patients with complete data.

Tables

Table 1.
Patients characteristic at treatment initiation.

On-line supplemental Table 1.
Follow up time in months. Values are given as **median ; mean** (standard deviation), range; number.

References

1. Kavanaugh A. Health economics: implications for novel antirheumatic therapies. *Ann Rheum Dis* 2005;64 Suppl 4:iv65-9.
2. Emery P. Review of health economics modelling in rheumatoid arthritis. *Pharmacoeconomics* 2004;22(2 Suppl 1):55-69.
3. Fleurence R, Spackman E. Cost-effectiveness of biologic agents for treatment of autoimmune disorders: structured review of the literature. *J Rheumatol* 2006;33(11):2124-31.
4. Kobelt G, Jonsson L, Young A, Eberhardt K. The cost-effectiveness of infliximab (Remicade) in the treatment of rheumatoid arthritis in Sweden and the United Kingdom based on the ATTRACT study. *Rheumatology (Oxford)* 2003;42(2):326-35.
5. Kobelt G, Lindgren P, Singh A, Klareskog L. Cost effectiveness of etanercept (Enbrel) in combination with methotrexate in the treatment of active rheumatoid arthritis based on the TEMPO trial. *Ann Rheum Dis* 2005;64(8):1174-9.
6. Kobelt G, Sobocki P, Sieper J, Braun J. Comparison of the cost-effectiveness of infliximab in the treatment of ankylosing spondylitis in the United Kingdom based on two different clinical trials. *Int J Technol Assess Health Care* 2007;23(3):368-75.
7. Kobelt G, Eberhardt K, Geborek P. TNF inhibitors in the treatment of rheumatoid arthritis in clinical practice: costs and outcomes in a follow up study of patients with RA treated with etanercept or infliximab in southern Sweden. *Ann Rheum Dis* 2004;63(1):4-10.
8. Kievit W, Adang EM, Fransen J, Kuper HH, van der Laar MA, Jansen TL, et al. The effectiveness and medication costs of three anti-TNF agents{alpha} in the treatment of rheumatoid arthritis from prospective clinical practice data. *Ann Rheum Dis* 2008.
9. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. *Arthritis Rheum* 1980;23(2):137-45.

10. Ekdahl C, Eberhardt K, Andersson SI, Svensson B. Assessing disability in patients with rheumatoid arthritis. Use of a Swedish version of the Stanford Health Assessment Questionnaire. *Scand J Rheumatol* 1988;17(4):263-71.
11. Jacobsson LT, Lindroth Y, Marsal L, Juran E, Bergstrom U, Kobelt G. Rheumatoid arthritis: what does it cost and what factors are driving those costs? Results of a survey in a community-derived population in Malmo, Sweden. *Scand J Rheumatol* 2007;36(3):179-83.
12. Kobelt G, Jonsson L, Lindgren P, Young A, Eberhardt K. Modeling the progression of rheumatoid arthritis: a two-country model to estimate costs and consequences of rheumatoid arthritis. *Arthritis Rheum* 2002;46(9):2310-9.
13. Hetland ML, Lindegaard HM, Hansen A, Podenphant J, Unkerskov J, Ringsdal VS, et al. Do changes in prescription practice in patients with rheumatoid arthritis treated with biologics affect treatment response and adherence to therapy? Results from the nationwide Danish Danbio Registry. *Ann Rheum Dis* 2008.
14. Söderlin MK GP. Changing pattern in the prescription of biological treatment in rheumatoid arthritis. A 7-year follow-up of 1839 patients in southern Sweden. *Ann Rheum Dis*;67:37-42.
15. Geborek P, Saxne T. Clinical protocol for monitoring of targeted therapies in rheumatoid arthritis. *Rheumatology (Oxford)* 2000;39(10):1159-61.
16. Geborek P, Crnkic M, Petersson IF, Saxne T. Etanercept, infliximab, and leflunomide in established rheumatoid arthritis: clinical experience using a structured follow up programme in southern Sweden. *Ann Rheum Dis* 2002;61(9):793-8.
17. Geborek P, Nitelius E, Noltorp S, Petri H, Jacobsson L, Larsson L, et al. Population based studies of biological antirheumatic drug use in southern Sweden: comparison with pharmaceutical sales. *Ann Rheum Dis* 2005;64(12):1805-7.
18. EuroQol--a new facility for the measurement of health-related quality of life. The EuroQol Group. *Health Policy* 1990;16(3):199-208.
19. Hurst NP, Kind P, Ruta D, Hunter M, Stubbings A. Measuring health-related quality of life in rheumatoid arthritis: validity, responsiveness and reliability of EuroQol (EQ-5D). *Br J Rheumatol* 1997;36(5):551-9.
20. Dolan P GC, Kind P, Williams A. A social tariff for EuroQol: Results from a UK population survey. York, UK: Centre for Health Economics, University of York, 1995.
21. Gülfe A, Kristensen LE, Geborek P. Six and twelve weeks response predicts continuation of TNF-blockade in rheumatoid arthritis. Observational cohort study from southern Sweden. *J Rheumatol* 2009, accepted for publication.
22. van der Linden S, Valkenburg HA, Cats A. Evaluation of diagnostic criteria for ankylosing spondylitis. A proposal for modification of the New York criteria. *Arthritis Rheum* 1984;27(4):361-8.
23. Dougados M, van der Linden S, Juhlin R, Huitfeldt B, Amor B, Calin A, et al. The European Spondylarthropathy Study Group preliminary criteria for the classification of spondylarthropathy. *Arthritis Rheum* 1991;34(10):1218-27.
24. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 1988;31(3):315-24.
25. Bennett PM WP. Population studies of the rheumatic diseases. *Amsterdam International Congress Series* 1966(148):148.
26. Kristensen LE, Saxne T, Geborek P. The LUNDEX, a new index of drug efficacy in clinical practice: results of a five-year observational study of treatment with infliximab and etanercept among rheumatoid arthritis patients in southern Sweden. *Arthritis Rheum* 2006;54(2):600-6.

27. McKenna SP, Doward LC, Whalley D, Tennant A, Emery P, Veale DJ. Development of the PsAQoL: a quality of life instrument specific to psoriatic arthritis. *Ann Rheum Dis* 2004;63(2):162-9.
28. Whalley D, McKenna SP, de Jong Z, van der Heijde D. Quality of life in rheumatoid arthritis. *Br J Rheumatol* 1997;36(8):884-8.
29. Kobelt G, Lindgren P, Lindroth Y, Jacobson L, Eberhardt K. Modelling the effect of function and disease activity on costs and quality of life in rheumatoid arthritis. *Rheumatology (Oxford)* 2005;44(9):1169-75.
30. Bansback NJ, Young A, Brennan A. The NICE reappraisal of biologics in 2005: what rheumatologists need to know. *Rheumatology (Oxford)* 2005;44(1):3-4.
31. Bansback NJ, Brennan A, Ghatnekar O. Cost effectiveness of adalimumab in the treatment of patients with moderate to severe rheumatoid arthritis in Sweden. *Ann Rheum Dis* 2005;64(7):995-1002.
32. 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(11):1478-83.
33. van Riel PL, Freundlich B, MacPeek D, Pedersen R, Foehl JR, Singh A. Patient-reported health outcomes in a trial of etanercept monotherapy versus combination therapy with etanercept and methotrexate for rheumatoid arthritis: the ADORE trial. *Ann Rheum Dis* 2008;67(8):1104-10.
34. Braun J, McHugh N, Singh A, Wajdula JS, Sato R. Improvement in patient-reported outcomes for patients with ankylosing spondylitis treated with etanercept 50 mg once-weekly and 25 mg twice-weekly. *Rheumatology (Oxford)* 2007;46(6):999-1004.
35. Olivieri I, de Portu S, Salvarani C, Cauli A, Lubrano E, Spadaro A, et al. The psoriatic arthritis cost evaluation study: a cost-of-illness study on tumour necrosis factor inhibitors in psoriatic arthritis patients with inadequate response to conventional therapy. *Rheumatology (Oxford)* 2008;47(11):1664-70.
36. Zink A, Listing J, Kary S, Ramlau P, Stoyanova-Scholz M, Babinsky K, et al. Treatment continuation in patients receiving biological agents or conventional DMARD therapy. *Ann Rheum Dis* 2005;64(9):1274-9.
37. Kristensen LE, Saxne T, Nilsson JA, Geborek P. Impact of concomitant DMARD therapy on adherence to treatment with etanercept and infliximab in rheumatoid arthritis. Results from a six-year observational study in southern Sweden. *Arthritis Res Ther* 2006;8(6):R174.

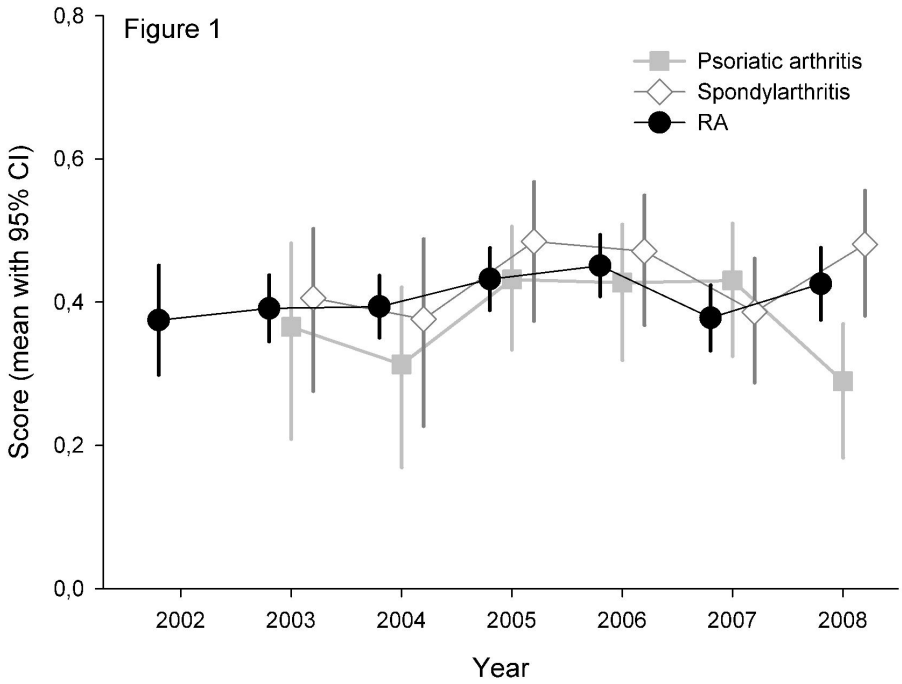
Table 1
Baseline characteristics in the different diagnostic and treatment groups

	RA			Psoriatic arthritis			Spondylarthritis		
	1st anti-TNF	2nd anti-TNF	>2 anti-TNF	1st anti-TNF	2nd anti-TNF	>2 anti-TNF	1st anti-TNF	2nd anti-TNF	>2 anti-TNF
Valid N	1584	742	228	401	135	38	430	117	39
Male/Female (%)	23/77	19/81	17/83	51/49	41/59	34/66	60/40	54/46	58/42
Age years (Mean 95% CI)	55.0 (54.2-55.8)	55.1 (53.8-56.4)	57.0 (54.6-59.4)	46.8 (45.6-48.0)	48.4 (46.1-50.7)	52.0 (48.1-56.0)	43.2 (42.0-44.3)	43.9 (41.7-46.2)	47.0 (43.1-50.8)
Disease duration (Mean 95% CI)	9.8 (9.2-10.4)	12.1 (11.0-13.1)	14.2 (12.2-16.3)	10.1 (9.3-11.0)	12.0 (10.5-13.4)	16.0 (12.6-19.4)	13.1 (12.0-14.2)	16.1 (13.9-18.2)	16.3 (12.6-20.0)
Previous DMARD number# (Mean 95% CI)	2.3 (2.2-2.3)	4.0 (3.8-4.2)	5.7 (5.3-6.0)	1.7 (1.6-1.8)	3.0 (2.8-3.1)	4.6 (4.0-5.2)	1.4 (1.3-1.5)	2.9 (2.7-3.1)	4.5 (4.0-4.9)
Ongoing DMARD number (Mean 95% CI)	1.1 (1.1-1.1)	1.0 (1.0-1.1)	1.0 (1.0-1.1)	0.7 (0.7-0.8)	0.6 (0.5-0.7)	0.5 (0.3-0.7)	0.6 (0.5-0.7)	0.5 (0.4-0.6)	0.5 (0.4-0.7)
Ongoing Methotrexate (%)	68.1	52.5	51.3	62.8	54.1	39.5	42.6	41.9	43.6
Etanercept (%)	46.8	51.8	35.1	48.4	57.8	31.6	40.7	43.6	17.9
Adalimumab (%)	26.6	40.7	49.6	23.2	37.0	52.6	20.5	45.3	61.5
Infliximab (%)	26.6	7.5	15.4	28.4	5.2	15.8	38.8	11.1	20.5
Clinical spondylitis ever (%)				29.4	29.6	47.4	64.2	69.2	69.2
Peripheral arthritis ever (%)				66.3	77.0	88.8	44.0	48.7	64.1
Definite ankylosing spondylitis (%)				2.7	1.5	0	46.3	51.3	48.7

#Includes previous biologics

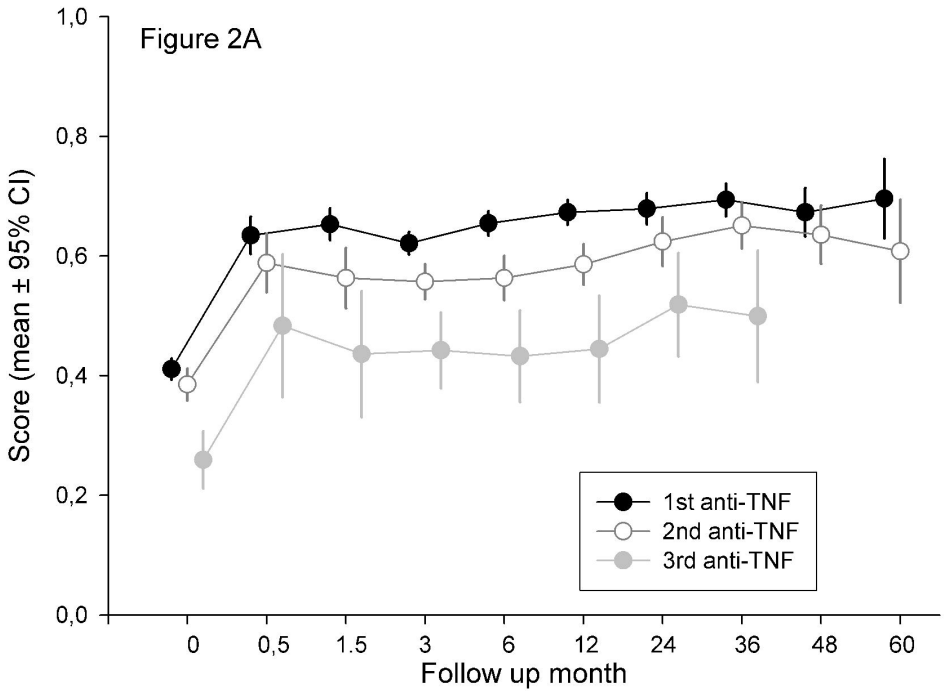
Table 2
Follow up time in months. Values are given as **median ; mean** (standard deviation) min-max ; number

	All treatments	Ongoing treatment	Stop any reason	Stop adverse event	Stop failure	Stop other reasons
RA						
1st anti-TNF	**18 ; 23.7** (20.1) 0-78 ; 1584	**30 ; 32.0** (20.5) 0-78 ; 915	**7 ; 12.2** (12.2) 0-63 ; 669	**6 ; 11.1** (12.0) 0-61 ; 311	**8 ; 12.2** (11.3) 0-59 ; 274	**11.5 ; 16.6** (14.8) 1-63 ; 84
2nd anti-TNF	**21.5 ; 25.1** (21.6) 0-79 ; 742	**35 ; 34.8** (21.5) 1-79 ; 429	**7 ; 11.7** (12.9) 0-67 ; 313	**6 ; 11.0** (12.9) 0-67 ; 139	**6 ; 11.0** (12.9) 0-67 ; 139	**14 ; 20.1** (18.0) 1-58 ; 23
>2 anti-TNF	**13 ; 19.7** (18.5) 0-69 ; 228	**30 ; 31.1** (18.6) 1-69 ; 113	**5 ; 8.6** (9.4) 0-47 ; 115	**4 ; 7.0** (7.6) 0-39 ; 53	**5 ; 10.1** (10.5) 1-47 ; 53	**5 ; 9.4** (11.7) 0-39 ; 9
Psoriatic arthritis						
1st anti-TNF	**15 ; 21.1** (17.7) 0-75 ; 401	**25 ; 27.0** (18.4) 0-75 ; 257	**7 ; 10.5** (10.0) 0-58 ; 144	**5 ; 9.4** (11.2) 0-58 ; 61	**7 ; 10.2** (8.6) 1-40 ; 62	**13 ; 17.7** (14.6) 0-65 ; 22
2nd anti-TNF	**13 ; 18.7** (17.8) 0-79 ; 135	**13 ; 18.7** (17.8) 0-79 ; 135	**6 ; 8.5** (7.35) 0-35 ; 50	**5.5 ; 8.4** (8.7) 0-35 ; 24	**6 ; 8.0** (5.4) 1-19 ; 23	**13 ; 18.1** (16.0) 4-47 ; 6
>2 anti-TNF	**6.5 ; 12.6** (14.5) 0-54 ; 38	**15 ; 18.6** (17.4) 1-54 ; 19	**4 ; 6.6** (7.14) 0-29 ; 19	**3 ; 4.5** (4.45) 1-15 ; 10	**6.5 ; 10.0** (8.9) 2-29 ; 8	**11.5 ; 12** (15.5) 1-23 ; 2
Spondylarthritis						
1st anti-TNF	**18 ; 22.0** (19.0) 0-77 ; 430	**24 ; 26.7** (19.7) 1-77 ; 302	**8 ; 10.9** (11.1) 0-65 ; 128	**6 ; 7.6** (9.0) 0-45 ; 49	**8 ; 11.2** (10.1) 1-42 ; 57	**15 ; 15.2** (9.6) 1-38 ; 21
2nd anti-TNF	**11 ; 18.5** (18.6) 0-69 ; 117	**20 ; 27.0** (19.9) 1-69 ; 64	**5 ; 8.2** (9.78) 0-47 ; 53	**4 ; 6.6** (7.7) 0-29 ; 17	**4 ; 7.1** (8.4) 0-42 ; 30	**15.5 ; 13.3** (9.5) 4-23 ; 3
>2 anti-TNF	**15 ; 18.3** (15.4) 0-56 ; 39	**20 ; 23.4** (15.9) 0-56 ; 24	**7 ; 10.2** (10.8) 0-35 ; 15	**10 ; 8.2** (10.9) 0-21 ; 6	**7 ; 11.5** (11.3) 2-35 ; 7	N=0



Number Initiated/Number with baseline EQ-5D values

RA	119/68	232/204	254/222	260/209	254/224	268/218	201/171
Spondylarthritis	24/10	45/35	47/42	59/54	79/73	89/77	93/79
Psoriatic arthritis	18/14	28/25	47/41	87/76	70/61	80/66	71/63

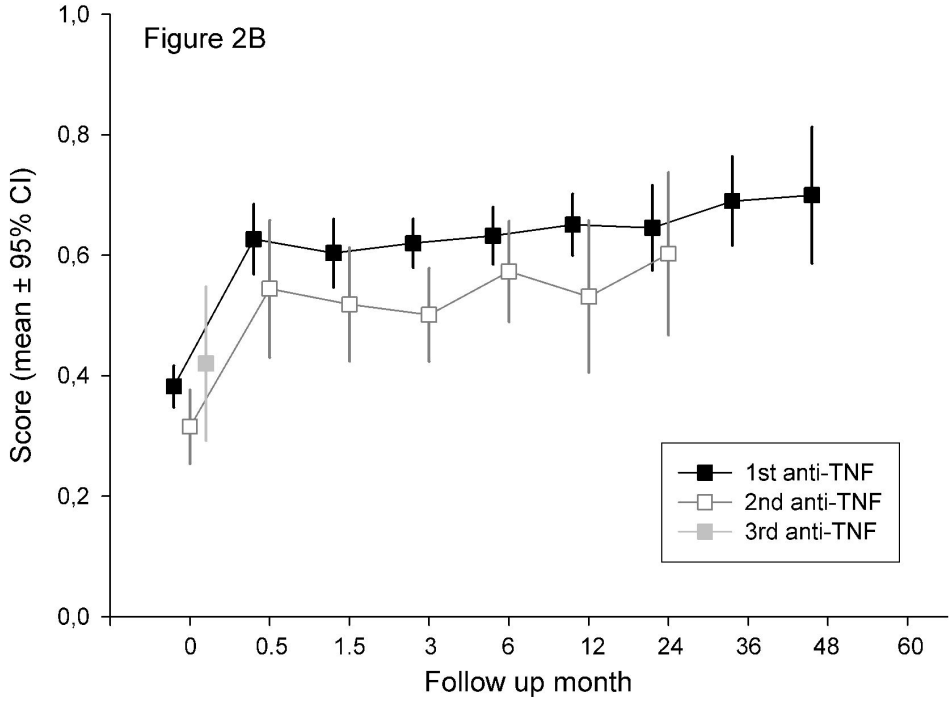


Valid N (number of EQ-5D values)

1st antiTNF	1321	237	335	859	607	532	346	264	150	59
2nd anti TNF	631	111	152	377	274	249	167	153	97	29
3rd anti-TNF	193	32	45	110	75	63	43	31	17	4

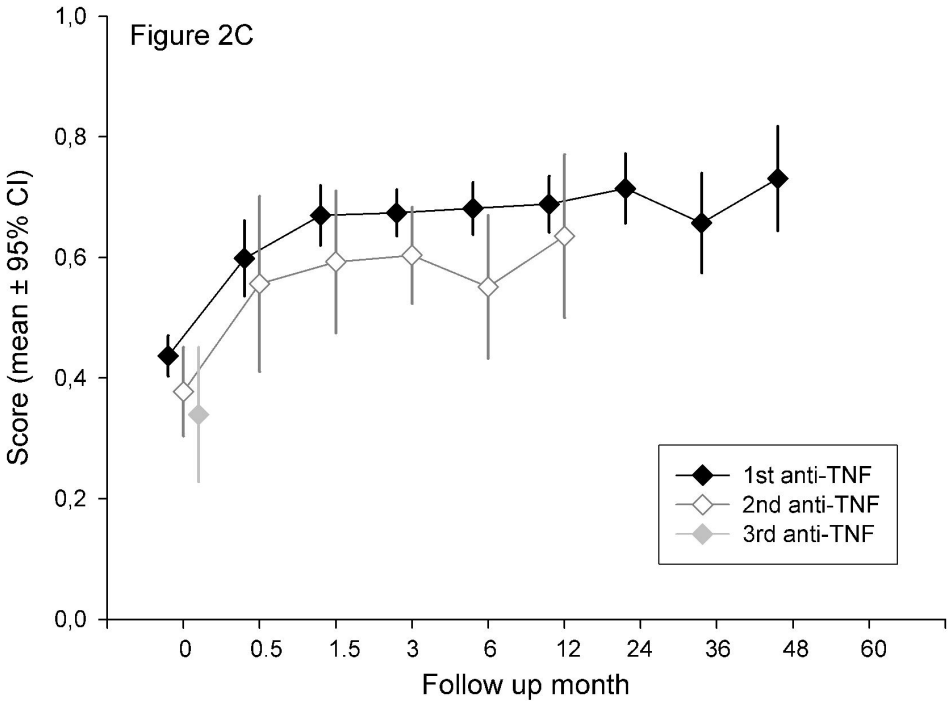
Theoretical proportion remaining on therapy (%) (Kaplan-Meyer estimates)

1st antiTNF	100	98.3	95.9	88.9	79.4	68.4	56.5	50.5	45.8	43.4
2nd anti TNF	100	93.9	90.3	85.0	78.7	68.3	59.5	54.6	50.4	47.7
3rd anti-TNF	100	92.3	88.7	80.3	67.6	57.7	48.6	45.4	41.6	



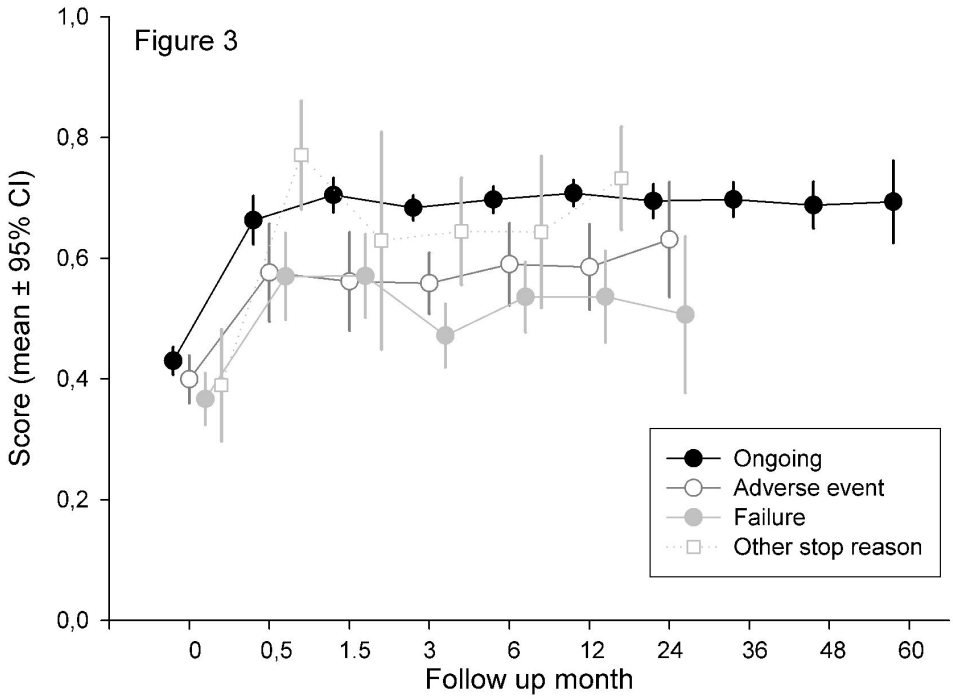
Valid N (number of EQ-5D values)										
1st antiTNF	344	72	99	216	144	136	80	58	22	6
2nd anti TNF	118	27	38	68	40	31	21	14	7	2
3rd anti-TNF	30	11	9	14	9	7	1	2	1	0

Theoretical proportion remaining on therapy (%) (Kaplan-Meyer estimates)										
1st antiTNF	100	99.2	94.9	90.2	81.3	69.9	59.8	54.7	52.6	50.1
2nd anti TNF	100	95.1	91.4	84.0	72.8	63.0	46.9	54.6		
3rd anti-TNF	100	94.7	73.7	63.2						



Valid N (number of EQ-5D values)										
1st antiTNF	372	87	113	230	150	139	80	53	21	8
2nd anti TNF	99	22	33	58	34	25	15	15	5	2
3rd anti-TNF	37	7	12	18	8	9	5	4	1	0

Theoretical proportion remaining on therapy (%) (Kaplan-Meyer estimates)										
1st antiTNF	100	98.1	93.6	91.9	88.4	76.8	68.3	61.4	59.1	59.1
2nd anti TNF	100	98.5	89.4	83.3	71.2	62.1	51.5			
3rd anti-TNF	100	95.2	90.5	85.7	85.7	52.4				



Valid N (number of EQ-5D values)										
Ongoing	772	138	204	515	401	382	288	234	140	57
Adverse event	255	41	55	148	79	66	24	13	3	1
Failure	231	47	64	157	97	64	23	10	2	0
Other	63	11	12	39	30	20	11	7	5	1

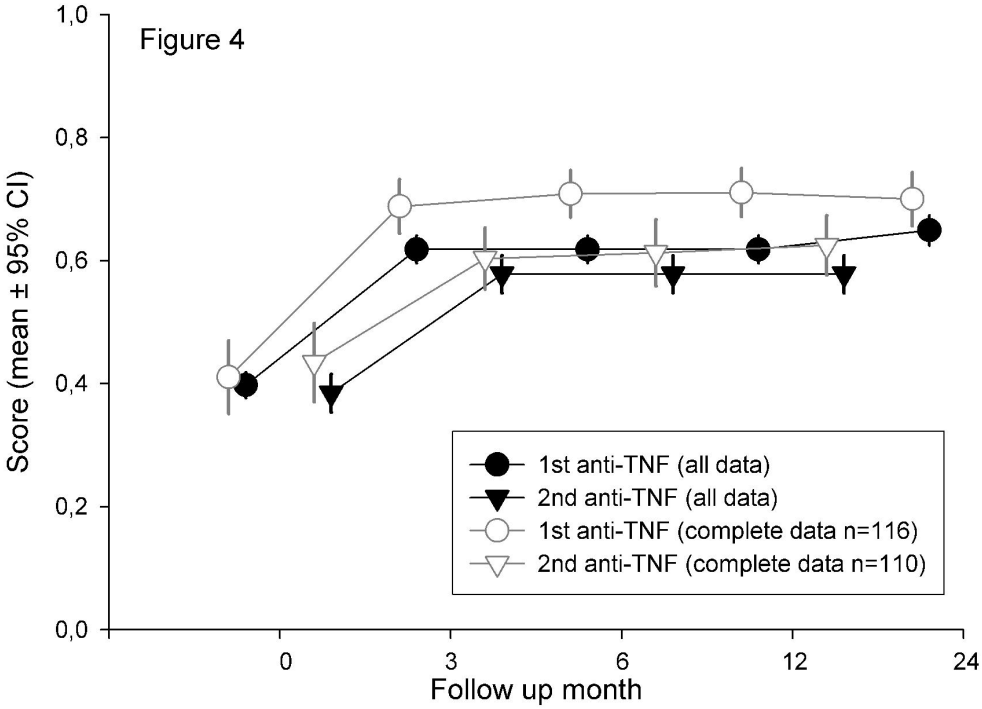


Figure 4
1,0
0,8
0,6
0,4
0,2
0,0
Score (mean ± 95% CI)
0
3
6
12
24
Follow up month
1st anti-TNF (all data)
2nd anti-TNF (all data)
1st anti-TNF (complete data n=116)
2nd anti-TNF (complete data n=110)

V

Utility-based outcomes made easy: The Number Needed per QALY gained (NNQ). Observational cohort study from southern Sweden of TNF blockade in inflammatory arthritis

Anders Gülfe, MD#, Lars Erik Kristensen, MD, PhD#, Tore Saxne, MD, PhD#, Lennart TH Jacobsson, MD, PhD##, Ingemar F Petersson, MD, PhD###, Pierre Geborek, MD, PhD#

Objective. To introduce a novel, simple, utility based outcome measure, the Number Needed per Quality adjusted life year (QALY) gained (NNQ), and to apply it in clinical practice in anti-TNF treated patients with rheumatoid arthritis (RA), psoriatic arthritis (PsA), and spondylarthritis (SpA).

Methods. The NNQ is the number of patients one has to treat in order to gain 1 QALY. It is calculated as the inverted value of the utility gain (area under curve) over 1 year in a cohort subjected to an intervention. EuroQoL-5-dimensions (EQ-5D) utility data from the South Sweden Arthritis Treatment Registry was used.

Results. 1001 RA, 241 PsA, and 255 SpA patients were eligible for the study. First, 2nd and 3rd treatment courses were studied. For RA, NNQ was 4.5, 6.4 and 5.2 for 1st, 2nd and 3rd courses, respectively. For PsA and SpA, NNQ was 4.2-4.5 irrespective of treatment order. Treatment groups with N<50 were not analysed. During the study period 2002-2007, there were no secular trends of utility gains.

Conclusion. We found NNQ to be a simple and easily understood group level, utility based outcome measure that worked well across 3 arthritis diagnoses. NNQ varied little over diagnoses and treatment course order, with a possible exception in 2nd treatment course in RA.

#Dept of Clinical Sciences, Lund, Section for Rheumatology, Lund University, Sweden
##Dept of Clinical Sciences. Malmö, Section for Rheumatology, Lund University, Sweden
###South Sweden Musculoskeletal Research Centre, Dept of Orthopedics, Lund University Hospital, Lund, Sweden

Manuscript

Abbreviations. AUC, area under curve; CDAI, clinical disease activity index; DAS, disease activity score; EQ-5D, EuroQoL-5-dimensions; HAQ, health assessment questionnaire; HRQoL, health related quality of life; HUI, health utility index; LOCF, last observation carried forward; NNQ, number needed (to treat) per QALY gained; NNT, number needed to treat; PAS, patient activity scale; PsA, psoriatic arthritis; QALY, quality adjusted life year; RA, rheumatoid arthritis; RAID, RA impact of disease score; RCT, randomised controlled trial; SDAI, simple disease activity index; SG, standard gamble; SpA, spondylarthritis; TNF, tumour necrosis factor; TTO, time trade off.

Introduction

In the current trends of new and costly modalities of arthritis treatment, interest in treatment evaluation from a health related quality of life (HRQoL) and economic point of view is increasing. Health economic studies generally involves gathering of real costs and complicated mathematical models, and they are seldom very transparent. There seems to be a need for a simple and intuitive measure for the extent to which an intervention is worth while. In an attempt to fill this need, we propose a new, utility based outcome measure, the number needed (to treat) per QALY gained, the NNQ.

A number of composite activity indices and response criteria have been devised to evaluate

treatment of inflammatory arthritis, some of them disease specific, others generic. A few, such as the Stanford Health Assessment Questionnaire (HAQ)(1) were developed for a specific diagnosis (rheumatoid arthritis, RA) but have also been applied in other diseases (2, 3) and by some it has even been suggested to represent a generic measure of function(4). Many activity indices, like the Disease Activity Score (DAS)(5) and Simple Disease Activity Index (SDAI)(6) consist of patient and evaluator derived measures as well as of a laboratory measure of inflammation, while others are comprised solely of patient derived data, such as the Patient Activity Scale (PAS) or Rheumatoid Arthritis Impact of Disease score (RAID)(7, 8). All of these are dependent both on inflammation and tissue damage (joint damage/erosions). In the case of HAQ disability index, the relative importance of inflammation versus damage has been quantified in an RA cohort using randomised controlled trial (RCT) data (9).

From a biologic and theoretical standpoint, the separation of inflammation and tissue damage is pivotal to the understanding of function in inflammatory arthritis. Clinically, however, as well as from the patient perspective, a broader and "softer" concept of perceived health, such as HRQoL(10), may reflect important aspects of the disease process not covered by the usual activity and function measures. HRQoL may be measured directly, by asking individuals to estimate their life quality relative to an ideal, perfect state by use of hypothetical or real scenarios like standard gamble (SG) and time trade-off (TTO), or a visual analogue scale (VAS) may be employed. SG and TTO appear to work better in situations like surgery or terminal illness than in chronic disease(11). In chronic diseases, indirect methods involving questionnaires with health related items are therefore widely used, for example the EuroQoL-5-dimensions (EQ-5D)(12-14), the Short Form-6-dimensions (SF-6D)(15) and the Health Utility Index (HUI)(16). All these instruments are generic. Whereas utility values derived from SG or TTO relate to the individual's perception of his or her health, the indirect instruments refer to a reference population; e. g. the general public.This may result in utilities better suited for health economic modelling. In a population sample, the questionnaire, e. g. EQ-5D, is administered together with one of the direct QoL instruments, and the various health states defined by the former are calibrated with the latter. This valuation yields a "social tariff" for the indirect instrument by way of an algorithm involving, among other features, weighting of the various items. The tariff thus describes each of the valuated health states as a utility value assigned by the direct HRQoL measurements in the reference population. In principle, indirect HRQoL measures should be validated in each disease studied and also in the relevant population, to account for cultural, socioeconomic and other differences(17). EQ-5D has been validated in a Swedish population sample(18), but there is no Swedish tariff. The weights of the UK tariff(19) employed in this study are displayed in Table 1A. An example of utility calculation from a health state is given in Table 1B. There are also disease-specific HRQoL instruments, but they are not used for calculation of utility and thus outside the scope of this article. Which one to choose of the many HRQoL instruments is largely dependent on the kind of investigation performed(20).

Utility may be regarded as a preference made by the patient (given a choice) scored between 0 (death) and 1 (perfect health). Health states worse than death may be assigned in the valuation process; EQ-5D utility by the UK tariff may score down to –0.59(17). By multiplying the time spent in a certain health state by its utility, one may calculate quality-adjusted life years (QALYs) after certain assumptions(21). One year spent in a state with 0,5 utility, for example, yields 0,5 QALY. The utility gained by some intervention, e g TNF blockade in inflammatory arthritis, is obtained as the difference between two time points, analogous to response for the activity indices. This difference, i e the delta-EQ-5D, multiplied by the time elapsed, yields the number of QALYs gained. A QALY may in turn be assigned a price that a funding source is considered willing to pay.

The aim of the present study was to introduce a new, utility based outcome measure, the Number Needed per QALY gained, NNQ, as outlined below, and to apply it to a cohort of patients with RA, psoriatic arthritis (PsA) and spondyloarthritis (SpA) treated with anti-TNF drugs in clinical practice. We also wanted to study possible secular trends in NNQ.

Table 1. Panel A: Weights for the various items of EQ-5D according to the UK tariff (N3 model). Adapted from(39).

Parameter	Definition	Estimate/preference value
Constant		0.081
M2	1 if mobility is level 2; otherwise 0	0.069
M3	1 if mobility is level 3; otherwise 0	0.314
S2	1 if self-care is level 2; otherwise 0	0.104
S3	1 if self-care is level 3; otherwise 0	0.214
U2	1 if usual activities is level 2; otherwise 0	0.036
U3	1 if usual activities is level 3; otherwise 0	0.094
P2	1 if pain/discomfort is level 2; otherwise 0	0.123
P3	1 if pain/discomfort is level 3; otherwise 0	0.386
A2	1 if anxiety/depression is level 2; otherwise 0	0.071
A3	1 if anxiety/depression is level 3; otherwise 0	0.236
N3	1 if any dimension is level 3; otherwise 0	0.269

Table 1. Panel B: Utility estimation of EQ-5D health state 11223 (item 1, mobility, scored at level 1, etc) using the UK valuation. Adapted from(19).

Full health = 1.000

Constant (any dimension >level 1)	-0.081
Mobility: level 1	-0
Self-care: level 1	-0
Usual activities: level 2	-0.036
Pain or discomfort: level 2	-0.123
Anxiety or depression: level 3	-0.236
N3 (level 3 occurs in at least 1 dimension)	-0.269
Estimated disutility for state 11223	-0.745
Utility for state 11223	0.255

Methods and patients

NNQ. We propose a new, simple measure, the NNQ, which is based on the utility and QALY concepts, as a group level estimate of the degree to which an intervention is worth while from a HRQoL perspective. NNQ is the number of patients that must be subjected to an intervention in order to gain 1 QALY, and it is calculated by multiplying the inverted value of utility gain (delta value) and the time during which this gain takes place.

Utility (u) may be expressed as a function of time (t); u0, u1, u1.5, u3 etc are utilities at baseline, time points 1, 1.5 and 3 months, etc, and $\Delta u=u—u0$. The QALY gain during 1 year is the area under the curve (AUC) of $\Delta u=f(t)$. If QALY gain occurs immediately and remains relatively stable for 1 year, as we have demonstrated to be the case for RA, PsA and SpA(22), NNQ may be calculated as

$$NNQ=1/(u12—u0);$$

If, however, the utility gain is not immediate but more gradual or fluctuating, the denominator is substituted for the AUC:

$$NNQ=1/\int_0^{12} f(t)dt;$$

t is time in months. Thus, NNQ is the inverted value of the amount of QALYs gained during 1 year. In practice, due to the immediate onset of treatment effect in TNF blockade in RA, PsA and SpA, NNQ in these diseases may be calculated from mean QALY gain for each patient, assuming no spontaneous

remissions and no effect on mortality over the time period under study. In other diagnoses or therapies, it may be more appropriate to use a formula that takes the distribution of utility measurements over time into account. The AUC thus is affected by the shape (slope) of the $\Delta u=f(t)$ curve.

Patients. Patients with inflammatory arthritides in southern Sweden, starting a course of biologic treatment, are entered into the South Swedish Arthritis Treatment Group (SSATG) registry as detailed in previous publications(23, 24). The ethics committee has concluded that no approval was needed because of the safety and quality surveillance character of the registry.

Demographic data, treatment and starting date were collected at baseline as well as core set outcome variables. EQ-5D excluding the 20 cm global VAS scale was administered. Data was collected at 0, 0.5, 1.5, 3, 6 and 12 months. At the start of each treatment course, a checklist with ACR 1987 diagnostic criteria for RA(25), modified New York AS(26) and ESSG (European Spondylarthritis Study Group) spondylarthropathy criteria(27) was to be ticked off.

To be eligible for the study, patients should have a diagnosis of RA, PsA or SpA as provided by the attending rheumatologist and start a treatment course with Adalimumab, Etanercept or Infliximab during the period January 2002 – December 2007. Data extraction was closed December 1, 2008. Complete data at baseline was mandatory. Treatment courses were assigned as being 1st, 2nd or 3rd. In general, 2nd and 3rd treatment courses refer to patients also included as 1st course. During the study period, availability of the 3 TNF blockers was variable, and they have thus been studied together. Because of limited numbers, secular trends were only studied in RA patients starting their 1st anti-TNF drug.

In the present study, utility gain was calculated as mean utility during the first year minus baseline utility for each treatment course. The AUC was also calculated in a subgroup with sufficient number of EQ-5D observations according to the trapezium rule. The UK tariff was used(19, 28). Utility gain and NNQ was calculated (i) for all eligible treatment courses together, (ii) for courses with duration of 1 year or more, and (iii) for courses of <1 year's duration, separately for each diagnosis and treatment course order. Furthermore, all utility gain and NNQ values were calculated with and without correction for the actual duration of treatment. For uncorrected values, AUC was calculated as utility gain multiplied by 1 year, irrespective of real treatment duration (last observation carried forward, LOCF); for the corrected values, utility gain was multiplied with the fraction of the year treatment was actually given.

Statistics. Descriptive statistics was used. Values are mean and 95% confidence intervals (CI) unless stated otherwise. The 2 different methods of AUC calculation were compared using Spearman's ρ. Only groups with >50 observations were used.

Results

Baseline characteristics of the patients at first anti-TNF treatment are shown in Table 2. For RA, there were 1001, 468 and 139 eligible 1st, 2nd and 3rd treatment courses, respectively. The corresponding figures for PsA were 241, 86 and 17, and for SpA they were 255, 63 and 26, respectively. There were several differences regarding age, sex, disease duration, medication and disease activity indices between the diagnostic entities. Also HAQ differed, while patients' global VAS and EQ-5D utilities were similar.

The mean NNQ values for the various diagnoses and treatment course numbers are given in Table 3. In the analysis of all courses with EQ-5D data, for RA NNQ was 4.5, 6.4 and 5.3 for 1st, 2nd and 3rd treatment course, respectively, without time correction. The time corrected NNQ values were slightly higher: 4.7, 6.7 and 5.7, respectively. For RA courses 1 year or longer, NNQ values were 3.8, 5.4 and 4.4 (time correction not needed). Uncorrected NNQ for RA courses <1 year where 10.4, 12.6 and 9.6, respectively, whereas the corresponding corrected values were 16.5, 20.6 and 17.0. NNQ data for PsA

and SpA are displayed in Table 3.

To assess the correctness of utility AUC based upon the mean of utility observations minus the baseline utility value, AUC was estimated by an alternative approach (the trapezium rule) in a subgroup (N=696) of 1st course RA treatment with sufficient number of observations. The methods yielded similar results, and the correlation coefficient was found to be 0.99 (Figure 1).

To study possible secular trends regarding utility gain (delta EQ-5D) during the study period, we also calculated NNQ per year of anti-TNF initiation (Figure 2). No obvious trends could be seen.

Table 2. Baseline characteristics at initiation of 1st anti-TNF treatment. Values are mean (95% CI) unless stated otherwise.

Valid N	RA 1001	SpA 255	PsA 241
Age (ys)	55,8 (55,0—56,7)	43,7 (42,2—45,1)	46,7 (45,1—48,3)
Disease duration (ys)	10,9 (10,2—11,6)	14,0 (12,5—15,4)	10,6 (9,47—11,8)
Nr of previous DMARDs	2,63 (2,54—2,72)	1,53 (1,41—1,64)	1,72 (1,59—1,85)
Nr of ongoing DMARDs	0,92 (0,89—0,96)	0,71 (0,63—0,78)	0,81 (0,73—0,88)
DAS28 (score 0-10)	5,37 (5,30—5,45)	3,76 (3,61—3,91)	4,51 (4,33—4,69)
CDAI (score 0-100)	29,4 (28,6—30,1)	15,8 (14,7—16,8)	22,5 (21,0—24,0)
VAS global (0-100)	61,6 (60,2—62,9)	61,6 (59,0—64,2)	61,5 (58,7—64,2)
Evaluators Global (Likert scale 0-4)	2,28 (2,24—2,32)	1,98 (1,89—2,06)	2,04 (1,96—2,12)
HAQ (score 0-3)	1,20 (1,16—1,23)	0,78 (0,70—0,85)	0,89 (0,82—0,96)
EQ-5D utility (-0.59-1)	0,40 (0,38—0,42)	0,44 (0,40—0,48)	0,40 (0,36—0,44)
Male (%)	22,4	60,4	52,3
Adalimumab (%)	25,5	16,1	16,6
Etanercept (%)	48,0	45,9	51,9
Infliximab (%)	26,6	38,0	31,5
Clinical signs of			
Spondyltis (%)		67,8	34,9
Peripheral disease (%)		50,2	75,1
Spondyltis+peripheral disease (%)		19,7	29,5

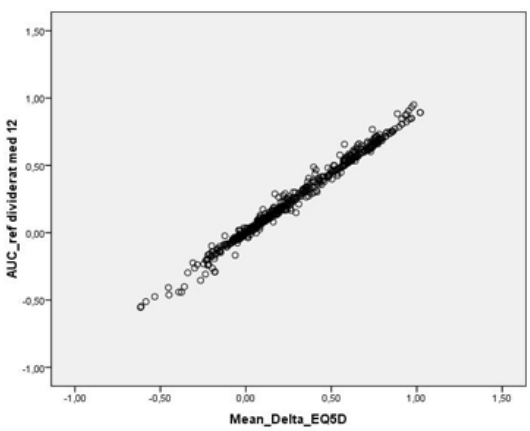


Figure 1. Utility AUC obtained by the trapezium rule as function of AUC based on mean utility value. N=696; Spearman's r=0.99.

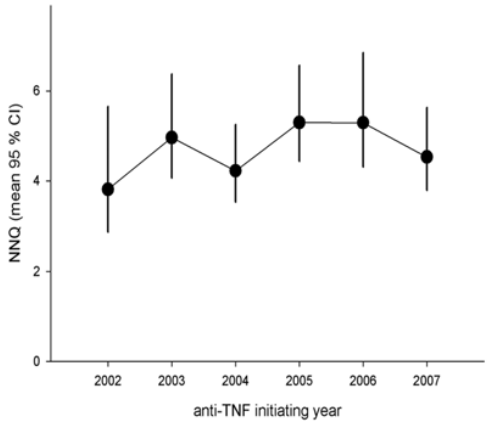


Figure 2. Utility gain (DEQ-5D) by year of treatment initiation 2002-2008. Bars are 95% CI.

Table 3. Utility gain (ΔEQ-5D) and NNQ for 1st, 2nd, and 3rd anti-TNF course for RA, PsA and SpA. Data not given for <50 observations. Values are mean (95% CI). Panel A: All treatment courses with EQ-5D utility data (LOCF approach). Panel B: Treatment courses with at least 1 year follow up. Panel C: Treatment courses with less than 1 year follow up.

A	RA		PsA		SpA	
Not time corrected	ΔEQ-5D(95%CI)	NNQ(95% CI)	ΔEQ-5D(95%CI)	NNQ(95% CI)	ΔEQ-5D(95%CI)	NNQ(95% CI)
1st course	0,21 (0,19—0,23)	4,5 (4,1—5,0)	0,21 (0,17—0,26)	4,5 (3,8—5,6)	0,23 (0,20—0,27)	4,1 (3,6—4,9)
2nd course	0,15 (0,12—0,18)	6,4 (5,3—7,9)	0,23 (0,16—0,31)	4,2 (3,2—6,2)	0,22 (0,14—0,31)	4,3 (3,2—6,8)
3rd course	0,18 (0,13—0,24)	5,2 (4,0—7,5)	0,01 (-0,1—0,18)		0,26 (0,11—0,42)	
Time corrected						
1st course	0,21 (0,19—0,22)	4,7 (4,3—5,2)	0,20 (0,16—0,24)	4,8 (4,0—5,9)	0,22 (0,18—0,26)	4,4 (3,8—5,3)
2nd course	0,14 (0,12—0,17)	6,7 (5,7—8,2)	0,20 (0,14—0,27)	4,7 (3,6—6,9)	0,22 (0,14—0,29)	4,5 (3,3—6,8)
3rd course	0,17 (0,12—0,22)	5,7 (4,4—7,9)	-0,0 (-0,1—0,11)		0,24 (0,09—0,39)	

B	RA		PsA		SpA	
	ΔEQ-5D(95%CI)	NNQ(95% CI)	ΔEQ-5D(95%CI)	NNQ(95% CI)	ΔEQ-5D(95%CI)	NNQ(95% CI)
1st course	0,25 (0,23—0,28)	3,8 (3,5—4,2)	0,26 (0,21—0,31)	3,7 (3,1—4,6)	0,25 (0,21—0,29)	3,9 (3,3—4,7)
2nd course	0,18 (0,14—0,21)	5,4 (4,6—6,6)	0,27 (0,18—0,35)	3,6 (2,8—5,3)	0,28 (0,19—0,37)	
3rd course	0,22 (0,16—0,28)	4,4 (3,4—6,1)	-0,0 (-0,3—0,22)		0,30 (0,14—0,47)	

C	RA		PsA		SpA	
Not time corrected	ΔEQ-5D(95%CI)	NNQ(95% CI)	ΔEQ-5D(95%CI)	NNQ(95% CI)	ΔEQ-5D(95%CI)	NNQ(95% CI)
1st course	0,09 (0,05—0,13)	10,4 (7,2—18,8)	0,09 (0,02—0,17)	10, (5,7—44,)	0,18 (0,09—0,27)	5,3 (3,6—10,)
2nd course	0,07 (0,01—0,14)	12,6 (6,8—83,3)	0,15 (0,00—0,31)		0,09 (-0,0—0,25)	
3rd course	0,10 (-0,0—0,21)		0,09 (-0,1—0,31)		0,09 (-0,4—0,65)	
Time corrected						
1st course	0,06 (0,03—0,08)	16,5 (11,6—28,8)	0,05 (0,00—0,09)	19, (10,—111)	0,11 (0,06—0,17)	8,4 (5,7—16,)
2nd course	0,04 (0,00—0,08)	20,6 (11,1—144,1)	0,07 (0,00—0,14)		0,06 (-0,0—0,16)	
3rd course	0,05 (-0,0—0,11)		0,01 (-0,0—0,09)		-0,0 (-0,3—0,33)	

Discussion

In the current study, we propose a new measure, the Number Needed per QALY gained, providing an easy to understand estimate of one aspect (HRQoL) of the degree to which a treatment is worth while in a population. We found that the NNQ provided an intuitive, HRQoL related outcome measure for anti-TNF treatment across 3 different arthritis diagnoses, which could be helpful in understanding the health and economic benefits of these and other therapies. NNQ is easy to determine and follow. It does not necessitate the gathering of actual costs, a burdensome and often precarious procedure including many assumptions. NNQ is thus a simplified way to summarize information in the observational setting, and it does not preclude health economic modelling.

The time corrected NNQ values must be considered more reflective of actual utility gain in the dataset under consideration. The optimal basis for utility gain calculation (all data, courses with complete data only or courses <1year) is not self-evident. In the observational setting, giving all three will provide the most complete information. Patients with >1 year treatment constitute a selection of responders, and courses <1 year the opposite. The higher NNQ values of these reflect the much smaller utility gain compared to those remaining on therapy. Time corrected NNQ based on all data may be considered to represent a tempered estimate of the treatment effect in the cohort as a whole.

When NNQ is corrected for the time of courses <1 year in duration, values become slightly higher, as expected. NNQ based upon courses at least 1 year of length is, not surprisingly, lower than for all courses; this group represents patients continuing treatment, whereas those terminating treatment before 1 year conceivably in many cases represent treatment failure or toxicity in patients gaining fewer QALYs. This is supported by the considerably higher NNQ values for those treated for shorter than 1 year, with a further increase after time correction, reflecting the small utility AUC gain in this group. It is difficult to account for the cost of adverse events in health economic calculations. Life table analysis reveal length of time spent on treatment, but withdrawal gives rise to costs represented by the smaller utility and QALY gain, reflected in higher NNQ values. NNQ includes information on the (utility) cost of adverse events, albeit not the whole truth(29). However, NNQ based on all data rises only slightly after time correction. Thus, the treatment courses <1 year, in spite of constituting a sizable proportion (25%) of all treatments, affect the "crude" NNQ only to a minor degree. However, including all utility data without compensating for actual time on drug tends to inflate utility gain. Of course, results are even more inflated when only courses >1 year are taken into account, similarly to the results of open label extensions of RCTs.

In general, the time corrected NNQ yielded very similar results with NNQ between 4 and 6 across diagnoses and treatment courses with overlapping 95 % confidence intervals. The only exception was a significantly higher NNQ of 6.7 for 2nd anti-TNF treatment course in RA patients, suggesting a selection of cases less prone to improve health utilities after switching to a 2nd anti-TNF drug. However, the pattern was neither reproduced in the other diagnostic entities nor in the 3rd time switchers, thus lending some doubt to the validity of this information. The results were more variable in the smaller groups, with wide confidence intervals, and we therefore chose to only study groups with numbers exceeding 50.

The lack of secular trends for NNQ over time (figure 2) was somewhat unexpected in view of clear trends regarding baseline characteristics in our setting during 1999-2007(30). However, we did not find such trends for baseline EQ-5D utilities for the current time period 2002-2008(22). Also, it must be remembered that NNQ represents change after intervention, which is not necessarily a function of baseline characteristics.

Utility development was studied for 1 year, but it is possible that the NNQ observed would remain valid for a longer period of time. Utility in those remaining on therapy tends to remain constant after the initial rise(22), and life table analysis in the SSATG registry has shown, that the number of patients terminating treatment tends to level out over time(31, 32).

The NNQ concept, like the concept of utility, is not limited to rheumatology, but should easily lend itself to a wide range of interventions and diagnostic entities. We have thus calculated a few examples of NNQ values based on published data: In a large British study of gastro-oesophageal reflux, surgery was found to produce a gain of 0.088 QALYs, equivalent to NNQ = 11.4(33), i.e. one has to operate 11-12 patients to gain 1 QALY. Surgery for a herniated lumbar disc was, in a Swedish study, associated with a QALY gain of 0.41, corresponding to NNQ = 2.4(34). Both of these studies utilized EQ-5D utilities. In a health economic evaluation of valsartan for heart failure in patients who had had a myocardial infarction and were not suitable for ACE inhibitors, the amount of incremental QALYs gained, based on literature-derived utilities weighted for various cardiovascular events, was found to be 0.5021, i.e. NNQ=1.99(35).

In a retrospective, observational study of anti-TNF treatment of RA, PsA and AS, the number needed to treat in order to gain at least the minimal clinically important change (MCID) in HAQ, was calculated(36). In fact, this method is not similar to the commonly used NNT statistic, as there is no contrast population(37). Furthermore, it seems to have less potential for use over different chronic diseases, as it is based on the HAQ rather than a generic utility measure.

The concept "NNT to gain 1 QALY" has been used in a health economic evaluation of orlistat in the treatment of overweight patients(38). No details regarding the calculation of this estimate are given, however, and the study is based on pooled data from 5 RCTs comparing calorie-reduced diet plus orlistat or placebo. The incremental cost per QALY gained is given, and it may thus be inferred, that "NNT to gain 1 QALY" here refers to the usual definition of NNT involving a placebo group. This is not the case with the NNQ described in the current study.

There are limitations to our study. The NNQ may be regarded as an over-simplification. Like other measures, it is no better than the data from which it is derived, but in addition it entails some approximations and assumptions that must be taken into account.

Firstly, mean utility gain assumes a constant health state for 1 year. This could be amended by more measurements during the observation period and basing AUC calculations on all these. The 2 AUC calculation methods employed in the present example, however, yielded very similar results. Baseline utility was also a single value, rather than based on 2 observations some time apart. Second and 3rd or more treatment courses in our setting, however, have baseline utilities roughly the same as 1st courses(22). This observation supports the reliability of the reported baseline utility values, since patients tend to return to their original utility level upon anti-TNF cessation.

Secondly, there are inherent drawbacks to the utility measures as compared to the HAQ and other scales; many of them lack robustness in at least some respect. The EQ-5D represents a compromise exhibiting feasibility, acceptable responsiveness and construct validity, but rather poor reliability(20). On the other hand, generic measures like EQ-5D seem to give more uniform results across diagnostic entities, than HAQ or disease activity measures, which are more dependent on inflammation. Furthermore, the transformation of the questionnaire raw scores into utility values has its pitfalls. We have used the UK valuation of the EQ-5D(28), which was made in the beginning of the 1990s, utilizing a British, general population sample. It is possible, that the preferences of a Swedish population 10 years later had been different. There are EQ-5D valuations for several other countries, including the US and Denmark, but not Sweden. The item weights and algorithms of these tariffs vary, and so do the utilities resulting from their respective application. By applying different tariffs in the same study, widely differing QALY gains and cost-effectiveness estimations may be arrived at(39, 40). It is not self evident, if emphasis should be put on relevance of the valuation population and algorithm to the cohort studied, or the comparability of absolute utility values, which should be facilitated by applying the same tariff to different cohorts. Comparing utility levels, gains and QALYs – and, consequently, NNQ – from various studies must be done carefully.

Finally, the term NNQ in itself may associate to Number Needed to Treat, the NNT statistic commonly used in RCTs(37). The NNT is the inverted value of the absolute risk reduction, i.e. the

difference between the absolute risk for a defined outcome in populations exposed to the intervention studied and a standard (or no) intervention, respectively. By contrast, the NNQ does not include a parallel contrast population. We assume that utility improvement is unlikely in RA, PsA and SpA patients not given TNF inhibitors. This may not be true in other diseases or therapies. Furthermore, in a clinical setting as ours, both the intervention (3 anti-TNF drugs) and the study population (non-randomised patients in day-to-day care) lack homogeneity, making it hard to relate the NNQ values found to a well defined treatment situation. The purpose of the study, however, was not just to investigate the effect of TNF blockers, but rather to present and test the feasibility and face validity of the NNQ concept.

The NNQ is a group level measure giving an idea of the extent to which the intervention studied is worth while in a given population. It may thus be regarded as a much simplified health economic measure, which does not include the gathering of real costs. The incremental cost is equal to the sum spent on anti-TNF drug, and the gain is represented by the QALYs.

The NNQ may help non-economists understand how interventions should be valued economically. Health economic studies are generally not very transparent, and they rarely show that the drug studied (from the sponsoring company) is not worth the expense. It should be possible to apply the NNQ in many settings. Validation in other cohorts, both in trials and in clinical practice, is called for, however, to determine its role in the armamentarium of outcome measures in rheumatology and other fields.

Authors' contributions

AG wrote the manuscript and helped plan the study. LEK, TS, LTHJ and IFP helped plan the study and draft the manuscript. PG conceived the study, handled the database and helped draft the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We are indebted to all colleagues and staff in the South Swedish Arthritis Treatment Group for cooperation and data supply, to Jan-Åke Nilsson for help with statistical calculations and to Martin Neovius for valuable suggestions. This study was supported by grants from Swedish Research Council, Österlund and Kock Foundations, King Gustav V 80 year fund, Lund University Hospital, Region Skåne, Faculty of Medicine, Lund University and Reumatikerförbundet.

References

1.Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. Arthritis Rheum 1980;23(2):137-45.

2.Blackmore MG, Gladman DD, Husted J, Long JA, Farewell VT. Measuring health status in psoriatic arthritis: the Health Assessment Questionnaire and its modification. J Rheumatol 1995;22(5):886-93.

3.Forejtova S, Mann H, Stolfa J, Vedral K, Fenclova I, Nemethova D, et al. Factors influencing health status and disability of patients with ankylosing spondylitis in the Czech Republic. Clin Rheumatol 2008;27(8):1005-13.

4.Bruce B, Fries JF. The Stanford Health Assessment Questionnaire: a review of its history, issues, progress, and documentation. J Rheumatol 2003;30(1):167-78.

5.www.DAS-score.nl. In.

6.Aletha D SJ. The Simplified Disease Activity Index (SDAI) and the Clinical Disease Activity Index (CDAI): A review of their usefulness and validity in rheumatoid arthritis. Clin Exp Rheumatol 2005;23(Suppl 39):S100-S108.

7.Gossec L, Dougados M, Rincheval N, Balanescu A, Boumpas DT, Canadelo S, et al. The elaboration of the preliminary Rheumatoid Arthritis Impact of Disease (RAID) score: a EULAR initiative. Ann Rheum Dis 2008.

8.Wolfe F, Michaud K, Pincus T. A composite disease activity scale for clinical practice, observational studies, and clinical trials: the patient activity scale (PAS/PAS-II). J Rheumatol 2005;32(12):2410-5.

9.Aletaha D, Smolen J, Ward MM. Measuring function in rheumatoid arthritis: Identifying reversible and irreversible components. Arthritis Rheum 2006;54(9):2784-92.

10.Sajid MS, Tonsi A, Baig MK. Health-related quality of life measurement. Int J Health Care Qual Assur 2008;21(4):365-73.

11.Ariza-Ariza R, Hernandez-Cruz B, Carmona L, Dolores Ruiz-Montesinos M, Ballina J, Navarro-Sarabia F. Assessing utility values in rheumatoid arthritis: a comparison between time trade-off and the EuroQol. Arthritis Rheum 2006;55(5):751-6.

12.Group TE-Q. EuroQol—a new facility for the measurement of health-related quality of life. The EuroQol Group. Health Policy 1990;16(3):199-208.

13.Hurst NP, Kind P, Ruta D, Hunter M, Stubbings A. Measuring health-related quality of life in rheumatoid arthritis: validity, responsiveness and reliability of EuroQol (EQ-5D). Br J Rheumatol 1997;36(5):551-9.

14.Brooks R. EuroQol: the current state of play. Health Policy 1996;37(1):53-72.

15.Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J Health Econ 2002;21(2):271-92.

16.Horsman J, Furlong W, Feeny D, Torrance G. The Health Utilities Index (HUI): concepts, measurement properties and applications. Health Qual Life Outcomes 2003;1:54.

17.Harrison MJ, Davies LM, Bansback NJ, Ingram M, Anis AH, Symmons DP. The validity and responsiveness of generic utility measures in rheumatoid arthritis: a review. J Rheumatol 2008;35(4):592-602.

18.Burstrom K, Johannesson M, Diderichsen F. A comparison of individual and social time trade-off values for health states in the general population. Health Policy 2006;76(3):359-70.

19.Dolan P. Modeling valuations for EuroQol health states. Med Care 1997;35(11):1095-108.

20.Linde L, Sorensen J, Ostergaard M, Horslev-Petersen K, Hetland ML. Health-related quality of life: validity, reliability, and responsiveness of SF-36, 15D, EQ-5D [corrected] RAQoL, and HAQ in patients with rheumatoid arthritis. J Rheumatol 2008;35(8):1528-37.

21.Rasanen P, Roine E, Sintonen H, Semberg-Konttinen V, Ryynanen OP, Roine R. Use of quality-adjusted life years for the estimation of effectiveness of health care: A systematic literature review. Int J Technol Assess Health Care 2006;22(2):235-41.

22.Gülfe A, Kristensen L, Saxne T, Jacobsson L, Peterson IF, Geborek P. Rapid and sustained health utility gain in anti-TNF treated inflammatory arthritis. Observational data during seven years in southern Sweden. Ann Rheum Dis 2009.

23.Geborek P, Crnkic M, Petersson IF, Saxne T. Etanercept, infliximab, and leflunomide in established rheumatoid arthritis: clinical experience using a structured follow up programme in southern Sweden. Ann Rheum Dis 2002;61(9):793-8.

24.Geborek P, Saxne T. Clinical protocol for monitoring of targeted therapies in rheumatoid arthritis. Rheumatology (Oxford) 2000;39(10):1159-61.

25.Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1988;31(3):315-24.

26.van der Linden S, Valkenburg HA, Cats A. Evaluation of diagnostic criteria for ankylosing spondylitis. A proposal for modification of the New York criteria. Arthritis Rheum 1984;27(4):361-8.

27.Dougados M, van der Linden S, Juhlin R, Huitfeldt B, Amor B, Calin A, et al. The European Spondylarthropathy Study Group preliminary criteria for the classification of spondylarthropathy. Arthritis Rheum 1991;34(10):1218-27.

28.Dolan P GC, Kind P, Williams A. A social tariff for EuroQol: Results from a UK population survey. York, UK: Centre for Health Economics, University of York; 1995.

29.Kobelt G, Geborek P, al. E. Costs and outcomes for patients with Rheumatoid Arthritis treated with biological drugs in Sweden. A model based on registry data. Scand J Rheumatol In press.

30.Soderlin MK, Geborek P. Changing pattern in the prescription of biological treatment in rheumatoid arthritis. A 7-year follow-up of 1839 patients in Southern Sweden. Ann Rheum Dis 2008;67:37-42.

31.Kristensen LE, Saxne T, Geborek P. The LUNDEX, a new index of drug efficacy in clinical practice: results of a five-year observational study of treatment with infliximab and etanercept among rheumatoid arthritis patients in southern Sweden. Arthritis Rheum 2006;54(2):600-6.

32.Kristensen LE, Saxne T, Nilsson JA, Geborek P. Impact of concomitant DMARD therapy on adherence to treatment with etanercept and infliximab in rheumatoid arthritis. Results from a six-year observational study in southern Sweden. Arthritis Res Ther 2006;8(6):R174.

33.Grant A, Wileman S, Ramsay C, Bojke L, Epstein D, Sculpher M, et al. The effectiveness and cost-effectiveness of minimal access surgery amongst people with gastro-oesophageal reflux disease - a UK collaborative study. The REFLUX trial. Health Technol Assess 2008;12(31):1-181, iii-iv.

34.Jansson KA, Nemeth G, Granath F, Jonsson B, Blomqvist P. Health-related quality of life in patients before and after surgery for a herniated lumbar disc. J Bone Joint Surg Br 2005;87(7):959-64.

35.Taylor M, Scuffham PA, Chaplin S, Papo NL. An Economic Evaluation of Valsartan for Post-MI Patients in the UK Who Are Not Suitable for Treatment with ACE Inhibitors. Value Health 2009.

36.Barra L, Pope JE, Payne M. Real-world anti-tumor necrosis factor treatment in rheumatoid arthritis, psoriatic arthritis, and ankylosing spondylitis: cost-effectiveness based on number needed to treat to improve health assessment questionnaire. J Rheumatol 2009;36(7):1421-8.

37.McHugh ML. Scientific inquiry: clinical statistics for primary care practitioners: Part II-absolute risk reduction, relative risk, relative risk reduction, and number needed to treat. J Spec Pediatr Nurs 2008;13(2):135-8.

38.Lacey LA, Wolf A, O'Shea D, Erny S, Ruof J. Cost-effectiveness of orlistat for the treatment of overweight and obese patients in Ireland. Int J Obes (Lond) 2005;29(8):975-82.

39.Nan L, Johnson JA, Shaw JW, Coons SJ. A comparison of EQ-5D index scores derived from the US and UK population-based scoring functions. Med Decis Making 2007;27(3):321-6.

40.Noyes K, Dick AW, Holloway RG. The implications of using US-specific EQ-5D preference weights for cost-effectiveness evaluation. Med Decis Making 2007;27(3):327-34.