Assessment of intra- and inter-individual variability of outcome measures in ankylosing spondylitis and the efficacy and adverse effects of anti-TNF therapy

Lara Jane Maxwell

Thesis submitted to the Faculty of Graduate and Postdoctoral Studies in partial fulfillment of the requirements for the M.Sc. in Epidemiology

Epidemiology and Community Medicine

Faculty of Medicine

University of Ottawa

ABSTRACT

Ankylosing spondylitis (AS) is a chronic, inflammatory rheumatic disease that has a highly variable disease course. Three biologic agents, adalimumab, etanercept, and infliximab, have been developed for the treatment of AS. We conducted three studies: 1) an exploratory analysis of a year-long longitudinal dataset to gain insight into the variability of disease activity, physical function, and well-being and to explore the relationship between these outcome measures; 2) a systematic review of the available evidence for the efficacy of biologic treatment; 3) a systematic review of potential adverse effects of this treatment. We found that repeated measures of disease activity, function and well-being fluctuate considerably between patients, with complex patterns occurring over time within patients. There was mostly high quality evidence that these biologics are efficacious against placebo. We did not find evidence of an increase in serious adverse events or serious infections from short-term randomized controlled trials.

EXECUTIVE SUMMARY

Ankylosing spondylitis (AS) is a chronic, inflammatory rheumatic disease characterized by inflammatory back pain due to sacroiliitis and spondylitis, enthesitis, and the formation of syndesmophytes leading to ankylosis. The course of AS is highly variable, and at least one third of patients with AS will progress to severe disabling disease. Few studies have explored the variability of disease activity, function, or well-being of AS patients over time. Describing and understanding this variability is of interest when determining therapeutic options for patients.

Treatment of AS requires a multidisciplinary approach and is usually managed with a combination of exercises, physiotherapy, and drug therapy. An exciting advance in treatment options is the development of biologic therapies which target specific elements of the immune system. Tumour necrosis factor (TNF)-alpha is a protein that the body produces during the inflammatory response. TNF-alpha promotes inflammation and subsequent pain, tenderness, swelling and fever in several inflammatory conditions including ankylosing spondylitis. Three anti-TNF agents, also known as TNF-inhibitors, have been developed to target the binding of this protein, thus reducing the pain, swelling, and inflammation associated with AS. The generic drug names are adalimumab, etanercept, and infliximab. While these biologics offer an important therapeutic advance by appearing to reduce disease activity and improve function and well-being of patients, it is important to understand and try to quantify not only the potential benefits of this treatment, but also the potential harms. Clinicians and patients need this information in order to make an informed decision about the trade-offs of using this treatment option.

As part of this thesis, three manuscripts were prepared. The first manuscript was based on an analysis of a longitudinal dataset of weekly measures of disease activity, function, and well-being gathered on 42 AS patients over the course of 52 weeks.

Descriptive statistical techniques and mixed effects modeling were employed to characterize the within- (intra) and between (inter)-patient variability of these three outcomes over time. We also investigated the relationship between these outcomes, assessed the effect of patient characteristics on change of these outcomes over a period of one year, and explored how the stability of these outcomes relates to current guidelines on the use of anti-TNF agents for AS. We found that the weekly measures of disease activity, function and well-being fluctuate considerably between patients, with complex patterns occurring over time within patients. Between-individual variability explained a greater proportion of the variance. The current guidelines for basing the use of anti-TNF agents on two assessments of disease activity four weeks apart fits well with the natural course of disease activity found in this cohort.

The second manuscript was based on a systematic review of the literature with the objective to summarize the efficacy and harms of adalimumab, etanercept, and infliximab. We included 15 randomized controlled trials (RCTs) and four non-RCTs (for the assessment of harms only) and found mainly high quality evidence for the short-term efficacy of these three TNF-inhibitors in improving disease activity, function, and inflammation as measured by MRI. There was no data on improvement of radiographic progression. One head-to-head RCT of etanercept versus infliximab over the course of two years found no difference between the two biologics in terms of important efficacy outcomes. Given the paucity of head-to-head studies comparing treatments of interest to practitioners who must make choices as to which treatment to prescribe to their patient, we performed indirect comparisons using three different methodological techniques. There was no statistically significant difference between the three TNF-inhibitors in terms of various disease activity improvement response measures. The adverse effects data is described in greater detail in the third manuscript.

The third manuscript was based on the same systematic review of the literature as described above, but with the objective of focusing on the potential adverse effects of anti-TNF treatment in AS over short and long time periods. Given the challenges of using RCTs to assess adverse effects, we broadened our inclusion criteria to include non-RCTs, especially for the assessment of rare or delayed adverse effects such as demyelinating diseases or malignancies. In RCTs up to six months in duration, withdrawals due to adverse events were greater in the etanercept group but not in adalimumab or infliximab groups, compared to placebo. Serious adverse events and serious infections were not statistically different between groups for any of the anti-TNF agents or for all three drugs pooled together, but there was a statistically significant increase in total infections. Four extension studies provided data for up to three years after the RCT period and we found no large signal of key safety outcomes such as serious infections and malignancies. However, these studies were judged to be at a high risk of bias. Indirect comparisons did not find differences between the three TNF-inhibitors for any of the adverse event outcomes.

In conclusion, based on the three manuscripts of this thesis:

1. Repeated measures of disease activity, function and well-being fluctuate considerably between patients, with complex patterns occurring over time within patients. The current guidelines for basing the use of anti-TNF agents on two assessments of disease activity four weeks apart fits well with the natural course of disease activity found in this cohort.

2. There is high quality evidence for the short-term efficacy of these three TNF-inhibitors in improving disease activity, function, and achieving partial remission when assessed against placebo. There does not appear to be a difference in efficacy between the three agents using network meta-analysis and other indirect comparison techniques to model head-to-head evaluations.

3. The short-term toxicity profile for the use of anti-TNF agents for the treatment of AS appears acceptable. Data up to three years of follow-up was judged to be at a high risk of bias, but there was no large safety signal in terms of serious infections and malignancies. Biologic registries with appropriate control patients as well as data from post-marketing surveillance are required to assess for longer-term harms.

CONTRIBUTION OF THE AUTHORS

Three manuscripts have been prepared for publication as part of this thesis. The student (LJM) is the first author on all three manuscripts as she had the primary responsibility for conceptualization of the manuscript series, data collection, analysis, interpretation, and writing. Her co-supervisors, Dr. Peter Tugwell and Dr. George Wells, as well as advisor to the thesis, Dr. Annelies Boonen are co-authors on these manuscripts. Drs. Robert Landewe and Desiree van der Heijde are co-authors on the first manuscript. Drs. Jane Zochling and Jasvinder Singh are co-authors on the second and third manuscripts.

ACKNOWLEGEMENTS

My supervisors, Peter Tugwell and George Wells. Thank you for your longstanding support and encouragement to pursue my M.Sc.

My co-authors. Your valuable feedback and guidance made the success of this project possible. A special thank-you to Annelies for suggesting the analysis of the dataset and for obtaining access to it for me.

Andre. Thank you for your friendship, your wisdom, and for always listening.

My family and friends. Thank you for all your love and support and for cheering me on every step of the way.

And to my Kevin. Thank you for your unwavering confidence in me, for inspiring me to advance my career and be happy, and for your patience in this long journey. I love you.

Isla and Keira. I dedicate this achievement to you, my darling girls.

TABLE OF CONTENTS

LIST OF TABLES

Chapter Three

Chapter Four

Chapter Five

LIST OF FIGURES

Chapter Three

Chapter Four

Chapter Five

CHAPTER ONE: INTRODUCTION

1.1 Rationale

AS is a disease characterized by periods of flare and remission of disease activity (1). However, little research has been conducted to describe the variation of disease activity, physical function, and well-being outcome measures of AS patients over periods of time. One small study (2) (n=22) conducted in 1991 studied disease activity on a monthly basis over the course of one year and found a high level of variability within each patient's profile as well as high between-individual heterogeneity in disease activity fluctuations. A second study (3) (N=24) published in 2008 assessed week-to-week variability in disease activity and function over 30 weeks with the objective of investigating whether the natural fluctuation in disease activity meant that several clinical assessments are needed before deciding to start or stop anti-TNF therapy. The results of this study also found high intra-individual variability over the course of the study. These interesting observations on small numbers of patients need confirmation on a larger dataset. Data on the variability of the course of this disease will be of interest to clinicians making treatment decisions.

A major breakthrough in the treatment of AS has been the development of TNF-blockers. The three biologic agents approved for use in AS are adalimumab, etanercept, and infliximab. While open-label studies and randomized controlled trials have demonstrated evidence of their efficacy against placebo, a systematic synthesis of the literature on anti-TNF use for AS should be conducted to confirm and quantify the benefits of this treatment. As well, a systematic assessment of their potential harms should be undertaken to allow clinicians and patients to make an informed decision about the trade-off between benefit and harm of this treatment.

1.2 Outline and objectives

This thesis has been formatted as a manuscript-based thesis. Three manuscripts have been prepared, in addition to a background chapter and an overall discussion and

conclusions chapter. The appendices include additional material related to outcome measures, administrative documents, and further details on the methodology used for chapter three that would not appear in the manuscript.

The following is an outline of the format of this thesis, with an overview of the objectives of each chapter:

Chapter One

An introductory chapter designed to orient the reader to the layout of the thesis and the objectives of each chapter.

Chapter Two

This chapter contains background information on the condition of AS, outcome measures used in AS, the mechanism of action of anti-TNF agents, and existing clinical guidelines on the use of anti-TNF agents in AS.

Chapter Three

A manuscript based on the analysis of a longitudinal dataset of weekly measures of disease activity, function, and well-being gathered on 42 AS patients over the course of 52 weeks. The objectives of this analysis were to:

1. describe the variability and patterns over time of disease activity, physical function, and well-being in a single cohort of AS patients over a period of one year.
2. determine the number of consecutive assessments that need to be averaged to provide a stable estimate of disease activity, specifically in the context of establishing baseline disease activity for entry to a randomized controlled trial.
3. describe the relationship (correlation) between disease activity and function, disease activity and well-being, and function and well-being outcomes within individual patients.
4. explore whether patient characteristics (age, sex, disease duration) may explain the change over time in disease activity, function, or well-being.

5. assess whether the variability in disease activity in a pre-biologic era (i.e. a historical control) group of patients has implications for the current guidelines on the use of anti-TNF agents.

Chapter Four

A manuscript based on a systematic review of the literature with the objective to summarize the efficacy and harms of adalimumab, etanercept, and infliximab for use in treating AS. The format of the manuscript is based on requirements for publication in the Cochrane Library.

Chapter Five

A manuscript based on the same systematic review of the literature mentioned above but with a specific focus on the potential harms of adalimumab, etanercept, and infliximab for use in treating AS. The format of the manuscript is based on requirements for co-publication of a Cochrane review in the Journal of Rheumatology.

Chapter Six

A summary chapter to discuss the implications and conclusions of the work contained in the three manuscripts.

1.3 References

(1) van der Linden S, Van Der Heijde D, Maksymowych W. Spondyloarthropathies: Ankylosing Spondylitis. In: Firestein G.S, Budd R.C., Harris JrED, McInnes I.B., Ruddy S, Sergent J.S., editors. Kelley's Textbook of Rheumatology. 8th ed. Philadelphia: Saunders Elsevier; 2009. p. 1169-89.

(2) Goodacre JA, Mander M, Dick WC. Patients with ankylosing spondylitis show individual patterns of variation in disease activity. Br J Rheumatol 1991 Oct;30(5):336-8.

(3) Berthelot JM, Tortellier L, Lavy-Bregeon D, le GB, Maugars Y. High intraindividual week-to-week variability in BASDAI and BASFI values: are several evaluations needed before starting or stopping TNFalpha antagonist therapy for spondyloarthropathies? Joint Bone Spine 2008 Mar;75(2):167-71.

CHAPTER TWO: BACKGROUND

2.1 Description of the condition

Ankylosing spondylitis (AS) is a chronic, inflammatory rheumatic disease characterized by inflammatory back pain due to sacroiliitis and spondylitis, enthesitis, and the formation of syndesmophytes leading to ankylosis. Extraspinal manifestations are common, including peripheral arthritis (25-50%), uveitis (25-40%), and inflammatory bowel disease (26%) and contribute to disease morbidity (1). The etiology of the disease is not yet fully understood but there is a strong association with the HLA-B27 gene. Studies have shown the prevalence of AS in the adult general population to vary from 0.4% (Alaska Eskimos) to 1.4% (Northern Norway) (2). More recent studies in France and Norway estimated the prevalence of clinical AS at 150 and 210 per 100,000 adults, respectively (3). The peak age of onset is between 20 and 30 years, although there is often a 5 to 6 year delay in diagnosis (2). Symptoms of AS appear to be two to three times more common in men than women. The course of AS is highly variable, and at least one third of patients with AS will progress to severe disabling disease (4). Clinical symptoms usually begin with back pain and stiffness in adolescence and early adulthood and can lead to impaired spinal mobility and/or chest expansion. The burden of disease in AS has been found to be similar to that of rheumatoid arthritis in terms of pain, disability and decreased well being (4). Additionally, compared to the general population, those with AS experience higher work disability and absence from work which can lead to substantial direct and indirect socio-economic costs (5;6).

2.2. Outcome measures used in ankylosing spondylitis

The ASsessment in Spondyloarthritis (formerly ASsessment in Ankylosing Spondylitis) (ASAS) Working Group (www.asas-group.org) has developed core sets of standardized outcome measures for use in clinical practice and trial settings. This work has been undertaken in conjunction with the OMERACT (Outcome Measures in Rheumatology, www.omeract.org) initiative which aims to establish standardized,

validated outcome measures for use in clinical trials in the field of rheumatology. Outcome measures are selected using data-driven, consensus-based methodology in which the outcomes are measured against three key concepts known as the 'OMERACT Filter': Truth (does the instrument measure what it was designed to measure? i.e. face validity), Discrimination (can the instrument detect change between groups? i.e. responsiveness), and Feasibility (is the instrument appropriate in terms of time, financial, and other constraints?) (7;8). Assessment tools developed by ASAS include both patient-reported and physician-measured. The three assessment tools selected by the researchers who conducted the year-long longitudinal study were the BASDAI (Bath Ankylosing Spondylitis Disease Activity Index) (9), the BASFI (Bath Ankylosing Functional Index) (10), and the BAS-G (Bath Ankylosing Spondylitis Global) (11). Details of all three are available on the ASAS website (http://www.asas-group.org/ under 'Assessment') and provided in Appendix A of this thesis. The BASDAI questionnaire assesses fatigue, pain, discomfort, and stiffness. The BASFI captures aspects of daily activities that may be impeded by AS; such as the ability to climb stairs, stand unsupported, or reach up high. The BAS-G asks about the effect of the disease on well-being over the past week. All three have been validated in AS patients and are self-reported (12).

The outcome measures selected for the systematic review were based on the disease-controlling anti-rheumatic treatment (DC-ART) Ankylosing Spondylitis Working Group Core Set (13) and the International ASAS consensus statement for the use of TNF- inhibitors in patients with AS (14). The outcomes selected for the summary of findings table were decided by consensus with five experts from the ASAS working group.

2.3 Description of TNF-inhibitors

As the result of research demonstrating that tumour necrosis factor alpha (TNF-alpha) is present in inflamed sacroiliac joints (15), drugs were developed to block TNF-alpha (TNF-alpha inhibitors). TNF-alpha is a protein that the body produces during the inflammatory response. TNF-alpha promotes inflammation and

subsequent pain, tenderness, swelling and fever in several inflammatory conditions including ankylosing spondylitis.

There are three biologic agents available that target TNF-alpha and are approved for use in AS: infliximab (Remicade) is a chimeric (mouse/human) monoclonal antibody of the IgG1κ isotype that binds with a high affinity to TNF-alpha, etanercept (Enbrel) is a receptor fusion protein that binds to TNF-alpha, thus competitively inhibiting the binding of TNF-alpha to the cell surface, and adalimumab (Humira) is a recombinant human IgG1 monoclonal antibody specific for human TNF-alpha. These three drugs prevent TNF-alpha from promoting inflammation and therefore reduce pain, tenderness and swelling of joints in patients with ankylosing spondylitis. Infliximab is given as an intravenous infusion over one to two hours, etanercept and adalimumab are given as sub-cutaneous injections. Recently a fourth anti-TNF biologic, golimumab, was approved for use in AS. However, as this agent was introduced on the market after this thesis work began, it is not included in this systematic review.

Various terms are used in the literature to describe these biologics. These include ‘anti-TNF-alpha agents’, ‘TNF-alpha blockers’, and ‘TNF-alpha inhibitors’. When the ‘alpha’ is not included after ‘TNF’, it is understood that we are referring to ‘TNF-alpha’.

2.4 Guidelines on the use of anti-TNF agents in clinical practice

The Ankylosing Spondylitis Working Group (ASAS) (14;16), the Spondyloarthritis Research Consortium of Canada (17;18), and the French Society for Rheumatology (19) have developed recommendations for the use of anti-TNF therapy. These recommendations include guidance on criteria to help clinicians decide when it is appropriate to start or stop anti-TNF therapy. A measure of high disease activity or ‘active disease’ is common to these guidelines in terms of deciding when to initiate anti-TNF therapy. Given the high cost of this therapy and the lack of evidence on long-term adverse events, the guidelines recommend that a continuous level of high

disease activity (though there is no consensus on the timing) should be present to ensure that therapy is not being prescribed for a temporary flare. Disease activity is also used to assess treatment response. If a reduction in active disease is not achieved (in twelve to sixteen weeks, depending on the guideline), then it is recommended that anti-TNF therapy is either switched to another anti-TNF agent or withdrawn.

In chapter 3, we explore how disease activity scores of patients in the longitudinal study relate to current guidelines on the use of anti-TNF agents for AS. We used the guidelines mentioned above to devise various scenarios of whether initiation and maintenance of anti-TNF therapy was warranted in this historical control cohort.

2.5 References

(1) Edmunds L, Elswood J, Kennedy LG, Calin A. Primary ankylosing spondylitis, psoriatic and enteropathic spondyloarthropathy: a controlled analysis. J Rheumatol 1991 May;18(5):696-8.

(2) Khan MA. Update on spondyloarthropathies. Ann Intern Med 2002 Jun 18;136(12):896-907.

(3) van der Linden S, Van Der Heijde D, Maksymowych W. Spondyloarthropathies: Ankylosing Spondylitis. In: Firestein G.S, Budd R.C., Harris JrED, McInnes I.B., Ruddy S, Sergent J.S., editors. Kelley's Textbook of Rheumatology. 8th ed. Philadelphia: Saunders Elsevier; 2009. p. 1169-89.

(4) Zink A, Braun J, Listing J, Wollenhaupt J. Disability and handicap in rheumatoid arthritis and ankylosing spondylitis--results from the German rheumatological database. German Collaborative Arthritis Centers. J Rheumatol 2000 Mar;27(3):613-22.

(5) Boonen A, Chorus A, Miedema H, Van Der Heijde D, van der Tempel H, van der Linden S. Employment, work disability, and work days lost in patients with ankylosing spondylitis: a cross sectional study of Dutch patients. Ann Rheum Dis 2001 Apr;60(4):353-8.

(6) Boonen A, Chorus A, Miedema H, Van Der Heijde D, Landewe R, Schouten H, et al. Withdrawal from labour force due to work disability in patients with ankylosing spondylitis. Ann Rheum Dis 2001 Nov;60(11):1033-9.

(7) Tugwell P, Boers M, Brooks P, Simon L, Strand V, Idzerda L. OMERACT: An international initiative to improve outcome measurement in rheumatology. Trials 2007;8(1):38.

(8) Boers M, Brooks P, Strand V, Tugwell P. The OMERACT Filter for outcome measures in rheumatology. J Rheumatol 1998;25:198-9.
(9) 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 Dec;21(12):2286-91.
(10) Calin A, Garrett S, Whitelock H, Kennedy LG, O'Hea J, Mallorie P, et al. A new approach to defining functional ability in ankylosing spondylitis: the development of the Bath Ankylosing Spondylitis Functional Index. J Rheumatol 1994 Dec;21(12):2281-5.
(11) Jones SD, Steiner A, Garrett SL, Calin A. The Bath Ankylosing Spondylitis Patient Global Score (BAS-G). Br J Rheumatol 1996 Jan;35(1):66-71.
(12) Zochling J, Braun J, Van Der Heijde D. Assessments in ankylosing spondylitis. Best Pract Res Clin Rheumatol 2006 Jun;20(3):521-37.
(13) Van Der Heijde D, Bellamy N, Calin A, Dougados M, Khan MA, van der Linden S. Preliminary core sets for endpoints in ankylosing spondylitis. Assessments in Ankylosing Spondylitis Working Group. J Rheumatol 1997 Nov;24(11):2225-9.
(14) Braun J, Pham T, Sieper J, Davis J, van der Linden S, Dougados M, et al. International ASAS consensus statement for the use of anti-tumour necrosis factor agents in patients with ankylosing spondylitis. Ann Rheum Dis 2003 Sep;62(9):817-24.
(15) Braun J, Bollow M, Neure L, Seipelt E, Seyrekbasan F, Herbst H, et al. Use of immunohistologic and in situ hybridization techniques in the examination of sacroiliac joint biopsy specimens from patients with ankylosing spondylitis. Arthritis Rheum 1995 Apr;38(4):499-505.
(16) Braun J, Davis J, Dougados M, Sieper J, van der Linden S, Van Der Heijde D. First update of the international ASAS consensus statement for the use of anti-TNF agents in patients with ankylosing spondylitis. Ann Rheum Dis 2006 Mar;65(3):316-20.
(17) Maksymowych WP, Inman RD, Gladman D, Thomson G, Stone M, Karsh J, et al. Canadian Rheumatology Association Consensus on the use of anti-tumor necrosis factor-alpha directed therapies in the treatment of spondyloarthritis. J Rheumatol 2003 Jun;30(6):1356-63.
(18) Maksymowych WP, Gladman D, Rahman P, Boonen A, Bykerk V, Choquette D, 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.
(19) Pham T, Fautrel B, Dernis E, Goupille P, Guillemin F, Le Loet X, et al. Recommendations of the French Society for Rheumatology regarding TNF[alpha]

antagonist therapy in patients with ankylosing spondylitis or psoriatic arthritis: 2007 update. Joint Bone Spine 2007 Dec;74(6):638-46.

CHAPTER THREE: AN ANALYSIS OF THE VARIABILITY OF DISEASE ACTIVITY, FUNCTION, AND WELL-BEING IN A YEAR-LONG LONGITUDINAL STUDY OF AS PATIENTS

Chapter overview

The following is a manuscript prepared for publication, based on the analysis of a longitudinal study of AS patients. The objectives of this analysis were to gain insight into the weekly variability of disease activity (BASDAI), physical function (BASFI) and well-being (BAS-G) outcomes in a cohort of ankylosing spondylitis (AS) patients, to explore the relationship between these outcomes, and to assess the effect of patient characteristics on change of these outcomes over a period of one year. The natural changes in disease activity scores and their relation to the current clinical guidelines on anti-TNF use were explored using scenarios of different timing of assessments to investigate the appropriateness of initiation and maintenance in this historical control cohort. This information is useful for clinicians when deciding whether to initiate anti-TNF therapy in their patients.

A copy of the letter of ethics approval from the University of Ottawa for the secondary analysis of the existing dataset is provided in Appendix B.

The copyright permission to reproduce Figure 1 in this thesis is provided in Appendix C.

Further details on the methodology used in this manuscript is provided in Appendix D.

This manuscript was co-authored by the student (LJM) and her co-supervisors, Dr. Peter Tugwell and Dr. George Wells, as well as advisors to the thesis, Dr. Annelies Boonen, Dr. Robert Landewe, and Dr. Desiree van der Heijde. The student is the first author on this manuscript as she had the primary responsibility for data analysis, interpretation, and writing. Drs Tugwell, Wells, Boonen, Heijde and Landewe provided valuable feedback throughout the process.

Assessment of the variability of disease activity, function, and well-being in ankylosing spondylitis in a year-long cohort study

Lara Maxwell, George Wells, Peter Tugwell, Annelies Boonen, Robert Landewé, Désirée van der Heijde

Abstract

Objectives: To gain insight into the weekly variability of disease activity (BASDAI), physical function (BASFI) and well-being (BAS-G) outcomes in a cohort of ankylosing spondylitis (AS) patients, to explore the relationship between these outcomes, and to assess the effect of patient characteristics on change of these outcomes over a period of one year.

Methods: We studied the variability of the main outcomes in a year-long study cohort using various descriptive techniques and employed correlation analysis to explore the relationship between the outcomes. We applied mixed effects modeling to investigate the general patterns of change over time and to explore the association of age, sex, and disease duration with these outcomes. To examine the stability of the estimate of disease activity and how it relates to current guidelines on the use of anti-TNF agents for AS, we used a moving average approach and also ran various scenarios based on different timings of the assessments.

Results: We found large variability in terms of BASDAI, BASFI and BAS-G both between and within subjects, with between-individual variability explaining a greater proportion of the variance. Complex and variable trends over time were evident for the main outcomes. There was a wide range of correlations in the three main outcomes within subjects, but in the majority, disease activity, function, and well-being were quite strongly and positively related. Statistically significant interactions were found between sex and NSAID-intake in explaining the change of BASDAI, BASFI and BAS-G over time. Increasing the number of assessments did not significantly reduce the variability associated with the mean and standard deviation of BASDAI score and very few patients switched categories when the different scenarios of the timing of assessments were applied.

Conclusion: Repeated measures of disease activity, function and well-being fluctuate considerably between patients, with complex patterns occurring over time within patients. The current guidelines for basing the use of anti-TNF agents on two assessments of disease activity four weeks apart fits well with the natural course of disease activity found in this cohort.

Key indexing terms: ankylosing spondylitis, outcome measures, disease activity

Word count=5414
Abstract Word count=350

Introduction

Ankylosing spondylitis (AS) is a chronic, inflammatory rheumatic disease that affects the sacro-iliac joints and often other parts of the spine. The disease course is highly variable and the onset of the disease is usually before age 45, although there is often a 5 to 6 year delay in diagnosis (1). Clinical symptoms usually begin with back pain and stiffness in adolescence and early adulthood and can lead to impaired spinal mobility and/or chest expansion that then lead to impaired ability to carry out activities of daily living.

The goals of treatment of AS are to relieve symptoms (pain, stiffness, joint swelling), improve physical function, and delay or avoid structural damage which leads to disabling deformities.

Although there is awareness among clinicians that disease activity, physical function, and well-being outcome measures of AS patients vary from day to day or frequently, little work has been done to investigate the variation of these outcomes over periods of time. One small study (n=22) conducted in 1991 studied disease activity on a monthly basis over the course of one year (2). A high level of variability was found within each patient's profile and the authors were unable to distinguish clear patterns because of the high individual heterogeneity in disease activity fluctuations. They noted that, "these observations raise the possibility that the natural history of disease in AS may be different for each individual patient, which, if true, would have important implications for the design of studies to test the efficacy of therapy and for approaches to investigating pathogenic mechanisms in patients with AS."

Another study published in March 2008 assessed the week-to-week variability in disease activity and function in 24 patients with AS (3). The results of this study also found high intra-individual variability over the 30-week course of the study. The large fluctuations in disease activity and function led the authors to conclude that the outcome measures used to assess disease activity and physical function in AS, the BASDAI and BASFI, are sensitive tools and encouraged " physicians to obtain

repeated measurements of both scores in order to evaluate disease activity" when deciding whether to initiate anti-TNF therapy. They recommended that the number and frequency of outcome measures needs further investigation. The Berthelot paper concluded that their results, "strongly suggest a role for spontaneous fluctuations in disease activity" and cautioned physicians against switching anti-TNF agents too quickly, given that an increase in BASDAI score at one time point may not necessarily mean a lack of efficacy of the therapy.

Our study aims to explore this intra- and inter-individual variability using a year-long longitudinal study of pre-biologic era patients with ankylosing spondylitis with frequent, that is, weekly, measurements of disease activity, function, and well-being. Building on the results of the studies described above, we will assess the pattern of disease activity in our pre-biologic era cohort. We will examine whether the fluctuations in disease activity were great enough to show if there was substantive natural improvement in disease activity to warrant discontinuation of therapy.

Objectives

The objective of this analysis was to gain insight into the variability of disease activity, physical function, and well-being outcomes over the course of one year and to explore the relationship between these outcome measures.
Specific research objectives were:
1. To describe the variability and patterns over time of disease activity, physical function, and well-being in a cohort of AS patients over a period of one year.
2. To determine the number of consecutive assessments that need to be averaged to provide a stable estimate of disease activity, specifically in the context of establishing baseline disease activity for entry to a randomized controlled trial.
3. To describe the relationship (correlation) between disease activity and function, disease activity and well-being, and function and well-being outcomes within individual patients.
4. To explore whether patient characteristics (age, sex, disease duration) may explain the change over time in disease activity, function, or well-being.

5. To assess whether the variability in disease activity in a pre-biologic era (i.e. a historical control) group of patients has implications for the current guidelines on the use of anti-TNF agents.

Methods

Patients with AS who were registered at the outpatient department of Rheumatology in the Atrium Hospital Heerlen, The Netherlands, were invited by letter to participate. Respondents were consecutively checked to see if they met the modified New York criteria for AS.

Exclusion criteria were:

1. Crohn's disease (present or past).
2. Pain disease not related to AS and which could interfere with disease activity.
3. History of alcoholism, drug abuse, psychological or emotional problems or severe co- morbidity that could result in informed consent not being reasonably accepted and/or could result in limited participation in the study.
4. Non-spinal disease such as peripheral arthritis, uveitis, cervicitis, balanitis or urethritis.
5. Participation in the OASIS (Outcome in Ankylosing Spondylitis International Study) cohort.

After patient consent was obtained, an appointment to visit the out-patient department was made. Data was collected by a weekly phone call and patient assessments included BASDAI, BASFI, BAS-G (last week), and drug use. All patients started the study in December 1999. Patient-important, validated outcome measures used to assess disease activity, physical function, and global well-being in AS include the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) (4), Bath Ankylosing Spondylitis Functional Index (BASFI) (5), and Bath Ankylosing Spondylitis patient Global Score (BAS-G) (6). These outcomes measures have been validated (7). The BASDAI assesses the severity of disease activity in terms of spinal and peripheral joint pain, fatigue, localized tenderness, and morning stiffness. The BASFI assesses functional capability with the aim of capturing one's ability to

cope with daily tasks. An overall sense of the effect of the disease on well-being is assessed with BAS-G.

Methods for each objective:

1. Describing variability in the main outcomes, BASDAI, BASFI, and BAS-G, in of this cohort of patients over the 52 weeks:

1.1. We used box plots and time plots to illustrate the variability per patient

1.2 Overall inter (between)- versus intra (within)-patient variability was assessed using a variance component model and intracluster correlation coefficient (ICC).

1.3 We employed mixed effects modeling to investigate the general patterns of change over time

1.4 We counted the number of weeks each patient exceeded the minimally clinical important difference (MCID) from their median score using the following MCID definitions: BASDAI = 1, BASFI = 0.7, and BAS-G = 1.5, all on a 0 to 10 scale (8)

1.5 We categorized each patient in this cohort into one of the graphical illustrations of flare and remission patterns of disease activity (Figure 1) recently described by Stone et al (9). In their study, a flare was defined as an "exacerbation of the disease that may have required additional treatment or necessitated a visit to a health care professional". Given the retrospective nature of our data, we used a change in 1 point or more in the BASDAI (since the BASDAI MCID =1) to define a flare in our cohort. The patterns are: 1. flares and remissions with periods of symptom-free remission; 2. flares with disease activity in between flares; 3. severe flare of long duration followed by long symptom-free remission; and 4. severe flare of long duration followed by continuing, constant symptoms. In this cohort, we also had some patients who did not have a flare over the course of the 52 weeks and so we created a fifth category to describe this lack of a flare.

We used the chi-square procedure to test for an association between the disease pattern groupings and sex. We used ANOVA and Kruskal-Wallis to test for differences between the disease pattern groupings, age, and disease duration for

normally distributed variables and non-normally distributed variables respectively. Statistical significance was defined as $p<0.05$.

2. We examined the concept of examining whether several clinical assessments provide a more stable estimate of disease activity and are advisable before entering patients into a clinical trial of anti-TNF therapy. We applied a moving-average approach to the first four observations in our dataset and evaluated whether this results in a large reduction in variability. Repeated assessments may be worthwhile if they result in significantly reduced standard deviations of our measurements.

3. We used the Pearson and Spearman correlation statistics to investigate the relationship between the three main outcomes of interest: disease activity (BASDAI) and function (BASFI) disease activity (BASDAI) and well-being (BAS-G), and function (BASFI) and well-being (BAS-G) for each patient.

4. We used mixed effects modeling to investigate the general patterns of change over time in disease activity, function, and well-being and to explore the association of age, sex, and disease duration with these outcomes. We modeled time as a random factor and used a random coefficient analysis which included both a random intercept and a random slope. We used $p<0.05$ as the cut-point for keeping variables in the model and used backward elimination to remove non-significant variables. In selecting the best fit model we used Akaike's and Schawarz's information criteria (AIC and BIC, respectively) where the model with the smaller information criteria is preferred. We entered the following variables in the maximum model: age, sex, disease duration, daily dose of NSAID, time modeled as linear, quadratic, and cubic functions, and interaction terms of time with the patient characteristic variables mentioned above. Any statistically significant interaction terms were further explored graphically in an attempt to provide an explanation for the interaction.

5. As this cohort of patients were assessed in 2000, before anti-TNF therapy was available as a treatment option, they can be considered a historical control that can be used to make decisions on when biologics are indicated or should be switched. Based on the ASAS recommendation of patient selection for initiation of anti-TNF therapy (10) of active disease for >4 weeks and BASDAI ≥ 4 (0–10 scale), we assumed that two consecutive BASDAI scores greater than four and measured four

weeks apart would be sufficient to initiate biologic therapy. We then assessed whether the BASDAI score every four weeks, from the point of initiation, either stayed above 4 or declined below 4 for 12 consecutive weeks, as per ASAS guidelines. Week one was taken as the arbitrary starting point of the assessment. We categorized patients into one of three categories:

Category 1: a level of disease activity that would have qualified a patient for biologic treatment (defined as a BASDAI ≥4 for 2 consecutive visits 4 weeks apart), and the natural course of their BASDAI suggests that initiation and continuation of biologic therapy would have been warranted (i.e. BASDAI was maintained ≥4 at 12 weeks follow-up);

Category 2: while two consecutive ≥4 BASDAI scores qualified a patient for biologic treatment, 12 weeks after this point there was a natural decline in disease activity and biologic therapy may not have been necessary (i.e. BASDAI was not ≥4 at 12 weeks follow-up);

Category 3: there was no point at which biologic treatment would have been warranted because there were no two consecutive BASDAI scores ≥4, measured 4 weeks apart.

We repeated the above analysis using different assessment scenarios of three assessments, four weeks apart and four assessments, four weeks apart, while keeping the same follow-up time of 12 weeks, to determine if the number of patients in each category changed significantly depending on the assessment scenario. We hypothesized that if many patients were switching categories in the different scenarios, then more than the two currently recommended assessments may be required.

Ethics approval was obtained from the University of Ottawa for the analysis of this dataset.

Results

Forty-two patients agreed to participate and were included in the study which involved an intense schedule of weekly assessments. The characteristics of the study cohort are provided in Table 1.

All 42 patients provided data on BASDAI, BASFI, and BAS-G for 52 weeks for a total of 2184 observations for each outcome. Data on the daily dose of NSAIDs per week were available for weeks 6 to 52. Only four patients were on DMARD therapy for the 52 weeks. No patients received anti-TNF or other biologic therapy.

Objective 1: To describe the variability and patterns over time of disease activity, physical function, and well-being in a cohort of AS patients over a period of one year.

1.1 Box plots and time plots (Figures 2, 3 and 4) showed that there was large variability both within-and between-patients over the 52 weeks, in terms of disease activity, function and well-being. The trend line in Figure 3 shows that there does not appear to be an increasing or decreasing linear trend, or a quadratic or cubic trend, for the mean BASDAI score of the sample over the year; it remained relatively constant around 3.8. Similar results were seen for function and well-being (data not shown).

Descriptive statistics for disease activity, function, and well-being outcomes for the entire cohort are found in Table 2.

1.2 We explored the inter- and intra- subject variability across the 42 subjects. We found that the inter-subject variance (3.11) contributes substantively more than the intra-subject variance (0.79) to the overall variability of disease activity. This may be explained by the fact that some patients have a fairly stable disease activity pattern over the year while others fluctuate greatly, so we would expect to see a larger variability between the patients compared to the variability of the pattern of disease activity within a patient. The intracluster correlation (ICC) coefficient was 0.80. This high ICC indicates that the proportion of the true total variation in the BASDAI can be attributed more to differences in the pattern of disease activity over the year between patients than differences within patients. Similarly high ICC were found for function and global well-being outcomes (0.89 and 0.74, respectively).

1.3 Using mixed effects modeling to model the effect of time on the three main outcomes of interest, we found there was significant quadratic and cubic trends for

the change in the population mean of the BASDAI, BASFI, and BAS-G scores over the year (Table 3). The average trend line (Figure 3) appears fairly stable over the course of the year, but the mixed effects modeling demonstrates that the changing patterns in this dataset are quite complex, with evidence of quadratic and cubic trends. When the individual patient time plots are assessed, it is clear that different patients experience different patterns over the year.

1.4 In another assessment of variability, we counted the number of patients who had a change (either positive or negative) from their median score that was larger than the minimally important clinical difference (MCID) of disease activity (BASDAI=1), function (BASFI=0.7), and global well-being (BAS-G=1.5). We counted the number of weeks each patient exceed the MCID (histogram) and found that all but 3 out of 42 patients had an absolute change in their BASDAI (from their median BASDAI score) greater than one point over the course of the 52 weeks. The majority of patients (24/42, 57%) exceeded the MCID between 1 and 10 times over the 52 weeks; 12/42 (29%) exceeded it between 11 and 20 times; and 7% (3/42) were highly variable and changed more than 20 times by more than 1 point from their median BASDAI .

Similarly, 39/42 patients had an absolute change in their BASFI score greater than 0.7. A change of more than 1.5 points on the BAS-G scale was found in 37/42 patients. This analysis demonstrates that the majority of patients' BASDAI, BASFI, and BAS-G scores fluctuated to a significant extent throughout the 52-week study. Figure 5 shows the number of weeks that each patient exceeded the minimal clinically important difference in BASDAI.

1.5 We classified participants into one of the five disease activity patterns identified by Stone et al. (9) as described in the methods section. The patient distribution into each of the patterns is described in Table 4.

Similar to the paper by Stone, the majority of patients in this cohort experienced a pattern of ongoing flares with disease activity in between the flares. None of the patients in this cohort experienced a pattern of long disease activity followed by a prolonged symptom-free state. Although a pattern of no flares was not identified by Stone, we found five patients (12%) who did not experience any flare over the 52 weeks.

Chi-square test showed we do not have evidence of an association of disease patterns and sex (p=0.80). Kruskal-Wallis tests showed there was no evidence of an association between the different disease patterns and disease duration (p=0.09). ANOVA, using Proc GLM to account for the unequal sample sizes in the different groups, showed there was also no evidence of an association between the different disease patterns and age (P=0.30). The Kruskal-Wallis test gave the same result indicating the robustness of the estimate.

Objective 2: To determine the number of consecutive assessments that need to be averaged to provide a stable estimate of disease activity, specifically in the context of establishing baseline disease activity for a randomized controlled trial.

We computed the sample mean and SD for the weekly BASDAI trend and for each moving average analysis using the first four observations in the dataset. As expected, the mean BASDAI was similar across the four analyses (Table 5). Similar results were found for the standard deviation, indicating that there does not appear to be compelling evidence that taking more than one repeated measurement will significantly reduce the variability of the disease activity outcome. The CV was similar, varying between 47.6% to 48.8%.

Objective 3: To describe the relationship (correlation) between disease activity and function, disease activity and well-being, and function and well-being outcomes within individual patients

The mean, median, minimum, and maximum, correlations from the individual patient results are given in Table 6. There was large variability in the correlation

results for these outcomes across the various individual patients. A graphical display of the correlation between the three outcomes of interest for two patients (patient ID= 25 and 26) demonstrates the extremes of the associations seen in individual patients (Figure 8). Patient 26 has a high correlation between BASDAI and BASFI (r=0.92), BASFI and BAS-G (r=0.94) and BASDAI and BAS-G (r=0.93). With patient 25, there is a weak, though positive, correlation between BASDAI and BASFI (r=0.32), no association between BASFI and BAS-G (r=-0.06) and a weak correlation between BASDAI and BAS-G (r=0.17). Results were almost identical when the Spearman correlation statistic was used.

Objective 4: To explore whether patient characteristics (age, sex, disease duration) may explain the change over time in disease activity, function, or well-being.
The best-fitting mixed models for BASDAI, BASFI and BAS-G along with associated P-values are described in Table 7. For all three outcomes, neither disease duration and age, nor their interaction terms, were statistically significant indicating they could be removed from the model. However, we kept age in as we thought was clinically important. We found statistically significant interactions with sex and time and daily dose of NSAID intake and various functions of time. We further explored these interactions by graphing the predicted mean of the outcome variable by males and females and by the following groupings of NSAID intake, measured in daily doses per week: 0, 1 to 6, 7, and >7.

Figures 6 and 7 show the graphical representation of the significant interactions between sex and NSAID use. Similar results were found for BASFI and BAS-G (data not provided). In women, predicted mean BASDAI and BAS-G scores showed a cubic pattern, showing a steeper drop in scores in the early weeks of the study before increasing again. In men, the BASDAI score remained fairly stable over the year. And for BASFI, the women's scores were consistently worse than the men's scores over the 52 weeks. In those patients taking less than a daily dose of NSAID per week, we see lower overall predicted mean values of BASDAI and BASFI (graph not shown) compared to those with a daily dose of NSAIDs or higher and a

quadratic trend in the lower NSAID-intake group while the higher groups have more of a cubic trend.

We undertook model diagnostics to assess the fit of our model (Appendix, Figures 9 & 10). The histogram of residuals had a normal distribution and the normal quantile plot of residuals looked close to a straight line with the exception of a few outliers. The plots of residuals versus predicted values and covariates showed the presence of some outliers in the data, but no discernible systematic trend. The line of the lowess smoothed curve was almost straight indicating no systematic trend. The distribution of random effects for intercept and week also appeared normal.

Objective 5: To assess whether the variability in disease activity in a pre-biologic era (i.e. a historical control) group of patients has implications for the current guidelines on the use of anti-TNF agents.

We categorized patients into one of three categories based on the current guidelines for initiation and maintenance of anti-TNF therapy, which is two assessments, four weeks apart. We then explored different scenarios of changing the number of assessments to three or four assessments, both four weeks apart, to determine the change in the number of patients assigned to each category. The percentage of patients who met each category in the three scenarios and the distribution of patients across categories when the different assessment scenarios are compared are provided in Table 8 and 9, respectively.

When assessing this cohort using the current guidelines for initiation and maintenance of anti-TNF therapy, treatment would have been warranted in 41% of the patients. Approximately 43% of the patients did not have two consecutive BASDAI scores ≥4 to warrant the initiation of biologic therapy. Of interest, in 16.7% of patients, two consecutive high BASDAI scores might have qualified them for biologic treatment, but after this point there was a natural decline in disease activity at the 12-week follow-up point. Therefore, treatment with biologic therapy may not have been necessary in these patients. Changing the scenario to three assessments,

four weeks apart, before initiation of therapy increased the number of patients who did not have three consecutive BASDAI scores >=4 to 52.4% and when four consecutive assessments were considered, 61.9% of patients did not meet the requirements for initiation of therapy. The percentage of patients in category 1 stayed quite consistent across the three scenarios. When the most stringent assessment scenario is applied, only one patient has a natural decline in BASDAI scores to suggest that after 12 weeks, the natural course of the disease improved to the point where anti-TNF therapy may not be required.

The number of patients who changed categories based on the different assessment scenarios is described in Table 9. It clearly shows that about 1/3 of patients (35%) have high disease activity throughout the year and stay in category 1, regardless of the timing of the assessments. Only two patients switch from category 1 to category 2 if the scenario is changed from two to three assessments and none switch between scenarios three to four; thus indicating that the majority of patients show stability over the timing of the assessments.

Discussion

In this year-long longitudinal study of 42 patients with ankylosing spondylitis, we found there was evidence of large variability both between and within patients in terms of the patterns of disease activity, function, and well-being as measured on weekly basis. Our descriptive statistics showed individual disease activity patterns ranging from very stable to largely fluctuating over the 52 weeks. Minimum and maximum BASDAI, BASFI and BAS-G scores cover almost the entire 0 to 10 scale. The three main outcomes of interest indicated that there were quadratic and cubic patterns occurring over time in this cohort. We also found that the majority of patients' BASDAI, BASFI, and BAS-G scores fluctuated in a clinically important way over the year, based on the number of times each patient exceeded the MCID from their median score. All but three patients exceed the BASDAI MCID of one point over the year, and at the other extreme, three patients were very highly variable and changed more than 20 times by more than one point from their median score.

We also found that more of the variation in BASDAI, BASFI, and BAS-G scores can be attributed to the differences found between patients, than variation of scores within patients. The different patterns of disease activity experienced by AS patients were recently explored by researchers in the UK and all of their patients experienced flares. While 12% of our cohort did not experience a flare in disease activity over the year they were followed, the majority (60%) experienced one or more flares with ongoing disease activity in between which is similar to the Stone et al. paper (9).

This descriptive analysis provides a picture of the considerable variability in clinical symptoms both across patients and within many patients and is consistent with previous studies which also found high inter- and intra-individual variability (2;3). Our results also demonstrate the ability of the BASDAI, BASFI and BAS-G instruments to detect frequent changes in these outcome measures.

This variability is also demonstrated in radiographic progression. Similar to the debate about the progression of clinical symptoms, there has been little data available on the natural radiographic progression of AS. However, a recent study (11) sought to describe the natural course of radiographic progression in a cohort of patients with AS. They characterized patients into three groups based on their rate of radiographic progression - fast, moderate, and slow - and also found evidence of substantial variation in radiographic progression between individual patients; from 23% of patients showing no progression to a 4-times greater rate of progression than the mean cohort progression in 43% of patients.

In terms of the number of assessments needed to determine whether a patient should begin anti-TNF therapy in the context of an RCT, we determined that increasing the number of assessments from one to four did not significantly reduce the variability associated with the mean and standard deviation of BASDAI scores. Although there is clearly a lot of fluctuation in disease activity over the course of a year in our cohort of patients, we do not have evidence to suggest that increasing the number of

repeated assessments for a baseline assessment over a short period of time (e.g. 4 weeks) is a worthwhile way to reduce the variability.

We explored the appropriate number of assessments necessary for initiation of anti-TNF therapy by creating different scenarios of the timing of assessments and determining whether many patients changed between the following categories: 1. initiation and maintenance of anti-TNF therapy was warranted; 2. a natural decline in disease activity to the point where anti-TNF therapy may not have been needed; or 3. no point at which consecutive disease activity scores were high enough to warrant initiation of therapy. Approximately 1/3 of patients had consistently high disease activity throughout the 52 weeks which warranted anti-TNF therapy initiation and maintenance. Of interest, very few patients switched from category 1 to 2 when the current guidelines of two assessments of disease activity four weeks apart fits well with the natural course of disease activity seen in this cohort different assessment scenarios were applied, indicating that, in this cohort of patients, taking more than two assessments of disease activity was not necessary. Together, these two findings demonstrate that the current guidelines of two assessments of disease activity four weeks apart fits well with the natural course of disease activity seen in this cohort. Similar to our conclusion, a recent review on the management of ankylosing spondylitis (12) commented that the Berthelot study "proposal to perform several evaluations before starting or stopping anti-tumour necrosis factor (TNF) therapy does not seem to be really feasible and necessary".

We found positive relationships between BASDAI, BASFI, and BAS-G, with disease activity and well-being being the most strongly correlated. This is understandable given that patients reporting higher disease activity would be likely to report a more severe effect of AS on their well-being. As expected, the correlation coefficients of the individual patients ranged widely for the relationship between disease activity and function. The relationship between physical function and disease activity and structural damage was recently investigated (13) and it was shown that disease activity and structural damage both independently predict

changes in function. BASDAI was found to correlate well with BASFI, which is similar to our result.

Although our study is comprised of a relatively small number of participants, the strength of this study lies in the frequent collection of data. Using mixed effects modeling allowed us to explore the relationship of patient characteristics to the change over time in disease activity, function and well-being, while taking into account the correlated nature of this data. We found evidence of both quadratic and cubic relationships over time, which reinforces the complexity of the patterns of disease activity, function, and well-being in AS. In addition, patterns in men tended to be more stable while women had stronger cubic and quadratic patterns. Over time, the trend lines of men and women intersected, demonstrating that women did not have consistently better disease activity than men. Thus, we do not have strong evidence from this cohort to support the theory that women have less severe disease than men as seen in other cohort studies (14;15). There was a significant interaction between NSAID-intake and each of the three main outcomes; higher doses of NSAIDs were associated with higher BASDAI, BASFI and BAS-G scores. It is not surprising that people with higher disease activity reported taking more NSAIDs.

This dataset is remarkably complete in terms of the main outcomes of interest; there was no missing data for BASDAI, BASFI, or BAS-G from the 42 participants over the 52 weeks of the study and all participants completed the study. However, there was a significant amount of missing data on the status of the genetic marker HLA-B27, so we were unable to use this as a covariate and data on NSAID use was not collected before week six. The frequency of the data collection allowed us to explore patterns in regular changes in patient's disease activity, function, and well-being over time. The repeated measures nature of the data meant that sophisticated modeling techniques were required to handle the intrinsic correlations within individuals and mixed effects modeling was an appropriate method to use. We would expect that recall bias would not be a significant problem for this study since patients were asked about their health status for the previous week, which is not a long period of

time. Participant recruitment was through a hospital outpatient rheumatology department which would be less likely to select for patients with severe disease compared to those admitted to hospital.

However, despite the strengths listed above, this study does have some limitations. While this data was collected prospectively with a specific research question in mind, the researchers conducting the original study did not obtain information on the patients' perception of when their disease activity was stable so that we could relate this back to the BASDAI, BASFI and BAS-G outcomes. Ankylosing spondylitis is usually diagnosed quite early in life and is a chronic disease; however, we only had data for one year, in effect a 'snapshot' of the disease course at which different patients were at different stages of disease duration. Yet the high frequency of the data collection gives us a unique view of the frequency of fluctuation of disease activity, function and well being in AS patients. The sample size was not very large, but was larger than two other prospective studies on this topic and the frequent repeated measures of the participants provided a great deal of information about the individual's change over time. In our analysis of how the variability in disease activity relates to the current guidelines on the use of anti-TNF agents, we could only use BASDAI scores while guidelines also recommend that 'expert opinion' be taken into account when making a treatment decision.

In conclusion, we found that frequent repeated measures of disease activity, function and well-being fluctuate considerably between patients, with complex patterns occurring over time within patients. The current guidelines on the use of anti-TNF agents of two assessments of disease activity four weeks apart fits well with the natural course of disease activity found in this cohort.

Acknowledgment

Thank you to M Curf who entered the data on NSAIDs and calculated the 'proportion of the full weekly NSAID dose' patients reported.

References

(1) Khan MA. Update on spondyloarthropathies. Ann Intern Med 2002 Jun 18;136(12):896-907.

(2) Goodacre JA, Mander M, Dick WC. Patients with ankylosing spondylitis show individual patterns of variation in disease activity. Br J Rheumatol 1991 Oct;30(5):336-8.

(3) Berthelot J, Tortellier L, Lavy-Bregeon D, Le Goff B, Maugars Y. High intraindividual week-to-week variability in BASDAI and BASFI values: Are several evaluations needed before starting or stopping TNF antagonist therapy for spondyloarthropathies? Joint Bone Spine 2008;75:167-71.

(4) 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 Dec;21(12):2286-91.

(5) Calin A, Garrett S, Whitelock H, Kennedy LG, O'Hea J, Mallorie P, et al. A new approach to defining functional ability in ankylosing spondylitis: the development of the Bath Ankylosing Spondylitis Functional Index. J Rheumatol 1994 Dec;21(12):2281-5.

(6) Jones SD, Steiner A, Garrett SL, Calin A. The Bath Ankylosing Spondylitis Patient Global Score (BAS-G). Br J Rheumatol 1996 Jan;35(1):66-71.

(7) Zochling J, Braun J, Van Der Heijde D. Assessments in ankylosing spondylitis. Best Pract Res Clin Rheumatol 2006 Jun;20(3):521-37.

(8) Pavy S, Brophy S, Calin A. Establishment of the minimum clinically important difference for the bath ankylosing spondylitis indices: a prospective study. J Rheumatol 2005 Jan;32(1):80-5.

(9) Stone MA, Pomeroy E, Keat A, Sengupta R, Hickey S, Dieppe P, et al. Assessment of the impact of flares in ankylosing spondylitis disease activity using the Flare Illustration. Rheumatology (Oxford) 2008 Aug;47(8):1213-8.

(10) Braun J, Davis J, Dougados M, Sieper J, van der Linden S, Van Der Heijde D. First update of the international ASAS consensus statement for the use of anti-TNF agents in patients with ankylosing spondylitis. Ann Rheum Dis 2006 Mar;65(3):316-20.

(11) Baraliakos X, Listing J, von der RA, Braun J. The natural course of radiographic progression in ankylosing spondylitis--evidence for major individual variations in a large proportion of patients. J Rheumatol 2009 May;36(5):997-1002.

(12) Braun J, Baraliakos X. Treatment of ankylosing spondylitis and other spondyloarthritides. Curr Opin Rheumatol 2009 Jul;21(4):324-34.

(13) Landewe R, Dougados M, Mielants H, van der Tempel H, Van Der Heijde D. Physical function in ankylosing spondylitis is independently determined

by both disease activity and radiographic damage of the spine. Ann Rheum Dis 2009 Jun;68(6):863-7.

(14) Jimenez-Balderas FJ, Mintz G. Ankylosing spondylitis: clinical course in women and men. J Rheumatol 1993 Dec;20(12):2069-72.

(15) Robertson LP, Davis MJ. A longitudinal study of disease activity and functional status in a hospital cohort of patients with ankylosing spondylitis. Rheumatology (Oxford) 2004 Dec;43(12):1565-8.

Tables and Figures

Table 1 Characteristics of study cohort

Total N	42
Male:Female	33:9
Mean, median age years (SD, range, IQR)	46.0, 43.0 (11.1, 21 to 69, 15)
Mean, median length of disease duration, yrs (SD, range, IQR)	11.3, 10.5 (6.4, , 2 to 35, 9)
HLA-B27 +, n/N (%)*	24/28 (85.7)

*HLA-B27+ status only available for 28 patients; IQR=inter-quartile range

Table 2: Descriptive statistics for disease activity (BASDAI), function (BASFI), and well-being (BAS-G) of the overall cohort over the 52 weeks

Outcome	Mean (SD)	Median	Minimum	Maximum	IQR
BASDAI (0 to 10 scale)	3.49 (1.96)	3.50	0	9.3	2.9
BASFI (0 to 10 scale)	3.46 (2.12)	3.40	0	9.6	3.6
BAS-G (0 to 10 scale)	3.74 (2.20)	4.00	0	10	3.0

Note: higher scores mean worsening disease activity, function, and well-being; IQR=interquartile range

Table 3: Results of mixed modeling for BASDAI, BASFI, BAS-G with various functions of time

	Linear trend, P-value	Quadratic trend, P-value	Cubic trend, P-value
BASDAI	0.915	0.028	0.0001
BASDFI	0.618	0.016	0.0001
BAS-G	0.558	0.013	0.0002

Table 4: Distribution of patients into the each of the disease activity patterns

	Disease pattern				
	Flares & remisssion	Flares & ongoing disease activity	Severe, lengthy flare, then no symptoms	Severe, lengthy flare, then ongoing disease activity	No flare
Number of patients	9	25	0	3	5

Table 5: Comparison of the BASDAI for the weekly trend and for two-week, three-week, and four-week moving average data (based on first four weeks in the dataset)

	Week 1	Moving average of 2 weeks	Moving average of 3 weeks	Moving average of 4 weeks
Mean BASDAI (95% CI)	3.74 (3.18 to 4.29)	3.81(3.23 to 4.39)	3.82 (3.24 to 4.39)	3.76(3.19 to 4.33)
SD (95% CI)	1.78 (1.47 to 2.27)	1.86 (1.53 to 2.37)	1.83 (1.51 to 2.34)	1.82 (1.50 to 2.32)
CV %	47.6	48.8	47.9	48.4

SD=standard deviation; CI= confidence interval; CV=coefficient of variation (SD/mean)

Table 6: Descriptive results from the correlation analysis of all patients (N=42)

Pearson correlation statistic	BASDAI:BASFI	BASFI:BAS-G	BASDAI:BAS-G
Mean	0.68	0.61	0.68
SD	0.27	0.30	0.20
Median	0.75	0.69	0.75
Minimum	-0.63	-0.29	-0.05
Maximum	0.94	0.96	0.94
IQR	0.20	0.26	0.24

Table 7: Best-fitting models of association between BASDAI, BASFI, BAS-G, and patient characteristic variables

BASDAI		**BASFI**		**BAS-G**	
Variable	**P-value**	**Variable**	**P-value**	**Variable**	**P-value**
Week	<.0001	Week	<.0001	Week	<.0001
Weeksq	<.0001	Weeksq	<.0001	Weeksq	<.0001
Weekcu	<.0001	Weekcu	<.0001	Weekcu	<.0001
Age	0.3018	Age	0.3860	Age	0.3527
Sex	0.0004	Sex	0.1467	Sex	0.0001
NSAID	0.0847	NSAID	0.0149	NSAID	0.0593
Week*Sex	<.0001	Week*Sex	<.0001	Week*Sex	<.0001
Week*NSAID	<.0001	Week*NSAID	<.0001	Week*NSAID	<.0001
Weeksq*Sex	<.0001	Weeksq*Sex	<.0001	Weeksq*Sex	<.0001
Weeksq*NSAID	<.0001	Weeksq*NSAID	<.0001	Weeksq*NSAID	<.0001
Weekcu*Sex	<.0001	Weekcu*Sex	<.0001	Weekcu*Sex	<.0001
Weekcu*NSAID	<.0001	Weekcu*NSAID	<.0001	Weekcu*NSAID	<.0001

Weeksq=week*week; weekcu=week*week*week

Table 8: Categorization of patients based on current guidelines for initiation and maintenance of anti-TNF therapy (2 assessments) and two other scenarios of assessments

	Category 1: Initiation of biologic therapy was warranted, followed by consistently high disease activity scores	Category 2: Initiation of biologic therapy was warranted, followed by a natural decline in disease activity	Category 3: No consecutive BASDAI scores >=4 to warrant initiation of biologic therapy
2 assessments, % patients	40.5	16.7	42.9
3 assessments, % patients	35.7	11.9	52.4
4 assessments, % patients	35.7	2.4	61.9

Table 9: Comparison of the number of patients assigned to each category based on different assessment scenarios

		3 assessments, 4 weeks apart		
		Category 1	Category 2	Category 3
2 assessments, 4 weeks apart	Category 1	15	2	0
	Category 2	0	3	4
	Category 3	0	0	18

		4 assessments, 4 weeks apart		
		Category 1	Category 2	Category 3
3 assessments, 4 weeks apart	Category 1	15	0	0
	Category 2	0	1	4
	Category 3	0	0	22

Category 1: Initiation of biologic therapy was warranted, followed by consistently high disease activity scores
Category 2: Initiation of biologic therapy was warranted, followed by a natural decline in disease activity
Category 3: No consecutive BASDAI scores ≥4 to warrant initiation of biologic therapy

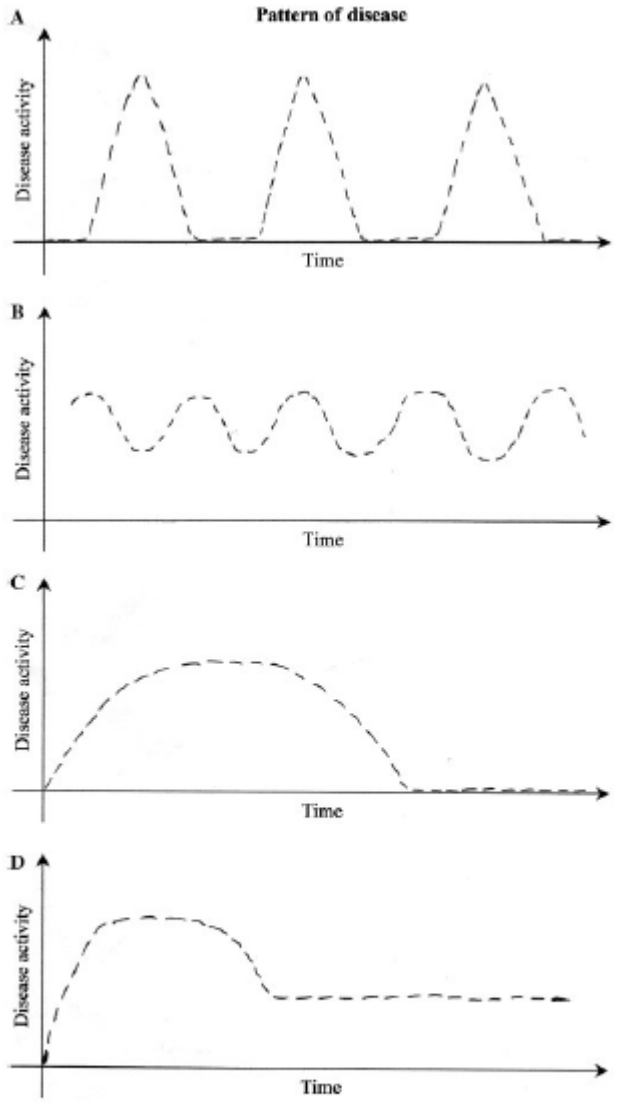


FIG. 1. The Flare Illustration shown to patients. The following patterns of disease activity are represented graphically. Pattern (**A**) Flares and remissions. During the remissions patients would be symptom free. Pattern (**B**) Flares with disease activity in between flares. Pattern (**C**) Severe flare lasting a long time and followed by return to a baseline for a long period of time without any symptoms. Pattern (**D**) This was similar to (C) except that patients did not return to a symptom-free baseline but had continued symptoms that were constant in nature over time thereafter.

Figure 1: Graphical representation of flare illustrations of patterns of disease in ankylosing spondylitis patients as identified by Stone et al (Stone MA, et al. Assessment of the impact of flares in ankylosing spondylitis disease activity using the Flare Illustration. Rheumatology. 2008,47:1213–1218. Pg. 1214, by copyright permission of Oxford University Press and the British Society for Rheumatology)

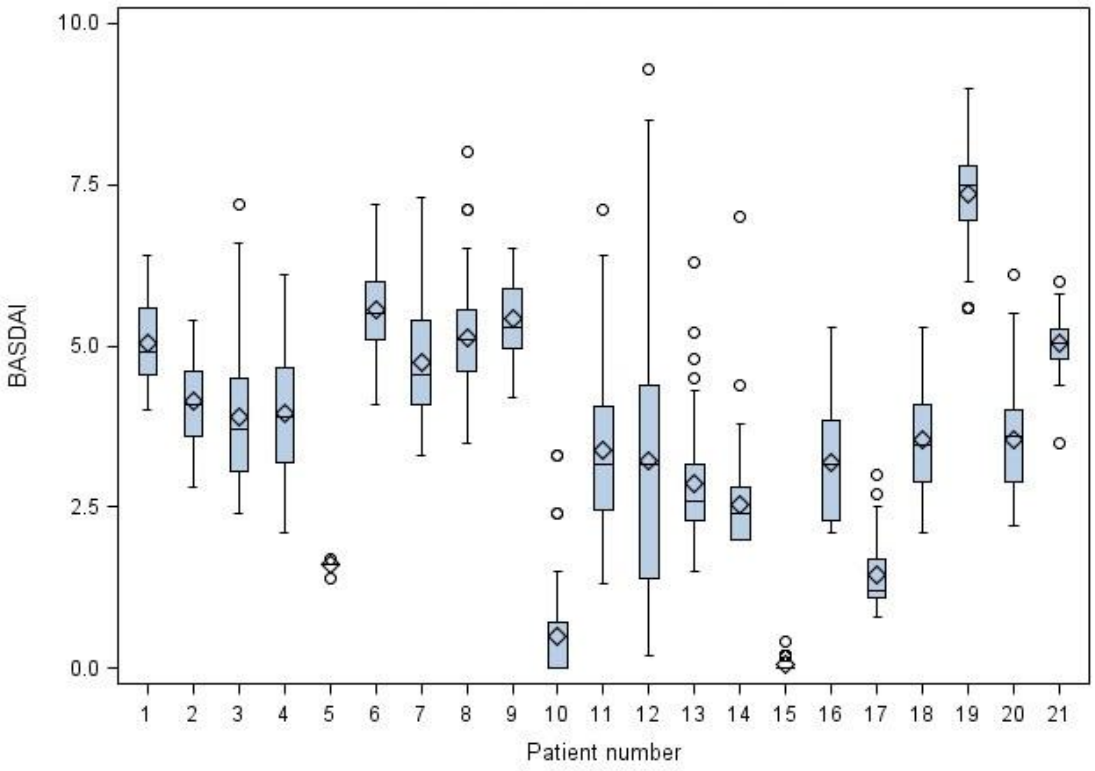


Figure 2: Boxplot of BASDAI scores for patients 1 to 21

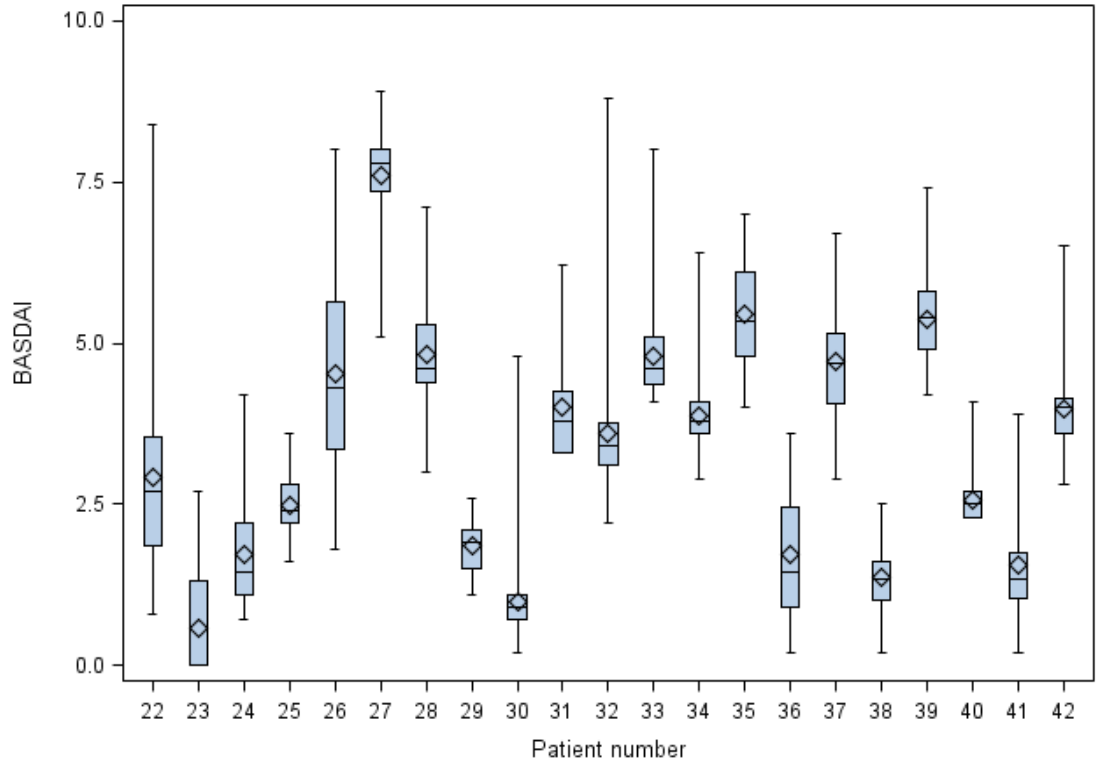


Figure 3: Boxplot of BASDAI scores for patients 22 to 42

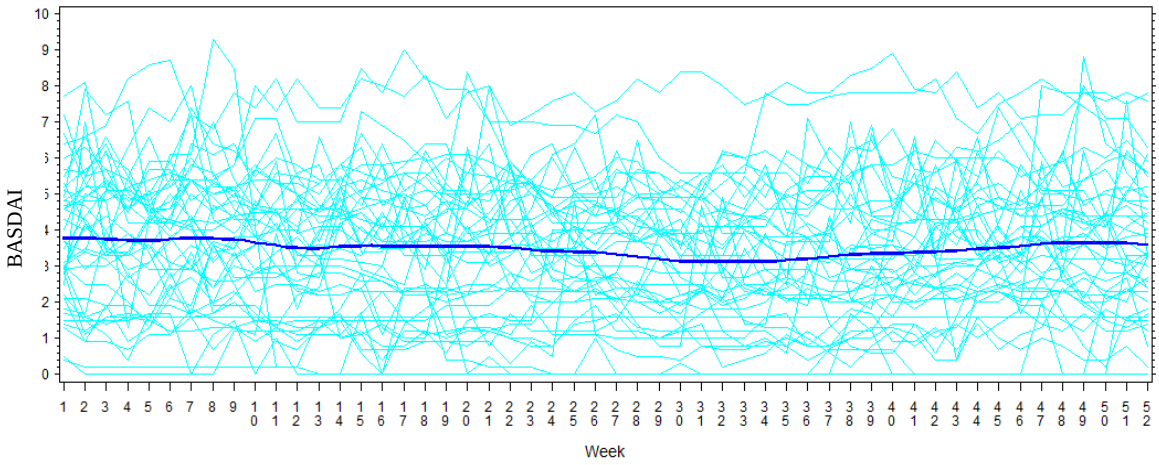


Figure 4: Time plot of BASDAI scores over 52 weeks with the average trend line for the 42 patients

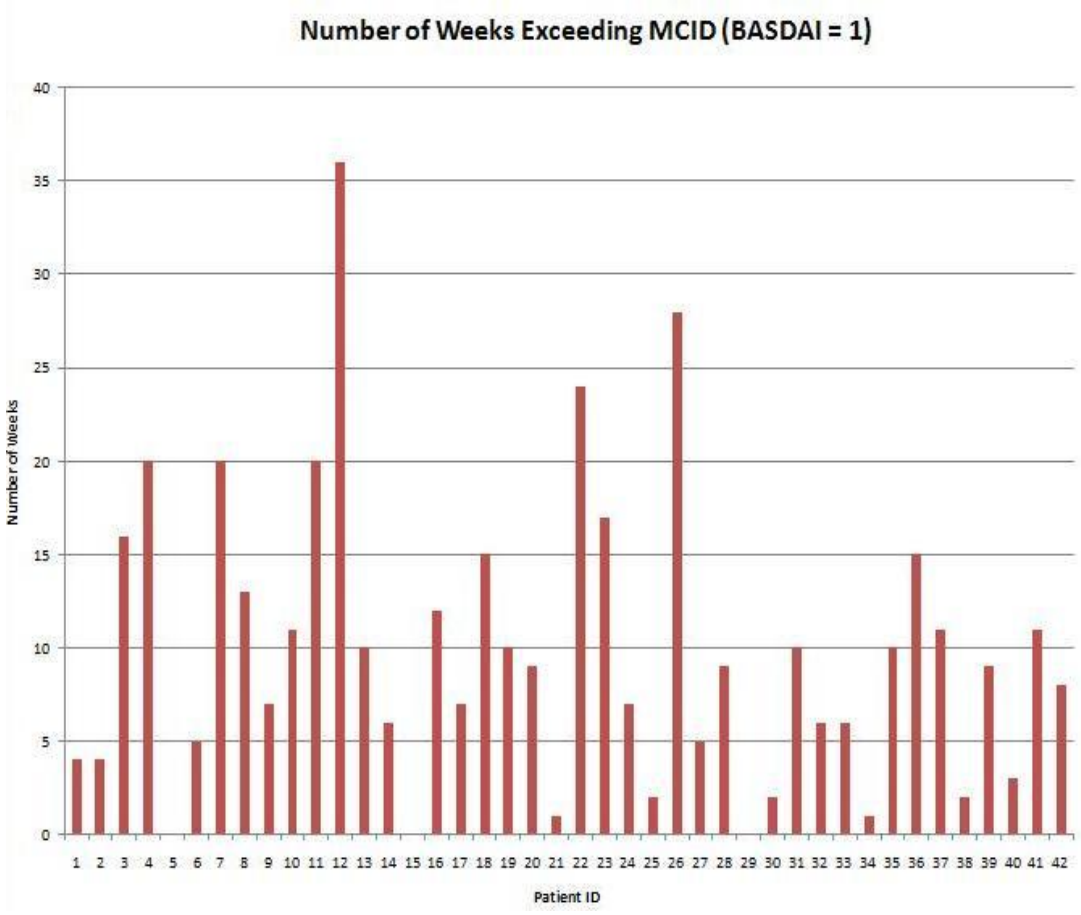


Figure 5: Histogram of the number of weeks that each patient exceeded the minimal clinically important difference in BASDAI (MCID = 1, on 0 to 10 scale)

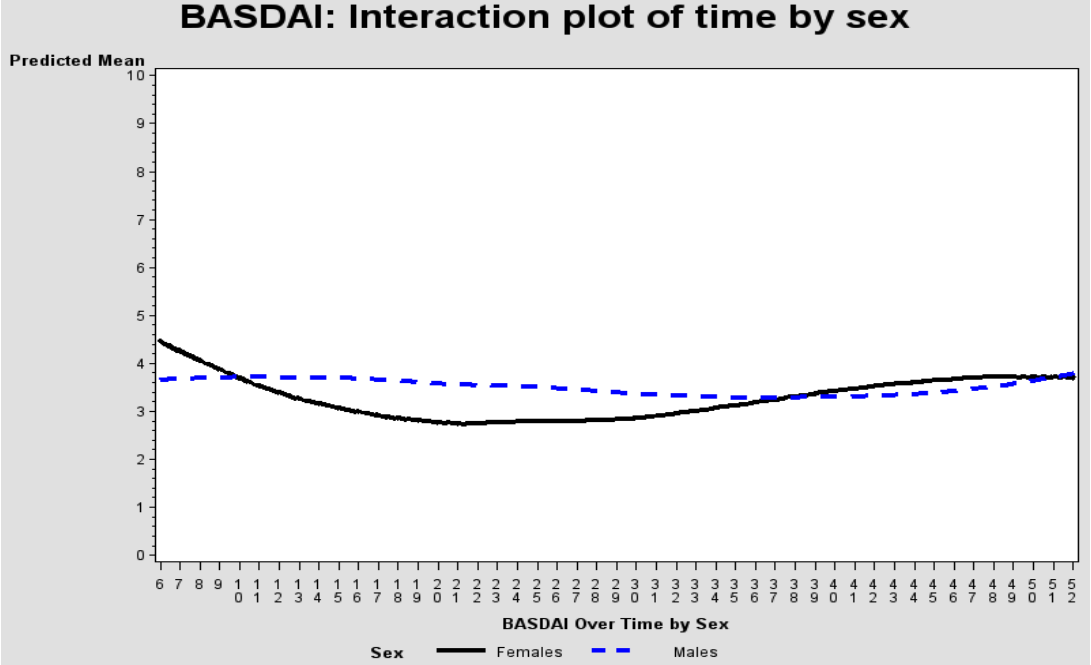
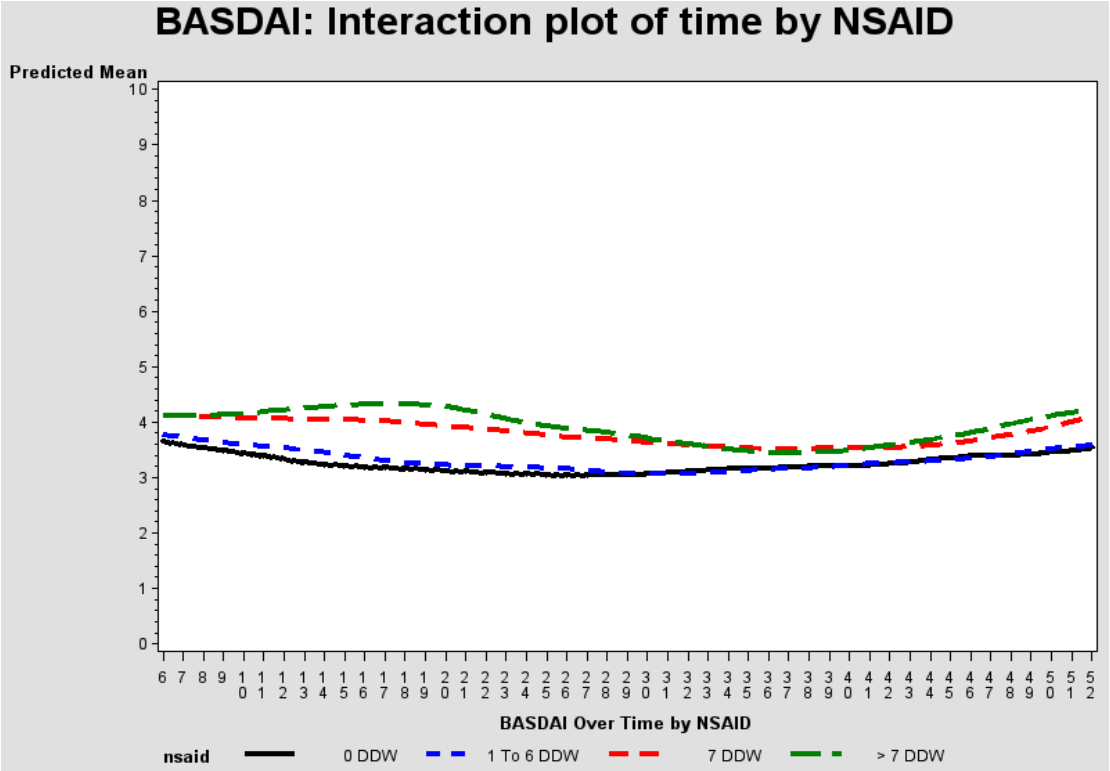


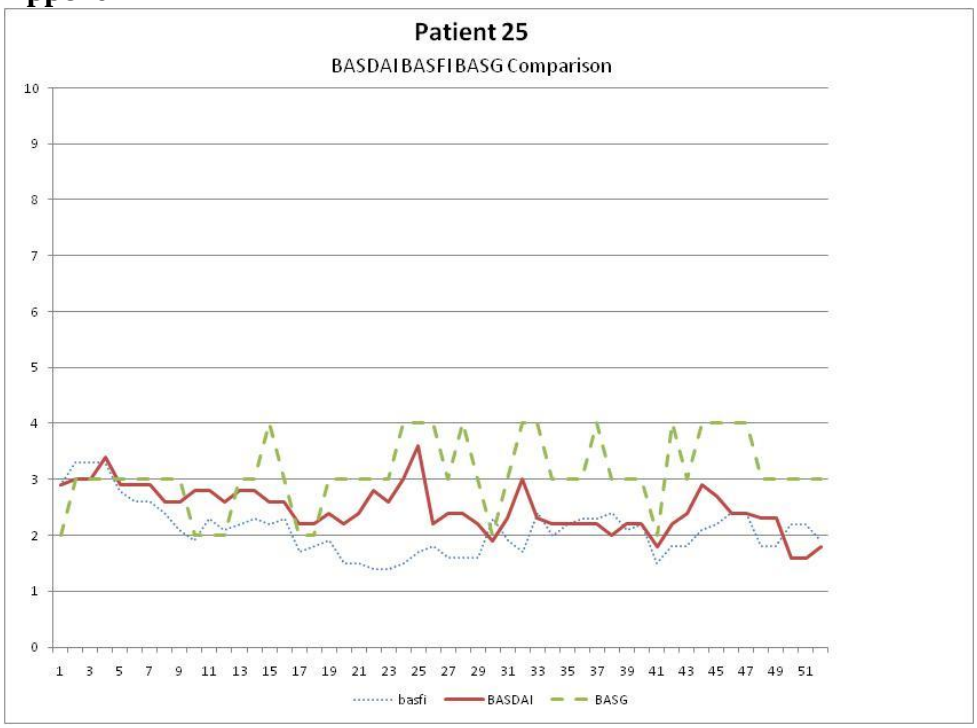
Figure 6: Plot of the predicted mean of BASDAI by sex from the mixed model analysis



DDW=daily dose per week

Figure 7: Plot of the predicted mean of BASDAI by daily dose per week of NSAID intake from the mixed model analysis

Appendix



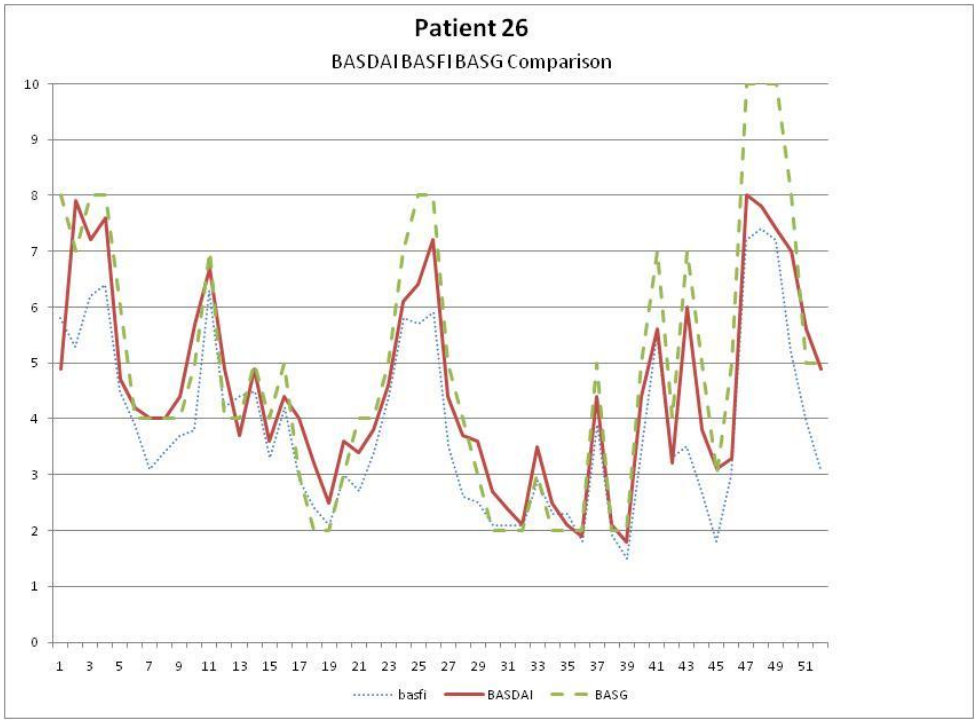


Figure 8. A graphical display of the correlation between BASDAI, BASFI, and BAS-G for two patients (patient ID= 25 and 26) demonstrates the extremes of the associations seen in individual patients (patient 25=weak correlation; patient 26=strong correlation)

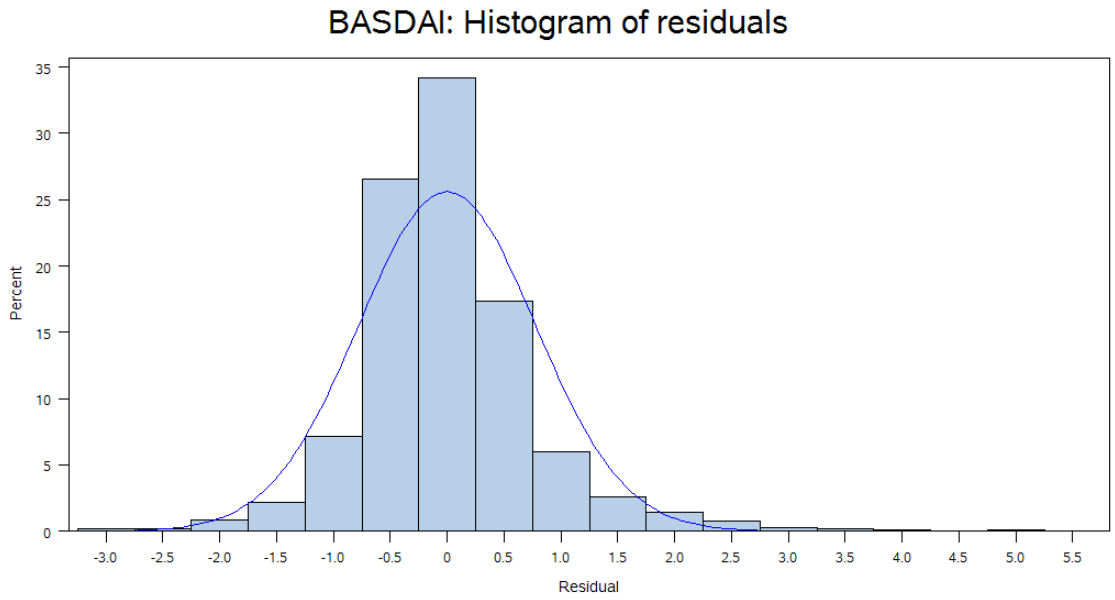


Figure 9. Normal distribution of residuals from mixed effect model analysis

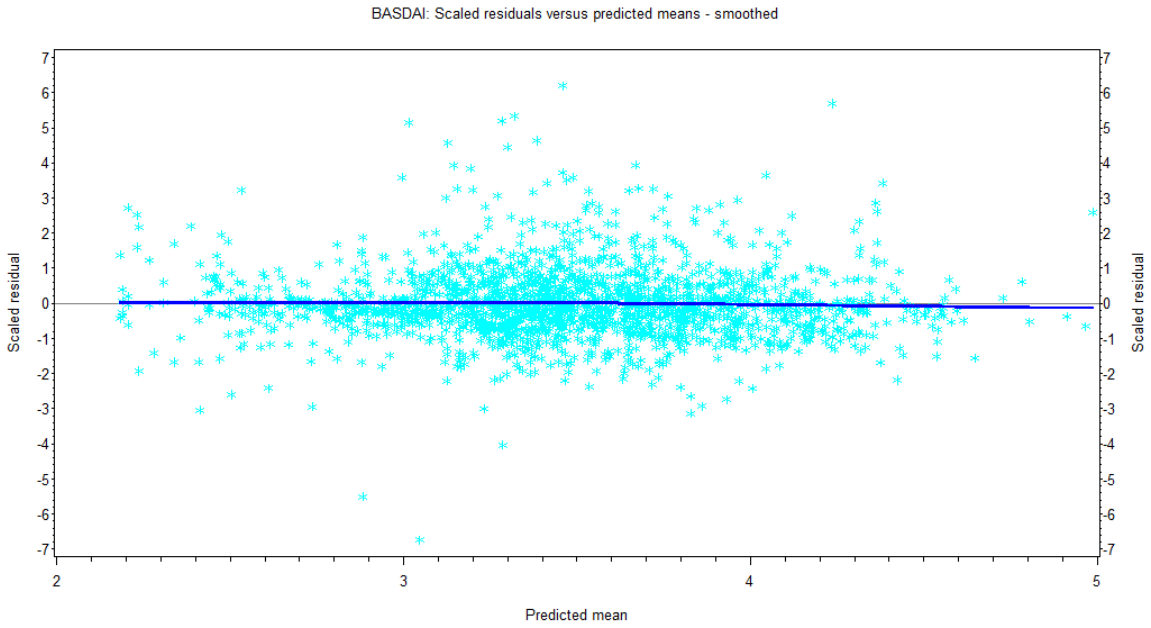


Figure 10. Plot of scaled residuals versus predicted value for BASDAI with lowess smoothed trend. Shows the presence of some outliers in the data, but no discernible systematic trend.

CHAPTER FOUR: A COCHRANE SYSTEMATIC REVIEW ON TNF-INHIBITORS (ADALIMUMAB, INFLIXIMAB, ETANERCEPT) FOR THE TREATMENT OF ANKYLOSING SPONDYLITIS

Chapter overview

The following is a manuscript prepared for publication, based on a systematic review of the literature on the efficacy and safety of TNF-inhibitors for treating AS. The previous manuscript provided evidence of the natural fluctuations in disease activity in a historical control group and how this relates to the current clinical guidelines on the treatment of AS. For clinicians and their patients who need to make a treatment decision on whether to initiate anti-TNF therapy, a systematically reviewed evidence base of the efficacy and harms of this therapy is useful in decision-making by quantifying the trade-off between benefit and harm.

This manuscript was co-authored by the student (LJM) and her co-supervisors, Dr. Peter Tugwell and Dr. George Wells, as well as advisor to the thesis, Dr. Annelies Boonen. Drs Singh and Zochling screened search results, performed data extraction, and provided comments on drafts of the review. The student is the first author on this manuscript as she had the primary responsibility for data collection, analysis, interpretation, and writing. Drs Singh, Zochling, Tugwell, Wells, and Boonen provided valuable feedback throughout the process.

TNF-alpha inhibitors for ankylosing spondylitis

Review information

Authors

Lara Maxwell, Jane Zochling, Annelies Boonen, Jasvinder Singh, Peter Tugwell, George A Wells

Citation example: Maxwell L, Zochling J, Boonen A, Singh J, Tugwell P, Wells GA. TNF-alpha inhibitors for ankylosing spondylitis. Cochrane Database of Systematic Reviews 2005, Issue 3. Art. No.: CD005468. DOI: 10.1002/14651858.CD005468.

Dates

Assessed as Up-to-date:	01 October 2010
Date of Search:	20 March 2010
Next Stage Expected:	01 July 2012
Protocol First Published:	Issue 3, 2005
Review First Published:	Not specified
Last Citation Issue:	Issue 3, 2005

Abstract
Background

Anti-TNF (tumour necrosis factor) agents block a key protein in the inflammatory chain reaction responsible for the joint inflammation, pain, and damage in ankylosing spondylitis.

Objectives

To assess the efficacy and safety of adalimumab, etanercept, and infliximab in reducing disease activity, pain, and improving function in people with ankylosing spondylitis (AS).

Search methods

We searched the following databases to January 26, 2009: MEDLINE (from 1966); EMBASE (from 1980); The Cochrane Library (Issue 4, 2008); ACP Journal Club; CINAHL (from 1982); ISI Web of Knowledge (from 1900). A separate electronic search specific to adverse events was conducted to May 2009 and we searched the websites of regulatory agencies to April 2010. An updated search to March 2010 was run in Ovid MEDLINE(R) and EMBASE. We searched conference proceedings from the American College of Rheumatology (ACR) and European Congress of Rheumatology (EULAR) meetings.

Selection criteria

Randomized controlled trials comparing adalimumab, etanercept, and infliximab to placebo, other drugs or usual care in patients with AS. For adverse effects, we included controlled clinical trials (CCTs), cohort studies (prospective (e.g. long-term extension of RCT) or retrospective), case-control studies, case-series, and published registry data as long as they reported at least one year of follow-up for AS patients on anti-TNF agents, had a sample size > 100 patients and reported an included outcome.

Data collection and analysis

Two authors independently assessed search results and risk of bias, and extracted data. We contacted authors of included studies for additional information. Dichotomous outcomes were reported as relative risk (RR) or Peto odds ratio (Peto OR) if a rare event. Continuous outcomes were reported as mean differences. Indirect comparisons were undertaken using Bucher's Indirect Treatment Comparison (ITC), Generalized Linear Mixed Models (GLIMMIX), and Mixed Treatment Comparisons (MTC) approaches.

Results

Fifteen RCTs met the inclusion criteria with a total of 2,459 participants. Compared with placebo, patients on an anti-TNF agent were 3 to 4.5 times more likely to achieve an ASAS 40 response by 6 months (adalimumab: RR 3.05, 95% confidence

interval (CI) 1.82 to 5.11; etanercept: RR 3.07, 95% CI 2.11 to 4.47; infliximab: RR 4.33, 95% CI 2.50 to 7.52) with a 26% to 39% absolute risk difference between treatment and placebo groups. The number needed to treat to achieve an ASAS 40 response on one of these agents ranged from 3 to 4. Significant improvements in physical function and the number of patients meeting the ASAS partial remission criteria were found in anti-TNF-treated patients compared to placebo. No data for radiographic progression was available. One RCT found adalimumab reduced spinal and sacroiliac MRI activity scores compared to placebo, although it is not clear what the clinical relevance of this difference may be. Withdrawals due to adverse events were greater in the etanercept group (Peto OR 3.93, 95% CI 1.40 to 11.02) but not in adalimumab or infliximab groups. Serious adverse events were not statistically different between groups for either drug or for all three drugs pooled together, but total infections were (RR 1.22, 95% CI 1.02, 1.45). Four extension studies were included for assessment of adverse events with data for up to three years after the RCT period and were judged to be at a high risk of bias. Most results were presented compared to the RCT period and large increases in rates of serious infections, malignancies, demyelinating diseases or death were not apparent. Using indirect comparison methodology and one head-to-head study of etanercept versus infliximab, there was no evidence for significant differences in important outcomes between anti-TNF agents.

Authors' conclusions

There is mostly high quality evidence demonstrating that the anti-TNF agents are efficacious against placebo in the treatment of AS. We did not find evidence of an increase in serious adverse events or serious infections from RCTs, though total infections were increased, and other AEs had low event rates but the short-term toxicity profile appears acceptable. Data included for assessment of adverse events was at high risk of bias and difficult to interpret, but there was no large signal of key safety outcomes such as serious infections and malignancies. Based on indirect comparison methodology, there does not appear to differences between anti-TNF agents in terms of the key efficacy or safety outcomes.

Plain language summary

Adalimumab, etanercept, and infliximab (anti-tumor necrosis factor (TNF) drugs) for treating ankylosing spondylitis

This summary of a Cochrane review presents what we know from research about the effect of anti-TNF drugs (adalimumab, etanercept, and infliximab) on ankylosing spondylitis:

The review shows that in people with ankylosing spondylitis:

- Anti-TNF drugs probably improve pain, function and other symptoms of ankylosing spondylitis.
- Anti-TNF drugs probably reduces disease activity
- Anti-TNF drugs probably increase the chance of acheiving partial remission of symptoms of ankylosing spondylitis.

We do not have precise information about side effects and complications, but in these short-term studies there was no evidence of any increase in serious adverse events. Possible side effects may include a serious infection or upper respiratory infection. Rare complications may include certain types of cancer.

What is ankylosing spondylitis and what are anti-TNF drugs?

Ankylosing spondylitis (AS) is a type of arthritis usually in the joints and ligaments of the spine. It may also affect shoulders, hips, or other joints and cause tendonitis. Pain and stiffness occurs and limits movement in the back and affected joints. It can come and go, last for long periods, and be quite severe.

Anti-TNF drugs target a protein called tumor necrosis factor that causes inflammation. These drugs suppress the immune system and reduce the inflammation in the joints. Even though suppressing the immune system can make it slightly harder to fight off infections, it also helps to stabilize an overactive immune system. By reducing the inflammation, the aim is to help prevent damage to the joints.

Best estimate of what happens to people with ankylosing spondylitis who take anti-TNF drugs:

ASAS 40 (40% improvement in pain, function, inflammation as measured by morning stiffness, and patient overall well-being)

- Among people who took adalimumab 40 people out of 100 experienced improvement in the signs of their ankylosing spondylitis compared to 13 people out of 100 who took a placebo (27% absolute improvement).
- Among people who took etanercept 57 people out of 100 experienced improvement in the signs of their ankylosing spondylitis compared to 19 people out of 100 who took a placebo (38% absolute improvement).
- Among people who took infliximab 45 people out of 100 experienced improvement in the signs of their ankylosing spondylitis compared to 10 people out of 100 who took a placebo (35% improvement).

Partial remission

- Among people who took adalimumab 20 people out of 100 experienced partial remission in the signs of their ankylosing spondylitis compared to 4 people out of 100 who took a placebo (16% absolute improvement).
- Among people who took etanercept 20 people out of 100 experienced partial remission in the signs of their ankylosing spondylitis compared to 5 people out of 100 who took a placebo (15% absolute improvement).
- Among people who took infliximab 23 people out of 100 experienced partial remission in the signs of their ankylosing spondylitis compared to 1 people out of 100 who took a placebo (22% improvement).

Physical function (lower score means better function)

- People who took adalimumab rated their function to be 18 points lower on a scale of 0 to 100 after 12 weeks (18% absolute improvement).

-People who took adalimumab rated their pain to be 34 on a scale of 0 to 100 after 12 weeks.

-People who took a placebo rated their pain to be 52 on a scale of 0 to 100.

- People who took etanercept rated their function to be 16 points lower on a scale of 0 to 100 after 24 weeks (16% absolute improvement).
-People who took etanercept rated their pain to be 39 on a scale of 0 to 100 after 24 weeks.
-People who took a placebo rated their pain to be 55 on a scale of 0 to 100.

- People who took infliximab rated their function to be 19 points lower on a scale of 0 to 100 after 24 weeks (19% absolute improvement).
-People who took infliximab rated their pain to be 41 on a scale of 0 to 100 after 24 weeks.
-People who took a placebo rated their pain to be 60 on a scale of 0 to 100.

X-rays of the joints

-There was no x-ray data in the studies we looked at.

Side effects

- Among people who took etanercept 16 people out of 1000 dropped out of the study because of the side effects compared to 4 people out of 1000 who took a placebo (3% absolute difference).

There may be little or no difference in people who dropped out because of side effects with adalimumab or infliximab compared to people who took a placebo (fake pill)

Background

Ankylosing spondylitis (AS) is a chronic, inflammatory rheumatic disease characterized by inflammatory back pain due to sacroiliitis and spondylitis, enthesitis, and the formation of syndesmophytes leading to ankylosis. Extraspinal manifestations are common, including peripheral arthritis (25-50%), uveitis (25-40%), and inflammatory bowel disease (26%) and contribute to disease morbidity (Edmunds 1991). The etiology of the disease is not yet fully understood but there is a

strong association with the HLA-B27 gene. Studies have shown the prevalence of AS in the adult general population to vary from 0.4% (Alaska Eskimos) to 1.4% (Northern Norway) (Khan 2002). The peak age of onset is between 20 and 30 years, although there is often a 5-6 year delay in diagnosis (Khan 2002). It is two to three times more common in men than women. The course of AS is highly variable, and at least one third of patients with AS will progress to severe disabling disease (Zink 2000). Clinical symptoms usually begin with back pain and stiffness in adolescence and early adulthood and can lead to impaired spinal mobility and/or chest expansion. The burden of disease in AS has been found to be similar to that of rheumatoid arthritis in terms of pain, disability and decreased well being (Zink 2000). Additionally, compared to the general population, those with AS experience higher work disability and absence from work which can lead to substantial direct and indirect socio-economic costs (Boonen 2001; Boonen 2001a).

The goals of treatment of AS are to relieve symptoms (pain, stiffness, joint swelling), improve physical function, and delay or avoid structural damage which leads to physical impairments and deformities. AS requires a multidisciplinary treatment approach and is usually managed with a combination of drug therapy and physiotherapy. Regular exercise is crucial for maintaining or improving spinal mobility and physical function (Dagfinrud 2004). Non-steroidal anti-inflammatory drugs (NSAIDs) are the mainstay of symptomatic drug therapy, reducing the pain and stiffness of inflammation. Although these interventions can alleviate the symptoms of the disease, it is not clear whether they are able to prevent or delay the structural damage leading to physical disability. Recent work suggests continuous NSAID therapy may have an effect on the spinal radiographic changes seen in AS (Wanders 2005). At least one third of patients respond insufficiently to NSAID therapy or experience serious side effects from NSAIDs and thus require disease controlling drugs in addition to symptom modifying treatment. In contrast to rheumatoid arthritis, there are no established disease modifying anti-rheumatic treatments (DMARDs) in AS, although sulphasalazine may be effective for peripheral joint symptoms but not for axial disease (Dougados 2002).

As the result of research demonstrating that tumour necrosis factor alpha (TNF-alpha) is present in inflamed sacroiliac joints (Braun 1995), biologic drugs were developed to block TNF-alpha (known as TNF-alpha inhibitors, anti-TNF agents, or TNF-alpha blockers). TNF-alpha is a protein that the body produces during the inflammatory response. TNF-alpha promotes inflammation and subsequent pain, tenderness, swelling and fever in several inflammatory conditions including ankylosing spondylitis. There are three main biologic agents currently available that target TNF-alpha: infliximab (Remicade) is a chimeric (mouse/human) monoclonal antibody of the IgG1κ isotype that binds with a high affinity to TNF-alpha; etanercept (Enbrel) is a receptor fusion protein that binds to TNF-alpha, thus competitively inhibiting the binding of TNF-alpha to the cell surface; and adalimumab (Humira) is a recombinant human IgG1 monoclonal antibody specific for human TNF-alpha. These three drugs prevent TNF-alpha from promoting inflammation and therefore reduce pain, tenderness and swelling of joints in patients with ankylosing spondylitis. Infliximab is given as an intravenous infusion over one to two hours, etanercept and adalimumab are given as sub-cutaneous injections.

Open label studies have shown biologics are efficacious in AS (Brandt 2000; Maksymowych 2002; Marzo-Ortega 2001; Stone 2001; Haibel 2004) and randomized controlled trials have shown them to be highly effective in improving disease activity, spinal mobility, function, and pain (Braun 2002; Van Den Bosch 2002; Gorman 2002). It has been suggested that biologic treatment might prevent or significantly decrease the rate of progression of structural damage and thus be the first disease controlling anti-rheumatic treatment (DC-ART) for AS; however, more evidence is required to determine if biologics really do have this effect. Recognized adverse effects of anti-TNF-alpha therapy include serious infections such as tuberculosis, allergic reactions and autoimmune reactions. It is still unclear whether there is an increased incidence of malignancy (lymphoma or solid tumors) associated with anti-TNF-alpha therapy.

The relatively high cost of treatment and possible serious side effects of anti-TNF-alpha therapy led the ASsessment in Ankylosing Spondylitis (ASAS) International

Working Group (Braun 2003) and the Spondyloarthritis Research Consortium of Canada (Maksymowych 2003) to develop recommendations for the use of TNF-alpha inhibitors.

At this time it is appropriate to conduct a systematic review of randomized controlled trials of TNF-alpha inhibitors to try to better quantify the benefits and potential harms of their use.

Objectives

To summarize the efficacy and safety of infliximab, etanercept and adalimumab (TNF-alpha inhibitors) for the short-term and long-term treatment of ankylosing spondylitis.

Methods

Criteria for considering studies for this review

Types of studies

To assess efficacy and adverse effects, we included randomized controlled trials (RCTs).

To further assess adverse effects, we included the following types of studies as long as they reported at least one year of follow-up for patients on anti-TNF agents, had a sample size > 100 patients (EMEA 2005; Gartlehner 2006), and reported an included outcome: controlled clinical trials (CCTs), cohort studies (prospective (e.g. long-term extension of RCT) or retrospective), case-control studies, case-series, and published registry data.

We defined "short-term" efficacy and adverse effects less than six months duration and "long-term" efficacy and adverse effects as longer than six months.

Types of participants

Studies of patients meeting the following AS classification criteria were included: 1961 Rome, 1966 New York, or modified 1984 New York. No additional restrictions in studies with regard to age of patients, past or present (co-)medication or AS-related comorbidity were applied. Studies on spondylarthropathies that mention AS patients as a subgroup were included as far as the subgroup was properly randomized and outcome measures were available specifically for the AS subgroup. Patients on other medications and with or without AS-related co-morbidity (e.g. peripheral joint impairment, IBD, psoriasis) were included. There was no age restriction.

Types of interventions

- Adalimumab versus placebo, other medications, or usual care
- Etanercept versus placebo, other medications, or usual care
- Infliximab versus placebo, other medications, or usual care

There were no restrictions with regard to dosage or concomitant treatments in the placebo group (for example, physical exercises and/or NSAIDs).

Types of outcome measures

Primary and secondary outcome measures were based on the disease-controlling anti-rheumatic treatment (DC-ART) Ankylosing Spondylitis Working Group Core Set (van der Heijde 1997) and the International ASAS consensus statement for the use of TNF-alpha inhibitors in patients with AS (Braun 2003). Secondary outcomes were the proportion of responders based on either the ASAS measures (ASAS 20 (Anderson 2001), ASAS 40 and/or ASAS 5 of 6 (Brandt 2004)), or on any alternative response criteria formulated by the authors. Additional outcomes including quality of life measures, other imaging outcomes and reduction of other medications were also recorded. Specific adverse effects of interest were collected (as listed below), as well as the total number of adverse events, total number of serious adverse events, total number of withdrawals, number of withdrawals due to adverse events and withdrawals due to inefficacy.

Primary outcomes:

- Disease activity (BASDAI)
- Physical function (BASFI)
- Spinal Pain
- Spinal stiffness (duration of morning stiffness)
- Spinal mobility
- Patient global assessment
- Peripheral joint/enthesis inflammation
- Changes in spine radiographs
- Fatigue
- Acute phase reactants (erthyrocyte sedimentation rate (ESR), C-reactive protein (CRP))

Adverse effects:

- Total withdrawals
- Withdrawals due to adverse effects
- Total adverse events
- Total serious adverse events (as defined by the authors)
- Specific adverse events: headache, diarrhea, nausea, infections
- Specific serious adverse events: serious infections, tuberculosis, lymphoma, hepatotoxicity, chronic or congestive heart failure, injection reactions, autoimmunity symptoms and disorders, demyelinating disease (Lee 2005)
- Mortality

Secondary outcomes:

- ASAS 20 response
- ASAS 40 response
- ASAS 5 out of 6 response
- Changes in hip radiograph
- Physician global assessement
- Quality of life
- MRI evidence of suppression of inflammation
- Reduction of steroid or NSAID use
- Withdrawals due to inefficacy

The criteria for an ASAS 20 response are: at least a 20% improvement with a minimum of 10 units (0 to 100 scale) in at least three of four domains (pain, function, inflammation as measured by morning stiffness, and patient global assessment), with no worsening in the fourth domain by 20% and 10 units. The ASAS 40 response represents improvement of at least 40% and absolute improvement of at least 20 units compared with baseline in at least 3 of the 4 domains of the ASAS 20 criteria, with no deterioration in the remaining domain. Partial remission is defined as a value of less than 2 on a 0-10 scale in each of the 4 domains of the ASAS 20.

ASAS 5/6 response criteria require at least 20% improvement in 5 of 6 domains: spinal mobility (according to the Bath Ankylosing Spondylitis Metrology Index [BASMI];other instruments may be used) and acute-phase reactants (the CRP concentration) in addition to the 4 domains included in the ASAS 20 response criteria. The BASMI is a composite index with a range of 0 to 10, based on a 0 to 2-point scale for each of 5 clinical measurements: tragus-to-wall distance (in cm), anterior lumbar flexion (modified Schober's test; in cm), lumbar side flexion (in cm), intermalleolar distance (in cm), and cervical rotation (in degrees).

The outcomes listed above were chosen in 2005 when the protocol for this review was published. Since then, further work on outcome measures has been conducted by the ASAS Working Group. In discussion with experts from the ASAS Working Group, the following outcomes were chosen to be included in the Summary of Findings (SoF) table: ASAS partial remission, BASFI, ASAS 40, MRI, serious adverse events, withdrawals due to adverse events, and radiographic progression. For etanercept, we combined the results from trials using administration schedules of 25 mg twice a week and 50 mg once a week.

Search methods for identification of studies

The Cochrane Musculoskeletal Group trial search coordinators assisted with the development of the search strategies. The following electronic databases were searched: MEDLINE (1966 to January 26, 2009); EMBASE (1980 to January 26,

2009); The Cochrane Library ((Issue 4, 2008) including the following databases: Cochrane Central Register of Controlled Trials (CENTRAL), Cochrane Database of Systematic Reviews (CDSR), Database of Abstracts of Reviews of Effects (DARE), Cochrane Library Health Technology Assessment Database (CLHTA), and NHS Economic Evaluation Database (NHS EED); ACP Journal Club (January 26, 2009); CINAHL (1982 to January 26, 2009); ISI Web of Knowledge (1900 to January 2009). An updated search from 2008 to March week 3, 2010 was run in Ovid MEDLINE(R) and EMBASE.

The MEDLINE search strategy is listed below. The other search strategies are available in Appendix 1.

Database: Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R) <1950 to Present>

1 exp spondylitis, ankylosing/
2 exp spondylarthropathies/
3 (ankylosing or spondyl$).tw.
4 (bekhterev or bechterew).tw.
5 4 or 1 or 3 or 2
6 exp Receptors, Tumor Necrosis Factor/
7 exp tumor necrosis factor/
8 exp Antibodies, Monoclonal/
9 anti-tumo?r necrosis factor$.sh,rn,tw.
10 antitumo?r necrosis factor$.sh,rn,tw.
11 anti-tnf.sh,rn,tw.
12 antitnf.sh,rn,tw.
13 etanercept.sh,rn,tw.
14 enbrel.sh,rn,tw.
15 infliximab.sh,rn,tw.
16 remicade.sh,rn,tw.
17 adalimumab.sh,rn,tw.
18 humira.sh,rn,tw.
19 or/6-18
20 5 and 19
21 randomized controlled trial.pt.
22 controlled clinical trial.pt.
23 randomized.ab.
24 placebo.ab.
25 drug therapy.fs.
26 randomly.ab.
27 trial.ab.

28 groups.ab.
29 27 or 25 or 28 or 21 or 26 or 22 or 24 or 23
30 (animals not (humans and animals)).sh.
31 29 not 30
32 20 and 31
33 from 32 keep 1-10

The reference sections of retrieved articles were reviewed. Authors of relevant papers and experts in the field were contacted regarding any further published or unpublished work. Conference proceedings from the American College of Rheumatology (ACR) and European Congress of Rheumatology meetings (EULAR, http://www.abstracts2view.com/eular/ textword search: ankylosing AND etanercept OR infliximab OR adalimumab) were hand searched (by a single author, LM) from 2005 to 2009 for both efficacy and harms. There was no language restriction.

An updated search was run from 2008 to March 31, 2010 (March week 3) of MEDLINE, Pre-MEDLINE and EMBASE (see search strategy in Appendix 2). References of review articles picked up in this search were screened to ensure no studies were missed. In addition, we searched specifically for full-text articles for any abstracts that were included from the earlier search.

A separate search was conducted for adverse events to May 2009. The search strategies are reported in Appendix 3.

For safety assessments, we searched the websites of the regulatory agencies (US Food and Drug Administration-MedWatch (http://www.fda.gov/Safety/MedWatch/default.htm) , European Medicines Evaluation Agency (http://www.emea.europa.eu), Australian Adverse Drug Reactions Bulletin (http://www.tga.gov.au/adr/aadrb.htm), and UK Medicines and Healthcare products Regulatory Agency (MHRA) pharmacovigilance and drug safety updates (http://www.mhra.gov.uk); (note 'Current Problems in Pharmacovigilance' was superseded by 'Drug Safety Update' in July 2007) using the terms “ankylosing spondylitis,” “adalimumab,” "humira", "etanercept", "enbrel", "infliximab", and "remicade” on April 1, 2010.

Published reports from known national biologic registers (British Society for Rheumatology Biologics Register, Danish Biologic Register (DANBIO), the Register of Biological Treatment in Finland (ROB-FIN), the Spanish Society of Rheumatology's BIOBADASER, and Australian Rheumatology Association Database for biologics (ARAD) were also considered for inclusion.

Data collection and analysis

Study selection

Results of the search strategy were independently reviewed by two people (JZ, JS, LM, BD). Abstracts were reviewed and if more information was required to determine whether the trial meets the inclusion criteria, the full text was obtained. Disagreement was resolved by a third author (AB).

Data extraction

Data was independently extracted from the included trials by two reviewers (JZ, LM) and entered into RevMan 5.0. Data extraction forms were pilot-tested on a selection of trials. The authors of the primary studies were contacted to obtain additional data.

The following data was extracted:

- General study information such as title, authors, contact address, publication source, publication year, country, study sponsor
- Characteristics of the study: design, study setting, inclusion/exclusion criteria, risk of bias criteria (e.g. randomization method, allocation procedure, blinding of patients, caregivers and outcome assessors)
- Characteristics of the study population and baseline characteristics of the intervention and control groups (age, sex, type of classification criteria, duration of disease, presence of co-morbidity and peripheral disease, concomitant treatments, BASDAI, BASFI, patient global assessment) and numbers in each group

- Characteristics of the intervention, such as treatment comparators, dose, method of administration, frequency of administration and duration of treatment
- Outcomes measures as noted above
- Results for the intention to treat population (where possible), outcome measures at the end of the placebo phase, and any summary measures with standard deviations, confidence intervals and p-values where given

Assessment of risk of bias: RCTs

The risk of bias of the included RCTs was assessed by two independent reviewers (LM, JZ). As recommended by the Cochrane Handbook (Higgins 2008), the following methodological domains were assessed:

1. Sequence generation - was the method used to generate the allocation sequence appropriate to produce comparable groups?
2. Allocation sequence concealment - was the method used to conceal the allocation sequence appropriate to prevent the allocation being known in advance of, or during, enrolment?
3. Blinding of participants, personnel and outcome assessors - were measures used to blind study participants, personnel, and outcome assessors from knowledge of which intervention a participant received?
4. Incomplete outcome data - how complete was the outcome data for the primary outcomes? Were dropout rates and reasons for withdrawal reported? Was missing data imputed appropriately? We considered an overall completion rate of 80% or higher as a low risk of bias. If completion rates were only provided by group, a less than 80% completion rate in the treatment group was considered a high risk of bias.
5. Selective outcome reporting - were appropriate outcomes reported and were any key outcomes missing?

The following two criteria are specific to assessment of adverse effects:

6. Ascertainment of outcome - did the researchers actively monitor for AEs (low risk of bias) or did they simply provide spontaneous reporting of AEs that arise (high risk of bias)?
7. Definition of adverse outcomes- were definitions provided for general 'adverse event' or 'serious adverse event'?

Each of these criteria was explicitly judged using: Yes=low risk of bias; No=high risk of bias; Unclear= either lack of information or uncertainty over the potential for bias.

Assessment of risk of bias: non-RCTs

For the non-RCTs included in this review, the risk of bias was assessed using the following criteria from the Newcastle-Ottawa Scale (NOS) (Wells) for cohort studies. Each study is rated as follows: a study can be awarded a maximum of one star for each numbered item within the Selection and Outcome categories. A maximum of two stars can be given for Comparability:

Selection

1. Representativeness of the exposed cohort
2. Selection of the non exposed cohort
3. Ascertainment of exposure
4. Demonstration that outcome of interest was not present at start of study

Comparability

1. Comparability of cohorts on the basis of the design or analysis

Outcome

1. Assessment of outcome
2. Was follow-up long enough for outcomes to occur
3. Adequacy of follow-up of cohorts

Since the included non-RCT studies were all extension studies from RCTs, there was no control group. We decided that the NOS criteria assessed the important potential risk of bias criteria concerning these studies with the exception of two criteria which were not applicable to a single-group cohort; namely 'selection of the non-exposed cohort' and 'comparability of cohorts on the basis of the design or analysis'.

Given the lack of a control group, one could consider these extension studies as having the design features of a case-series study. We searched for risk of bias tools to assess case-series and found the following guidance on criteria for assessing case-series in the CRD guidance for systematic reviewers (Khan 2001):

Case series

1. Is the study based on a representative sample selected from a relevant population?
2. Are the criteria for inclusion explicit?
3. Did all individuals enter the survey at a similar point in their disease progression?
4. Was follow-up long enough for important events to occur?
5. Were outcomes assessed using objective criteria or was blinding used?
6. If comparisons of sub-series are being made, was there sufficient description of the series and the distribution of prognostic factors?

Since these concepts are covered in the NOS, which is a validated scale, we decided to use the NOS to assess the extension studies included in this review.

Data analysis

Results from the three biologics was analyzed separately. One analysis pooled adverse effects data from all three anti-TNF agents to assess for a class-effect of TNF-blockers. The results of the studies were analyzed using RevMan 5.0. Data was summarised using meta-analysis when it was sufficiently homogeneous, both clinically and statistically. Continuous data is expressed as weighted mean difference (WMD) or standarized mean difference (SMD) (depending on similarity of scales). Dichotomous data is expressed as relative risk (RR), unless the data consisted of rare events (<10%) and then the Peto OR was used. The Peto OR is not recommended when there is a severe imbalance in the number of people in the treatment groups

(Bradburn 2007), and we used the rule of thumb as suggested in *Systematic Reviews in Health Care* (p.295, Egger 2001), that this is a concern when there is four times or more as many participants in one group as another. We made a post-hoc decision to check the robustness of estimates obtained by Peto OR by also using the Mantel-Haenzsal OR with the standard continuity correction of 0.5.

Heterogeneity of the data was tested for using the I-squared statistic (Higgins 2003). A value greater than 50% may be considered substantial heterogeneity. In the absence of significant heterogeneity, a fixed effects model was used, otherwise a random effects model was used for analysis. Where available, the analyses was based on intention-to-treat data from the individual studies. We planned to undertake visual inspection of funnel plots to assess publication bias; however, the few studies combined for individual outcomes made this rather difficult.

Summary of Findings tables were compiled to improve the readability of the review. For dichotomous outcomes, such as complications, the number needed to treat was calculated from the control group event rate and the relative risk using the Visual Rx NNT calculator (Cates 2003). For continuous outcomes, the absolute benefit was calculated as the improvement in the intervention group minus the improvement in the control group, in the original units. Relative difference in the change from baseline was calculated as the absolute benefit divided by the baseline mean of the control group. The NNT for continuous measures was also calculated using the Wells calculator (available at the CMSG Editorial office). The outcomes included in the SoF table are: ASAS partial remission, BASFI, ASAS 40, MRI, serious adverse events, withdrawals due to adverse events, and radiographic progression. These were decided upon in consultation with five experts in AS. For etanercept, we combined the results from trials using administration schedules of 25 mg twice a week and 50 mg once a week.

The following sub-group analyses were planned *a priori* in order to explore possible effect size differences:

1. Intervention - different dosage, trial duration
2. Characteristics of participants - different AS classification criteria; severity of

baseline disease (based on BASDAI, BASFI); age; disease duration; sex; with or without peripheral joint involvement.

Sensitivity analysis were pre-specified to assess the effect of study quality (proper generation of randomization sequence, adequate allocation concealment and blinding) on the overall estimates of effect.

We made a post-hoc decision to present the results both separately and combined for etanercept trials which used either 25 mg adminstered twice a week or 50 mg administered once a week. The results from the van der Heijde 2006b study showed these doses to be equivalent in both efficacy and safety.

Transformations

The following transformations were performed for the purpose of entering data into RevMan: when the median change from baseline and interquartile range (IQR) change from baseline were reported (as in van der Heijde 2005), the median change was assumed to be the mean change and the standard deviation was calculated as the IQR at baseline divided by 2 and this SD assumed for the end of study score.

Where the variance for the change from baseline was provided in the in the study report as the standard error of the change, the end of study score was calculated and the baseline SD was assumed for the end of study SD (van der Heijde 2006a).

In Davis 2003, the standard error of the mean (SEM) was transformed to SD by the calculation SD=SEM *sqrt(N). In Gorman 2002, the median was assumed for the mean for continuous efficacy outcomes.

Indirect comparisons

A recent report for the Canadian Agency for Drugs and Technology in Health (CADTH) (Wells 2009) provides an overview of various methods used to provide indirect evidence of treatment effects in the comparison of two different treatments when direct evidence is not available.

Three different methods of indirect comparisons were used to obtain summary estimates for comparisons of etanercept versus infliximab, adalimumab versus etanercept; and adalimumab versus infliximab, and to compare and contrast the results from these different methods:

i) the Bucher Adjusted Indirect Treatment Comparison (ITC) approach (Bucher 1997), using the ITC software designed for the Wells 2009 report;

ii) the Generalized Linear Mixed Models (GLIMMIX) approach;

iii) the Mixed Treatment Comparisons (MTC) approach (also referred to as "network meta-analysis")

The GLIMMIX and MTC methods also provide refined placebo-estimates.

All three methods utilize techniques to preserve the randomization inherent in the RCTs. That is, they avoid the “naïve” method of pooling the results across trials from the different treatment arms of interest and then comparing the results of treatment A versus treatment B versus treatment C. This naïve method ignores the randomization that was present in the original RCTs and introduces biases expected in an observational cohort (i.e. potential confounders are no longer likely to be randomly distributed between the treatment groups) (Bucher 1997;Wells 2009).

In the Bucher model, the magnitude of the effect measures that result from the RCTs comparing treatment A to placebo and treatment B to placebo are compared to provide an adjusted indirect comparison of treatment A to treatment B.

$OR_{(AB)}/OR_{(CB)}=OR_{(AC\ INDIRECT)}$

Using a Bayesian framework, the MTC method provides a refined estimate of the treatment effect by combining the information from the direct and indirect data to strengthen the precision of the estimate of effect.

An extension to the generalized linear model is the "generalized linear mixed model (GLMMIX)". In this model, the linear predictor contains random effects in addition

to the usual fixed effects and these random effects are usually assumed to have a normal distribution.

Assumptions assessed were: i. homogeneity; ii. trial similiarity; iii. consistency of evidence

We used the back-calculation method (Dias 2010) to assess the consistency between direct and indirect estimates. The indirect estimate is compared to the direct estimate and this measure of discrepancy can then be applied to a standard normal distribution test statistic to answer the null hypothesis that there is no difference between the direct and indirect estimates.

Results

Description of studies

Results of the search

The search of the electronic databases listed in the methods section for RCTs resulted in 2445 records. After de-duplication, there were 1644 records left to screen. 60 records were assessed more in-depth to see if they met the inclusion criteria. Two additional articles were included from handsearching EULAR abstracts on their website.

There were some abstracts from conference proceedings that were later published as full-text articles. In this case, only the full-text article was included; however, if the abstract provided additional important information that was not provided in the full-text article, then the data from the abstract was also included. Some trials had more than one publication with the secondary publications reporting on other outcomes such as health-related quality of life, patient-reported outcomes, or MRI data.

After assessing all the records, we included 15 trials, with 24 published articles (either abstracts or full-text articles) related to those trials. The additional articles relating to a trial are listed as secondary references in the reference section. Data from seven abstracts from conference proceedings were included in this review.

A flow chart of the RCT search results is provided in Figure 9.

A separate electronic search was conducted for non-RCTs to assess adverse events. There were 2053 records identified and 550 duplicates removed which left 1503 records for screening. Eighty-six records were assessed in more detail to determine if they met the review's inclusion criteria. Four full-text articles met the review's inclusion criteria. Two additional abstracts were obtained through handsearching the ACR and EULAR conference proceedings. All were single-cohort extension studies from RCTs included in this review.

A flow chart of the non-RCT search results is provided in Figure 10.

Included studies

Further details on each included study are available in the Characteristics of included studies table.

Included studies

Fifteen RCTs met the inclusion criteria with a total of 2,459 participants. 303 people received inflixmab (Braun 2002; Inman 2010; van der Heijde 2005) (28 in combination with methotrexate (Marzo-Ortega 2005); 995 received etanercept (Barkham 2008; Brandt 2003;Braun 2008; Calin 2004;Davis 2003; Gorman 2002; Huang 2008; van der Heijde 2006b) and 246 received adalimumab (Lambert 2007;van der Heijde 2006a). One study was an open-label head to head study of etanercept (N=25) and infliximab (N=25) (Giardina 2009).

Additional data

Additional data was received from the trial authors for the following studies: Brandt 2003;Braun 2002 Calin 2004; Davis 2003 (though we were unable to use data from Davis 2003 since variance was not provided in the additional information received). This was mainly to obtain data on clinical endpoints where the published results for continous outcomes had been reported as a statistic different from the mean and SD which is required for entry into RevMan. Additional details were also sought to clarify risk of bias items from some authors (Davis 2003; van der Heijde 2006a).

Non-RCT data for assessment of adverse effects

Open-label extension studies occurred at the end of the RCTs noted below, in which all patients who had entered the study (i.e. both the treatment and placebo groups) were eligible to receive the study drug during a follow-up study.

Open-label extension-etanercept

There were two published follow up studies of the Davis 2003 trial. In 2005, Davis published an article on 257 people who were followed up for 72 weeks after the 24 week RCT (total 96 weeks) and in 2008 a full-text article was published on the 168 week results of this open-label extension (total 192 weeks).

Extension and open-label extension- infliximab

Two-year results from the follow-up of the van der Heijde 2005 RCT were published in a full-text article (Braun 2008). Patients remained blinded during the extension phase of this trial. Two abstracts were found on the follow up of the EASIC cohort, at one and two years, which was an extension of the ASSERT trial (Heldmann 2008 and Heldmann 2009). However, the 2008 abstract did not report separately on those patients who received continuous infliximab treatment in the 1.3 years between the end of ASSERT and start of EASIC. Therefore, we did not include data from this abstract.

Open-label extension- adalimumab

Two- and three-year open-label extension data from the van der Heijde 2006a trial were reported in a full-text article (van der Heijde 2009) and an abstract (van der Heijde 2008).

Participants

The majority of participants were Caucasian males in their early 40s. The percent of male participants in the treatment groups ranged between 65% to 80%, and 74% to 100% in the control groups. The mean age ranged from 38 to 45 years in the treatment groups and 39 to 47 years in the control groups. Between 75% and 98% of the participants in the treatment groups were white with a similar distribution in the control groups (70% to 97%).

The mean disease duration in the treatment groups ranged from 8 to 16 years, and 10 to 17 years in the control groups.

Interventions

Etanercept

Four RCTs, assessed etanercept (Brandt 2003; Calin 2004; Davis 2003; Gorman 2002) at a dosage of 25mg twice weekly delivered subcutaneously against placebo. The dosage in Huang 2008 was 50mg once weekly versus. van der Heijde 2006b assessed 50mg once weekly versus 25mg twice weekly versus placebo. Braun 2008 compared 50mg once weekly to 3g daily of sulphasalazine. The dosage was not

reported in the Barkham 2008 abstract. The length of treatment ranged from 6 weeks (Brandt 2003 and Huang 2008) to 24 weeks (Davis 2003).

Concomitant therapy

Brandt 2003 allowed NSAIDs at the same or less dosage at baseline; Calin 2004 allowed pre-study physiotherapy; Davis 2003; Gorman 2002, and van der Heijde 2006b allowed stable dosages of DMARDs, NSAIDs, and oral corticosteroids; and Huang 2008 allowed stable DMARDs dosages.

Infliximab

Braun 2002 assessed infliximab at 5 mg/kg intraveneously at weeks 0, 2, and 6. van der Heijde 2005 delivered this same dosage of infliximab at weeks 0, 2, 6, 12 and 18 weeks. Inman 2010 evaluated infliximab at 3 mg/kg delivered at weeks 0, 2, and 6. Marzo-Ortega 2005 assessed infliximab (5 mg/kg) in combination with methotrexate against placebo plus methotrexate.

Concomitant therapy

In both Braun 2002 and van der Heijde 2005, patients were allowed to continue on stable doses of NSAIDs. It appears concomitant therapy of NSAIDs, corticosteroids, analgesics, and DMARDs were allowed as long as doses remained stable in the Inman 2010 study. Marzo-Ortega 2005 allowed concomitant use of NSAIDs or oral corticosteroids.

Adalimumab

Two studies assessed adalimumab: Lambert 2007 and van der Heijde 2006a at 40 mg every other week for a 24 week period.

Concomitant therapy

Lambert 2007 did not mention concomitant therapy. In van der Heijde 2006a, patients were allowed to continue sulfasalazine (3 gm/day), methotrexate (25 mg/week), hydroxychloroquine (400 mg/day), prednisone or prednisone equivalent

(10 mg/day), and NSAIDs, if the dosage had remained stable for at least 4 weeks before the baseline visit.

Outcomes

All studies used the outcomes recommended by the ASAS working group. The primary outcome in two studies was the BASDAI >=50% (Brandt 2003; Braun 2002) and the ASAS 20 in ten studies (Braun 2008; Calin 2004; Davis 2003; Gorman 2002; Huang 2008; Inman 2010; Lambert 2007; van der Heijde 2005; van der Heijde 2006a; van der Heijde 2006b). The change in BASDAI score was the primary outcome in Marzo-Ortega 2005.

In Barkham 2008, the primary outcome was a change in the work instability of patients after 3 months as measured by the AS WIS scale.

In the abstract of Giardina 2009, the primary outcome was stated to be the proportion of patients achieving a 50% BASDAI response at week 102; Secondary: ASAS 50; BASFI, back pain, morning stiffness, CRP, spinal mobility. However, in the full-text article, the outcome defined as primary is not stated, and the 50% BASDAI response is not reported. ASAS 20 and 40, BASDAI, BASFI and AE were reported.

Source of funding

van der Heijde 2005 was supported by Centocor. Braun 2002 was funded by a grant from the German Minstry of Research and by Essex Pharma who provided the study drug. Inman 2010 did not report the funding source in the abstracts but the trial protocol states the study was sponsored by Schering-Plough. Marzo-Ortega 2005 reported that the study was supported by a grant in aid from Schering-Plough, UK.

Brandt 2003 was supported by a grant from the German Minstry of Research and by Wyeth Pharma who provided the study drug. Calin 2004 was funded by Wyeth Research. Gorman 2002 was funded by the National Institute of Arthritis and Musculoskeletal and Skin Diseases and Immunex. The trial report states that Immunex was "not involved in the study design, data collection, statistical analysis, or manuscript preparation". Davis 2003 was supported by Immunex Corporation. van der Heijde 2006b was supported by Wyeth Pharmaceuticals (study drug and grants to

investigational sites). Barkham 2008 and Huang 2008 did not report the source of funding in their abstract.

van der Heijde 2006a and Lambert 2007 were sponsored by Abbott Laboratories.

Braun 2008 and Giardina 2009 did not list any source of funding.

Excluded studies

Six studies were excluded after assessing the full-text article. The Characteristics of excluded studies table provides more details for the exclusions. Briefly, the participants in three studies (Barkham 2008b; Breban 2008; Haibel 2008) did not meet the review's inclusion criteria; the intervention in Li 2008 assesses the effect of methotrexate, not infliximab; the outcome in Visvanathan 2008 is not of interest to this review; and there is no separate information provided for anklyosing spondylitis patients in Van den Bosch 2002 (and we were unable to obtain this from the author).

Risk of bias in included studies

Additional information was received from authors on methodology and data for Davis 2003; Gorman 2002; and van der Heijde 2006a.

The following RCTs were reported as abstracts and did not provide enough information in the abstracts to make a judgement about risk of bias. They were all judged to be at 'unclear' risk of bias. Barkham 2008; Braun 2008; and Huang 2008.

Sequence generation

Calin 2004; Giardina 2009; Inman 2010; Lambert 2007; van der Heijde 2005; and van der Heijde 2006b did not provide any information regarding sequence generation so the judgement was 'unclear'. The six other studies provided evidence of appropriate generation of the randomization sequence.

Allocation

Calin 2004, Giardina 2009; Inman 2010; Lambert 2007; van der Heijde 2005; and van der Heijde 2006b did not provide information regarding the method of allocation

concealment. The six other studies provided evidence of appropriate concealment of allocation of the randomization sequence.

Blinding of patient assessed outcomes

Braun 2002, Brandt 2003; Calin 2004; Davis 2003; Gorman 2002; and van der Heijde 2005 reported the patient was blinded. We were unclear about the methods of blinding in Inman 2010;Lambert 2007; Marzo-Ortega 2005; van der Heijde 2006a; and van der Heijde 2006b which were reported as "double blind". There was no blinding in Giardina 2009 which places it at a higher risk of bias.

Blinding of physician reported outcomes

Braun 2002, Brandt 2003; Davis 2003; Gorman 2002; Lambert 2007 ; van der Heijde 2005; and van der Heijde 2006a reported investigator was blinded. Calin 2004 did not specify who other than the patient was blinded and physician blinding was unclear in Inman 2010; Marzo-Ortega 2005 and van der Heijde 2006b. There was no blinding in Giardina 2009 which places it at a higher risk of bias.

Incomplete outcome data

All trials were judged to be at low risk of incomplete outcome data bias for efficacy outcomes. Inman 2010 may be at a higher risk of bias because safety data wasn't provided for the 12 week RCT portion, only for the combined RCT and open-label phase. Lambert 2007 was judged to be unclear since safety data was not reported in the main article and only briefly mentioned in an abstract.

Selective outcome reporting

Most of the trials were judged to be at low risk of selective outcome reporting bias, with the exception of Giardina 2009. In Giardina 2009, the abstract we found first for this trial had the primary outcome listed in the abstract as the proportion of people achieving a 50% response in BASDAI. However, the full-text article does not report this outcome. We could not find a protocol for this trial. In terms of risk of bias for selective adverse event reporting, Inman 2010 and Lambert 2007 were judged as ‘unclear’ given the lack of specifics provided on harms data.

Method of adverse event monitoring

The following studies stated that the patients were monitored (though few details on the specifics of the monitoring were provided) for adverse events. These were judged to be at low risk of bias: Calin 2004; Davis 2003; Giardina 2009; Gorman 2002; Huang 2008; and van der Heijde 2006a. The rest of the studies did not mention how the patients were monitored for adverse events and were judged as an 'unclear' risk of bias.

Definition of serious adverse event provided

The following studies used a common grading system, though the specific definition of SAE was not provided in the articles: Davis 2003 Gorman 2002; Inman 2010. These were judged to be at low risk of bias. van der Heijde 2005 and van der Heijde 2006a did not provide general SAE definition, but each SAE was clearly explained in the published report. The other studies did not report their definition of 'SAE' and were judged to be at an unclear risk of bias.

Risk of bias for non-RCTs

There were six non-RCTs articles (Davis 2005 and 2008 from the Davis 2003 RCT), (van der Heijde 2008 and 2009 from the van der Heijde 2006a RCT) and Braun 2008 and Heldmann 2009 from the van der Heijde 2005 RCT) included in this review to assess the potential harms of anti-TNF treatment in ankylosing spondylitis. They were all follow-on, or extension studies of RCTs. Four were reported in full-text articles and two were conference abstracts. All but Braun 2008, were judged to be at an overall high risk of bias. Braun 2008 was judged to be moderate, mainly due to the continued blinding of patients in the extension period. Figure 11 summarizes each of the judgments of the non-RCTs using the Newcastle-Ottawa cohort scale. Table 1 provides the definitions for serious adverse events reported in the articles and the methods of ascertainment.

The items of 'Selection of the Non-Exposed Cohort' and 'Comparability of Cohorts on the Basis of the Design or Analysis' were marked as not-applicable since there was only a single cohort in these extension studies (i.e. no control group).

Selection

1. Representativeness of the exposed cohort: no study received a point for this item. Participants entering the RCT phase of the study were already a highly select group of participants, with various co-morbidities excluded (see Characteristics of included studies for more details). Those who then continued in the extension phase would have been participants who did not have a significant adverse event during the RCT-phase. Therefore, the participants exposed to anti-TNF agents for a follow-up period were judged to be at a high risk of bias how representative they may be to those in the community who may be taking this treatment.

2. Ascertainment of exposure: none of the published articles stated how they ascertained that participants in the extension study were taking the anti-TNF therapy as prescribed, therefore this criterion was judged to be at a high risk of bias.

3. Demonstration that outcome of interest was not present at start of study: Given the strict exclusion criteria for entering the RCT phase, we judged that the specific adverse events of interest to us such as tuberculosis, infection, congestive heart failure, and malignancy probably would not have been present at the start of the RCT. As well, if such an event occurred during the preceeding RCT phase, it would have been captured. The Heldmann 2009 study was more difficult to assess because there was a lag period of 1.3 +/-0.9 years (on average) between the end of the ASSERT RCT and the start of the EASIC cohort study so this was judged as not meeting the criteria and therefore at a higher risk of bias.

Outcome

1. Assessment of outcome: in all extension studies, except for the Braun 2008 extension study, patients and investigators were not blinded to the intervention. Although some diagnostic bias may exist (i.e. physicians may look more closely for

adverse events such as malignancy given the existing concern about this adverse event), we judged that given the nature of the adverse effects of interest, and the broad awareness of the existing concern of these adverse events with anti-TNF therapy, that the risk of bias is low.

2. Was follow-up long enough for outcomes to occur: we judged all studies to have adequate length of follow up (approximately 1.5 to 5 years) to assess most of our side effects of interest; though for some events like malignancies, this follow-up time is still quite short.

3. Adequacy of follow up of cohorts: in van der Heijde 2009, 83.9% received adalimumab treatment to 2 years and in Braun 2008, 82.6% completed 102 weeks on infliximab. Heldmann 2009 did not report the number of people completing 5 years in the continuous-infliximab group. We gave a point to those studies with a follow-up rate of 80% or greater.

Effects of interventions

The primary outcomes for this review were based on the disease-controlling anti-rheumatic treatment (DC-ART) Ankylosing Spondylitis Working Group Core Set (van der Heijde 1997). The review prespecified that outcomes measured at six months or less would assess short-term results and greater than six months would assess long-term results; however, all outcomes in the placebo-controlled trials were reported at six months or less. Note that the definitions of "serious adverse event" as defined in the trials are listed in Table 1.

A summary of findings table provides an overview of the results for important outcomes for all three biologics (Figure 12)

Etanercept (25mg twice weekly) versus placebo

Efficacy

Five studies assessed the effect of etanercept 25mg twice weekly versus placebo: Barkham 2008; Brandt 2003;Calin 2004;Davis 2003; Gorman 2002).

There was a statistically significant reduction in the BASDAI (0 to 100 scale; higher scores mean more disease activity) in the etanercept group compared to placebo in a pooled analysis of two studies at 6 and 24 weeks with a mean difference (MD) -20.9, 95% confidence interval (CI) -26.2 to -15.7 and the BASFI (0 to 100 scale; none to severe limitations), MD -16.1, 95% CI -20.1 to -11.2. Spinal pain on a 0 to 100 scale was also significantly reduced by 24 weeks (MD -21.7, 95% CI -26.8 to -16.7). Other outcome measures that were statistically significantly improved by 12 weeks included fatigue, entheseal pain, chest expansion, spinal mobility (as measured by modified Schober and occiput to wall tests), and morning stiffness. The BASMI, a measure of spinal mobility, and peripheral joint pain were not statistically significantly different between groups at 6 weeks, though they clearly favoured the etanercept group (MD -0.89, 95% CI -2.38 to 0.60 and MD -1.70 (95% CI, -3.60 to 0.20), respectively).

Patient and physician global assessments were also statistically significantly better in the etanercept versus placebo group over 16 to 24 weeks (patient global: standardized mean difference (SMD) -0.79, 95% CI -1.02 to -0.56). Brandt 2003 reported that there was a statistically significant difference (P 0.026) between SF-36 physical component scores at week 6 between treatment and control groups, but no details were provided. However, in terms of the SF-36 mental component score, there was no improvement in either group by 6 weeks.

There was a statistically significant greater number of patients who reached a >50% improvement in BASDAI at six and twelve weeks (relative risk (RR) 3.32, 95% CI 1.93 to 5.72).

A statistically significantly greater proportion of patients in the treatment group reached the ASAS 20, ASAS 50, ASAS 50, or ASAS 70 outcomes compared to placebo by 24 weeks. For the ASAS 40 (based on van der Heijde 2006b), the RR was 2.47 (95% CI, 1.43 to 4.26) and the RR for ASAS 5/6 was 7.71 (4.64 to 12.83), both in favour of etanercept. In a pooled analysis of 3 studies (Brandt 2003; Calin 2004; Davis 2003), the RR was 4.38 (95% CI 2.81, 6.83) for the ASAS 50 up to 24

weeks. The RR for achieving partial remission in the etanercept group compared to placebo was 3.86, 95% CI (1.93 to 7.70).

A significant proportion of patients in the etanercept group reduced their use of NSAIDs by 50% at 6 weeks (Brandt 2003), but the confidence interval is wide (RR 10.29, 95% CI 1.48 to 71.40) and there was no difference between groups in those that stopped use of NSAIDs.

Laboratory outcomes (CRP and ESR) were both statistically significant in favour of etanercept at 24 weeks. MRI outcomes were not assessed in these studies. Radiographic data was also not reported; which is not unexpected given the short time frame of these RCTs.

The results for Barkham 2008 were reported in an abstract (dosage not reported and number of patients in each group not clear so we were unable to include in the meta-analysis) and etanercept was stated to be statistically significantly better than placebo in terms of BASDAI, BASFI, ASQoL, and gait velocity at 3 months.

Adverse effects

Adverse effects from RCTs:

Withdrawals due to adverse events (Peto OR 4.03, 95% CI 1.43 to 11.30) and injection site reactions (RR 3.60, 95% CI 2.41 to 5.38) were significantly more frequent in the etanercept group. Total numbers of adverse events were not reported in these studies.

Infections (RR 1.18, 95% CI 0.89 to1.57), serious adverse events (Peto OR 2.00, 95% CI 0.71 to 5.64) and serious infections (Peto OR 2.25, 95% CI 0.30 to 17.13) were not significantly different between groups, but the point estimate favoured placebo.

Withdrawals due to inefficacy and total withdrawals were not significantly different between etanercept and placebo groups, but the point estimates favoured etanercept.

There was no significant between-group difference in headaches, diarrhoea, and nausea.

Most of the studies did not report on tuberculosis, lymphoma, demyelinating disease, autoimmune disorders, chronic or congestive heart failure, or hepatotoxicity. However, Davis 2003 reported no tuberculosis and van der Heijde 2006b reported no lymphoma, demyelinating disease, or autoimmune disorders. Calin 2004 and Gorman 2002 reported some elevated hepatic enzyme levels, but none were classed as clinical liver problems.

Adverse effects from open-label extensions (Table 2):

Two open-label extension studies met our inclusion criteria: Davis 2005 and Davis 2008 are extension studies of the Davis 2003 RCT.

At two years, 20.6% of patients experienced an injection site reaction. Five percent of the participants withdrew due to adverse events and 6.6% experienced serious adverse events, though only 1.6% withdrew due to serious adverse events. 2% (5/257) had a serious infection and one person was PPD+ but had no clinical symptoms of tuberculosis. No deaths or cases of lymphoma or demyelination occurred. A total of 57/257 (22.2%) withdrew from the 72-week open label extension.

Davis 2008 reported results at 168 weeks of the OLE. 131/257 (49%) dropped out of the extension period and 8.2% of those were withdrawals due to adverse events. A total of 234/257 (91%) participants experienced an adverse event and 12.8% had a serious adverse event. 22.2% had an injection site reaction. 72.8% of the participants had an infection, of which 6 (2.3%) were classified as serious. There was one case of tuberculosis. No deaths or cases of lymphoma or demyelination occurred.

Etanercept (50mg once weekly) versus placebo

Efficacy

Two studies assessed the effect of 50mg of etanercept once weekly versus placebo: Huang 2008 at six weeks and and van der Heijde 2006b at twelve weeks. Both showed a statistically significant response in the etanercept group compared to placebo in terms of ASAS 20, ASAS 40, and ASAS 5/6. With respect to partial remission, the pooled RR was 4.41 95% CI (1.99 to 9.76) and the ASAS 40 was RR 3.11, 95% CI 2.13 to 4.53.

Adverse effects

Adverse effects from RCTs

There was no significant difference between the 50mg etanercept and placebo groups in terms of the total number of adverse events, serious adverse events, withdrawals due to adverse events, number of infections or serious infections, headaches, diarrhea, or nausea. There was a significant increase in injection site reactions in the etanercept group compared to placebo (RR 5.41, 95% CI 2.40 to 12.19).

Both Huang 2008 and van der Heijde 2006b reported more elevated hepatic enzyme levels in the etanercept group, but none were classed as clinical liver problems. van der Heijde 2006b reported no malignancies, demyelinating disease, or lupus. Heart failure was not reported.

Etanercept 50mg weekly versus etanercept 25 mg twice weekly

van der Heijde 2006b performed a double-blind, placebo-controlled non-inferiority trial with 356 patients to compare the efficacy of the two methods of administration of etanercept. Both dosing regimens were found to be significantly better than placebo in terms of ASAS 20, ASAS 5/6, ASAS 40, BASDAI, BASFI, and other clinical measures. The study also showed that 50 mg once weekly was not inferior to the usual standard of 25 mg twice weekly in terms of the primary outcome of ASAS 20 response at twelve weeks. As well, patient-reported outcomes such as fatigue, EuroQOL-5D, and SF-36 scores were similar between the two doses.

Injection site reactions were similar in the 50 mg once weekly and 25 mg twice weekly groups (20.7% versus 22.7%). Infections were also similar between the two

groups (22.6% and 22.0%). The percentage of non-infectious serious adverse events was 5.2% and 4.0% in the 50mg once- and 25mg twice-per-week groups, respectively. One serious infection occurred in each group.

The authors concluded that "the efficacy and safety of etanercept 50 mg once weekly was comparable with that of the standard regimen of 25 mg twice weekly in patients with ankylosing spondylitis."

Combined etanercept 50mg once weekly or etanercept 25 mg twice weekly versus placebo

Given the results of the van der Heijde 2006b study showing that the efficacy and safety of the two administrative regimes of etanercept were comparable, we combined studies with these two regimes for the following outcomes included in the summary of findings table: ASAS 40, partial remission, withdrawals due to adverse events and serious adverse events. The overall results did not change significantly; but the confidence intervals were more precise. Data analyses 9.1 to 9.5 show these results.

Etanercept 50mg weekly versus sulphasalazine (3g) daily

The results for Braun 2008 were reported in an abstract. Statistically significant results were found for ASAS 20 (RR 1.47 (95% CI 1.26 to 1.71), ASAS 40 (RR 1.85 (95% CI 1.48 to 2.31), and ASAS 5/6 (RR 2.13 (95% CI 1.57 to 2.89) at 16 weeks in the etanercept compared to sulphasalazine group. BASDAI, BASFI, BASMI, nocturnal back pain, and modified Schober's response were all significantly improved in the etanercept versus the sulphasalazine groups. A second abstract reported on patient-reported outcomes. Outcomes such as health-related quality of life (as measured by EQ-5D, SF-36 or ASQoL) were all statistically significantly improved in the etanercept group compared to the sulphasalazine group at 16 weeks. In terms of fatigue, the percent change from baseline was 47.1% in the etanercept group compared to 26.4% in the sulphasalazine group (P<0.001).

The risk of serious adverse events was no different between etanercept and sulphasalazine (Peto OR 1.17 (95% CI 0.33 to 4.14).

Infliximab versus placebo (pooled 3mg/kg and 5mg/kg)

Efficacy

Two studies assessed the effect of infliximab 5mg/kg versus placebo; Braun 2002 at 12 weeks and van der Heijde 2005 at 24 weeks. Inman 2010 evaluated a 12-week RCT of a lower dose of 3 mg/kg against placebo. There was no significant heterogenetity of including the 3mg/kg dose with the 5mg/kg dose results so they were pooled together for dichotomous outcomes (no variance was provided for continuous outcomes). Separate details for the lower dose are also reported below.

There was a statistically significant reduction in the BASDAI in the infliximab group compared to placebo in a pooled analysis of two studies (Braun 2002; van der Heijde 2005) for 12 and 24 weeks with a MD -2.4, 95% CI -2.57 to -2.13) and the BASFI, MD -1.91, 95% CI -2.28 to -1.54. Spinal pain on a 0 to 10 scale was also significantly reduced by 12 weeks (MD -2.82, 95% CI -3.82 to -1.82). Other outcome measures that were statistically significantly reduced by 12 weeks included BASMI, morning stiffness and peripheral joint pain. Fatigue was not statistically significantly different between groups at 12 weeks in the Braun 2002 study.

Patient and physician global assessments were also statistically significantly better in the infliximab versus placebo group over 12 weeks (patient global, 0 to 10 scale, higher worse: MD -3.00, 95% CI -4.12 to -1.88).

There was a statistically significant greater number of patients who reached a >50% improvement in BASDAI, pooled at 12 and 24 weeks (relative risk (RR) 5.16, 95% CI 2.90 to 9.20).

A statistically significantly greater proportion of patients in the treatment group reached the ASAS 20, ASAS 50, ASAS 50, or partial remission outcomes compared to placebo by 24 weeks. For the ASAS 20 (based on Inman 2010 and van der Heijde 2005), the RR was 2.69, 95% CI 1.96 to 3.71 and for ASAS 40, RR 4.01, 95% CI,

2.50 to 7.52 and the RR for ASAS 5/6 was 7.62, 95% CI 3.73 to 15.56, all in favour of infliximab. For partial remission, the relative risk was 17.47, 95% CI 3.42 to 89.14.

For assessment of quality of life, the SF-36 physical functioning score was significantly better in the infliximab group in a pooled analysis over 12 and 24 weeks (MD 8.19, 95% CI 6.76 to 9.62) (1-100 scale). There was no difference between groups in the SF-36 mental health score.

Two abstracts reported on pain and fatigue from the ASSERT trial. The median percent change in fatigue in terms of VAS scores was 29.4% in the infliximab group compared to 1.8% in the placebo group ($p<0.001$) and for pain it was 57.4% versus 6.7% ($p<0.001$).

A significant proportion of patients in the infliximab group reduced their use of NSAIDs by 50% at 12 weeks (van der Heijde 2005), (RR 3.09, 95% CI 1.40 to 6.83) and the result was similar for the outcome of stopped use of NSAIDs. CRP (in mg/dl) was significantly reduced in the infliximab group compared to placebo (MD -2.20, 95% CI -2.54 to -1.86)

Radiological changes were assessed using the Bath ankylosing spondylitis radiology index (BASRI) in Braun 2002 though data was not shown. They found that the "initial degree of radiological axial changes assessed by the BASRIs was similar in both groups." Interestingly, the study stated that there was "no less benefit in patients with higher BASRI scores than those with lower scores"; indicating that the amount of ankylosis did not impact the benefit of infliximab. Radiographic outcome data was not reported in van der Heijde 2005 and Inman 2010.

MRI data was reported in a secondary publication of van der Heijde 2005 (Braun 2006). The MRI Activity Score was used to assess spinal inflammation as detected by MRI. The reduction in MRI Activity Score from baseline to week 24 was significantly greater in the infliximab-treated group (MD -4.42, 95% CI -5.59 to -3.25 on a 0 to 138 scale). As well, evidence of ‘some’ (defined as a MRI Activity

score >1) spinal inflammation by week 24 was 37% in the infliximab group compared to 73.6% in the placebo group (P<0.001)

Adverse effects

Adverse effects from RCTs:

Withdrawals due to adverse events were not significant between groups but the point estimate favours placebo (Peto OR 2.76, 95% CI 0.50 to 15.22). Similar results were found for results pooled at 12 and 24 weeks: total adverse events (RR 1.14, 95% CI 0.98 to 1.33); serious adverse events (Peto OR 2.47, 95% CI 0.75 to 8.14); and serious infections (Peto OR 5.10, 95% CI 0.44 to 58.76). However, at 12 weeks, there was a significant increase in serious adverse events, Peto OR 7.80, 95% CI 1.07 to 56.65, in data pooled from Braun 2002 and Inman 2010. There was one event of tuberculosis in the infliximab group in Braun 2002.

There was no statistically significant difference between groups in terms of infusion reactions, any infection, headache, or diarrhea.

There were no cases of TB, lymphoma, or death in van der Heijde 2005. Liver enzymes were reported as increased in the infliximab group, but no clinical liver problems occurred. These outcomes were not reported in Braun 2002 or Inman 2010, nor did any study report on chronic or congestive heart failure.

*Adverse effects from open-label extensions (*Table 2*):*

Results for participants from van der Heijde 2005 were reported in the Braun 2008 article after being followed for an average of 70 weeks of exposure to infliximab in the original placebo group and 94 weeks of exposure in the original treatment group. Data was presented separately for the two aforementioned groups. We chose to present on those that were originally assigned to treatment (N=201) so the total treatment period covers 102 weeks. 97.5% of patients reported any adverse event and 34/201 (16.9%) reported a serious adverse event. 78.6% of patients had an infection, and 8 patients (4%) had a serious infection, with pneumonia the most frequent. Two patients had a malignancy (lung cancer and breast cancer) and 2 patients experienced

a lupus syndrome. There were no serious infusion reactions, tuberculosis or serious opportunistic infections, and no deaths.

The EASIC study (Heldmann 2009) was a cohort (N=103) of European patients who completed the van der Heijde 2005 RCT. There was a 1.3+/-0.9 year lag period after the ASSERT study finished until the EASIC study began. Fourteen patients did not take infliximab in this interim period. Data at four years was not provided separately for those who had been continuously on infliximab versus those who were not, so we did not include this data. During the two-year EASIC study, 75/89 patients 84% of patients reported they had experienced any adverse event and there were 230 reported infections, of which 3 were considered serious infections. There were 4 infusion reactions, and no cases of tuberculosis or malignancy.

Low dose infliximab (3mg/kg) versus placebo

Efficacy

Inman 2010 conducted a study on 76 patients to compare the efficacy of infliximab at 3mg/kg versus placebo. The first twelve weeks were a double-blind placebo phase and then there was an open-label phase where placebo patients switched to 3mg/kg infliximab and received infusions at weeks 14, 16, 22 and every 8 weeks afterwards. Patients were eligible for a dose escalation to 5mg/kg at weeks 22 or 38 if they were not responding adequately.

Significantly more participants in the infliximab group than placebo achieved an ASAS 20 response at 12 weeks (RR 1.81, 95% CI 1.02 to 3.22). As well, the ASAS 40, 50, 70, and ASAS 5/6 responses were also significantly better in the infliximab group versus placebo at week 12. A separate publication (Maksymowych 2010) reported on spinal inflammation as measured by the SPARCC MRI method found a large treatment effect in favour of infliximab (mean per cent change based on evaluation of the most severely affected discovertebral units (6 DVU score) in the infliximab group for was -55.1% compared to +5.8% in the placebo group, P<0.001). When the evaluation was based on the entire spine (23 DVU score), the infliximab

group had a mean reduction of 57.2% compared to 3.4% in the placebo group ($p < 0.001$).

However, after the 12 week double-blind placebo phase, 68% of patients in the 3 mg/kg infliximab group switched to the 5 mg/kg by 38 weeks because of lack of efficacy.

Adverse effects

Adverse event data was not presented separately for the placebo-controlled, 12-week phase, with the exception of one serious adverse event of arthralgia in an infliximab-treated patient.

Infliximab + methotrexate versus placebo + methotrexate

Efficacy

Marzo-Ortega 2005 assessed the efficacy of adding infliximab (5mg/kg) or placebo to methotrexate therapy. The final infliximab or placebo treatment was given at 22 weeks. At 30 weeks, neither the change in BASDAI (MD -1.14, 95% CI -2.76 to 0.48), nor the ASAS 20 response (RR 2.33, 95% CI 0.80 to 6.80) or a 50% improvement in BASDAI (RR 2.50, 95% CI 0.87 to 7.22) were significantly different between the two groups. There was a significant difference at 10 weeks, but this did not extend to 30 weeks. The last dosing of infliximab was at 22 weeks, so the authors concluded that the addition of methotrexate to the treatment regimen did not lengthen the efficacious period of infliximab. There was a significantly greater reduction in the number of lesions in the sacroiliac joints and spine resolving completely in the combination group versus the methotrexate monotherapy group, as assessed by MRI.

Adverse effects

There was no significant difference between the two groups in terms of the relative risk of infections (RR 1.50, 95% CI 0.35 to 6.50). Two mild infusion reactions

occurred in the combination group and none in the MTX group. No severe adverse events were seen in either group.

Adalimumab versus placebo

Efficacy

There was a statistically significant reduction in the BASDAI and BASFI scores in the adalimumab group compared to placebo in a pooled analysis of two studies (Lambert 2007 (efficacy results reported in Maksymowych 2005 and 2008) ; van der Heijde 2006a) at 12 weeks; BASDAI: MD -1.79, 95% CI -2.16 to -1.41 and BASFI MD -1.80, 95% CI -2.27 to -1.33; results were similar at 24 weeks.

Total back pain on a 0 to 10 scale was also significantly reduced by 12 and 24 weeks (MD -2.29 95% CI, -2.75 to -1.82). Other outcome measures that were statistically significantly reduced by 24 weeks included morning stiffness (mean intensity and duration, 0 to 10 scale, MD -2.00, 95% CI -2.45 to -1.55), fatigue (MD -2.00, 95% CI -2.45 to -1.55), entheseal pain (MD -1.90 95% CI -3.60 to -0.20) and BASMI (MD -1.00, 95% CI -1.50 to -0.50).

Patient and physician global assessments were also statisticially significantly better in the adalimumab versus placebo group over 12 to 24 weeks (patient global (0 to 10 scale): MD -3.20, 95% CI -3.68 to -2.72).

There was a statistically significant greater number of patients who reached a >50% improvement in BASDAI at 12 and 24 weeks (RR 2.83, 95% CI 1.75 to 4.57).

A statistically significantly greater proportion of patients in the adalimumab group reached the ASAS 20, ASAS 50, ASAS 50, ASAS 70 and partial remission outcomes compared to placebo at both 12 and 24 weeks. For the ASAS 20 (based on Lambert 2007, van der Heijde 2006a), the RR was 2.53, 95% CI 1.82 to 3.51 and the RR for ASAS 5/6 was 3.71, 95% CI 2.23 to 6.17, both in favour of adalimumab at 12 weeks. In Lambert 2007, the RR was 5.79, 95% C1.81 to 18.49 for the ASAS 50 at 12 weeks. At 12 weeks, the risk of partial remission was in favour of adalimumab (RR 3.94, 95% CI 1.74 to 8.94).

The proportion of people reducing their use of NSAIDs was not reported.

CRP was statistically significant in favour of adalimumab at 24 weeks.

Spinal and sacroiliac (SI) joint inflammation

Lambert 2007 used MRIs to assess the effect of adalimumab compared to placebo in reducing spinal and SI joint inflammation using the SPARCC scoring method. MRIs were obtained for all participants (N=82) at baseline and week 12. SPARCC scores for the spine can range from 0 to 108 and SPARCC SI joint scores can range from 0 to 72.There was a statistically significantly greater reduction in the mean spine SPARCC score of adalimumab-treated patients (median change 6.3, range 34.0 to 2.0) compared with placebo-treated patients (median change 0.5, range 26.0 to 13.5) (P< 0.001). The mean SI joint SPARCC score also decreased significantly between the adalimumab (median change 0.5, range 22.5 to 2.5) and placebo groups (median change 0.0, range 13.5 to 16.0) (P <0.001). In terms of percent change from baseline, placebo-treated patients had a 9.4% mean increase in spine SPARCC scores compared to a 53.6% mean reduction in scores in adalimumab-treated patients. There was also a significant difference in the mean percent reduction in adalimumab (52.9%) and placebo-treated (12.7%) patients in the SI joint SPARCC score.

Adverse effects

Adverse effects from RCTs:

Lambert 2007 did not report adverse events in the published article. An abstract by Maksymowych (2005) on the same study reported briefly on adverse events.
There was a significant increase between adalimumab and placebo in terms of total adverse events (RR 1.25, 95% CI 1.05 to 1.49) and any infection (RR 1.58, 95% CI 1.10 to 2.27). There was only one serious infection reported, and that was in the placebo group, and there was also one event of hepatoxicity requiring hospitalization in the infliximab group in van der Heijde 2006a.

Withdrawals due to adverse events, headache, and serious adverse events were not significantly different between adalimumab and placebo groups.

van der Heijde 2006a stated that there were "no cases of tuberculosis/granulomatous infections, demyelination, drug-induced lupus, congestive heart failure, malignancies, and no deaths occurred during the 24-week period of the study." These outcomes were not reported in Lambert 2007.

Adverse effects from open label extensions (Table 2):

Two- and three-year open label extension data was available from van der Heijde 2006a in a full-text article (van der Heijde 2009) and an abstract (van der Heijde 2008). Data was presented as the rate in the double-blind RCT period versus the extension period per 100 patient-years. After two years, with respect to withdrawals due to adverse events, the rate was 4.5 versus 3.8/100 patient-years, 10.5 versus 10.2/100 patient-years for serious adverse events and 0.0 versus 0.9 for malignancies (four people developed cancer). Serious infections were 1.1 versus 0/100 patient-years. No cases of tuberculosis, demyelinating disease, congestive heart failure or death were reported.
After three years, withdrawals were 3.6 versus 3.8/100 patient-years and serious adverse events were 11.1 versus 10.2/100 patient-years. Serious infections were slightly increased: 0 to 1.4/100 patient-years. Malignancies increased from 0.0 to 0.7/100 patient-years and there was one death out of 277 people who had received at least three years of adalimumab treatment. Other adverse event outcomes of interest were not reported.

Etanercept versus infliximab

Efficacy

One RCT (Giardina 2009) compared the efficacy and safety of etanercept compared to infliximab over a two-year period. The trial was reported first in a conference abstract and later a full-text article was published.

Fifty patients were enrolled in the trial and the abstract states that "no patients discontinued therapy". To calculate the proportion of responders reported in the abstract for entry into RevMan, the number of participants in each treatment group

was assumed to be the denominators at the start of the study. However, there is a slight discrepancy in the number of people in each group; the full-text article states that there were 25 people in each group, while the abstract states there were 26 people in the etanercept group and 24 people in the infliximab group. We used the numbers from the full-text article.

The abstract reports no statistically significant differences in ASAS 50 or the proportion of patients achieving a 50% BASDAI response between the two interventions at week 102.

The full-text article reports no statistically significant difference between etanercept and infliximab in ASAS 20 and ASAS 40 responders at weeks 12 or 104. At week 12 there was a statistically significantly higher reduction in BASDAI and BASFI scores in the infliximab compared to etanercept group; however, by week 104, there was no significant difference between groups.

Adverse effects

Adverse effects from RCT

With respect to adverse events, the incidence of adverse events in terms of serious infections (Peto OR 0.50, 95% CI 0.05 to 5.03), headaches and diarrhea were similar among etancercept and infliximab groups. There was a difference in injection site/infusion reactions between the groups (25% in the etanercept group versus 4% in the inflixmab group) which the authors state was statistically significantly different ($p<0.005$), but which in RevMan gives a non-significant difference. There were no withdrawals due to adverse events in either group. As well, there were "no cases of cases of opportunistic infections, tuberculosis, congestive heart disease, demyelinating disorders, lupus-like syndromes, and malignancy."

Adverse events - pooled results from all three anti-TNF agents versus placebo

We pooled the results from etanercept, infliximab, and adalimumab to assess for a class-effect of adverse effects of TNF-inhibitors. Given our interest in adverse events was pre-specified in short (<6 months) and long term (>6 months) periods, and that

all time points reported were six months or less, we pooled all trials together for an assessment of short-term effects.

There was a statistically significantly greater total number of adverse events and withdrawals due to adverse events in the anti-TNF versus placebo group (RR 1.24, 95% CI 1.10 to 1.40 and Peto OR 2.78, 95% CI 1.27 to 6.10, respectively). The risk of injection or infusion site reactions was significantly greater in the anti-TNF group (RR 3.02, 95% CI 2.12 to 4.31). The risk of any infection was also statistically increased in the anti-TNF group, RR 1.22, 95% CI 1.02 to 1.45.

Serious adverse events were not statistically significant between the two groups (Peto OR 1.83, 95% CI 0.93 to 3.62), though the results favoured placebo. Most studies did not provide a definition of a 'serious adverse event'. Two studies (Davis 2003 and Gorman 2002) stated that the National Cancer Institute Common Toxicity criteria scale was used to grade adverse events and abnormal laboratory values and Inman 2010 used MedDRA ver9.

Serious infections were also not statistically different between the two groups (Peto OR 1.89, 95% CI 0.45 to 7.96; with a risk difference of 0.00, 95%CI -0.01 to 0.02). The nature of the serious infections were: etanercept: one staphylococcal cellulitis after a spider bite and one wound infection after a cat bite (Davis 2003) and one erysipela and one streptococcus pyogenes at an insulin catheter site (van der Heijde 2006b); infliximab: one systemic tuberculosis in the lymph node and spleen (Braun 2002) and 1 cholecystis and 1 pneumonia (van der Heijde 2005). No serious infections were reported with adalimumab.

There was one case of tuberculosis reported in Braun 2002 in the infliximab-treated group compared with none in the placebo group (Peto OR 7.61, 95% CI 0.15, 383.66). No cases of tuberculosis were reported in Davis 2003; Giardina 2009;van der Heijde 2005;van der Heijde 2006a. The other trials did not mention tuberculosis as an outcome.

The trials of Giardina 2009;van der Heijde 2005; van der Heijde 2006a; van der Heijde 2006b reported no cases of malignancies or lymphoma. Lymphoma was not reported as an outcome in the other trials.

Similarly, three studies (Giardina 2009; van der Heijde 2006a;van der Heijde 2006b) reported on cases of demyelinating or autoimmune disease or congestive heart failure, but no events occurred. Five studies reported on liver function tests, but only one person required hospitalization for hepatoxicity (van der Heijde 2006a) (Peto OR 7.35, 95% CI 0.15 to 370.60).

Adverse effect warnings from regulatory websites:

Table 3 summarizes the warnings on the TNF-inhibitors from the websites of FDA MedWatch, European Medicines Evaluation Agency, Australian Adverse Drug Reactions Bulletin, and UK Medicines and Healthcare products Regulatory Agency (MHRA) Drug Safety Updates. A recent warning (August 31, 2009) of interest on the FDA MedWatch website stated the FDA had "completed its analysis of tumor necrosis factor (TNF) blockers and has concluded that there is an increased risk of lymphoma and other cancers associated with the use of these drugs in children and adolescents". Most of the warnings reported on outcomes that were part of the assessment of outcomes in this review, with the exception of concerns about reactivation of hepatitis B, skin reactions: new or worsening psoriasis, and haematological reactions such as pancytopenia and aplastic anaemia.

Sensitivity analyses

Etanercept

Three studies, Brandt 2003; Davis 2003; Gorman 2002, were judged to be at low risk of bias in terms of adequate generation of the randomization sequence, allocation concealment, and blinding. Sensitivity analyses of BASDAI, BASFI, spinal pain, patient global assessment, ASAS 50, withdrawals due to AE, and all infections did not result in significant changes from the overall pooled effect when only the lower risk of bias studies were pooled. For example, BASFI changed from a MD of -16.09, 95% CI (-20.05 to -11.23) to a MD of -16.51 (-21.90 to -11.11) when Calin 2004 was

removed from the overall estimate. When the study with an 'unclear' rating (van der Heijde 2006b) was removed from the pooled estimate for >50% improvement in BASDAI, there was a large change in the effect estimate (RR 3.32, 95% CI 1.93 to 5.72) to a RR of 9.14 (95% CI 1.3 to 64.34). However, this large RR has very wide confidence limits and is based on a small sample size (N=30).

Infliximab

Braun 2002 is judged to be at a low risk of bias and van der Heijde 2005 and Inman 2010 did not provide enough information in the published papers so they were rated as 'unclear' for adequate sequence generation, allocation concealment, and blinding. There was very little change in the BASDAI, BASFI and >50%improvement in BASDAI when those with unclear risk of bias were removed from the pooled analysis.

Adalimumab

Lambert 2007 did not provide enough information in the published articles so is judged to be at a 'unclear' risk of bias while van der Heijde 2006a is judged to be at low risk of bias. However, removing Lambert 2007 from the pooled analyses did not change the result significantly for BASDAI, BASFI, total back pain, patient global, or ASAS 20.

Mantel-Haenzsal versus Peto OR for rare events

Although not specified *a priori*, we decided to perform a sensitivity analysis using the Mantel-Haenszel OR method with a standard continuity correction of 0.5 on those meta-analyses in which we had used the Peto OR to check the robustness of our results (as recommended by Sweeting 2004). Although using the Mantel-Haenszel method changed the width of some confidence intervals, none of the point estimates changed significantly, nor did any estimates change in their statistical significance. Therefore, we feel our estimates of outcomes with rare events are robust.

Subgroup analyses

As there are not many included studies per outcome by drug, any subgroup analyses must be viewed with caution.

Dosage

We pooled the 5mg/kg dose and 3mg/kg dose of infliximab for the efficacy analyses of ASAS 20, ASAS 40, ASAS 50, ASAS 5/6, BASDAI 50%. There was very little heterogeneity when these doses were pooled. We investigated the effect size in the low dose and high dose groups separately and found them to be quite similar.

Trial duration

All RCT portions of the studies were less than six months duration and we had specified *a priori* that trials less than six months would be used to assess short-term efficacy.

Patient characteristics

Etanercept

Overall, the studies were quite homogeneous. Barkham 2008 and Huang 2008 were reported in abstracts so details of patient characteristics are not available. All seven studies included patients meeting the modified New York classification criteria. Patients were similar across trials in terms of age (range in treatment group 38 to 45 years). Calin 2004 included the oldest patients but the effect estimates were similar to the other studies for spinal pain, BASFI, ASAS 20, and ASAS 50. Mostly men were enrolled in these studies (%male in treatment group range: 65% to 80%). BASDAI and BASFI baseline scores were similar across studies (BASDAI range: 5.8 to 6.6 and BASFI range: 5.3 to 6.2). Disease duration was similar, ranging from 10 to 15 years in the treatment groups across studies. Only Brandt 2003 did not allow concomitant DMARD therapy but in terms of the BASDAI, the result from this study was very similar to the other study it was pooled with (van der Heijde 2006b). Almost all included studies excluded patients with complete ankylosis (fusion)

(exclusion criteria not reported in Lambert 2007 and van der Heijde 2006a restricted a priori to <10% of recruited patients could have complete ankylosis).

Infliximab

All patients fulfilled the modified New York classification criteria. Patients were similar across the three trials in terms of age (range in treatment group 40 to 42.9 years), percent of males enrolled (treatment group range: 68% to 82%), and disease duration (ranging from 10 to 16 years in the treatment groups across studies). BASDAI and BASFI scores were similar at baseline in Braun 2002 and van der Heijde 2005 but not reported in Inman 2010. Complete ankylosis was not mentioned as an exclusion criteria in Inman 2010. Given the homogeneity of the studies, no subgroups were conducted by patient characteristics.

Adalimumab

Both studies (Lambert 2007; van der Heijde 2006a) included patients with the modified New York classification criteria. The patients were of similar age (41.9 and 41.7 years average in the treatment groups), sex distribution (76% in the treatment groups) and disease duration (14.5 and 11.3 years). They were also very similar at baseline in terms of BASDAI and BASFI. No subgroup analyses were undertaken by patient characteristics due to the homogeneity of the studies.

Indirect comparisons

Indirect comparisons of one treatment versus another are useful when there is no, or limited, direct evidence from head-to-head randomized controlled trials comparing treatments of interest to practitioners who must make choices as to which treatment to prescribe to their patient. In the case of the three anti-TNF agents for use in ankylosing spondylitis, the majority of RCTs assess each of adalimumab, etanercept, and infliximab against placebo and we also found one unblinded RCT comparing etanercept to infliximab over a two year period (Giardina 2009). Figure 3 describes the relationship of trials in this systematic review. It is a network with one closed loop (Wells 2009).

We used the direct evidence from the Giardina 2009 study in the MTC analysis to give refined estimates of the comparison of etanercept versus infliximab and refined estimates of the three anti-TNF agents versus placebo. As well, the MTC analysis provided us with two new indirect estimates of adalimumab versus etanercept and adalimumab versus infliximab. We performed indirect comparisons on the following outcomes: ASAS 5/6, ASAS 40, ASAS 50, >50% BASDAI, withdrawals due to AE, all infections, and serious infections. Giardina 2009 was able to contribute data to all the outcomes except ASAS 5/6 and withdrawals due to AE. The outcomes of withdrawals due to AE and serious infections had rare events, so in the MTC analysis, a random effects model was used and the zero events were adjusted for by adding 0.5.

Figure 5 provides details on the refined GLIMMIX and MTC placebo estimates and also shows the original odds ratio obtained from the RevMan forest plots. There was general consensus in terms of whether statistical significance is reached and which comparator is favoured, but some of the confidence intervals are much wider for the modeled results. For the MTC refined placebo estimates: both etanercept and infliximab remained statistically significant versus placebo for ASAS 5/6, ASAS 50, and >50% BASDAI. Adalimumab was statistically significant compared to placebo in the original estimates for the efficacy outcomes listed above, but did not achieve statistical significance in the refined MTC estimate (though the point estimates were still in favour of adalimumab). For withdrawals due to AE, the refined estimates for all three anti-TNF agents were not statistically significant (though the point estimate favoured placebo more strongly). In the original estimate of etanercept versus placebo, the OR was significant in favour of placebo. In terms of infections, there was no statistically significant difference between anti-TNF and placebo groups, though the adalimumab versus placebo original estimate was statistically significant. For serious infections, there was again no statistically significant result, though the point estimate was in favour of adalimumab while for the other two drugs it was in favour of placebo.

As already noted, we had direct estimates of the efficacy and safety of etanercept versus infliximab from the Giardina 2009 study. We found the indirect MTC estimate and direct estimates to be fairly similar. For example, for ASAS 50, the MTC estimate was OR 0.37, 95% CI 0.12 to 1.90 compared to the direct estimate of OR 0.50, 95% CI 0.15 to 1.71. For >50% BASDAI response, the MTC estimate was OR 0.58, 95% CI 0.21 to 2.57 compared to the direct estimate of OR 0.62, 95% CI 0.19 to 1.99. The point estimate for serious infections was different, but the confidence interval was similar: OR 0.18, 95% CI 0.03 to 5.31 (MTC estimate) compared to 0.48, 95% CI 0.04 to 5.65 for the direct estimate (Figure 6).

Two new estimates adalimumab versus etanercept and adalimumab versus infliximab were obtained from the MTC analysis (Figure 6). For adalimumab versus etanercept, the results were not statistically significant in terms of ASAS 5/6, ASAS 40, ASAS 50, and >50% BASDAI, though the point estimate favoured etanercept. The three safety outcomes were also not statistically significant and there was no consistency of which drug was favoured.

For adalimumab versus infliximab, none of the four efficacy outcomes were statistically significant, though the point estimates favoured infliximab. The three safety outcomes were also not statistically significant and there was no consistency of which drug was favoured.

We assessed the similarity of results across the three methodologies: ITC, GLIMMIX, and MTC (see Figure 6 for details) and found that for the majority of comparisons, the results agreed in terms of which drug the point estimate favoured and whether statistical significance was achieved or not. There were three examples in which, depending on the methodology, the point estimate fell on different sides of the null line, though none of the results reached statistical significance (noted as NC* in the figure). For example, in the etanercept versus infliximab ASAS 40 outcome, the point estimate (OR) was 0.69 for GLIMMIX, 0.20 for MTC, and 1.04 for ITC, though none were statistically significant. The adalimumab versus etanercept ASAS 5/6 outcome provides an interesting example in which the ITC method gives a statistically significant result (OR 0.23, 95% CI 0.09, 0.59) while the GLIMMIX and

MTC results do not (OR 0.32, 95% CI 0.02, 5.78 and OR 0.29, 95% CI 0.07, 4.22, respectively). We investigated potential reasons for this discrepancy, but could not determine a plausible reason for it. A similar discrepancy was found with the outcome of serious infections for the adalimumab versus etanercept and adalimumab versus infliximab comparisons.

As outlined in the papers by Bucher 1997 and Song 2009, there are several key assumptions that must be met when undertaking indirect comparisons. Song breaks these assumptions into three components: i.homogeneity; ii. similarity of trial; iii. consistency of evidence. Homogeneity refers to the standard assumptions used for pooling studies in a meta-analysis; i.e. trials comparing two treatments must be both clinically and methodologically similar to be combined. Trial similarity is comprised of clinical similarity and methodological similarity and the similarity of the bridging treatment. By 'bridging treatment' we mean the common comparator (ie. in a trial of A versus C and B versus C, the bridging treatment is 'C'). The assumption of consistency means that the results of direct and indirect evidence should not be heterogeneous.

With respect to the homogeneity assumption, statistical heterogeneity (as assessed by I^2 values) for the individual meta-analyses (e.g. adalimumab versus placebo, etanercept versus placebo, and infliximab versus placebo) used in the indirect comparison meta-analyses were all very low, with the exception of the outcome of infections in two infliximab trials, which had an I^2=60% when pooled at 12 and 24 weeks. As well, we assessed the trials in terms of their PICO: P-population; I-intervention; C-comparator; O-outcomes and determined they were similar enough to be pooled.

For the assumption of trial similarity, Figure 4 shows the important patient characteristics for the studies included in the network meta-analysis. The mean age in the head-to-head Giardina study was younger (31.9 years) compared to the placebo-controlled studies (range: 38 to 45.3 years); however, the disease duration of the Giardina study was towards the high end of the range (15.4, range 10 to 16.4). The majority of participants were male (range 65% to 80%). The baseline BASDAI

and BASFI scores were also fairly similar across studies (range 5.8 to 6.6 and 4.5 to 6.2, respectively). Therefore, it seems that the demographic and clinical characteristics of the patient populations are quite similar. As noted in the paper by Hochberg 2003 et al, "another method of assessing the validity of this assumption is to compare the proportion of patients randomly allocated to receive placebo who develop the study outcome". For the ASAS 5/6 outcome, the placebo groups across the different trials had a response rate of 8% to 12%; for ASAS 40, the range of responses was 12% to 21%; for ASAS 50, the range of responses was 6% to 13%; for >50% BASDAI response, the range was 6% to 20%. Thus, it seems that it is likely that the patients in these trials are drawn from similar patient populations.

With respect to methodological risk of bias, most of the studies are similar in terms of sequence generation and allocation concealment being either at low or unclear risk of bias due to lack of details in the published reports. However, all but the etanercept versus infliximab Giardina study were reported to be "double-blind", though details of the methods of blinding were not always clearly reported in the other studies. The Giardina was an unblinded study, presumably because it was difficult to blind treatments with different routes of administration; i.e. subcutaneous versus infusion. This puts it at a higher risk of bias than the other studies. As well, the length of outcome assessment was quite different in the Giardina study; an abstract reported 6-week results, and the full-text article reported results at 104 weeks. All other studies reported results at 24 weeks or less.

As described in the methods section we used back calculation to check for the consistency of direct and indirect evidence. Overall, there was no significant inconsistency for the GLIMMIX analysis and only three significant inconsistency estimates found for the MTC method of analysis.

Assessment of publication bias

We had planned to assess publication bias by visual inspection of funnel plots. However, it was very difficult to assess plots with few studies in them. According to the Cochrane Handbook; Chapter 10 (Higgins 2008), when assessing publication

bias using statistical tests for funnel plot asymmetry, at least 10 studies should be included in the meta-analysis, otherwise the power of the test is too low. There is no guidance regarding the minimum number of studies needed for visual inspection. We decided to look at plots with more than five studies and there were two analyses that met this criteria (see Figure 7; Figure 8). Both appear to have an empty area in the bottom left of the plot. This indicates there may be publication bias for smaller 'negative' studies with high standard errors and little effect on the relative risk. Figure 8 is an efficacy outcome (ASAS 20), so the lack of studies in the bottom left of the plot indicates that small studies showing little effect may not be published. Figure 7 is a safety outcome (injection site reactions), so the interpretation here is that small studies showing no safety concerns may not be published. This result is not surprising and also interesting in that it may mean that larger studies with safety concerns are being published, while smaller studies with no safety concerns are not being published, potentially leading to the problem of overstating safety concerns. However, there is evidence of some heterogeneity in this meta-analysis and given the few studies in the plots, it difficult to be sure that publication bias is the reason for the asymmetry in the plots.

Discussion

Summary of main results

Fifteen RCTs with a total of 2,459 participants met the inclusion criteria for this review. The primary efficacy outcomes were based on the disease-controlling anti-rheumatic treatment (DC-ART) Ankylosing Spondylitis Working Group Core Set.

Mainly high quality evidence was available for the efficacy outcomes. All three anti-TNF agents were statistically significantly better than placebo in terms of reducing disease activity (BASDAI, 0-100 scale; etanercept: MD -20.94 (95% CI, -26.21 to -15.67); infliximab (0-10 scale, MD -2.40 (95% CI, -2.67 to -2.13); adalimumab (0 to 10 scale; MD -1.80 (95% CI, -2.20 to -1.40). The other core set measures: physical function, morning stiffness, measures of spinal mobility, spinal pain (in etanercept and infliximab), peripheral joint pain, entheseal pain, fatigue, and patient global

assessment were also statistically significantly in favour of each of the anti-TNF agents compared to placebo.

In terms of the absolute risk difference between treated and control groups, there was a -16% absolute difference (95% CI -21% to -11%) for a reduction in BASFI for etanercept and similar results for the other two drugs. The number needed to treat (NNT) in order to achieve a change in BASFI of 0.7 points was 3 (95% CI 2 to 3) for infliximab and similar for etanercept and adalimumab.

The ASAS response criteria were developed for use in clinical trials and the ASAS 40 and ASAS 5/6 response criteria have both shown good discrimination in anti-TNF studies (Brandt 2004). Etanercept, infliximab and adalimumab all showed statistically significant improvements in the ASAS 40 and ASAS 5/6 responder criteria compared to placebo. Participants in these treatment groups were between 2.5 and 4.5 times more likely to have an ASAS 40 response (the ASAS 40 by 24 weeks had a 39% absolute difference (95% CI 29% to 48%) for etanercept) and 3.7 to 7.7 times more likely to have an ASAS 5/6 response. A 50% improvement in BASDAI was 2.8 to 4.6 times more likely with an anti-TNF agent. The NNT for an ASAS 40 response was 3 for etanercept and infliximab and 4 for adalimumab.

Assessment of reduction in spinal and sacroiliac inflammation using MRI was assessed in two studies, one on infliximab and one on adalimumab and both found significantly greater reductions in the anti-TNF treated group.

(Inman 2010) compared doses of 3mg/kg and 5mg/kg of infliximab and found they similar treatment responses to week 12 . In the open label phase where placebo patients switched to infliximab, and at the end of week 50, there were no significant difference between the two doses in terms of ASAS or BASDAI response. Patients who had not responded by weeks 22 or 38 (in terms of reaching a BASDAI improvement >50% or BASDAI <3) were allowed to escalate their dose to 5mg/kg. The majority (68%) of patients treated with 3mg/kg escalated their dose to 5mg/kg by week 38.

One small (N=50) head-to-head study of biologics (Giardina 2009) was found in the literature search. This unblinded RCT compared etanercept to infliximab for two years and did not find a statistically significant difference between the two anti-TNF agents at 104 weeks for efficacy outcomes of BASDAI, BASFI, ASAS 20 and ASAS 40, though there was a "more rapid clinical improvement in the infliximab treated group" up to 12 weeks.

One RCT (Braun 2008) assessed etanercept (50 mg once a week) compared to sulphasalazine. Etanercept was found to have a statistically significantly better response in terms of ASAS 20, ASAS 40, and ASAS 5/6 at 16 weeks. As well, BASDAI, BASFI, BASMI, nocturnal back pain, and Modified Schobers response were all greater in the etanercept versus the sulphasalazine groups.

Infliximab in combination with methotrexate versus placebo plus methotrexate was assessed in Marzo-Ortega 2005. At 30 weeks, there was not a statistically significant difference between the two groups in terms of ASAS 20, >50% BASDAI response or change in BASDAI, though there had been at 10 weeks. The 30 week assessment was 8 weeks after the last dose of infliximab had been given and the authors concluded that the addition of methotrexate did not confer a longer benefit of infliximab.

We included both RCTs and non-RCTs in our assessment of the adverse effects of anti-TNF therapy for ankylosing spondylitis.

The etanercept RCT data showed statistically significantly more withdrawals due to AE and injection site reactions in the etanercept group. None of the other outcomes were statistically significantly different, though the point estimates usually favoured placebo.

From the RCT data for infliximab, most outcomes assessing AE were not statistically significant. At 12 weeks, the number of serious adverse events was statistically significantly greater in infliximab compared to placebo group (Peto OR, 7.80 (1.07, 56.65) but when those two studies were pooled with 24 week data, the overall pooled

result was no longer significant (Peto OR 2.47, 95% CI 0.75 to 8.14). There was one event of tuberculosis at 12 weeks.

There was a significant increase in total adverse events and total infections in the adalimumab group compared to placebo (RR 1.25, 95% CI 1.05 to 1.49 and RR 1.58, 95% CI 1.10 to 2.27) respectively. There was no significant difference with respect to the other adverse event outcomes. Adalimumab was the only anti-TNF drug to show a significant increase in the number of infections.

We pooled adverse event data from adalimumab, etanercept, and infliximab to investigate a class effect of the TNF-inhibitors. We found a statistically increased risk of infections in the anti-TNF group versus placebo (RR 1.22, 95% CI 1.02 to 1.45). However, the risk of serious infections or serious adverse events was not statistically different between the two groups. Also significantly increased were the total number of adverse events (RR 1.24, 95% CI 1.10 to 1.40 and Peto OR 2.78, 95% CI 1.27 to 6.10, respectively), withdrawals due to adverse events, and injection or infusion site reactions.

For the specific adverse effects of interest such as serious infections, lymphoma, demyelinating diseases, hepatotoxicity, and heart failure, the RCTs did not show significant differences between the treatment and placebo groups, but that is probably because we did not have enough power to detect such a difference in these short-term studies, and also, some of the studies did not report on these outcomes.

The adverse event data from the non-RCT data is more difficult to interpret. Only open-label extension studies met our inclusion criteria and these were judged to be at a high risk of bias and lacked a control group. Some studies presented the data only for the open-label extension phase, and others for the both the RCT and open-label extension periods. As well, for some, the data is cumulative; e.g. the open-label results for week 168 include the adverse event data from the 70-week open-label results, and others report just on the open-label period; e.g. the EASIC cohort includes people who have been exposed to infliximab for almost five years, but only reports on events from the last two years of the cohort study. Lastly, some studies

report separately those who were originally randomized to placebo and those originally randomized to treatment, and other studies report them together. We do not recommend making comparisons across anti-TNF agents based on these results and any interpretations must be made with caution given the details outlined above.

In terms of serious infections and serious adverse events, in an open-label extension study on etanercept, at 168 weeks (~ 3 years) (Davis 2008), 2.3% (6/257) had a serious infection and 12.8% a serious adverse event . In an extension study from the ASSERT RCT on infliximab (Braun 2008), 4% had a serious infection and 16.9% a serious adverse event. In the EASIC (infliximab) cohort study (Heldmann 2009), data on adverse events was reported during the two-year EASIC study. 89 patients entering the study had completed ASSERT, had infliximab treatment for 1.3 years +/-0.9, and then entered EASIC. There were three serious infections during this two year period (total serious adverse events not reported). With adalimumab, there were 6/311 serious infections (1.9%) and 48/311 (15.4%) serious adverse events over approximately two years. There were no events of malignancy or lymphoma during the etanercept extension studies and the Heldmann 2009 infliximab cohort. The Braun 2008 (infliximab) and adalimumab studies had rates of about 1% for malignancies over two years.

Indirect comparisons can be useful when there is limited direct evidence for a clinically-important question. Appropriate methods should be employed to undertake indirect comparisons; for example, “naïve” pooling methods should be avoided, and in this analysis, we have undertaken three different types of suitable methodologies (ITC, GLIMM, and MTC) for indirect comparisons. The estimates were fairly robust across the three methodologies with a couple of exceptions which we were, unfortunately, unable to explain.

The direct estimates from the one head-to-head study of etanercept versus infliximab were comparable with the refined estimates obtained from the MTC analysis, indicating that the indirect comparison estimates were robust. The efficacy and safety outcomes of ASAS 40, ASAS 50, >50% BASDAI, withdrawals due to AE and serious infections were not statistically significant, so we can conclude, as the

Giardina article did, that there does not appear to be a difference between the efficacy and safety of etanercept and infliximab at the end of two years. The Giardina article did note that infliximab appeared to have a faster onset of action, but by the end of the study (two years), there was no difference in the response rates between the two groups. In our analysis, although there was no statistical significance reached, the point estimates did favour infliximab.

The new estimates obtained using the MTC analysis for the efficacy and safety outcomes of ASAS 5/6, ASAS 40, ASAS 50, >50% BASDAI, withdrawals due to AE, all infections, and serious infections for adalimumab versus etanercept and adalimumab versus infliximab were not statistically significant, therefore, one treatment does not appear to confer more benefit than the other.

Other factors must be taken into account when making decisions, including patient preferences and cost-effectiveness. The NHS technology assessment report on these three biologics for use in ankylosing spondylitis (NICE 2008) found that under various simulations, infliximab was much less cost-effective than the other two agents and therefore did not recommend its use based on this.

Overall completeness and applicability of evidence

In total, fifteen published trials addressed the use of etanercept (8 trials; one versus sulphasalazine, the rest versus placebo), infliximab (4 trials, one assessed the efficacy of infliximab in combination with methotrexate), or adalimumab (two trials), and one trial of etanercept versus infliximab, for the treatment of ankylosing spondylitis. The RCTs were between six and 24 weeks duration, so all data about efficacy is based on short-term studies.

Participants in the included studies had high disease activity (entry criteria was a BASDAI ≥4) and the length of disease duration ranged between 10 and 16 years. However, the high levels of disease activity seen in the patients included in these trials may not be typical of patients seen in daily clinical practice. In addition, patients selected for RCTs generally have few major comorbidities. Almost all studies excluded patients with complete ankylosis of the spine, and many excluded

patients with conditions related to the concerns of potential harms with biologics, i.e. recent serious infections, history of infectious diseases or malignancies in last five years, and signs of severe renal, cardiac, hepatic, demyelinating, or other diseases. This may impact the generalizability of these results to clinical practice.

The length of disease duration of the participants enrolled in these studies was mainly ten years and longer. The applicability of this evidence to those with shorter disease duration is unclear. However, in Rudwaleit 2004, a shorter disease duration (≤10 years) and a lower BASFI (< 4.5) were both found to be strong predictors of a major clinical response (as assessed by >50% BASDAI response). Therefore, it is of interest to conduct trials of anti-TNF agents in populations with early disease to determine if there is indeed a better response and any effect on the progression of ankylosis in this population and to determine when it is appropriate to start anti-TNF therapy (Haibel 2010).

Most trials used a standard dosage for each of the biologics. One trial (van der Heijde 2006b) assessed the efficacy of giving 50 mg of etanercept once a week versus 25 mg twice a week and concluded that the efficacy and safety outcomes were similar between the two groups.

Another trial (Inman 2010) evaluated the effect of a lower dosage of infliximab (3 mg/kg) compared to the standard dose of 5 mg/kg. Although a previous open-label study of lower dose infliximab showed sustained efficacy in the majority of patients, in the extension phase of this RCT (the double-blind phase was to 12 weeks), by week 38, most participants (68%) required an increase to 5 mg/kg to maintain efficacy.

A limitation of this review is that we did not include non-RCT data to assess changes in the radiological progression of patients exposed to anti-TNF therapy. It is unlikely that changes in radiological progression will be seen in RCTs of six months or less duration.

Appropriate outcomes based on OMERACT recommendations were assessed to establish short-term efficacy of the anti-TNF agents against placebo. An editorial on

the study of abatacept in rheumatoid arthritis (Boers 2006) highlights the desire of clinicians for active comparator trials, once efficacy of a treatment has been established against placebo. As described above in the summary of results section, only one head-to-head study of anti-TNF agents (Giardina 2009) was found in our systematic search. We therefore performed indirect comparisons using various methodologies to test the robustness of the results. However, additional head-to-head studies would be helpful to provide clinicians with a stronger evidence-base regarding differences between the various biologics.

We included non-randomized studies to assess important rare events which have been highlighted as concerns with the use of anti-TNF agents. The use of meta-analysis of RCTs to assess rare adverse events allows us to increase power by increasing the sample size while maintaining the benefits of randomization and a control group. However, the RCTs included in this review for assessment of efficacy were all shorter than six months duration. Our interest in events such as malignancies, heart failure, and autoimmune and demyelinating diseases, means we needed to assess patients over a longer time period. Our inclusion criteria for assessing these rare, delayed effects included a broad range of study designs, but with the restriction that the study had to have followed at least 100 people (Gartlehner 2006) for a minimum of one year. Only six articles from four studies met this criteria and they were all open-label (unblinded) extensions of RCTs, with the exception of one extension study in which the patients remained blinded to year two. The main problem with open-label extensions is that the people who enter them are highly selected in that they originally met strict RCT inclusion criteria and then did not experience adverse events that were serious enough to make them withdraw early from the trial before the extension phase began. Therefore, the results from these open-label extension studies are not generalizable to the population who are eligible for anti-TNF therapy. As well, since there is no control group, it is difficult to assess whether the results of important adverse events of interest such as serious infections, malignancies or death are much increased compared to what happens in a similar population. Studies based on larger numbers of participants in biologic registry

databases with a matched control group will be of more use to help us answer these questions.

Quality of the evidence

The Ankylosing Spondylitis International Working Group (ASAS, www.asas-group.org) and Outcome Measures in Rheumatology (OMERACT; www.omeract.org) groups have had great success in standardizing outcomes that should be measured in trials of interventions for ankylosing spondylitis. The trials included in this systematic review usually reported outcome measures as recommended by ASAS and OMERACT for trials on ankylosing spondylitis patients. There is little risk of bias due to selective reporting in these trials in terms of efficacy outcomes; adverse events were less clearly reported.

When combining studies it is important that the outcome measures are comparable. Of note, different definitions of serious adverse events were used in the assessment of these events across the trials. Most trials did not provide a specific definition; a couple stated they used the national cancer institute common toxicity criteria, and another MedDRA version 9. We assumed for the purpose of this review that the definitions were similar enough to warrant combining.

Adequate allocation concealment can avoid selection bias in controlled trials and there is evidence that inadequate allocation concealment leads to an overestimation of the treatment effect (Schultz 1995). Additional information was obtained from the authors for some studies to clarify the method of allocation. Still, there were three abstracts and five full-text studies which did not provide enough information on this domain and it had to be marked as unclear.

Blinding of participants was clearly reported in six studies; again the three abstracts did not provide enough information. Five studies were marked as unclear. Giardina 2009 was clearly not blinded; this two-year study compared etanercept (given subcutaneously) with infliximab (given as an infusion) so it is evident why the patients were not blinded. Given that the primary outcome measured in most trials was the BASDAI or ASAS 20, both self-reported measures, it is necessary for

patients to be blinded to ensure detection bias has not been introduced in these studies.

Completion rates were not reported in the three abstracts, but were greater than 80% in all but one study, Marzo-Ortega 2005, which had a large imbalance in withdrawals in the treatment and placebo arms, mostly due to inefficacy in the placebo group (93% versus 64%). Last observation carried forward data was provided for an ITT analysis - this study was marked as at 'unclear' risk of bias for the incomplete outcome domain. All the trials reported the numbers of patients who dropped out in the treatment and placebo groups. The drop-out rates were generally higher in the placebo group than the treatment group in all trials and there was a much higher rate of withdrawal due to lack of efficacy in the placebo groups. In most trials the missing data were imputed using last observation carried forward analysis for continuous data and 'non-responders' for dichotomous outcomes like ASAS 20. Most trials reported a proper intention-to-treat analysis. The other trials (Braun 2002; Inman 2010; van der Heijde 2006b) reported a 'modified' intention-to-treat analysis as one defined by those subjects who received at least one infusion of study medication. Although fewer than 3% of participants were affected, it is of interest to know why patients who were randomized did not receive the study drug.

The non-randomized studies included in this review were mostly open-label extension studies; in one extension study from a randomized double-blind period, patients continued to be blinded. All but one were judged to be at a high risk of bias; the other was moderate because patients continued to be blinded during the extension phase. The conclusions from open-label extension studies lead us to believe that they are providing us with more confidence about the rates of important adverse events, but given that there is a high selection bias, lack of blinding of participants, and no control group, it is difficult to assess what meaningful information these studies can provide us (Day 2007). As well, Taylor 2005, found that there was evidence that many open-label extension are never published so we should be concerned about publication bias of these types of studies. Therefore, the evidence we found on

adverse events from a broad search of the literature for non-randomized studies is at a high risk of bias.

Only two outcomes had enough studies in which to visually inspect a funnel plot for publication bias. These plots indicated a potential lack of publication of smaller, 'negative' studies but overall we do not have much evidence whether publication bias is an issue in this systematic review.

With regards to detecting adverse events in RCTs, an interesting paper (Yazici 2008) recently highlighted the fact that an inadequate sample size (Type II error) is a possible reason that a significant difference in the number of adverse events between treatment and placebo groups is often not observed. All the included studies termed themselves 'efficacy and safety' studies. But in none of the trials was there a discussion of necessary sample sizes to detect adverse events.

We concluded that there is 'high' level evidence for the short-term efficacy outcomes of ASAS 40 response, achievement of partial remission, physical function, and disease activity. Longer-term adverse event data was judged to be of 'low' level quality as it was based on observational data (extensions of RCTs).

Potential biases in the review process

We undertook a systematic, thorough search of the electronic literature and hand searched key conference proceedings to identify all studies meeting the inclusion criteria for this review. However, we did not approach pharmaceutical companies for additional data and it is possible that additional data from this source could contribute to this review. Study selection and data extraction were done in duplicate and independently and we reached consensus by discussing any discrepancies.

We conducted a specific search for *a priori* specified adverse events identified to be of concern with the use of anti-TNF agents and included study designs other than RCTs to identify evidence for rare and delayed adverse events. We limited our inclusion criteria for non-RCTs to studies that followed a minimum of 100 people for at least one year. A previous systematic review assessing adverse events from anti-

TNF use in rheumatoid arthritis also used a minimum sample size of 100 (Gartlehner 2006). While the choice of this particular sample size and follow-up period is rather arbitrary, we wanted to have an adequate sample size to find these rare events and an appropriate follow-up time to find potential delayed adverse events.

The Fouque-Aubert review performed a power calculation and suggests that, "According to the differences in risks of serious infections assessed in the meta-analysis, for the difference of risks to become significant, it would be necessary to add in the meta-analysis one balanced study with a total sample size of 5,500 patients." The EMEA guidance document on 'Clinical investigation of medicinal products for the treatment of ankylosing spondylitis' (EMEA 2009 pg.11) states, "AS is a prevalent chronic disease and medicinal products will need to be approved for long term treatment. The need for specific long term trials to demonstrate efficacy and the effect on structural changes may provide an adequate safety data base, likely to go beyond the minimum requirement of the CHMP/ICH guidance requesting 300-600 patients treated for 6 months or 100 patients treated for 12 months. Although this may depend on the characteristics of the product, in general, safety data from periods longer than one year are recommended." Therefore, we feel that the choice of a minimum sample size of 100 participants for a minimum of one year follow-up is justified. However, further evidence based on large registry data would be most useful. Unfortunately, none of the articles based on registry data that we screened provided data separately on ankylosing spondylitis patients for a minimum of 12 months.

For the assessment of the risk of bias of non-RCTs included in this review for the assessment of adverse events, we used the Newcastle-Ottawa Scale (NOS) for cohorts (Wells). It is designed to assess a controlled cohort, but our studies were a single cohort. However, we feel confident that the critical risk of bias criteria needed to assess such a study were captured in our use of the NOS.

Published trial reports did not provide enough details to adequately assess risk of bias and some variance measures necessary for meta-analysis were missing from the

report. We contacted some authors for further information and while some of the requested data was provided, it is a limitation of this review that not all the data were available. We had to undertake transformations and assumptions in order to enter continuous data into RevMan for some key trials (ATLAS, ASSERT) which may reduce the accuracy of our estimates.

Some adverse event outcomes consisted of sparse data, with few events in the groups. In cases where either there were no events in either study arm (zero total event studies), the study did not contribute to the meta-analysis. The rationale for this method is that no information on the magnitude of the treatment effect can be obtained from these studies. An investigation by Sweeting 2004 confirmed that "zero total event studies do not contribute to a fixed meta-analysis" and so we felt it appropriate not to include these in our meta-analysis. The Peto OR was shown to be one of the least biased estimators of a treatment effect when using sparse data. However, in the case of unbalanced study arms with four times as many participants in one arm as another, there is concern that the estimate will be biased (Bradburn 2007). Given that none of our included studies were unbalanced by more than 4:1, we decided the Peto OR was a suitable method. When analyzing rare data in a meta-analysis, it is recommended that various methods are used to determine the treatment effect estimate and sensitivity analyses are undertaken (Sweeting 2004). For those studies where there were zero events in one arm, or an event rate <10%, we used the Peto OR. We also performed the Mantel-Haenszel method with the standard continuity correction of 0.5 to check the robustness of our results and found that results did not change significantly. We did not perform other sensitivity analyses with other continuity corrections, but given that there was consistency of results with the standard continuity correction, we are confident that our estimates are robust.

For the indirect comparison results to be considered robust, it must be investigated whether the trials included in each standard meta-analysis comparison (e.g. A-B and C-B) are homogeneous; whether the effect of the linking treatment is similar across comparisons; and whether there is consistency of evidence. To judge the likelihood of these assumptions being true, the comparability of the linking treatment, the

patient population, the methodological risk of bias, the study design, and the date of publication must be assessed (Song 2009; Wells 2009 p.90). We did assess the studies included in terms of the criteria noted above. As noted in the Results section, the patient populations were quite similar in terms of length of disease duration, severity of disease, inclusion/exclusion trial criteria, trial duration and date of publication (2002 to 2008). The one head-to head study of etanercept versus infliximab (Giardina 2009) did not blind the participants (given the difference in route of administration of the two biologics) and it was measured at a longer follow-up time than the other included studies (2 years versus 6 months). However, since the treatment effect observed for etanercept and infliximab was similar to that seen in other placebo-controlled studies, and including this study allowed us to have a closed loop in our network meta-analysis for evaluating the consistency of the direct and indirect evidence, we decided it was appropriate to include this in our indirect comparison analysis.

However, it must be cautioned that while we assessed the necessary assumptions for undertaking indirect comparisons, and performed three methods (GLIMMIX, MTC, and ITC) to test the robustness of our results, indirect comparisons may not provide the same strength of validity of results that a well-conducted, head-to-head, RCT may.

A protocol was published for this review (Zochling 2005) and analyses were specified a priori. However, the original published Cochrane protocol did not describe the detailed assessment of adverse effects beyond the inclusion of RCTs. This was outlined in the thesis proposal.

Agreements and disagreements with other studies or reviews

The National Institute for Health and Clinical Excellence (NICE) undertook a technology appraisal report to provide guidance on the use of adalimumab, etanercept and infliximab for ankylosing spondylitis for the UK National Health Service (NICE 2008). The report was issued in May 2008 and the results are very similar to this review. This Cochrane review includes some newer information on the

efficacy of low-dose infliximab and use of 50mg once per week of etanercept versus 25 mg twice per week. We included non-randomized evidence for assessment of longer-term adverse events while the NICE report focused on evidence from RCTs. The NICE report also conducted indirect comparisons of the anti-TNF inhibitors and as this review did, found no statistically significant difference between the three biologics in terms of ASAS response rates.

The NICE review assessed cost-effectiveness of the three biologics in detail and recommended that the high cost of infliximab precluded its recommendation for use in people with ankylosing spondylitis. Both adalimumab and etanercept were recommended for use.

Another systematic review and meta-analysis was recently performed to assess the risk of serious infections (Fouque-Aubert 2009) in patients taking anti-TNFagents and compared them to estimates of risk in groups of patients taking placebo and NSAIDs. All the included studies in this review were also included in the Fouque-Aubert review. The Fouque-Aubert review reached the same conclusion as this one: there does not appear to be a statistically significant increase in serious infections in ankylosing spondylitis patients exposed to anti-TNF agents. The absolute and relative estimates from both reviews were not statistically significant (RD 0.00, 95% CI, -0.01, 0.02 and Peto OR 1.89, 95% CI 0.45, 7.96). The Fouque-Aubert review outlined three possible reasons for the lower incidence of serious infections in AS patients compared to those found in rheumatoid arthritis patients on anti-TNF therapy: i) the younger average age of ankylosing spondylitis patients means fewer co-morbidities; ii) the pathophysiology of the two diseases are different with RA characterized by inflammatory synovitis and AS characterized by new bone formation along with inflammation; and iii) the different concomitant treatments, NSAIDs in AS patients versus steroids with RA patients, making RA patients more susceptible to infections.

Of interest, a recent study based on the British Society for Rheumatology Biologic Register (BSRBR) assessed the use of anti-TNF therapy for ankylosing spondylitis in the UK in routine care. They found the mean improvement in BASDAI was 3.6 (0 to

10 scale), 52% of patients achieved a BASDAI >50 and concluded that "routine clinical use improves disease activity and functional impairment in patients with AS" (Lord 2010).

Authors' conclusions

Implications for practice

That there is high level evidence for short-term efficacy of adalimumab, etanercept, and infliximab compared to placebo in improving disease activity and function and achieving partial remission in ankylosing spondylitis. The addition of methotrexate to infliximab does not appear to confer any additional benefit. There was no significant difference in efficacy or adverse effect outcomes using network meta-analysis to model head-to-head studies, nor did one RCT find a difference when etanercept and infliximab were compared directly. Based on RCTs, there was a statistically significantly greater total number of adverse events, withdrawals due to adverse events, injection or infusion site reactions, and total number of infections when the three TNF-inhibitors were pooled versus placebo. No statistically significant increase in serious infections or serious adverse events were found in RCTs, however the number of total events was low. Given the high level of bias in the extension studies included in this review, we were unable to draw conclusions on the long-term adverse effect data, but there did not appear to be large signal of key safety outcomes such as serious infections and malignancies, though malignancies may only emerge years later. We did not assess evidence on the long-term efficacy or effectiveness of anti-TNF therapy for ankylosing spondylitis.

Implications for research

Future trials of anti-TNF agents should focus on populations with early disease to determine if there is indeed a better response and any effect on the progression of ankylosis in this population and to determine when it is appropriate to start anti-TNF therapy. Given the paucity of head-to-head study data we performed a network meta-analysis to estimate effects of one anti-TNF agent versus another; however, we recommend that well-conducted randomized trials of anti-TNF agents versus each

other are needed to clearly address issues of effectiveness among these agents. Studies based on biologic registry data on ankylosing spondylitis patients will be useful to assess rare and delayed adverse effects.

Acknowledgements

Many thanks to Louise Falzon, Jessie McGowan and Nancy Santesso for their assistance with the literature searching and to Elizabeth Ghogomu for her assistance with the summary of findings table. Thanks to Bharbhoor Dhaliwal for assistance with screening the searches. Thank you to the following authors of primary trials who provided us with additional information: John Davis and Desiree van der Heijde.

Contributions of authors

LM conceptualized the idea, performed screening and data extraction, entered data into RevMan, performed analyses, and wrote the review
JZ conceptualized the idea, performed screening and data extraction, entered data into RevMan, and commented on drafts of the review
AB provided clinical expertise and commented on drafts of the review
GW conceptualized the idea, provided statistical expertise, and commented on drafts of the review
PT conceptualized the idea, provided clinical expertise, and commented on drafts of the review

Declarations of interest

LM: none known

AB: research grants: Amgen, Pfizer, Abbott; consultant fees: Abbott, Pfizer

JZ:

JS: speaker honoraria from Abbott; research grants from AMGEN, Allergan, Takeda, Savient; consultant fee from Savient, URL Pharma, Novartis

GW - research grant and consultant fee from Bristol-Myers Squibb and Abbott Inc.

PT - grants/honoraria from Bristol Myers, Chiltern International, and UCB

Differences between protocol and review

- outcomes for specific AE more clearly defined

- included other study designs beyond RCTs to assess adverse effects

- included indirect comparisons

References

Included studies

Barkham 2008

Barkham N, Coates L, Keen H, Hensor E, Fraser A, Redmond A, et al. A double blind placebo controlled trial of etanercept in the prevention of work disability in ankylosing spondylitis. Rheumatology 2008;47:ii72.

Brandt 2003

Published and unpublished data

* Brandt J, Khariouzov A, Listing J, Haibel H, Sorensen H, Grassnickel L. Six-month results of a double-blind, placebo-controlled trial of etanercept treatment in patients with active ankylosing spondylitis. Arthritis & Rheumatism 2003;48:1667-75.

Braun 2002

Published and unpublished data

* Braun J, Brandt J, Listing J, Zink A, Alten R, Krause A, et al. Treatment of active ankylosing spondylitis with infliximab, a double-blind placebo controlled multicenter trial. Lancet 2002;359:1187-93.

Braun 2008

* Braun J, Burgos-Vargas R, van der Horst-Bruinsma I.E., Freundlich B, Vlahos B, Koenig A.S. et al. Assessment of clinical efficacy in a randomized, double-blind study of etanercept and sulphasalazine in patients with ankylosing spondylitis. In: Arthritis and Rheumatism. Vol. 58. 2008:S415.

Sieper J, Moots R, Taylor A, Sato R, Singh A, Roberson D, et al. Assessment of Patient-Reported Outcomes from a randomized, double-blind study of etanercept and sulphasalazine in patients with ankylosing spondylitis. In: Arthritis and Rheumatism. Vol. 58 (9). 2008:S580.

Calin 2004

Published and unpublished data

* Calin A, Dijkmans B, Emery P, Hakala M, Kalden J, Leirsalo-Repo M et al. Outcomes of a multicentre randomised clinical trial of etanercept to treat ankylosing spondylitis. Annals of the rheumatic diseases 2004;63:1595-1600.

Davis 2003

Published data only (unpublished sought but not used)

* Davis J, van der Heijde D, Braun J, Dougados M, Cush J, Clegg D et al. Recombinant Human Tumor Necrosis Factor Receptor (Etanercept) for Treating Ankylosing Spondylitis. Arthritis & Rheumatism 2003;48(11):3230-6.

Davis J, van der Heijde, Braun J, Dougados M, Clegg D.O., Kivits A, et al. Effiacy and safety of up to 192 weeks of etanercept therapy in patients with ankylosing spondylitis. In: Annals of the rheumatic diseases. Vol. 67. 2008:346-52.

Davis JC, van der Heijde DM, Braun J, Dougados M, Cush J, Clegg D, Inman RD, Kivitz A, Zhou L, Solinger A, Tsuji W.. Sustained durability and tolerability of

etanercept in ankylosing spondylitis for 96 weeks. Annals of the Rheumatic Diseases 2005;64:1557-1562.
Giardina 2009
* Giardina AR, Ferrante A, Ciccia F, Impastato R, Miceli M, Principato A, et al. A 2-year comparative open label randomized study of efficacy and safety of etanercept and infliximab in patients with ankylosing spondylitis. Rheumatology International 2009;online e-pub. [DOI: 10.1007/s00296-009-1157-3]
Impastato R, Ciccia F, Ferrante A, Principato A, Giardina A, Cadelo M, et al. Infliximab and etanercept are both effective and safe in patients with ankylosing spondylitis. A two-year randomised study. Annals of the rheumatic diseases 2007;66(Suppl II):399.
Gorman 2002
* Gorman JD, Sack KE, Davis JC. Treatment of ankylosing spondylitis by inhibition of tumor necrosis factor alpha. New England Journal of Medicine 2002;346:1349-56.
Huang 2008
Huang F, MacPeek D, Shunle C, Li Z, Gu J, Wu D, et al. Etanercept is effective and well tolerated in Chinese patients with ankylosing spondylitis: 6-week placebo-controlled trial followed by 6-week open-label treatment. In: Annals of the Rheumatic Diseases. Vol. 67. 2008:Suppl II: 521.
Inman 2010
* Inman R, Maksymowych W, for the CANDLE group. A double-blind, placebo-controlled trial of low dose infliximab in ankylosing spondylitis. Journal of Rheumatology 2010;37(6):1-8. [DOI: doi:10.3899/jrheum.091042]
Inman R, Shojania K, Keystone E, Haraoui B, Mitchell C, Maksymowych W. CANaDian evaluation of low DosE infliximab in ankylosing spondylitis (CANDLE). Journal of Rheumatology 2008; 35(6):1213-14.
Maksymowych W, Salonen D, Inman R, Paul F.L, Lambert R.G. CANaDian evaluation of Low DosE Infliximab in Ankylosing Spondylitis (CANDLE) - 12 week Magnetic Resonance Imaging Evaluation of Spinal Inflammation with the SPARCC MRI Method. In: Journal of Rheumatology. Vol. 35. 2008:1211-12.
Lambert 2007
* Lambert RGW, Salonen D, Rahman P, Inman RD, Wong RL, Einstein SG, et al. Adalimumab significantly reduces both spinal and sacroiliac joint inflammation in patients with ankylosing spondylitis. Arthritis and rheumatism 2007;56(12):4005-14.
Maksymowych W, Rahman P, Keystone E, Wong R, Inman R for the M03-606 Study Group. Efficacy of adalimumab in active AS: Results from the Canadian AS study. In: Arthritis and Rheumatism. 2005:S217.
Maksymowych W, Rahman P, Shojania K, Olszynski W, Thomson G, Ballal S, et al. Beneficial effects of adalimumab on biomarkers reflecting structural damage in patients with ankylosing spondylitis. Journal of Rheumatology 2008;35(10):2030-7.

Marzo-Ortega 2005

Marzo-Ortega H, McGonagle D, Jarrett S, Haugeberg G, Hensor E, O'Connor P, et al. Infliximab in combination with methotrexate in active ankylosing spondylitis: A clinical and imaging study. Annals of the Rheumatic Diseases 2005;64(11):1568-75.

van der Heijde 2005

Braun J, Deodhar A, Dijkmans B, Geusens P, Sieper J, Williamson P et al. Efficacy and safety of infliximab in patients with ankylosing spondylitis over a two-year period. Arthritis & Rheumatism 2008;59:1270-8.

Braun J, Landewe R, Hermann KGA, Han J, Yan S, Williamson P, et al. Major reduction in spinal inflammation in patients with ankylosing spondylitis after treatment with infliximab - Results of a multicenter, randomized, double-blind, placebo-controlled magnetic resonance imaging study. Arthritis and rheumatism 2006;54(5):1646-52.

Heldmann F, Brandt J, Braun J, on behalf of EASIC Study Group. European ankylosing spondylitis infliximab cohort (EASIC): safety of longerm therapy with infliximab in patients with anklyosing spondylitis. In: Annals of the rheumatic diseases. Vol. 67. 2008:Suppl II:386.

Heldmann F, Brandt J, van der Horst-Bruinsma, Landewe R, Sieper J, Burmester G et al. European Ankylosing Spondylitis Infliximab Cohort (EASIC): Safety of longterm therapy in patients with ankylosing spondylitis. In: Annals of the Rheumatic Diseases. Vol. 68 (Supple 3). 2009:626.

* van der Heijde D, Dijkmans B, Geusens P, Sieper J, DeWoody K, Williamson P et al. Efficacy and Safety of Infliximab in Patients with Ankylosing Spondylitis. Results of a Randomized, Placebo-Controlled Trial (ASSERT). Arthritis & Rheumatism 2005;52(2):582-91.

van der Heijde D, Han C, Bala M, Williamson P, Han J, Braun J. Infliximab improves fatigue and pain in patients with ankylosing spondylitis: Results of a randomized, placebo-controlled trial (ASSERT). In: Annals of the Rheumatic Diseases. Vol. 64. 2005:318-9.

van der Heijde D, Han C, Bala M, Williamson P, Han J, Braun J. Infliximab improves fatigue and pain in patients with ankylosing spondylitis: Results of a randomized, placebo-controlled trial. In: Arthritis and Rheumatism. Vol. 52 (9). 2005:s35.

van der Heijde 2006a

Published and unpublished data

Davis Jr.JC, Revicki D, van der Heijde DMF, Rentz AM, Wong RL, Kupper H, et al. Health-related quality of life outcomes in patients with active ankylosing spondylitis treated with adalimumab: Results from a randomized controlled study. Arthritis Care & Research 2007;57(6):1050-7. [DOI: 10.1002/art.22887]

Revicki DA, Luo MP, Wordsworth P, Wong RL, Chen N, Davis Jr J. Adalimumab reduces pain, fatigue, and stiffness in patients with ankylosing spondylitis: Results from the adalimumab trial evaluating long-term safety and efficacy for ankylosing spondylitis (ATLAS). Journal of Rheumatology 2008;35(7):1346-53.
van der Heijde D, Dijkmans B, Schiff M, Kivitz A, de Vlam K, Wordsworth P et al. Maintenance of reduction in disease activity and partial remission in patients with ankylosing spondylitis (AS) - 3 year results from the adalimumab (Humira) trial evaluating long-term efficacy and safety in AS (ATLAS). In: Arthritis and Rheumatism. 2008:S579.
* van der Heijde D, Kivitz A, Schiff MH, Sieper J, Dijkmans BAC, Braun J, Dougados M, Reveille JD, Wong RL, Kupper H, Davis JC Jr, for the ATLAS study group. Efficacy and safety of adalimumab in patients with ankylosing spondylitis. Arthritis & Rheumatism 2006;54(7):2136-2146.
van der Heijde D, Schiff M, Sieper J, Kivitz A, Wong R, Kupper H, et al. Adalimumab effectiveness for the treatment of ankylosing spondylitis is maintained for up to 2 years: long-term results from the ATLAS trial. Annals of the Rheumatic Diseases 2009;68:922-9.
van der Heijde 2006b
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 2007;46(6):999-1004.
* van der Heijde D, Da Silva JC, Dougados M, Geher P, Van DH-B, Juanola X, et al. Etanercept 50 mg once weekly is as effective as 25 mg twice weekly in patients with ankylosing spondylitis. Annals of the Rheumatic Diseases 2006;65(12):1572-7.
Excluded studies
Barkham 2008b
Barkham N, Keen HI, Coates P, O'Connor EMA, Hensor AD, Fraser LS, et al. A randomized controlled trial of infliximab shows clinical and MRI efficacy in HLA B27 positive very early ankylosing spondylitis. In: Annals of the Rheumatic Diseases. Vol. 67. 2008:Suppl II:382.
Breban 2008
Breban M, Ravaud P, Claudepierre P, Baron G, Henry Y-D, Hudry C, et al. Maintenance of infliximab treatment in ankylosing spondylitis: Results of a one-year randomized controlled trial comparing systematic versus on-demand treatment. Arthritis & Rheumatism 2008;58(1):88-97. [DOI: 10.1002/art.23167]
Haibel 2008
Haibel H, Rudwaleit M, Listing J, Heldmann F, Wong RL, Kupper H, et al. Efficacy of adalimumab in the treatment of axial spondylarthritis without radiographically defined sacroiliitis: Results of a twelve-week randomized, double-blind, placebo-

controlled trial followed by an open-label extension up to week fifty-two. Arthritis & Rheumatism 2008;58(7):1981-91. [DOI: 10.1002/art.23606]
Hamza 2007-NRS
Hamza S, Chon Y, Hooper M, MacPeek D, Lin S. Rates of serious infectious events and opportunistic infections in patients with rheumatoid arthritis, psoriatic arthritis, and anklyosing spondylitis treated with etanercept or placebo in clinical trials. In: Annals of the rheumatic diseases. Vol. 66. 2007:Suppl II: 171.
Li 2008
Li EK, Griffth JF, Lee VW, Wang YX, Li TK, Lee KK et al. Short-term effiacy of combination methotrexate and infliximab in patients with ankylosing spondylitis: a clinical and magnetic resonance imaging correlation. Rheumatology 2008;47(9):1358-63.
Van den Bosch 2002
Kruithof E, Van den Bosch F, Baeten D, Herssens A, De Keyser F, Mielants H, Veys EM. Repeated infusions of infliximab, a chimeric anti-TNF alpha monoclonal antibody, in patients with active spondyloarthropathy: one year follow up. Annals of Rheumatic Diseases 2002;61:207-212.
* Van den Bosch F, Kruithof E, Baeten D, Herssens A, de Keyser F, Mielants H, Veys EM. Randomized double-blind comparison of chimeric monoclonal antibody to tumor necrosis factor alpha (infliximab) versus placebo in active spondyloarthropathy. Arthritis and Rheumatism 2002;46(3):755-765.
Visvanathan 2008
Visvanathan S, Wagner C, Marini JC, Baker D, Gathany T, Han J, et al. Inflammatory biomarkers, disease activity and spinal disease measures in patients with ankylosing spondylitis after treatment with infliximab. Annals of the Rheumatic Diseases 2008;67(4):511-7.
Studies awaiting classification
Zhang 2009
Zhang J., Zhang YM., Zhang J.-L.,Deng X.-H., Huang,F. Efficacy of etanercept in patients with ankylosing spondylitis: A double-blind, randomized, placebo controlled trial. [Chinese]. Chinese Journal of New Drugs 2009;18(19):1846.
Zhou 2009
Zhao F-T, Zhao H.,Wang Y-L. Efficacy of etanercept on ankylosing spondylitis. [Chinese]. Journal of Shanghai Jiaotong University (Medical Science) 2009;29(12):1506-8.
Ongoing studies
Other references
Additional references
Anderson 2001

Anderson JJ, Baron G, van der Heijde D, Felson DT, Dougados M. Ankylosing Spondylitis Assessment group preliminary definition of short-term improvement in ankylosing spondylitis. Arthritis and Rheumatism 2001;44(8):1876-1886.
Boers 2006
Boers M. Abatacept in rheumatoid arthritis: a new branch on the biologics tree (editorial). Annals of the Rheumatic Diseases 2006;144:933-5.
Boonen 2001
Boonen A, Chorus A, Miedema H, van der Heidje D, van der Temple H, van der Linden S. Employment, work disability, and work days lost in patients with ankylosing spondylitis: a cross sectional study of Dutch patients. Annals of Rheumatic Diseases 2001;60:353-8.
Boonen 2001a
Boonen A, Chorus A, Miedema H, van der Heijde D, Landewé R, Schouten H et al. Withdrawal from labour force due to work disability in patients with ankylosing spondylitis. Annals of Rheumatic Diseases 2001;60:1033-9.
Bradburn 2007
Bradburn MJ, Deeks JJ, Berlin JA, Russell Localio A. Much ado about nothing: a comparison of the performance of meta-analytical methods with rare events. Statistics in Medicine 2007;26(1):53-77.
Brandt 2000
Brandt J, Haibel H, Cornley D et al. Successful treatment of active ankylosing spondylitis with the anti-tumor necrosis factor alpha monoclonal antibody infliximab. Arthritis and Rheumatism 2000;43:1346-52.
Brandt 2004
Brandt J, Listing J, Sieper J, Rudwaleit M, van der Heijde D, Braun J. Development and preselection of criteria for short term improvement after anti-TNF alpha treatment in ankylosing spondylitis. Ann Rheum Dis 2004;63(11):1438-1444.
Braun 1995
Braun J, Bollow M, Neure L, Seipelt E, Seyrekbasan F, Herbst H et al. Use of immunohistologic and in situ hybridization techniques in the examination of sacroiliac joint biopsy speciments from patients with ankylosing spondylitis. Arthritis Rheum 1995;38:499.
Braun 2003
Braun J, Pham T, Sieper J, Davis J, van der Linden S, Dougados M et al. International ASAS consensus statement for the use of anti-tumor necrosis factor agents in patients with ankylosing spondylitis. Annals of Rheumatic Diseases 2003;62:817-24.
Bucher 1997

Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. Journal of Clinical Epidemiology 1997;50(6):683-91.
Cates 2003
EBM website www.nntonline.com [Computer program]. Dr. Chris Cates. Chris Cates, 2003.
Dagfinrud 2004
Dagfinrud H, Kvien T K, Hagen K B. Physiotherapy interventions for ankylosing spondylitis. The Cochrane Database of Systematic Reviews 2004, Issue 4.
Day 2007
Day R.O, Williams K.M. Open-Label Extension Studies: Do They Provide Meaningful Information on the Safety of New Drugs? Drug Safety 2007;30(2):93-105.
Dias 2010
Dias S, Welton N.J. , Caldwell D.M. Ades A.E. Checking consistency in mixed treatment comparison meta-analysis. Statistics in Medicine 2010;29:932-44. [DOI: DOI: 10.1002/sim.3767]
Dougados 2002
Dougados M, Kijkmans B, Khan M, Maksymowych W, van der Linden S, Brandt J. Conventional treatments for ankylosing spondylitis. Annals of Rheumatic Diseases 2002;61 (Suppl iii):iii40-iii50.
Edmunds 1991
Edmunds L, Elswood J, Kennedy LG, Calin A. Primary ankylosing spondylitis, psoriatic and enteropathic spondyloarthropathy: a controlled analysis.. J Rheumatol 1991;18(5):696-698.
Egger 2001
Egger M, Smith GD, Altman D eds. Systematic Reviews in Health Care. BMJ Books, 2001.
EMEA 2005
European Medicines Agency: Committee for medicinal products for human use (CHMP). (DRAFT) Guideline on clinical investigation of medicinal products for the treatment of ankylosing spondylitis
http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003425.pdf;2005.
EMEA 2009
European Medicines Agency: Committee for medicinal products for human use (CHMP). Guideline on clinical investigation of medicinal products for the treatment of ankylosing spondylitis
http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003424.pdf;2009.

Fouque-Aubert 2009
Fouque-Aubert A, Jette-Paulin L, Combescure C, Basch A, Tebib J, et al. Serious infections in patients with ankylosing spondylitis with and without TNF blockers: A systematic review and meta-analysis of randomized placebo-controlled trials. Annals of the Rheumatic Diseases 2009;e-pub published online 28 Jul 2009. [DOI: 10.1136/ard.2008.098822]
Gartlehner 2006
Gartlehner G, Hansen RA, Jonas BL, Thieda P, Lohr KN. The Comparative Efficacy and Safety of Biologics forthe Treatment of Rheumatoid Arthritis: A Systematic Review and Metaanalysis. Journal of Rheumatology 2006;33:2398–408.
Haibel 2004
Haibel H, Brandt HC, Rudwaleit M, Listing J, Braun J, Kupper H, Sieper J.. Efficacy and safety of adalimumab in the treatment of active ankylosing spondylitis: Preliminary results of an open-label, 20-week trial. Arthritis Rheum 2004;50(Suppl):S217.
Haibel 2010
Haibel H, Sieper J. Editorial review: how early should ankylosing spondylitis be treated with a tumor necrosis factor-blocker? Current opinion in rheumatology 2010;22:388-92.
Higgins 2003
Higgins J, Thompson S, Deeks J, Altman D. Measuring inconsistency in meta-analysis. BMJ 2003;327:557-60.
Higgins 2008
Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.0 [updated February 2008].. The Cochrane Collaboration (Available from www.cochrane-handbook.org) 2008.
Hochberg 2003
Hochberg M C, Tracy J K, Hawkins-Holt M, Flores R H. Comparison of the efficacy of the tumour necrosis factor {alpha} blocking agents adalimumab, etanercept, and infliximab when added to methotrexate in patients with active rheumatoid arthritis. Annals of the Rheumatic Diseases 2003;62(Suppl 2):ii13-ii16.
Khan 2001
Khan Khalid S, Riet Gerben ter, Popay Jennie, Nixon John, Kleijnen Jos. Centre for Reviews and Dissemination (CRD) guidance on systematic reviews: STAGE II Conducting the review; PHASE 5: Study quality assessment. http://www.medicine.mcgill.ca/rtamblyn/Readings%5CUK%20Guide%20-%20Study%20Quality%20Assessment.pdf (Accessed April 19, 2010).
Khan 2002
Khan M. Update on Spondyloarthropathies. Annals of Internal Medicine 2002;136:896-907.

Lee 2005
Lee S.J, Kavanaugh A. Biologic agents in rheumatology: safety considerations. Rheumatic disease clinics of North America 2005;Supplement 1:3-10.
Lord 2010
Lord P.A.C, Farragher T.M., Lunt M., Watson K.D, Symmons D.P.M, Hyrich K.L. Predictors of response to anti-TNF therapy in ankylosing spondylitis: results from the British Society for Rheumatology Biologcs Register. Rheumatology 2010;49:563-70.
Maksymowych 2002
Maksymowych W, Jhangri G, Lambert R, Mallon C, Buenviaje H, Pedryca E et al. Infliximab in ankylosing spondylitis: a prospective observational inception cohort analysis of efficacy and safety. Journal of Rheumatology 2002;29:959-65.
Maksymowych 2003
Maksymowych W, Inman R, Gladman D, Thompson G, Stone M, Karsh J et al. Canadian Rheumatolgy Association Consensus on the use of anti-tumor necrosis factor-alpha directed therapies in the treatment of spondyloarthritis. Journal of Rheumatology 2003;30:1356-63.
Marzo-Ortega 2001
Marzo-Ortega H, McGonagle D, O'Connor P, et al. Efficacy of etanercept in the treatment of the entheseal pathology in resistant spondylarthropathy: a clinical and magnetic resonance imaging study. Arthritis and Rheumatism 2001;44:2112-7.
NICE 2008
National Institute for Health and Clinical Excellence. Adalimumab, etanercept and infliximab for ankylosing spondylitis. NICE technology appraisal guidance 143. www.nice.org.uk/TA143 (accessed January 2010) May 2008.
Rudwaleit 2004
Rudwaleit M, Listing J, Brandt J, Braun J, Sieper J. Prediction of a major clinical response (BASDAI 50) to TNF- blockers in ankylosing spondylitis? Annals of the rheumatic diseases 2004;63:665-70.
Schultz 1995
Schultz K, Chalmers I, Hayes R, Altman D. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 1995;273:408-12.
Song 2003
Song F, Altman DG, Glenny AM, Deeks JJ. Validity of indirect comparison for estimating efficacy of competing interventions:empirical evidence from published meta-analyses. BMJ 2003;326:472-5.
Song 2009
Song F, Loke YK, Walsh T, Glenny A-M, Eastwood AJ, Altman DG. Methodological problems in the use of indirect comparisons for evaluating

healthcare interventions: survey of published systematic reviews. BMJ 2009;338:b1147. [DOI: 10.1136/bmj.b1147]
Stone 2001
Stone M, Salonen D, Lax M, Payne U, Lapp V, Inman R. Clinical and imaging correlates of response to treatment with infliximab in patients with ankylosing spondylitis. Journal of Rheumatology 2001;28:1605-14.
Sweeting 2004
Sweeting M.J, Sutton A.J, Lambert P.C. What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. Statistics in Medicine 2004;23:1351-75.
Taylor 2005
Taylor GJ, Wainwright P. Open label extension studies: research or marketing? BMJ 2005;331:572-4.
Tugwell 2004
Tugwell P, Shea B, Boers M, Brooks P, Simon L, Strand V, Wells G (editors). Evidence Based Rheumatology. BMJ Books, 2004.
Van Den Bosch 2002
Van Den Bosch F, Kruithof E, Baeten D, Herssens A, de Keyser F, Mielants H et al. Randomized double-blind comparison of chimeric monoclonal antibody to tumor necrosis factor alpha (infliximab) versuss placebo in active spondylarthropathy. Arthritis and Rheumatism 2002;46:755-65.
van der Heijde 1997
van der Heijde D, Bellamy N, Calin A, Dougados M, Khan MA, van der Linden S. Preliminary core sets for endpoints in ankylosing spondylitis. Journal of Rheumatology 1997;24:2225-9.
Verhagen 1998
Verhagen A, de Vet H, de Bie R, Kessels A, Boers M, Bouter L et al. The Delphi List: A criteria list for quality assesment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. Journal of Clinical Epidemiology 1998;51(12):1235-41.
Wanders 2005
Wanders A, van der Heijde D, Landewe R, Behier JM, Calin A, Olivieri I, Ziedler H, Dougados M. Nonsteroidal antiinflammatory drugs reduce radiographic progression in patients with ankylosing spondylitis.. Arthritis and Rheumatism 2005;in press.
Wells
Wells GA , Shea B, O'Connell D, Peterson J, Welch V, Losos M, Tugwell P. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. http://www.ohri.ca/programs/clinical_epidemiology/oxford.htm (accessed April 20, 2010).
Wells 2009

Wells GA, Sultan SA, Chen L, Khan M, Coyle D. Indirect Evidence: Indirect Treatment Comparisons in Meta-Analysis. Ottawa: Canadian Agency for Drugs and Technologies in Health 2009.

Yazici 2008

Yazici Y, Adler NM, Yazici H. Most tumour necrosis factor inhibitor trials in rheumatology are undeservedly called 'efficacy and safety' trials: a survey of power considerations. Rheumatology 2008;47(7):1054-7.

Zink 2000

Zink A, Braun J, Listing J, Wollenhaupt J. Disability and handicap in rheumatoid arthritis and ankylosing spondylitis - results from the German rheumatological database. Journal of Rheumatology 2000;27:613-32.

Zochling 2005

Zochling J, Maxwell L, Beardmore J, Boonen A. TNF-alpha inhibitors for ankylosing spondylitis [Protocol]. Cochrane Database of Systematic Reviews 2005, Issue 3. Art. No.: CD005468. DOI: 10.1002/14651858.CD005468 .

Tables and Figures
Characteristics of studies

Table 1: Characteristics of included studies

Barkham 2008

Methods	Reported as an abstract from a conference proceeding. "Double blind, placebo controlled trial". States that participants were randomized to etanercept or placebo in a 1 to 1 ratio.
Participants	N=40. 32 (80%) male; mean age 40.1 years (range 20-61 years); mean duration of symptoms 17 years.
Interventions	Etanercept (25mg) or placebo twice weekly for 3 months.
Outcomes	Primary outcome was a change in the work instability of patients after 3 months as measured by the AS WIS scale. Secondary outcomes included changes in AS WIS score after 6 months, clinical response (BASDAI, BASFI, ASQoL) and gait analysis at 3 and 6 months.
Notes	Reported as an abstract so methodological details are not available. Not enough information to enter data into RevMan. Funding source not reported.

Risk of bias table

Item	Judgement	Description
Adequate sequence generation?	Unclear	not reported in abstract
Allocation concealment?	Unclear	not reported in abstract
Blinding? (Patient assessed outcomes)	Unclear	"double blind"
Blinding? (Physician assessed outcomes)	Unclear	"double blind"
Incomplete outcome data addressed? (Efficacy outcomes)	Unclear	not reported in abstract
Incomplete outcome data addressed? (Safety outcomes)	Unclear	not reported in abstract
Free of selective reporting?	Unclear	not reported in abstract
Method of AE monitoring	Unclear	not reported in abstract
SAE definitions provided?	Unclear	not reported in abstract

Brandt 2003

Methods	Multicenter, randomized, placebo-controlled trial
Participants	N=14 etanercept ; N=16 placebo Age (mean years): Treatment group - 40; Control group - 32 % male: Treatment group - 71%; Control group - 75% % white: Not reported Disease duration (years): Treatment group - 15; Control group -11 Patients fulfilled the modified NY criteria for AS and had active disease defined by BASDAI >=4 and spinal pain of >=4 on 0-10 scale.

	Excluded: active TB in past 3 years, serious infection in past 2 months, malignancies in past 5 years, MS or related disorder, current signs of severe disease. DMARDS and corticosteroids withdrawn at least 4 weeks prior to screening. Also, widespread ankylosis. NSAIDs at same or less dosage at baseline were allowed.
Interventions	Etanercept 25 mg twice weekly subcutaneously vs placebo for 6 weeks
Outcomes	Primary: BASDAI >=50% by week 6. Others: BASDAI, BASFI, BASMI, ASAS 20%, SF-36, BASRI-s, adverse event
Notes	"Supported by a grant (Kompentenznetz Rheuma) from the German Ministry of Research and by Wyeth Pharma who provided the study drug"

Risk of bias table

Item	Judgement	Description
Adequate sequence generation?	Yes	"Initials and sex of the 33 remaining patients were reported to a central independent registration office by fax. Patients were randomly allocated to one of the treatment groups."
Allocation concealment?	Yes	"Initials and sex of the 33 remaining patients were reported to a central independent registration office by fax. Patients were randomly allocated to one of the treatment groups."
Blinding? (Patient assessed outcomes)	Yes	"The pharmacist at each center prepared the medication, which was delivered in a blinded manner." "The placebo solution containing bacteriostatic water was supplied and administered identically." "Investigators and patients remained blinded until week 12, 6 weeks after the placebo controlled phase had finished."
Blinding? (Physician assessed outcomes)	Yes	"Investigators and patients remained blinded until week 12, 6 weeks after the placebo controlled phase had finished."
Incomplete outcome data addressed? (Efficacy outcomes)	Yes	Reasons for withdrawal provided. >80% follow up in both treatment and placebo groups.
Incomplete outcome data addressed? (Safety outcomes)	Yes	Reasons for withdrawal provided. >80% follow up in both treatment and placebo groups.
Free of selective reporting?	Yes	Appropriate outcomes reported. "As the primary end point of the study, an improvement in disease activity of 50% between baseline and week 6, measured by the BASDAI, was chosen. The secondary outcome parameters analyzed were improvements in numeric rating scale for spinal pain, BASFI, BASMI, SF-36, the ASAS response criteria, serum C-reactive protein (CRP) level, and erythrocyte sedimentation rate (ESR)."
Method of AE monitoring	Unclear	method not reported; or efficacy, clinical questionnaires filled out every 3 weeks

SAE definitions provided?	Unclear	SAE definition not provided

Braun 2002

Methods	Multicenter, randomized placebo-controlled trial
Participants	N=34 infliximab; N=35 placebo Age (mean years): Treatment group - 41; Control group - 39 % male: Treatment group - 68%; Control group - 63% % white: Treatment group - not reported Disease duration (years): Treatment group - 16; Control group -15 Patients fulfilled the modified NY criteria for AS and had active disease defined by BASDAI >=4 and spinal pain of >=4 on 0-10 scale. Excluded: The main reasons for exclusion were severe comorbidity, insufficient disease activity, complete ankylosis, incorrect diagnosis,and DMARD therapy. Also, active TB in past 3 years, specific changes in radiograph of chest at baseline, serious infection in past 2 months, malignancies in past 5 years, signs of severe renal, hepatic, haematological, gastrointestinal, endocrine, pulmonary, cardiac, neurological, cerebral disease. DMARDS and corticosteroids withdrawn at least 4 weeks prior to screening. Patients allowed NSAIDS but dose could not increase from baseline dose.
Interventions	Infliximab 5mg/kg iv vs placebo administered at 0, 2, 6 weeks for 12 weeks
Outcomes	Primary: BASDAI >=50% by week 12. Others BASDAI, BASFI, BASMI, ASAS 20%, SF-36, spinal pain, CRP, ESR, adverse events
Notes	"Funded by a grant (Kompentenznetz Rheuma) from the German Ministry of Research and by Essex Pharma, Munich, who provided the study drug"

Risk of bias table

Item	Judgement	Description
Adequate sequence generation?	Yes	"The allocation schedule was generated by computer-generated random numbers, done in blocks of four for every centre. Thus, within each group of patients enrolled by a single centre, two were randomly assigned to placebo, and two to infliximab."
Allocation concealment?	Yes	"The allocation schedule was generated by computer-generated random numbers, done in blocks of four for every centre. Thus, within each group of patients enrolled by a single centre, two were randomly assigned to placebo, and two to infliximab." "Investigators were informed by fax about the randomisation, and were provided with the trial number of the patient." This information was kept in a sealed envelope that was only opened in case of a serious adverse event.
Blinding? (Patient assessed outcomes)	Yes	"Investigators and patients were unaware of treatment status until all case report forms had been completed."
Blinding? (Physician assessed outcomes)	Yes	"Investigators and patients were unaware of treatment status until all case report forms had been completed." "The information had to be sent back once the patient had

		completed the trial."(referring to the sealed envelope with the group assignment)
Incomplete outcome data addressed? (Efficacy outcomes)	Yes	Greater than 80% follow-up. Reasons for withdrawal were provided.The last observation carried forward method was applied to the four infliximab group members that withdrew.
Incomplete outcome data addressed? (Safety outcomes)	Yes	Greater than 80% follow-up. Reasons for withdrawal were provided.The last observation carried forward method was applied to the four infliximab group members that withdrew.
Free of selective reporting?	Yes	Appropriate outcomes were reported. "The primary endpoint was improvement of disease activity by 50% between baseline and week 12, measured by BASDAI."
Method of AE monitoring	Unclear	not reported
SAE definitions provided?	Unclear	not provided

Braun 2008

Methods	RCT. Double blind.
Participants	N=566. AS patients had active disease based on BASDAI VAS>=30; morning stiffness VAS>=30; VAS>=30 for 2 of the following:patient global assessment of disease activity, pain, BASFI, and be a candidate for SSZ or ETN. All patients had failed >=1 NSAID for >=3 months. Exclusion: complete ankylosis of the spine; previous ETN treatment; SSZ treatment within 6 months of screening. Mean age=41 years; 74% male; average disease duration=7.5 years.
Interventions	Etanercept 50mg once weekly (N=379). Sulphasalazine (SSZ) 3 g daily (N=187)
Outcomes	Primary outcome: proportion of patients achieving ASAS 20 at 16 weeks
Notes	Reported in an abstract. Funding source not reported

Risk of bias table

Item	**Judgement**	**Description**
Adequate sequence generation?	Unclear	not reported in abstract
Allocation concealment?	Unclear	not reported in abstract
Blinding? (Patient assessed outcomes)	Unclear	double-blind; no more details reported in abstract
Blinding? (Physician assessed outcomes)	Unclear	double-blind; no more details reported in abstract
Incomplete outcome data addressed? (Efficacy outcomes)	Unclear	not reported
Incomplete outcome data addressed? (Safety outcomes)	Unclear	not reported
Free of selective reporting?	Unclear	not reported
Method of AE monitoring	Unclear	not reported

SAE definitions provided?	Unclear	not reported

Calin 2004

Methods	Multicenter, randomized, placebo-controlled trial
Participants	N=45 etanercept; N=39 placebo Age (mean, years): Treatment group - 45; Control group - 41 % male: Treatment group - 80%; Control group - 77% % white: Treatment group - 93%; Control group - 95% Disease duration (years): Treatment group - 15; Control group -10 Patients fulfilled the modified NY criteria for AS and had active disease defined by score >=30 on VAS 0-100 for spinal inflammation and a score of >=30 on at least 2 of the 3 domains:back pain, patient global assessment and physical function. Patients were excluded if they had complete ankylosis (fusion) of the spine; previously used TNFa inhibitors, including etanercept; used DMARDs other than hydroxychloroquine, sulfasalazine, or methotrexate within 4 weeks of baseline; used multiple NSAIDs; used >10 mg prednisone daily; or changed doses of NSAIDs or prednisone within 2 weeks of baseline. Patients were permitted to continue pre-study physiotherapy.
Interventions	Etanercept 25 mg twice weekly subcutaneously vs placebo for 12 weeks.
Outcomes	Primary: ASAS 20 by week 12. Others: BASDAI, BASFI, ASAS 50, ASAS 70, spinal inflammation, nocturnal and total pain, spinal mobility, CRP, ESR, adverse events
Notes	"Trial was funded by Wyeth Research"

Risk of bias table

Item	Judgement	Description
Adequate sequence generation?	Unclear	no details provided on sequence generation
Allocation concealment?	Unclear	no details provided on allocation concealment
Blinding? (Patient assessed outcomes)	Yes	"To preserve the integrity of the blind study, placebo and etanercept supplies were similar in appearance."
Blinding? (Physician assessed outcomes)	Unclear	not reported
Incomplete outcome data addressed? (Efficacy outcomes)	Yes	Reasons for withdrawal reported. Follow-up was greater than 80%. "Disease activity and safety analyses were based on the intention to treat population and included all patients who received at least one dose of the ‘‘blinded’’ test article. The last observation carried forward technique was used to handle missing data for continuous and ordinal end points."
Incomplete outcome data addressed? (Safety outcomes)	Yes	Reasons for withdrawal reported. Follow-up was greater than 80%. "Disease activity and safety analyses were based on the intention to treat population and included all patients who received at least one dose of the ‘‘blinded’’ test article. The last observation carried forward technique was used to

		handle missing data for continuous and ordinal end points."
Free of selective reporting?	Yes	Appropriate outcomes reported. "The primary efficacy end point was the percentage of ASAS 20 responders after 12 weeks of treatment."
Method of AE monitoring	Yes	"Patients were monitored for adverse events" over the course of the study
SAE definitions provided?	Unclear	SAE definition not provided

Davis 2003

Methods	Multicenter, randomized, placebo-controlled trial
Participants	N=138 etanercept; N=139 placebo Age (mean, years): Treatment group - 42; Control group - 42 % male: Treatment group - 76%; Control group - 76% % white: Treatment group - 94%; Control group - 91% Disease duration (years): Treatment group - 10; Control group -10 Patients fulfilled the modified NY criteria for AS and had active disease defined by score >=30 on VAS 0-100 for morning stiffness and a score of >=30 on at least 2 of the 3 domains:back pain, patient global assessment and BASFI. Patients were excluded if they had complete ankylosis (fusion) of the spine based on radiographic assessment; previously used TNFa inhibitors, had a serious infection (requiring hospitalization or IV antibiotics) within 4 weeks of screening or were pregnant. Patients were allowed to continue receiving hydroxychloroquine, sulfasalazine, or methotrexate at stable dosages during the study but were excluded if they had received any other DMARDs within 4 weeks of baseline. Also allowed to continue on stable NSAIDs, prednisone, and analgesics.
Interventions	Etanercept 25 mg twice weekly subcutaneously vs placebo for 24 weeks.
Outcomes	Primary: ASAS 20 by week 12 and 24. Others: ASAS 50, ASAS 70, partial remission (defined as value <20mm (0-100mm scale) in each of 4 ASAS domains (patient global assessment, pain, BASFI, inflammation)). BASDAI, spinal mobility, peripheral joint count, CRP, ESR, assessor global assessment, adverse events
Notes	"Supported by Immunex Corporation"

Risk of bias table

Item	Judgement	Description
Adequate sequence generation?	Yes	"Patients, investigators, assessors, other study site personnel, and representatives of the sponsor were blinded to the randomization schedule and to treatment assignment until completion of the trial"
Allocation concealment?	Yes	"Patients, investigators, assessors, other study site personnel, and representatives of the sponsor were blinded to the randomization schedule and to treatment assignment until completion of the trial"
Blinding? (Patient assessed outcomes)	Yes	"Patients, investigators, assesors, other study site personnel, and representatives of the sponsor were blinded to the randomization schedule and to treatment

		assignment until completion of the trial"
Blinding? (Physician assessed outcomes)	Yes	"Patients, investigators, assessors, other study site personnel, and representatives of the sponsor were blinded to the randomization schedule and to treatment assignment until completion of the trial"
Incomplete outcome data addressed? (Efficacy outcomes)	Yes	Completed study: 86% in placebo and 91% in treatment; 1 loss to follow-up in placebo and 2 loss to follow-up in treatment. LOCF used for missing data
Incomplete outcome data addressed? (Safety outcomes)	Yes	Completed study: 86% in placebo and 91% in treatment; 1 loss to follow-up in placebo and 2 loss to follow-up in treatment. LOCF used for missing data
Free of selective reporting?	Yes	ASAS 20 was primary outcome; not stated that it was prespecified in the protocol, but it is an appropriate outcome
Method of AE monitoring	Yes	Patients used a diary to record presence of adverse events
SAE definitions provided?	Yes	Adverse events graded on a scale derived from the national cancer institute common toxicity criteria

Giardina 2009

Methods	"Two year randomized study"
Participants	N=25 (abstract had N=26) etanercept; N=25 (abstract had N=24 infliximab Inclusion criteria: active disease for >3 months; BASDAI >4; VAS for spinal pain >4 Age (mean, years): Etanercept group - 32.6 SD 6.8; Infliximab group - 31.9 SD 9.2 % male: Etanercept group - 80%; Infliximab group - 76% Disease duration (years): Etanercept group - 15.7 SD6.5; Infliximab group -15.4 SD 10.6
Interventions	Etanercept 25mg twice weekly or 5mg/kg infliximab at week 0, 2, 6, and then every 6 weeks for a period of 102 weeks
Outcomes	In the abstract, the primary outcome was stated to be the proportion of patients achieving a 50% BASDAI response at week 102; Secondary: ASAS 50; BASFI, back pain, morning stiffness, CRP, spinal mobility. However, in the full-text article, the outcome defined as primary is not stated, and the 50% BASDAI response is not reported. ASAS 20 and 40, BASDAI, BASFI and AE were reported.
Notes	Reported as a full-text and an abstract from a conference. Funding source not reported.

Risk of bias table

Item	**Judgement**	**Description**
Adequate sequence generation?	Unclear	"Patients were randomised to receive alternatively etanercept or infliximab with a ratio of 1:1."
Allocation concealment?	Unclear	no details on concealment of allocation
Blinding? (Patient assessed outcomes)	No	"open-label"
Blinding? (Physician	No	"open-label"

assessed outcomes)		
Incomplete outcome data addressed? (Efficacy outcomes)	Yes	"No patients discontinued therapy"
Incomplete outcome data addressed? (Safety outcomes)	Yes	"No patients discontinued therapy"
Free of selective reporting?	No	primary outcome listed in abstract: proportion of people achieving a 50% response in BASDAI; full-text article does not state the primary outcome, but 50% BASDAI response not reported
Method of AE monitoring	Yes	Patients were monitored for adverse events and abnormal lab values over the course of the study.
SAE definitions provided?	Unclear	no SAE definition given

Gorman 2002

Methods	Randomized, placebo-controlled trial
Participants	N=20 (etanercept); N=20 (placebo) Age (median, years): Treatment group - 38; Control group - 39 % male: Treatment group - 65%; Control group - 90% % white: Treatment group - 75%; Control group - 70% Disease duration (years): Treatment group - 15; Control group -12 Patients >=18 years of age and classified as having definite AS based on the modified New York criteria. Active spondylitis was defined as the presence of inflammatory back pain (stiffness and pain that worsened with rest and improved with exercise), morning stiffness for at least 45 minutes, and at least moderate disease activity as assessed by the patient and the physician. The patient's global assessment of disease activity was based on a five-point scale (1,none; 2, mild; 3, moderate; 4, severe; and 5, very severe). The physician's assessment was measured with the use of a VAS (0 mm absence of disease activity and 100 mm very severe activity); a moderate or higher level of disease activity was defined by the placement of a vertical line at 40 mm or higher. Patients were excluded if they had a spondylitis other than AS, clinical or radiographic evidence of complete spinal ankylosis, a history of recurrent infections or cancer, or a serious liver, renal, hematologic, or neurologic disorder. Patients continued to take drugs that had already been prescribed for ankylosing spondylitis if the doses had not been changed for at least four weeks before randomization and if they remained unchanged throughout the trial. Acceptable medications included NSAIDs,oral corticosteroids («10 mg per day), gold injections («50 mg per month), methotrexate («20 mg per week), and sulfasalazine(«3 g per day).
Interventions	Twice-weekly subcutaneous injections of etanercept (25 mg) versus placebo for four months.

Outcomes	Primary: ASAS 20. Secondary outcomes: physician's global assessment of disease activity, measures of spinal mobility, the scores for enthesitis and peripheral-joint tenderness, ESR, CRP, adverse events.
Notes	"The majority of funding for the study was provided by the National Institute of Arthritis and Musculoskeletal and Skin Diseases. Immunex, the pharmaceutical funding source, supplied etanercept and placebo and provided partial funding. Immunex was not involved in the study design, data collection, statistical analysis, or manuscript preparation; these tasks were performed by the authors."

Risk of bias table

Item	Judgement	Description
Adequate sequence generation?	Yes	"A statistician not otherwise involved with the study randomly assigned patients to the study groups, using computer-generated, random blocks of two and four."
Allocation concealment?	Yes	"Cards with the group assignments were placed in sequentially numbered envelopes that were opened by the study pharmacist as each patient was enrolled."
Blinding? (Patient assessed outcomes)	Yes	"The patients and study investigators were unaware of the group assignments."
Blinding? (Physician assessed outcomes)	Yes	"The patients and study investigators were unaware of the group assignments."
Incomplete outcome data addressed? (Efficacy outcomes)	Yes	Reasons for withdrawal were provided. Follow-up was greater than 80% in both groups. ITT analyses were performed.
Incomplete outcome data addressed? (Safety outcomes)	Yes	Reasons for withdrawal were provided. Follow-up was greater than 80% in both groups.
Free of selective reporting?	Yes	"The primary outcome measure was a prespecified, composite treatment response, defined as 20 percent or greater improvement in at least three of five measures of disease activity, as recommended by the Assessments in Ankylosing Spondylitis Working Group". Corresponds to the protocol on clinicaltrials.gov.
Method of AE monitoring	Yes	"Side effects monitored at each clinic visit by means of open ended questions..."
SAE definitions provided?	Yes	Adverse events graded on a scale derived from the national cancer institute common toxicity criteria

Huang 2008

Methods	Randomized, double-blind, placebo-controlled study for 6 weeks with 6 week open label afterwards
Participants	N=74 etanercept; N=78 placebo Adult patients with AS; patients receiving hydrochloroquine, sulfasalazine, or methotrexate at screening continued on the

	medication
Interventions	Etanercept 50mg once weekly for 6 weeks or placebo subcutaneously
Outcomes	Primary endpoint: ASAS 20 week6. Secondary ASAS 40, ASAS 5/6, adverse events
Notes	Abstract from conference proceeding. Funding source not reported.

Risk of bias table

Item	**Judgement**	**Description**
Adequate sequence generation?	Unclear	not reported
Allocation concealment?	Unclear	not reported
Blinding? (Patient assessed outcomes)	Unclear	"double blind"
Blinding? (Physician assessed outcomes)	Unclear	"double blind"
Incomplete outcome data addressed? (Efficacy outcomes)	Unclear	2/152 withdrew, no group or reason given
Incomplete outcome data addressed? (Safety outcomes)	Unclear	2/152 withdrew, no group or reason given
Free of selective reporting?	Yes	primary outcome ASAS 20 reported and appropriate
Method of AE monitoring	Yes	"Adverse event and routine lab monitoring"
SAE definitions provided?	Unclear	No SAE definition provided

Inman 2010

Methods	Randomized, double-blind, placebo-controlled, multi-centre trial.
Participants	N=76; N=39 infliximab, N=37 placebo Note: one abstract assessing spinal inflammation with MRI indicates N=32 (16 infiximab and 16 placebo). The study protocol from clinicaltrials.gov indicates the study is complete and the sample is 76 participants. Age (mean, years): Treatment group - 42.9 (10.4); Control group - 39.3 (9) % male: Treatment group - 82%; Control group - 78% % white: Treatment group - 87%; Control group - 89% Disease duration (years): Treatment group - 11.7 (10.6); Control group -11.1 (10.3) Positive for HLA-B27: infliximab=72%; placebo=73% Inclusion: Adults (>18 years) with active AS (BASDAI >=4). In those patients taking NSAIDs, DMARDs, analgesics, or corticosteroids, the dose must have been stable for at least 14 days (30 days for DMARD) prior to the first infusion of study drug. Patients were excluded from the study if they had a history of chronic/recurrent infectious disease, including tuberculosis, hepatitis B, or HIV, and/or a diagnosis of malignancy or lymphoproliferative disease currently or within the past 5 years.
Interventions	Infliximab 3mg/kg or placebo intravenously at weeks 0, 2, and 6. An open-label phase followed after week 12 which lasted 46 weeks and the placebo group crossed over to receive infusions of IFX 3 mg/kg at Weeks 14, 16, and 22, and every 8 weeks thereafter. "All patients could receive dose-escalation of IFX to 5 mg/kg at Weeks 22 or 38 if

	the patient had an absolute BASDAI score > 3 and and a relative decrease of < 50% in BASDAI from baseline". Follow up was for 52 weeks.
Outcomes	Primary outcome listed in study protocol was ASAS 20 at week 12. Other clinical outcomes in the protocol: BASDAI, BASFI, BASGI, BASMI, ASAS 40/50/70 and ASAS 5/6 and MRI at week 12 were reported in the abstracts (though the number of participants is unclear).
Notes	Study known as "CANaDian evaluation of Low DosE infliximab (CANDLE)." Reported in one full-text article and 3 abstracts from conferences. "Supported by Schering-Plough, Canada"

Risk of bias table

Item	Judgement	Description
Adequate sequence generation?	Unclear	not reported
Allocation concealment?	Unclear	not reported
Blinding? (Patient assessed outcomes)	Unclear	"double-blind"; no further details reported
Blinding? (Physician assessed outcomes)	Unclear	"double-blind"; no further details reported
Incomplete outcome data addressed? (Efficacy outcomes)	Yes	Fig.1 shows the details of patient flow throughout the study. The number of people completing the study at week 12 (RCT phase) was not clearly reported, but the number of drop outs was low. ITT analysis performed, defined as those who received one dose of the study drug.
Incomplete outcome data addressed? (Safety outcomes)	No	Safety outcomes reported for the combined RCT and open-label phase; not for RCT phase separately
Free of selective reporting?	Unclear	Primary outcome appropriate but details on adverse events only mentioned briefly in an abstract
Method of AE monitoring	Unclear	"Safety and tolerability were assessed by the incidence of treatment emergent adverse events"
SAE definitions provided?	Yes	All adverse events were coded using the MedDRA dictionary of terms (version 9.0).

Lambert 2007

Methods	Randomized, multicenter, double-blind, placebo-controlled study
Participants	N=38 adalimumab, N=44 placebo Age (mean (SD), years): Treatment group - 41.9(11.1); Control group - 40 (10.9) % male: Treatment group - 76.3%; Control group - 81.8% % white: Treatment group - not reported Disease duration (years (SD)): Treatment group - 14.5 (9); Control group -12.1 (8.7) Inclusion:Patients were adults (18 years of age) diagnosed as having AS as defined by the modified New York criteria, who had been treated unsuccessfully (nonresponse or lack of tolerance) with 1

	NSAIDs. Patients who had failed to respond to 1 DMARD (e.g., methotrexate,sulfasalazine) were also allowed to enroll. Active AS at baseline was defined by fulfillment of 2 of the following 3 criteria: a BASDAI score 4, total back pain visual analog scale score 40, or morning stiffness of 1 hour in duration. Patients could continue taking sulfasalazine (3 gm/day), methotrexate (25 mg/week), hydroxychloroquine (400 mg/day), prednisone and/or prednisone equivalents (10 mg/day), and/or NSAIDs as long as these doses had remained stable for 4 weeks before baseline.
Interventions	40mg adalimumab or placebo every other week for 24 weeks (double-blind phase). Study visits occurred at baseline, week 2, week 4, every 4 weeks through week 24
Outcomes	Primary endpoint was ASAS 20 at 12 weeks but results were not provided in full-text Lambert article, but were reported in a conference abstract (Maksymowych 2005) Secondary outcomes of MRI of spine and SI joints scored using SPARCC methodology at week 12 were reported in Lambert2007. Another publication from this trial, Maksymowych 2008, which was focused on biomarkers for structural damage, also reported BASDAI, total back pain, patient global, BASFI, BASMI.
Notes	NCT00195819; M03-606 study group "ROLE OF THE STUDY SPONSOR An advisory committee, including authors from academic institutions and Abbott Laboratories, and members of the Abbott Laboratories clinical trial team designed the study, which was conducted at 11 centers in Canada. Clinical data were collected and analyzed by Abbott Laboratories. Data analyses were reviewed by members of the advisory committee. All authors reviewed and assisted in the manuscript preparation during its development, agreed to submit the manuscript, and approved the content of the submitted manuscript."

Risk of bias table

Item	Judgement	Description
Adequate sequence generation?	Unclear	not reported
Allocation concealment?	Unclear	not reported
Blinding? (Patient assessed outcomes)	Unclear	trial protocol states "Double Blind (Subject, Investigator)"
Blinding? (Physician assessed outcomes)	Yes	"readers were qualified, trained radiologists who were blinded to the patients' identities, treatments, and imaging time points"
Incomplete outcome data addressed? (Efficacy outcomes)	Yes	"at baseline and week 12, all 44 patients in the placebo group and 38 in the adalimumab group had evaluable MRIs"
Incomplete outcome data addressed? (Safety outcomes)	Unclear	safety data not reported in primary publication

Free of selective reporting?	Yes	primary outcome reported in abstract and other appropriate outcomes in full text articles
Method of AE monitoring	Unclear	not reported
SAE definitions provided?	Unclear	not reported

Marzo-Ortega 2005

Methods	30 week, single-centre, randomised, double blind placebo controlled trial
Participants	N=28 infliximab + methotrexate (MTX), N=14 placebo + MTX Age (mean (range), years): Treatment group - 41 (28–74); Control group - 39 (30–56) % male: Treatment group - 82%; Control group - 79% % white: - not reported Disease duration (median years (range)): Treatment group - 8 (0-41) ; Control group - 10 (0-35) Inclusion:fulfil the modified New York criteria for AS,14 be older than 18 years of age, and have active spinal disease. This was defined as persistent inflammatory back pain (defined as 3 cm or more on a 10 cm visual analogue scale (VAS)) and a raised inflammatory response in serum as shown by a C reactive protein (CRP) value of more than 10 mg/l despite treatment with conventional agents such as an optimal dosage of non-steroidal anti-inflammatory drugs (NSAIDs) or DMARDs. Exclusion: any history of tuberculosis, active infection, demyelinating disease, previous lymphoproliferative or malignant disorder, pregnancy, breast feeding, or uncontrolled concomitant disease in the opinion of the investigator.
Interventions	Infusions of infliximab (5 mg/kg in 250 ml 0.9% NaCl) + MTX or placebo + MTX.The infusion regimen was weeks 0, 2, 6, 14, and 22. All subjects also received a dose of 7.5 mg with folic acid cover (5 mg twice a week), which was eventually increased to 10 mg a week.
Outcomes	The primary outcome was evaluation of change in the BASDAI score at weeks 4, 10, and 30. Secondary outcomes were ASAS 20 and BASDAI 50% response. MRI also assessed.
Notes	This study was supported by a grant in aid from Schering-Plough, UK.

Risk of bias table

Item	Judgement	Description
Adequate sequence generation?	Yes	"A 2:1 randomisation list was generated by a statistician (who was unconnected with the final analysis of results)"
Allocation concealment?	Yes	"Study participants, clinical observers, and metrologists were unaware of the randomisation code, which was kept in the hospital pharmacy"
Blinding? (Patient assessed outcomes)	Unclear	"double-blind"
Blinding? (Physician assessed outcomes)	Unclear	"double-blind"

Incomplete outcome data addressed? (Efficacy outcomes)	Unclear	93% completed treatment, 64% completed in control group; an ITT analysis performed with LOCF imputation for missing data
Incomplete outcome data addressed? (Safety outcomes)	Unclear	93% completed treatment, 64% completed in control group; an ITT analysis performed with LOCF imputation for missing data
Free of selective reporting?	Yes	appropriate outcomes measured
Method of AE monitoring	Unclear	not reported
SAE definitions provided?	Unclear	not reported

van der Heijde 2005

Methods	Multicenter, randomized, placebo-controlled study for 24 weeks
Participants	N=201 infliximab N=78 placebo. Age (mean, years): Treatment group - 40; Control group - 41 % male: Treatment group - 78%; Control group - 87% % white: Treatment group - 98%; Control group - 97% Disease duration (years): Treatment group - 8; Control group -13 Patient fulfilling the modified New York criteria for at least 3 months prior to screening, with BASDAI >=4 (range 0–10) and spinal pain assessment score >=4 on a VAS (0–10 cm) were eligible for the study. Patients were also required to have a normal chest radiograph within 3 months prior to randomization and either a negative TB test. Exclusion: total ankylosis of the spine (defined by syndesmophytes present on the lateral views of spinal radiographs at all intervertebral levels from T6 through S1), any other inflammatory rheumatic disease, fibromyalgia, a serious infection within 2 months prior to randomization, TB (active or latent) or recent contact with a person with active TB, an opportunistic infection within 6 months of screening, hepatitis, human immunodeficiency virus, a transplanted organ, malignancy, multiple sclerosis, or congestive heart failure. Patients were allowed to receive concurrent stable doses of NSAIDs, acetaminophen (paracetamol), or tramadol during the study. Patients were not permitted to receive sulfasalazine or methotrexate within 2 weeks prior to screening, systemic corticosteroids within 1 month prior to screening, anti-TNF therapy other than infliximab within 3 months prior to screening, infliximab at any time prior to screening, DMARDs other than sulfasalazine or methotrexate within 6 months prior to screening, or cytotoxic drugs within 12 months prior to screening.
Interventions	5 mg/kg infliximab at weeks 0, 2, 6, 12, and 18 vs placebo
Outcomes	Primary end point: ASAS 20 responders at week 24. Other: BASDAI, night pain, patient’s global assessment, BASF, BASMI, chest expansion, the Mander enthesis index, the total swollen joint index, the C-reactive protein level, SF-36, adverse events.
Notes	"Supported by Centocor Inc"

Risk of bias table

Item	Judgement	Description
Adequate sequence generation?	Unclear	"Randomly assigned in 3:8 ratio"

Allocation concealment?	Unclear	Unclear "Patients were allocated to treatment groups using an adaptive treatment allocation stratified by investigational site and C-reactive protein level..."
Blinding? (Patient assessed outcomes)	Yes	"double blind"; "both infliximab and placebo were supplied as sterile, white, lyophilized powders in single-use 20-ml vials"
Blinding? (Physician assessed outcomes)	Yes	"double blind"; probably yes given the blinding of the study drug
Incomplete outcome data addressed? (Efficacy outcomes)	Yes	fairly complete follow up; 75/78 completed in placebo and 199/201 completed in infliximab group. ITT analysis.
Incomplete outcome data addressed? (Safety outcomes)	Yes	fairly complete follow up; 75/78 completed in placebo and 199/201 completed in infliximab group. ITT analysis.
Free of selective reporting?	Yes	Appropriate outcomes reported.
Method of AE monitoring	Unclear	"Safety assessments included adverse events, infections, infusion reactions, premature discontinuations, and lab tests"
SAE definitions provided?	Unclear	no general SAE definition provided; but all SAEs that occurred were explained

van der Heijde 2006a

Methods	Multicenter, randomized (2:1 ratio) placebo-controlled study
Participants	N=208 adalimumab; N=107 placebo Age (mean, years): Treatment group - 42; Control group - 43 % male: Treatment group - 76%; Control group - 74% % white: Treatment group - 97%; Control group - 93% Disease duration (years): Treatment group - 11; Control group -10 Patients >=18 years of age and classified as having definite AS based on the modified New York criteria. All had active disease, defined as fulfillment of at least 2 of the following 3 criteria: BASDAI >=4, a total back pain score >4 (VAS 0–10 cm), or a duration of morning stiffness >1 hour. Patients with stable and well-controlled psoriasis, uveitis, inflammatory bowel disease (i.e., ulcerative colitis,Crohn's disease), and reactive arthritis were allowed to participate. Inadequate response or intolerance to 1 or more NSAIDs was defined by the investigators. Patients in whom 1 or more DMARDs had failed were also allowed to participate. Patients were allowed to continue any of the following medications if the dosage had remained stable for at least 4 weeks before the baseline visit: sulfasalazine (<=3gm/day), methotrexate (<=25 mg/week), hydroxychloroquine(<=400 mg/day), prednisone or prednisone equivalent (<=10mg/day), and NSAIDs. Exclusions:"previously received anti-TNF therapy, cyclosporine, azathioprine, or DMARDs (other than the medications and dosages listed above) at any time and patients who had received intraarticular injection(s) with corticosteroids within 4 weeks prior to baseline. Patients with latent TB were allowed to participate in the study if a documented history of treatment was available or if treatment for

	latent TB was initiated before the first dose of study medication. Patients with clinically active TB were excluded from the study. History of any recent infections requiring antibiotic treatment; hepatitis or human immunodeficiency virus; a significant history of cardiac, renal, neurologic, psychiatric, endocrinologic, metabolic, or hepatic disease; and a history of demyelinating disease or multiple sclerosis. History of cancer or lymphoproliferative disease other than a successfully treated nonmetastatic squamous cell or basal cell carcinoma and/or localized carcinoma in situ of the cervix."
Interventions	Adalimumab 40 mg every other week or placebo for a 24-week period.
Outcomes	Primary efficacy outcome: percentage of ASAS 20 responders at week 12. Secondary: ASAS 5/6, ASAS 40.
Notes	"Supported by Abbott Laboratories"

Risk of bias table

Item	Judgement	Description
Adequate sequence generation?	Yes	From study author: "A random number generator was used to generate the randomization numbers. All patients were centrally randomized using an Interactive Voice Response System (IVRS). Randomization occurred within each site."
Allocation concealment?	Yes	From study author: "The patient, sponsor and the study sites were blinded to treatment allocation." "The treatment allocation for each patient was provided to the site in a sealed envelope, to be opened in the case of an emergency in which the investigator believed that knowledge of study drug treatment was required. However, no patient was unblinded during the course of the double blind period."
Blinding? (Patient assessed outcomes)	Unclear	From study author: "Yes, patients and assessors were blinded. In particular, the assessor who performed the tender and swollen joint counts, MASES, and the physical examination was blinded to information from the patient reported questionnaires at all visits." However, the double-blind period appears to have lasted only 12 weeks; patients who had not achieved an ASAS 20 response at 12 weeks could enter the early escape period to receive open-label adalimumab.
Blinding? (Physician assessed outcomes)	Yes	From study author: "Yes, patients and assessors were blinded. In particular, the assessor who performed the tender and swollen joint counts, MASES, and the physical examination was blinded to information from the patient reported questionnaires at all visits."
Incomplete outcome data addressed? (Efficacy outcomes)	Yes	95% of participants completed 24 weeks. Intention-to-treat analysis performed using non-responder imputation.
Incomplete outcome data	Yes	95% of participants completed 24 weeks.

addressed? (Safety outcomes)		
Free of selective reporting?	Yes	Outcomes reported as prespecified in trial protocol (ASAS 20 at week 12). Appropriate outcomes were reported.
Method of AE monitoring	Yes	Adverse events and vital signs assessed at every visit
SAE definitions provided?	Unclear	no general SAE definition provided; but each SAE that occurred was detailed in the report.

van der Heijde 2006b

Methods	12-week, randomized, placebo-controlled, double-blind, multicentre study with three treatment groups in a 3:3:1 ratio (etanercept 50mg once weekly: etanercept 25mg twice weekly: placebo). Study carried out in 38 centres in 11 European countries.
Participants	Etanercept 50mg once weekly:N=155; mean age (SD)=41.5 (11.0); 69.7% male; disease duration, years (SD)=9.0 (8.7). Etanercept 25mg twice weekly:N=150; mean age (SD)=39.8 (10.7); 76% male; disease duration, years (SD)=10.0 (9.1). Placebo:N=51; mean age (SD)=40.1 (10.9); 78.4% male; disease duration, years (SD)=8.5 (6.8). Inclusion - age 18 to 70 years with active AS based on the Modified New York Criteria for AS. Active AS defined by VAS >=30 for duration and intensity of morning stiffness and two or more of the following:patient global assessment of disease activit VAS >=30; mean of nocturnal and total pain VAS scores >=30. "Concomitant oral non-steroidal anti-inflammatory drugs and oral corticosteroids ((10 mg/day), if stable for >2 weeks before randomisation, and disease-modifying antirheumatic drugs (hydroxychloroquine, sulfasalazine and methotrexate), if stable for >4 weeks before randomisation, were permitted." Exclusion: "Patients previously treated with TNFa inhibitors, including etanercept or other biological agents, or disease-modifying antirheumatic drugs (other than hydrochloroquine, sulfasalazine and methotrexate) less than 4 weeks before baseline, were not eligible. Other important exclusion criteria included complete ankylosis (fusion) of the spine based on radiographic assessment and concurrent medical events, such as uncontrolled hypertension, unstable angina pectoris, congestive heart failure, severe pulmonary disease, cancer, demyelinating diseases of the central nervous system and serious infections."
Interventions	Etanercept 50mg once weekly versus etanercept 25mg twice weekly versus placebo.
Outcomes	Non-inferiority design to compare etanercept 50mg once weekly to 25mg twice weekly. Primary outcome: ASAS 20 at week 12. Secondary outcomes: Secondary outcomes:ASAS 40 and ASAS 5/620 criteria at all time points.
Notes	"Study was supported by Wyeth Pharmaceuticals, Collegeville, Pennsylvania, USA (study drug and grants to investigational sites).

Risk of bias table

Item	**Judgement**	**Description**

Adequate sequence generation?	Unclear	not reported in article
Allocation concealment?	Unclear	not reported in article
Blinding? (Patient assessed outcomes)	Unclear	"double-blind", no further details reported
Blinding? (Physician assessed outcomes)	Unclear	"double-blind", no further details reported
Incomplete outcome data addressed? (Efficacy outcomes)	Yes	Reasons for withdrawal were provided. Follow-up was greater than 80% in all groups. Modified ITT analyses were performed in which "all participants who received at least one dose of the test drugs" were included in the analyses (356/361 randomized). "A last-observation-carried-forward approach was used to impute missing data in the modified intent-to-treat population analysis."
Incomplete outcome data addressed? (Safety outcomes)	Yes	Same as efficacy outcomes.
Free of selective reporting?	Yes	All appropriate outcomes were assessed.
Method of AE monitoring	Unclear	Safety assessments based on reports of AE, routine physicals, lab tests
SAE definitions provided?	Unclear	refers to "non-infectious serious adverse events" but no definition provided

Footnotes

AS:Ankylosing Spondylitis
BASDAI - Bath Ankylosing Spondylitis Disease Activity Index
BASFI: Bath Ankylosing Spondylitis Functional Index
BASMI: Bath Ankylosing Spondylitis Metrology index
BASRI-s: Bath Ankylosing Spondylitis Radiology Index for the spine
DFI:Dougados Functional Index
AID: articular index according to Dougados
ESR - erythrocyte sedimentation rate
CRP - C-reactive protein level.
DMARD -disease-modifying antirheumatic drugs
NSAID- non-steroidal anti-inflammatory drugs
VAS: Visual analogue scale
SF-36: short form 36. A health-related assessment of quality of life

Table 2: Characteristics of excluded studies

Barkham 2008b

Reason for exclusion	Trial participant inclusion criteria does not meet review participant inclusion criteria (inflammatory back pain by Calin criteria <3yrs)

Breban 2008

Reason for exclusion	Trial participants did not meet complete modified NY criteria for AS

	and the intervention was systematic versus on-demand treatment using infliximab.

Haibel 2008

Reason for exclusion	The population included in this study "axial spondylarthritis without radiographically defined sacroiliitis" does not meet the review's inclusion criteria. This study is in early AS patients.

Hamza 2007-NRS

Reason for exclusion	Results are reported for AS, but the exposure is less than one patient-year.

Li 2008

Reason for exclusion	The intervention of infliximab+methotrexate vs infliximab + placebo assesses effect of methotrexate and does not meet the review's intervention inclusion criteria

Van den Bosch 2002

Reason for exclusion	Patients included AS, reactive arthritis, psoriatic arthritis, and undifferentiated SpA. Results for AS patients only were not available

Visvanathan 2008

Reason for exclusion	not an outcome of interest (biomarker response); clinical outcomes already reported in van der Heijde 2005 (ASSERT)

Table 3: Summary of findings table

TNF-alpha inhibitors for ankylosing spondylitis							
Outcome	**Intervention and comparison**	**Illustrative comparative risks**		**Relative effect (95% CI)**	**Number of participants (studies)**	**Quality of the evidence (GRADE)**	**Comments**
		Assumed risk with comparator	**Corres-ponding risk with interventi on (95% CI)**				
		Placebo	**TNF-alpha inhibitor**				
ASAS 40							
	Etanercept (25 mg twice weekly or 50 mg once weekly) versus placebo	**19 per 100**	**57 per 100** (39 to 83)	**RR 3.07** (2.11 to 4.47)	508 (3 studies)	⊕⊕⊕⊕ **high**[1]	Absolute increased benefit %= 38% (29% to 48%); Relative % change= 207%(111% to 347%); NNT= 3 (2 to 4)
	Infliximab versus placebo	**10 per 100**	**45 per 100**	**RR 4.33** (2.5 to	355 (2 studies)	⊕⊕⊕⊕ **high**[2]	Absolute increased

			(26 to 78)	7.52)			benefit %= 35% (27% to44%); Relative % change= 333%(150% to 652%); NNT= 3 (2 to 6)
	Adalimumab versus placebo	**13 per 100**	**40 per 100** (24 to 67)	**RR 3.05** (1.82 to 5.11)	315 (1 study)	⊕⊕⊕⊕ **high**	Absolute increased benefit %= 27% (17% to 36%); Relative % change= 205%(82% to 411%); NNT= 4 (3 to 8)
BASFI							
	Etanercept (25 mg twice weekly or 50 mg once weekly) versus placebo	The mean basfi in the control groups was **54.7 points**	The mean BASFI in the interventi on groups was **16.09 lower** (20.95 to 11.23 lower)		431 (4 studies)	⊕⊕⊕⊕ **high**	Absolute increased benefit % =-16% (-21% to -11%); Relative % change from baseline = -28.6% (-37.2% to -19.9%); NNT to achieve the MCID of 0.7 points= 4 (3 to 6)
	Infliximab versus placebo	The mean basfi in the control groups was **6 points**	The mean BASFI in the interventi on groups was **1.91 lower** (2.28 to 1.54 lower)		348 (2 studies)	⊕⊕⊕⊕ **high**	Absolute increased benefit %= -19.1% (-22.8% to -15.4%); Relative % change from baseline= -31.8% (-38% to -25.6%); NNT to achieve the MCID of 0.7 points= 3 (2 to 3)
	Adalimumab versus placebo	The mean basfi in the control	The mean BASFI in the		397 (2 studies)	⊕⊕⊕⊕ **high**	Absolute increased benefit %= -

		groups was **5.2 points**	intervention groups was **1.8 lower** (2.27 to 1.33 lower)				18% (-22.7% to -13.3%); Relative % change from baseline= -32.1% (-40.5% to -27.5%); NNT to achieve the MCID of 0.7 points= 3 (3 to 4)[3]
ASAS partial remission							
	Etanercept (25 mg twice weekly or 50 mg once weekly) versus placebo	**5 per 100**	**20 per 100** (11 to 35)	**RR 3.99** (2.21 to 7.2)	785 (4 studies)	⊕⊕⊕⊕ **high**[1]	Absolute increased benefit %= 15% (10% to20%); Relative % change = 299% (121% to 620%); NNT = 7 (4 to15)
	Infliximab versus placebo	**1 per 100**	**23 per 100** (3 to 100)	**RR 17.46** (2.45 to 124.51)	279 (1 study)	⊕⊕⊕⊖ **moderate**[4]	Absolute increased benefit %= 22% (15% to 27%); Relative % change = 1646% (145% to 12,351%); NNT = 5 (2 to 40)
	Adalimumab versus placebo	**4 per 100**	**20 per 100** (8 to 56)	**RR 5.53** (2.04 to 15)	315 (1 study)	⊕⊕⊕⊕ **high**	Absolute increased benefit %=16% (10% to 24%); Relative % change = 453% (104% to 1400%); NNT = 5 (3 to 16)
Withdrawals due to adverse events							
	Etanercept (25 mg twice weekly or 50 mg once weekly) versus placebo	**4 per 1000**	**16 per 1000** (6 to 42)	**OR 3.93** (1.4 to 11.02)	942 (5 studies[5])	⊕⊕⊕⊖ **moderate**[4]	Absolute increased harm % = 3%(1% to 6%); Relative %

							change = 305%(1% to 1523%); NNT = 87 (27 to 629)
	Infliximab versus placebo	**9 per 1000**	**22 per 1000** (4 to 113)	**OR 2.42** (0.42 to 14.02)	346 (2 studies)	⊕⊕⊕⊖ **moderate**[4]	Absolute increased harm % = 2% (-2% to 5%); Relative % change = 137%(-53% to 1110%); NNT = n/a
	Adalimumab versus placebo	**19 per 1000**	**20 per 1000** (4 to 105)	**RR 1.03** (0.19 to 5.53)	315 (1 study)	⊕⊕⊕⊖ **moderate**[4]	Absolute increased harm %= 0% (-3% to 3%); Relative%change=3%(-81% to 453%); NNT = n/a
Serious adverse events							
	Etanercept (25 mg twice weekly or 50 mg once weekly) versus placebo	**26 per 1000**	**44 per 1000** (20 to 94)	**OR 1.71** (0.75 to 3.9)	787 (5 studies[6])	⊕⊕⊕⊖ **moderate**[4]	Absolute increased harm %= 2% (-1% to 5%); Relative % change = 67% (-27% to 282%); NNT = n/a
	Infliximab versus placebo	**18 per 1000**	**39 per 1000** (12 to 125)	**OR 2.22** (0.64 to 7.77)	346 (2 studies)	⊕⊕⊕⊖ **moderate**[4]	Absolute increased harm %= 3% (-1% to 6%); Relative % change= 124% (-32% to 639%); NNT = n/a
	Adalimumab versus placebo	**28 per 1000**	**29 per 1000** (7 to 113)	**RR 1.03** (0.26 to 4.03)	315 (1 study)	⊕⊕⊕⊖ **moderate**[4]	Absolute increased harm %= 0% (-4% to 4%); Relative % change= 3% (-74% to 303%); NNT = n/a
MRI of spine or sacroiliac joints							
	Etanercept (25 mg twice weekly or 50	See comment	See comment	Not estimable	0 (0)	See comment	No studies assessed this outcome

	mg once weekly) versus placebo						
	Infliximab versus placebo	See comment	See comment	Not estimable	0 (0)	See comment	Data refers to spinal MRI Activity Score > 1. Also, reduction in MRI Activity Score from baseline to week 24 (MD -4.42, 95% CI -5.59 to -3.25 on 0 to 138 scale)
	Adalimumab versus placebo	See comment	See comment	Not estimable	0 (0)	See comment	Lambert '07: % change from baseline in SPARCC score: 1. spine: ADA gp = 53.6% mean decrease, PL gp = 9.4% mean increase; 2.SI joint mean decrease: ADA gp = 52.9%, PL = 12.7%[7]
Radiographic progression							
	Etanercept (25 mg twice weekly or 50 mg once weekly) versus placebo	See comment	See comment	Not estimable	0 (0)	See comment	No studies measured this outcome
	Infliximab versus placebo	See comment	See comment	Not estimable	0 (0)	See comment	Braun 2002 used the Bath Ankylosing Spondylitis Radiology Index (BASRI) to measure radiographic progression but data was not provided.
	Adalimumab	See	See	Not	0	See	No studies

	versus placebo	comment	comment	estimable	(0)	comment	assessed this outcome

Footnotes
[1] Data from Huang 2006 from an abstract so few details to assewss risk of bias.
[2] Inman 2010 and van der Heijde 2005 were unclear for allocation concealment, and Inman also for blinding.
[3] Used van der Heijde 2006a as most representative study for calculations.
[4] Few events and wide confidence interval.
[5] 3 studies (Brandt 2003, Gorman 2002, Calin 2004) did not have any events in either treatment or control groups
[6] Brandt 2003 and Gorman 2002 did not have any events in either treatment or control groups.
[7] ADA = adalimumab; PL = placebo; SI = sacroiliac joint

Additional tables

Table 4: Serious adverse event definitions and methods of AE assessment by study

Study ID	Definition of serious adverse event (SAE)	Method of assessment of adverse event (AE)
Etanercept		
Barkham 2008 (AB)	not reported	not reported
Braun 2008b (AB)	not reported	not reported
Brandt 2003 (FT)	not reported	not reported
Calin 2004 (FT)	not reported	patients were "monitored for AE and abnormal laboratory tests over the course of the study"
Davis 2003 (FT)	Adverse events graded on a scale derived from National Cancer Institute Common Toxicity Criteria.	From original study "the types and intensities of adverse events and infections reported during the study were recorded"
(Davis 2005 LTE) (FT)	Adverse events graded on a scale derived from National Cancer Institute Common Toxicity Criteria. SAE defined as "any events that resulted in death, a life threatening adverse experience, inpatient hospitalisation or prolongation of existing hospitalisation, a persistent or major disability/incapacity, or a congenital anomaly/birth defect."	From LTE: "The following variables were assessed: results of a physical examination, vital signs, haematology and chemistry profiles, urine analysis, antibodies against etanercept, premature discontinuation, adverse events, serious adverse events and death."
Davis 2007 LTE (AB)	No definition	Not reported
Gorman 2002 (FT)	"Adverse events and changes in laboratory values were graded on a scale derived from the Common	"At each visit, the study nurse asked the patients about reactions at the injection site..." "Side effects were

	Toxicity Criteria of the National Cancer Insititute"	monitored at each clinic visit by means of open-ended questions about any problems that had occurred since the previous visit"
van der Heijde 2006b (FT)	SAE not specifically defined but article stated "Medically important infections required hospitalisation or parenteral antimicrobial agents." Non infectious and infectious SAE reported separately.	"safety assessments were based on reports of AE, routine physical examinations and laboratory test results"
Huang 2008 (AB)	"not reported	"safety evaluation included adverse event and routine lab monitoring
Infliximab		
Braun 2002	not reported	not reported
Heldmann 2008 (AB; LTE)	not reported	"patients were asked and examined every 6-8 weeks"
Inman 2010 (FT)	"all adverse events were coded using the MedDRA dictionary of terms (version 9.0)	"safety assessments included incidence and severity of AE, and an evaluation of the relationship of AE to the study drug as assessed by the individual investigator."
vdH 2005 (FT)	no general SAE definition provided, but details on each SAE provided	??"safety assessments included adverse events, infections, infusion reactions, premature discontinuation, and laboratory tests"
Braun 2008 (FT; LTE)	no general SAE definition provided	"safety assessments included adverse events, infections, infusion reactions were made at each visit"(every 6 weeks)
Adalimumab		
Lambert 2007 (FT)	AE not reported in article; SAE definition not provided in abstract	not reported
vdH 2006a (FT)	no general SAE definition, but details provided for each SAE that occurred	adverse events and vital signs assessed at every visit
vdH 2009 (FT; LTE)	no SAE definition	"safety assessments were based on observed data reported following any adalimumab exposure"
vdH 2008 (AB; LTE)	no SAE definition	not reported

Footnotes:
AB=abstract; FT=full text; LTE=long term extension study

Table 5: Adverse events with long-term anti-TNF therapy: results from observational studies

Adverse event	Etanercept: Davis 2005-72 week OLE	Infliximab: Braun2008-72 week	Adalimumab:vdH2009-80 week OLE	

	of 24 week RCT (128 pts continued ETN and 129 from placebo entered OLE: results combined; Results reported for just 72 week OLE)		blinded extension of 24 week RCT (results for N=201 entering extension from original INF group; Results reported for cumulative RCT and OLE: average 94.2 weeks follow-up)		from 24 week RCT; appears N=311 entered OLE; combined original treatment and placebo groups; results reported for cumulative RCT and OLE period for up to 2 years of adalimumab treatment)	
Type of AE - approximately 2yr exposure	Events	# pts/# events	Events	# pts/# events	Events	# pts/# events
Total withdrawals	57/257 (22.2%)	# patients	47/201 (23.4%)	# patients	56/311 (18%)	# patients
Withdrawals due to AE	12/257 (5%)	# patients	NR	-	NR	
Total adverse events	NR	-	196/201 (97.5%)	# patients	293/311 (94.2%)	# patients
Nausea	17/257 (6.6%)	# AE	NR	-	27/311 (8.7%)	# patients
Total serious AE	17/257 (6.6%)	# patients	34/201 (16.9%)	# patients	48/311 (15.4%)	# patients
Any infection	NR	-	158/201 (78.6%)	# patients	213/311 (68.5%)	# patients
Serious infections	5/257 (2%)	# patients	8/201 (4%)	# patients	6/311 (1.9%)	# patients
Tuberculosis	1/257 (0.4%) (PPD+ but no symptoms)	# patients	2/201 (1%) (PPD+ but no symptoms)	# patients	0/311 (0%)	-
Malignancies or lymphoma	0/257 (0%)	-	2/201 (1%) (0 lymphoma)	# patients	4/311 (1.3%) (1 non-Hodgkin's lymphoma)	# patients
Abnormal (clinical) liver function	NR	-	0/201 (0%)	-	0/311 (0%)	-
Injection site/infusion reaction	53/257 (20.6%)	#patients	42/201 (21.4%)	# patients	42/311(13.5%)	# patients
Chronic or Congestive heart failure	NR	-	NR	-	0/311 (0%)	-
Demyelination	0/257 (0%)	-	NR	-	0/311 (0%)	-

Autoimmunity symptoms/ disorders	NR	-	2/201 (1%)	# patients	0/311 (0%)	-
Death	0/257 (0%)	-	0/201 (0%)	-	0/311 (0%)	-
Type of AE - approximately 3 year exposure	Etanercept: Davis 2008-168 week OLE of 24 week RCT (128 pts continued ETN and 129 from placebo entered OLE: results combined; Results reported for just 168 week OLE)				Adalimumab:van der Heijde 2008; OLE from 24 week RCT; appears N=311 entered OLE; combined original treatment and placebo groups; results reported for cumulative RCT and OLE period for up to 2 years of adalimumab treatment; reported as AE rate/100 pt years during 24-week double blind period versus 3 years adalimumab exposure)	
Total withdrawals	131/257 (49%)	# patients			NR	
Withdrawals due to AE	21/257 (8.2%)	# patients			3.6 vs 3.8 /100 pt yrs	
Total adverse events	234/257 (91.1%)	# patients			NR	
Nausea	NR	-			NR	
Total serious AE	33/257 (12.8%)	# patients			10.2 vs 11.1 /100 pt yrs	
Any infection	187/257 (72.8%)	# patients			NR	
Serious infections	6/257 (2.3%)	# patients			0.0 vs 1.4 /100 pt yrs	
Tuberculosis	1/257 (0.4%)	# patients			NR	
Malignancy or lymphoma	0	-			0.0 vs 0.7 /100 pt yrs	
Abnormal liver function	NR	-			NR	
Injection site reaction	57/257 (22.2%)	# patients			NR	
Chronic or congestive heart failure	NR	-			NR	

Demyelination	0	-			NR	
Autoimmunity symptoms/ disorders	0	-			NR	
Death	0/257 (0%)	-			1/311 (0.3%)	
Type of AE - approximately 5 year total exposure			Infliximab: Heldmann 2009: EASIC, a 2 year OLE after 2 yr ASSERT +1.3yr +lag phase; 89 entered EASIC after infliximab treatment during lag phase Results presented for the 2 years of the EASIC study			
Total withdrawals			NR	-		
Withdrawals due to AE			NR	-		
Total adverse events			75/89 (84.3%)	# patients		
Nausea			NR	-		
Total serious AE			NR	-		
Any infection			230 total	# events		
Serious infections			3	# events		
Tuberculosis			0	-		
Lymphoma			0	-		
Abnormal liver function			NR	-		
Injection site/infusion reaction			4/89 (4.5%)	# patients		
Chronic or congestive heart failure			NR	-		
Demyelination			NR	-		
Autoimmunity symptoms/ disorders			NR	-		
Death			NR	-		

Footnotes
Table 1 provides definitions for serious adverse events (SAE); OLE=open-label extension; NR=not reported

Table 6: Summary of warnings on the TNF-inhibitors from regulatory agency websites (April 2010)

Regulatory agency/source	# relevant/ # hits	Summary of Warnings and Conclusions
MedWatch: The FDA Safety Information and Adverse Event Reporting Program	11/16	- Recent warning of increased risk of lymphoma and other cancers associated with the use of TNF blockers in children and adolescents; as well specific warning of leukemia (malignancies already exists). - Label warnings added since 2000 for infliximab: hepatotoxicity; infections (pneumonia specifically added), lymphoma, tuberculosis, and other serious opportunistic infections including histoplasmosis, listeriosis, and pneumocystosis, malignancies. - Label warnings added since 2000 for etanercept: serious infections leading to hospitalization or death, including bacterial sepsis and tuberculosis; recommendation to screen for latent tuberculosis infection before beginning Enbrel; lymphoma and other malignancies, including acute and chronic leukemia - Label warnings added since 2000 for adalimumab: lymphoma and other malignancies; skin reactions: new or worsening psoriasis (all sub-types including pustular and palmoplantar) ; serious infections with the combined use of Humira (adalimumab) and anakinra, hypersensitivity reactions, including anaphylaxis, and hematologic events, including pancytopenia and aplastic anemia.
European Medicines Agency (EMEA)	5/73	Since 2000: - Public statements on Infliximab (Remicade) on the increased incidence of mortality and hospitalisation for worsening Congestive Heart Failure and increased tuberculosis infections - Humira should not be used in people who may be hypersensitive (allergic) to adalimumab or any of the other ingredients. Humira must not be used in patients with tuberculosis, other severe infections, or moderate to severe heart failure (an inability of the heart to pump enough blood around the body). -Summary for public: Enbrel: risk of serious infections and should not be used in people who may be hypersensitive (allergic) to etanercept or

		any of the other ingredients. Enbrel must not be used in patients who have or are at risk of sepsis (when bacteria and toxins circulate in the blood and start to damage the organs), or in patients with infections. Before using Enbrel, doctors must check that the patient is free of infections including tuberculosis.
UK MHRA Medicines and Healthcare products Regulatory Agency: Drug Safety Updates (formerly Current Problems in Pharmacovigilance)	4/292	- 2001: states infliximab patients should be screened for TB - July 2008: letter to health care providers re: reports of hepatosplenic T-cell lymphoma in patients treated with HUMIRA® (adalimumab) - congestive cardiac failure, cardiomyopathy, the frequency of blood dyscrasias, demyelination, infections, adult respiratory distress syndrome and TB should be kept under close monitoring by the MA (marketing authorization) Holder
Australian Adverse Drug Reactions Bulletin	2/72	- Drug-induced lupus erythematosus (June 2009): An emerging association with TNF inhibitors - TNA-alpha inhibitors (Dec. 2006): While extremely effective, TNFalpha inhibitors are associated with several serious reactions. These include: · Hypersensitivity reactions - immediately post-injection or delayed; · Serious and life-threatening infection and sepsis; · Recrudescence of tuberculosis and other granulomatous diseases; · Reactivation of hepatitis B; · Malignancy, including lymphoma; · Haematological reactions such as pancytopenia and aplastic anaemia; · Autoimmunity - eg, drug-induced lupus; · CNS reactions, including demyelinating disorders and seizures; · New-onset heart failure or worsening of advanced heart failure.

Figures

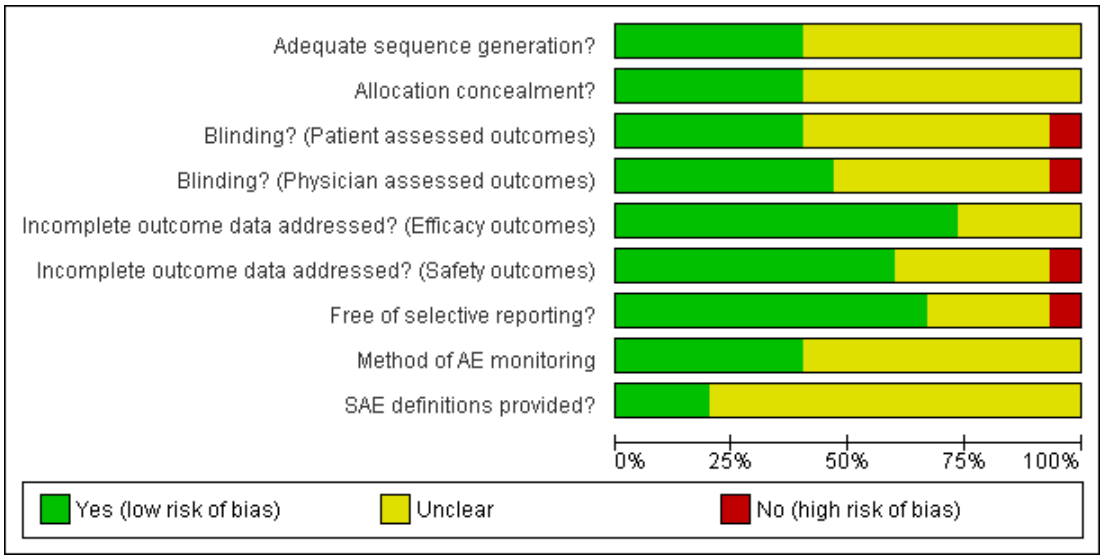


Figure 1: Methodological quality graph: review authors' judgements about each methodological quality item presented as percentages across all included studies.

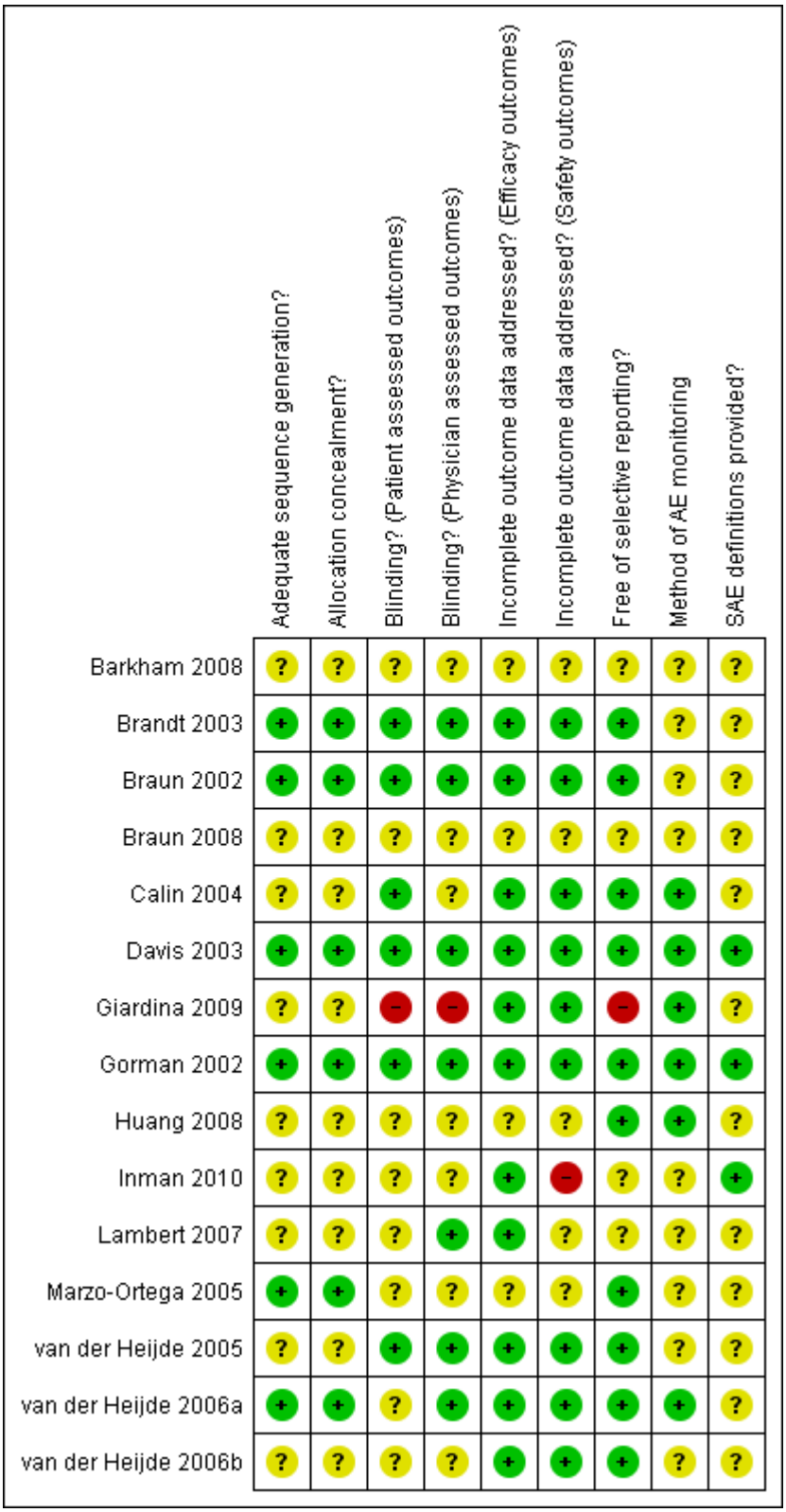


Figure 2: Methodological quality summary: review authors' judgements about each methodological quality item for each included study.

Relationship of comparators in anti-TNF for ankylosing spondylitis systematic review

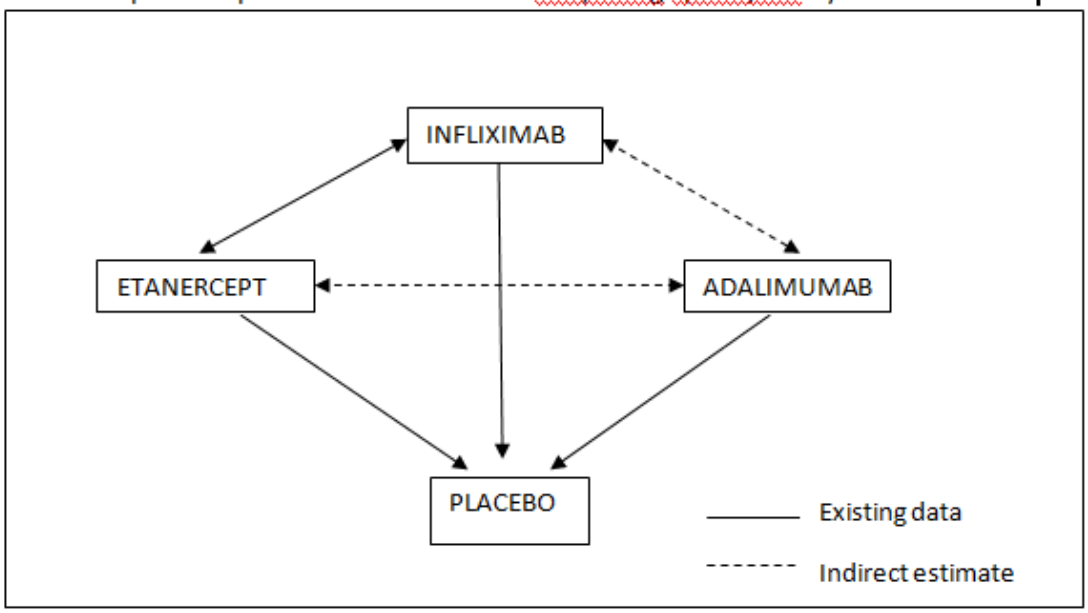


Figure 3: Relationship of comparators included in systematic review

Demographic and clinical characteristics of studies included in the indirect comparisons analysis

Study ID	# of patients		Endpoint, weeks	Age† (yrs, SD)	% male†	Disease duration† (yrs, SD)	Baseline BASDAI (SD)†	Baseline BASFI (SD)†
	Tx	Pl						
ETANERCEPT VS PLACEBO								
Brandt03	14	16	6	39.8 (9.1)	71.4	14.9 (8.3)	6.5(1.2)	6.2(1.8)
Calin04	45	39	12	45.3 (9.5)	80	15(8.8)	61	60.2
Davis03	138	139	24	42.1	76	10.1	58.1	51.7
VDH2006b	150	51	12,24	39.8(10.7)	76	10 (9.1)	59.4(16.7)	57.7 (20.1)
Gorman02	20	20	16	38(10)	65	15(10)	NR	4.5(2.1)
INFLIXIMAB VS PLACEBO								
Braun2002	34	35	12	40.6 (8)	68	16.4 (8.3)	6.5(1.2)	5.1 (2.2)
VDH2005	201	78	24	40 (32,47)*	78	10.1	6.6 (5.3, 7.6)*	5.7(4.5, 7.1)*
ADALIMUMAB VS PLACEBO								
VDH2006a	208	107	12,24	41.7(11.69)	75.5	11.3 (10)	6.3(1.7)	5.2(2.2)
Lambert07	38	40	12	41.9 (11.1)	76	14.5 (9)	6.2 (1.7)	5.3(2.0)
ETANERCEPT VS INFLIXIMAB								
Giardina09	25	25	6, 104	31.9 (9.2)	76	15.4 (10.6)	6.5 (1.2)	6.1 (0.9)

Note:Almost all of trials excluded those with complete fusion of the spine (VDH2006 a restricted to <10% of recruited patients; not stated Lambert2007)

Mean values reported except *=median (IQR)
† values in the treatment group
Tx=treatment; Pl=placebo
NR=not reported
IQR=interquartile range
VDH=van der Heijde

Figure 4: Demographic and clinical characteristics of studies included in indirect comparison analysis

Outcomes	DRUG1	DRUG2	GLIMMIX			MTC			ORIGINAL (OR, RE)			Favours
			OR	LCL	UCL	OR	LCL	UCL	OR	LCL	UCL	
ASAS 5/6	ADALIMUM	Placebo	8.039	0.446	145.0	6.320	0.655	24.700	5.85	3.08	11.10	ADA
	ETANERCE	Placebo	25.563	0.991	659.532	8.339	2.812	20.100	24.98	12.97	48.11	ETN
	INFLIXIMAB	Placebo	9.456	0.567	157.700	22.190	3.789	78.610	14.09	5.52	36.02	INF
ASAS 50	ADALIMUM	Placebo	4.243	1.629	11.046	5.914	0.738	22.530	3.60	1.13	11.43	ADA
	ETANERCE	Placebo	6.786	3.856	11.943	8.058	3.013	18.050	7.13	4.15	12.27	ETN
	INFLIXIMAB	Placebo	7.190	3.260	15.856	14.53	3.868	63.920	14.67	3.03	71.08	INF
>50% BASDAI	ADALIMUM	Placebo	4.628	2.493	8.591	6.127	0.873	21.380	4.17	2.29	7.59	ADA
	ETANERCE	Placebo	6.875	3.632	13.012	9.070	2.767	26.460	6.60	2.94	14.83	ETN
	INFLIXIMAB	Placebo	7.717	4.286	13.894	12.590	4.063	33.250	9.53	4.84	18.78	INF
ASAS 40	ADALIMUM	Placebo	3.932	1.943	7.957	9.651	0.448	41.040	4.32	2.31	8.09	ADA
	ETANERCE	Placebo	5.694	2.691	12.046	9.220	0.858	41.950	4.16	1.98	8.71	ETN
	INFLIXIMAB	Placebo	5.692	2.804	11.553	9.261	0.761	33.290	6.60	3.13	13.95	INF
Withdrawals due to AE	ADALIMUM	Placebo	1.573	0.109	22.715	3.358	0.066	19.510	1.03	0.19	5.65	PL (NS)
	ETANERCE	Placebo	4.125	0.507	33.576	7.516	0.914	154.1	4.03	1.43	11.30	PL (NS)
	INFLIXIMAB	Placebo	2.545	0.348	18.629	8.225	0.305	39.120	2.76	0.50	15.22	PL(NS)
All Infections	ADALIMUM	Placebo	1.708	0.731	3.992	2.176	0.884	4.749	1.87	1.16	3.03	PL (NS)
	ETANERCE	Placebo	1.215	0.666	2.215	1.320	0.638	2.320	1.27	0.84	1.91	PL (NS)
	INFLIXIMAB	Placebo	1.114	0.478	2.597	1.028	0.371	2.081	0.90	0.36	2.22	PL (NS)
Serious Infections	ADALIMUM	Placebo	0.216	0.006	7.512	0.759	0.000	5.346	0.05	0.00	3.30	ADA (NS)
	ETANERCE	Placebo	1.247	0.246	6.320	2.663	0.194	12.800	2.25	0.30	17.13	PL (NS)
	INFLIXIMAB	Placebo	2.693	0.517	14.022	7.368	0.339	39.170	5.10	0.44	58.76	PL (NS)

ADA=adalimumab; ETN=etanercept; INF=infliximab; PL=placebo; NS=not significant; OR=odds ratio; LCL=lower confidence limit; UCL=upper confidence limit; RE=random effects

Figure 5: Indirect comparison estimates using GLIMMIX and MTC approaches, as well as the original estimate, for placebo comparisons

Outcomes	DRUG1	DRUG2	GLIMMIX			MTC			ITC			Favours
			OR	LCL	UCL	OR	LCL	UCL	OR	LCL	UCL	
ASAS 5/6	ADALIMUM	ETANERCE	0.315	0.017	5.784	0.287	0.071	4.223	0.23	0.09	0.59	ETN
	ADALIMUM	INFLIXIMAB	0.850	0.076	9.514	0.104	0.022	2.714	0.42	0.13	1.29	INF
	ETANERCE	INFLIXIMAB	2.703	0.159	45.991	0.336	0.099	2.008	1.77	0.56	5.56	NC*
ASAS 50	ADALIMUM	ETANERCE	0.625	0.210	1.865	0.311	0.080	4.037	0.51	0.14	1.81	ETN
	ADALIMUM	INFLIXIMAB	0.590	0.192	1.813	0.123	0.028	2.597	0.25	0.04	1.74	INF
	ETANERCE	INFLIXIMAB	0.944	0.425	2.096	0.372	0.121	1.899	0.49	0.09	2.58	INF
	DIRECT ETN	*INF*				*0.50*	*0.15*	*1.71*				INF
>50% BASDAI	ADALIMUM	ETANERCE	0.673	0.309	1.466	0.317	0.070	3.476	0.63	0.23	1.73	ETN
	ADALIMUM	INFLIXIMAB	0.600	0.285	1.262	0.235	0.056	2.313	0.44	0.18	1.08	INF
	ETANERCE	INFLIXIMAB	0.891	0.454	1.749	0.576	0.209	2.567	0.69	0.24	1.99	INF
	DIRECT ETN	*INF*				*0.62*	*0.19*	*1.99*				INF
ASAS 40	ADALIMUM	ETANERCE	0.691	0.133	3.579	0.198	0.038	14.144	1.04	0.39	2.74	NC*
	ADALIMUM	INFLIXIMAB	0.691	0.124	3.841	0.229	0.047	16.377	0.66	0.25	1.74	INF
	ETANERCE	INFLIXIMAB	1.000	0.325	3.076	0.698	0.168	8.425	0.63	0.22	1.81	INF
	DIRECT ETN	*INF*				*0.62*	*0.20*	*1.89*				INF
Withdrawals due to AE	ADALIMUM	ETANERCE	0.381	0.008	18.315	0.013	0.002	4.484	0.25	0.03	1.87	ADA
	ADALIMUM	INFLIXIMAB	0.618	0.024	15.641	0.024	0.009	14.465	0.37	0.03	4.14	ADA
	ETANERCE	INFLIXIMAB	1.621	0.086	30.506	0.597	0.110	101.143	1.46	0.20	10.75	NC*
All Infections	ADALIMUM	ETANERCE	1.406	0.533	3.712	1.369	0.575	4.822	1.16	0.55	2.46	ETN
	ADALIMUM	INFLIXIMAB	1.534	0.487	4.835	1.759	0.702	7.886	2.08	0.74	5.81	INF
	ETANERCE	INFLIXIMAB	1.091	0.413	2.877	1.151	0.469	4.119	1.79	0.61	5.25	INF
Serious Infections	ADALIMUM	ETANERCE	0.173	0.005	6.359	0.062	0.000	7.849	0.02	0.00	0.76	ADA
	ADALIMUM	INFLIXIMAB	0.080	0.002	2.932	0.032	0.000	3.843	0.01	0.00	0.44	ADA
	ETANERCE	INFLIXIMAB	0.463	0.083	2.574	0.180	0.030	5.305	0.44	0.02	10.55	ETN
	DIRECT ETN	*INF*				*0.50*	*0.05*	*5.03*				ETN

ADA=adalimumab; ETN=etanercept; INF=infliximab; OR=odds ratio; LCL=lower confidence limit; UCL=upper confidence limit; NC*=not consistent

Figure 6: Indirect comparison estimates using GLIMMIX, MTC and ITC approaches, as well as the direct estimate

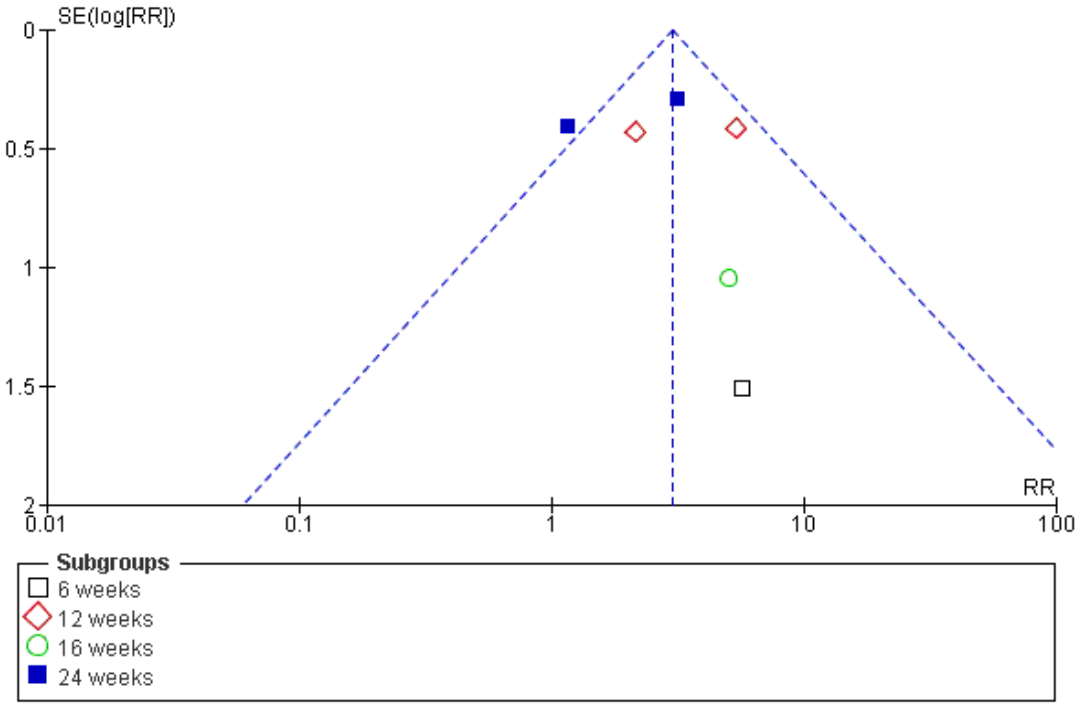


Figure 7 (Analysis 8.3): Funnel plot of comparison: TNF-inhibitors versus placebo, outcome: 8.3 Injection/Infusion site reaction.

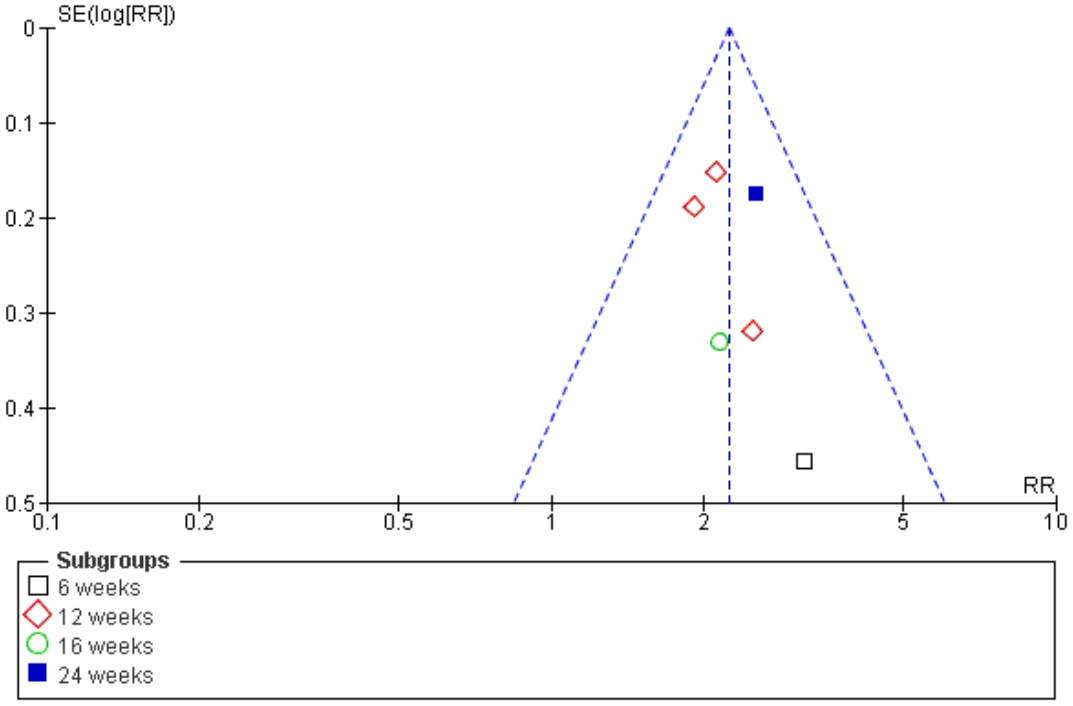


Figure 8 (Analysis 1.17): Funnel plot of comparison: Etanercept (25 mg twice weekly) versus placebo, outcome: 1.17 ASAS 20.

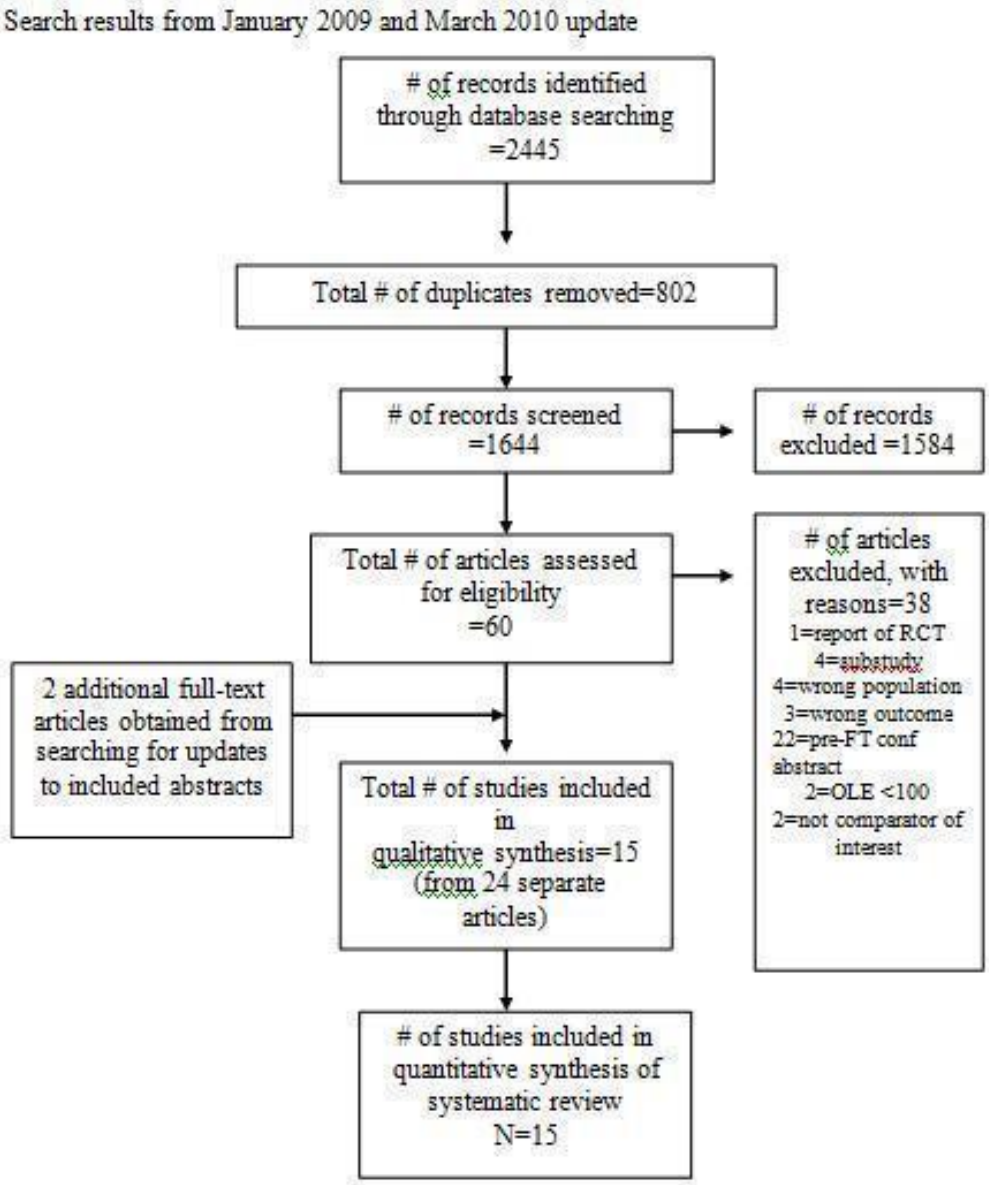


Figure 9: Flowchart of search for RCTs

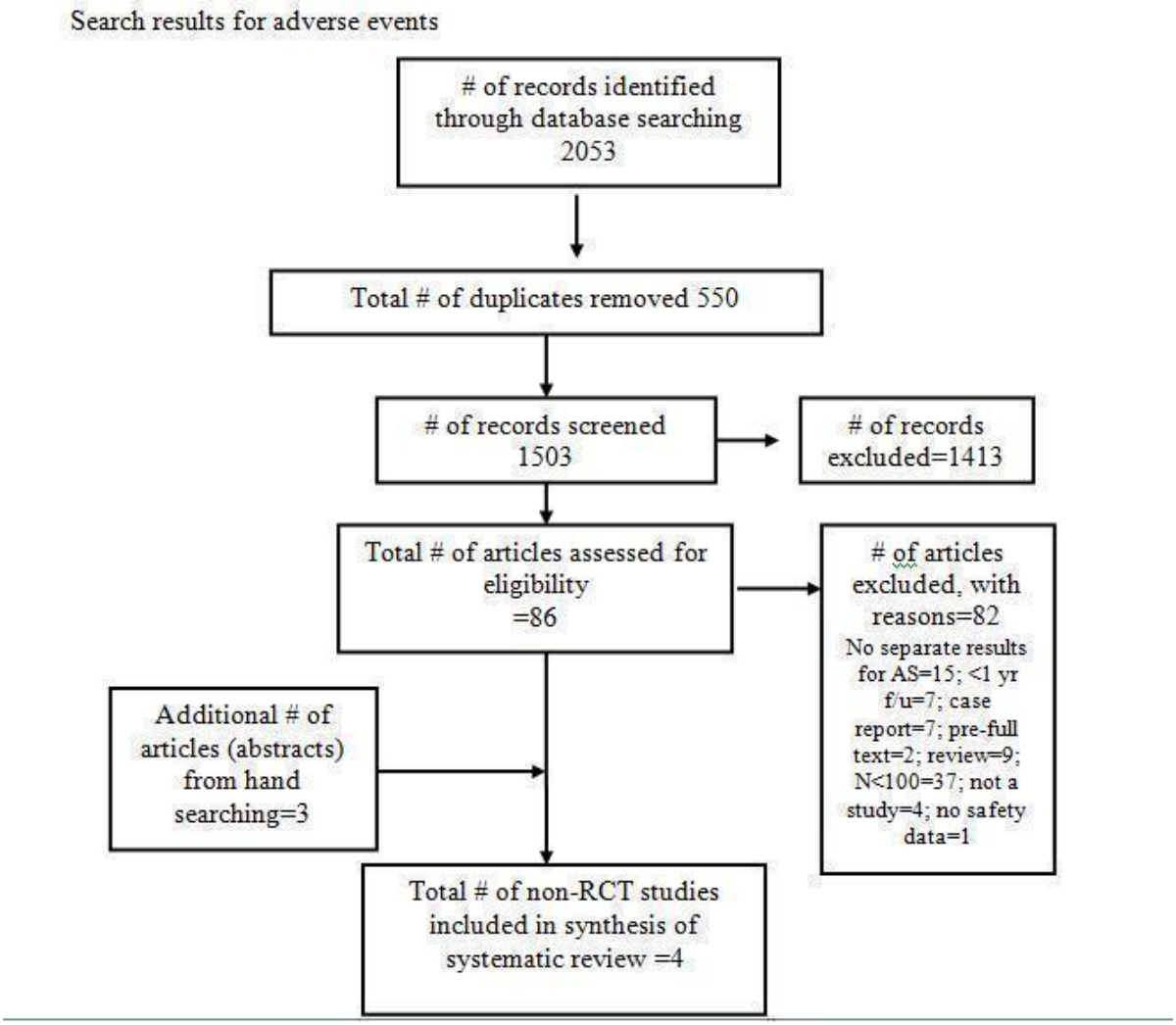


Figure 10: Flowchart for adverse events search for non-RCTs

Table: Assessment of risk of bias in non-RCTs using the Newcastle-Ottawa Scale for cohort studies

NOS Criteria	Study	Judgment:	Study	Judgment	Study	Judgment	Study	Judgment	Study	Judgment	Study	Judgment
Selection	vdH09		vdH08		Braun 08		Held-mann09		Davis 05		Davis08	
Representativeness of the exposed cohort	0	Highly selected for RCT; if a significant AE in RCT phase would not continue	0	Highly selected for RCT; if a significant AE in RCT phase would not continue	0	Highly selected for RCT; if a significant AE in RCT phase would not continue	0	Highly selected for RCT; if a significant AE in RCT phase would not continue	0	Highly selected for RCT; if a significant AE in RCT phase would not continue	0	Highly selected for RCT; if a significant AE in RCT phase would not continue
Selection of the non exposed cohort	N/A	Single cohort; all exposed	N/A	Single cohort; all exposed	N/A	Single cohort; all exposed	NA	Single cohort; all exposed	NA	Single cohort; all exposed	NA	Single cohort; all exposed
Ascertainment of exposure	0	none of the published articles stated how they ascertained that participants in the extension	0	none of the published articles stated how they ascertained that participants in the extension study were taking the anti-TNF therapy as	0	none of the published articles stated how they ascertained that participants in the extension study were taking the anti-TNF therapy as	0	none of the published articles stated how they ascertained that participants in the extension	0	none of the published articles stated how they ascertained that participants in the extension	0	none of the published articles stated how they ascertained that participants in the extension
Demonstration that outcome of interest was not present at start of study	*	Strict exclusion criteria for RCT (e.g TB, history of cardiac, renal, neurologic, psychiatric, endocrinologic metabolic, or hepatic disease; and a history of demyelinating disease, multiple sclerosis, or malignancy)	*	Strict exclusion criteria for RCT (e.g TB, history of cardiac, renal, neurologic, psychiatric, endocrinologic metabolic, or hepatic disease; and a history of demyelinating disease, multiple sclerosis, or malignancy)	*	Strict exclusion criteria for RCT (e.g TB, infection, CHF, multiple sclerosis, malignancy)	0	There was a 1.3 year lag after ASSERT and before starting EASIC so unclear if the outcome wss present at start of EASIC	*	Exclusion criteria for RCT(e.g had a serious infection (requiring hospitalization or IV antibiotics))	*	Exclusion criteria for RCT (e.g had a serious infection (requiring hospitalization or IV antibiotics))
Comparability												
Comparability of cohorts on the basis of the design or analysis	N/A	Nothing controlled for; single cohort	N/A	Nothing controlled for; single cohort	N/A	Nothing controlled for; single cohort	NA	Nothing controlled for; single cohort	NA	Nothing controlled for; single cohort	NA	Nothing controlled for; single cohort
Outcome												
Assessment of outcome	*	No blinding, but AE of interest had objective diagnostics	*	No blinding, but AE of interest had objective diagnostics	*	Only 1 patient unblinded over the 2 years (Fig1) and AE of interest had objective diagnostics	*	No blinding, but AE of interest had objective diagnostics	*	No blinding, but AE of interest had objective diagnostics	*	No blinding, but AE of interest had objective diagnostics
Was follow-up long enough for outcomes to occur	*	Yes, 2 yrs infections and malignancies	*	Yes, 2 yrs infections and malignancies	*	Yes, 2 yrs infections and malignancies	*	Yes, 5 years for infections and malignancies	*	Yes, 1.5 years for infections and malignancies	*	Yes, 3.5 years for infections and malignancies
Adequacy of follow up of cohorts	*	261/311 83.9%received adalimumab treatment to 2 years	0	227/311 (73%) received adalimumab for at least 3 years	*	61/76 (78.2%) and 166/201 (82.6%) completed 102 weeks	0	Not reported	0	200/257 (78%) completed 72 weeks of OLE	0	126/257 (49%) completed 168 weeks of OLE

Figure 11: Assessment of risk of bias in non-RCTs using Newcastle-Ottawa Scale

Appendices
1 Additional search strategies (January 2009 search)

Database: EMBASE <1980 to 2009 Week 04>
Search Strategy:
---
1 exp Spondylitis/ (10156)
2 (ankylosing or spondyl$).mp. (21570)
3 (bekhterev or bechterew).mp. (201)
4 or/1-3 (21620)
5 exp Tumor Necrosis Factor Receptor/ (5586)
6 exp Tumor Necrosis Factor/ (19676)
7 exp Monoclonal Antibody/ (170176)
8 anti-tumo?r necrosis factor$.mp. (1439)
9 anti-tnf.mp. (3283)
10 etanercept.mp. (7693)
11 enbrel.mp. (1723)
12 infliximab.mp. (11164)
13 remicade.mp. (2218)
14 exp adalimumab/ (3580)
15 humira.sh,rn,tw. (966)
16 trudexa.sh,rn,tw. (10)
17 Monoclonal antibody D2e7.rn,tw. (4)
18 or/5-17 (195408)
19 4 and 18 (1592)
20 random$.ti,ab. (388596)
21 factorial$.ti,ab. (8106)
22 (crossover$ or cross over$ or cross-over$).ti,ab. (39123)
23 placebo$.ti,ab. (108834)
24 (doubl$ adj blind$).ti,ab. (84015)
25 (singl$ adj blind$).ti,ab. (7379)
26 assign$.ti,ab. (107363)
27 allocat$.ti,ab. (33957)
28 volunteer$.ti,ab. (98245)
29 crossover procedure.sh. (20893)
30 double blind procedure.sh. (71118)
31 randomized controlled trial.sh. (164870)
32 single blind procedure.sh. (7912)
33 or/20-32 (652278)
34 exp animal/ or nonhuman/ or exp animal experiment/ (3407483)

35 exp Spondylitis/ (10156)
36 (ankylosing or spondyl$).mp. (21570)
37 (bekhterev or bechterew).mp. (201)
38 or/35-37 (21620)
39 exp Tumor Necrosis Factor Receptor/ (5586)
40 exp Tumor Necrosis Factor/ (19676)
41 exp Monoclonal Antibody/ (170176)
42 anti-tumo?r necrosis factor$.mp. (1439)
43 anti-tnf.mp. (3283)
44 etanercept.mp. (7693)
45 enbrel.mp. (1723)
46 infliximab.mp. (11164)
47 remicade.mp. (2218)
48 exp adalimumab/ (3580)
49 humira.sh,rn,tw. (966)
50 trudexa.sh,rn,tw. (10)
51 Monoclonal antibody D2e7.rn,tw. (4)
52 or/39-51 (195408)
53 38 and 52 (1592)
54 random$.ti,ab. (388596)
55 factorial$.ti,ab. (8106)
56 (crossover$ or cross over$ or cross-over$).ti,ab. (39123)
57 placebo$.ti,ab. (108834)
58 (doubl$ adj blind$).ti,ab. (84015)
59 (singl$ adj blind$).ti,ab. (7379)
60 assign$.ti,ab. (107363)
61 allocat$.ti,ab. (33957)
62 volunteer$.ti,ab. (98245)
63 crossover procedure.sh. (20893)
64 double blind procedure.sh. (71118)
65 randomized controlled trial.sh. (164870)
66 single blind procedure.sh. (7912)
67 or/54-66 (652278)
68 53 and 67 (230)
69 from 68 keep 1 (1)

Database: EBM Reviews - Cochrane Central Register of Controlled Trials <4th Quarter 2008> Search Strategy:
---
1 exp Spondylarthropathies/ (363)
2 (ankylosing or spondyl$).ti,ab. (668)
3 (bekhterev or bechterew).ti,ab. (4)
4 or/1-3 (790)
5 exp Receptors, Tumor Necrosis Factor/ (417)
6 exp Tumor Necrosis Factors/ (1445)
7 exp Antibodies, Monoclonal/ (1920)
8 anti-tumor necrosis factor$.ti,ab. (88)

9 antitumor necrosis factor$.ti,ab. (2)
10 anti-tnf.ti,ab. (121)
11 antitnf.ti,ab. (1)
12 etanercept.ti,ab. (224)
13 enbrel.ti,ab. (25)
14 infliximab.ti,ab. (256)
15 remicade.ti,ab. (10)
16 adalimumab.ti,ab. (73)
17 humira.ti,ab. (4)
18 6 or 11 or 7 or 9 or 17 or 12 or 15 or 14 or 8 or 16 or 10 or 13 or 5 (3597)
19 4 and 18 (89)
20 from 19 keep 1 (1)

Database: CLEED, CDSR, ACP Journal Club, DARE, CLHTA Search Strategy:
---
1 exp Spondylarthropathies/ (40)
2 (ankylosing or spondyl$).ti,ab. (61)
3 (bekhterev or bechterew).ti,ab. (0)
4 or/1-3 (74)
5 exp Receptors, Tumor Necrosis Factor/ (43)
6 exp Tumor Necrosis Factors/ (57)
7 exp Antibodies, Monoclonal/ (237)
8 anti-tumor necrosis factor$.ti,ab. (4)
9 antitumor necrosis factor$.ti,ab. (1)
10 anti-tnf.ti,ab. (9)
11 antitnf.ti,ab. (0)
12 etanercept.ti,ab. (70)
13 enbrel.ti,ab. (2)
14 infliximab.ti,ab. (91)
15 remicade.ti,ab. (5)
16 adalimumab.ti,ab. (30)
17 humira.ti,ab. (4)
18 6 or 11 or 7 or 9 or 17 or 12 or 15 or 14 or 8 or 16 or 10 or 13 or 5 (363)
19 4 and 18 (26)
20 from 19 keep 1-26 (26)
21 from 20 keep 1 (1)

Web of Knowledge 1900-2008

#5 #4 AND #3

#4 TS=(trial* or random* or placebo* or control* or double or treble or triple or blind* or mask* or allocat* or prospective* or volunteer*or comparative or evaluation or follow-up or followup)

#3 #2 AND #1

#2 TS=(tumor necrosis factor* or tumour necrosis factor* or monoclonal antibod* or anti-tnf or antitnf or etanercept or enbrel or infliximab or remicade)

#1 TS=(ankylos* or spondyl*)

CINAHL

S12	S4 and S11
S11	S5 or S6 or S7 or S8 or S9 or S10
S10	TX etanercept or TX enbrel or TX infliximab or TX remicade or TX adalimumab or TX humira
S9	(MH "etanercept+")
S8	TX "anti-tnf" or TX antitnf
S7	TX "anti-tumor necrosis factor*" or TX "anti-tumour necrosis factor*" or TX "antitumor necrosis factor*" or TX "antitumour necrosis factor*"
S6	(MH "Antibodies, Monoclonal")
S5	(MH "Tumor Necrosis Factor")
S4	(S1 or S2 or S3)
S3	TX (bekhterev or bechterew)
S2	TX (ankylosing or spondyl*)
S1	(MH "Spondylarthropathies+")

2 Updated search strategy (2008 to March 2010)

Database: Ovid MEDLINE(R) <1996 to March Week 3 2010> Search Strategy:

---

1 Spondylitis, Ankylosing/ (3321)
2 (bekhterev* or bechterew*).tw. (109)
3 (ankylosing adj (spondyl$ or rheumatoid)).tw. (3373)
4 or/1-3 (4158)
5 etanercept.sh,rn,tw. (2284)
6 enbrel.sh,rn,tw. (132)
7 infliximab.sh,rn,tw. (5159)
8 remicade.sh,rn,tw. (158)
9 adalimumab.sh,rn,tw. (1371)
10 humira.sh,rn,tw. (68)
11 or/5-10 (6982)
12 4 and 11 (561)
13 meta-analysis.mp,pt. (33987)
14 search.tw. (80522)
15 systematic review.tw. (16760)
16 medline.tw. (29002)
17 cochrane database of systematic reviews.jn. (6097

18 or/13-17 (121940)
19 randomized controlled trial.pt. (186954)
20 (randomized or placebo).mp. (304855)
21 clinical trial.pt. (253172)
22 or/19-21 (433732)
23 12 and 18 (21)
24 12 and 22 (196)
25 limit 23 to yr=2008-2010 (7)
26 limit 24 to yr=2008-2010 (52)
27 25 or 26 (56)

Database: EMBASE Search Strategy:

---

1 (bekhterev* or bechterew*).tw. (280)
2 (ankylosing adj (spondyl$ or rheumatoid)).tw. (7621)
3 ankylosing spondylitis/ (9435)
4 or/1-3 (10693)
5 etanercept/ (9466)
6 (enbrel or etanercept).tw. (5389)
7 infliximab/ (13822)
8 (remicade or infliximab).tw. (8690)
9 adalimumab/ (5039)
10 (adalimumab or humira).tw. (3001)
11 or/5-10 (21310)
12 4 and 11 (1702)
13 randomized.tw. (265932)
14 meta-analysis.tw. (31569)
15 systematic review.tw. (26786)
16 or/13-15 (304374)
17 12 and 16 (130)
18 limit 17 to yr=2008-2010 (57)
19 Remove duplicates from MEDLINE (41)

3 Adverse event search strategies

MEDLINE

Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R) 1950 to May Week 2, 2009

1. exp Spondylarthropathies/
2. (ankylosing or spondyl$).ti,ab.
3. (bekhterev or bechterew).ti,ab.
4. or/1-3
5. exp Receptors, Tumor Necrosis Factor/

6. exp Tumor Necrosis Factors/
7. exp Antibodies, Monoclonal/
8. anti-tumor necrosis factor$.ti,ab.
9. antitumor necrosis factor$.ti,ab.
10. anti-tnf.ti,ab.
11. antitnf.ti,ab.
12. etanercept.ti,ab.
13. enbrel.ti,ab.
14. infliximab.ti,ab.
15. remicade.ti,ab.
16. adalimumab.ti,ab.
17. humira.ti,ab.
18. or/5-17
19. 4 and 18
20. (ae or co or to or po or de).fs.
21. (adverse adj (effect$ or reaction$ or event$ or incident$)).tw.
22. toxic$.tw.
23. ((injurious or undesirable) adj (effect$ or reaction$ or event$ or incident$)).tw.
24. safety.tw.
25. ((drug or chemical$) adj induced).tw.
26. or/20-25
27. 19 and 26

Embase 1980 to 2009 Week 20

1. exp Spondylitis/
2. (ankylosing or spondyl$).mp.
3. (bekhterev or bechterew).mp.
4. or/1-3
5. exp Tumor Necrosis Factor Receptor/
6. exp Tumor Necrosis Factor/
7. exp Monoclonal Antibody/
8. anti-tumo?r necrosis factor$.mp.
9. anti-tnf.mp.
10. etanercept.mp.
11. enbrel.mp.
12. infliximab.mp.
13. remicade.mp.
14. exp adalimumab/
15. humira.sh,rn,tw.
16. trudexa.sh,rn,tw.
17. Monoclonal antibody D2e7.rn,tw.
18. or/5-17
19. 4 and 18
20. (ae or si or to or co).fs.
21. (adverse adj (effect$ or reaction$ or event$ or incident$)).tw.

22. toxic$.tw.
23. ((injurious or undesirable) adj (effect$ or reaction$ or event$ or incident$)).tw.
24. safety.tw.
25. ((drug or chemical$) adj induced).tw.
26. or/20-25
27. 19 and 26

Web of Knowledge

#1 ankylos* or spondyl*

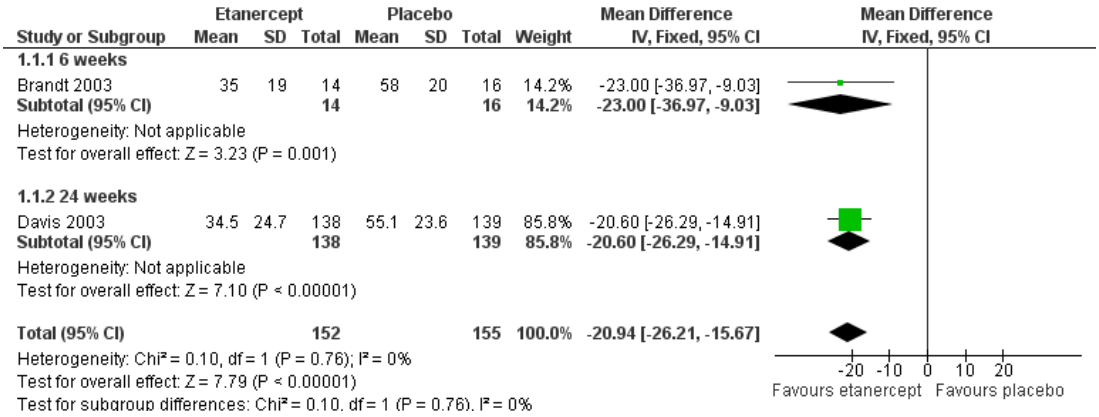
#2 tumor necrosis factor* or tumour necrosis factor* or monoclonal antibod* or anti-tnf or antitnf or etanercept or enbrel or infliximab or remicade

#3 adverse or reaction* or event* or incident* or toxic* or injurious or undesirable or safety or "drug induced" or "chemical* induced"

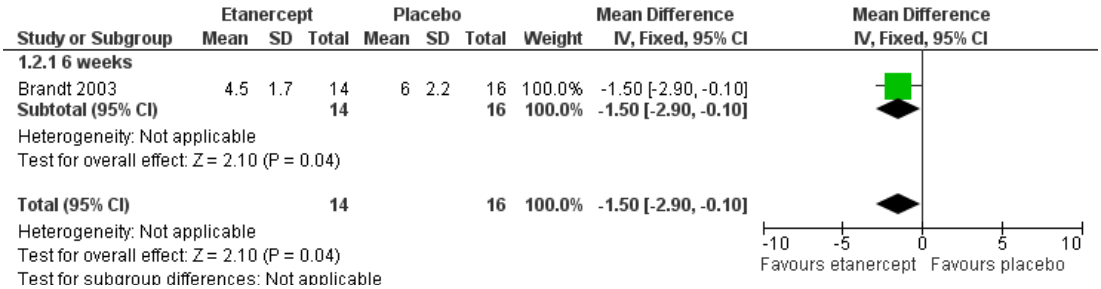
#4 #1 AND #2 AND #3

Forest Plots

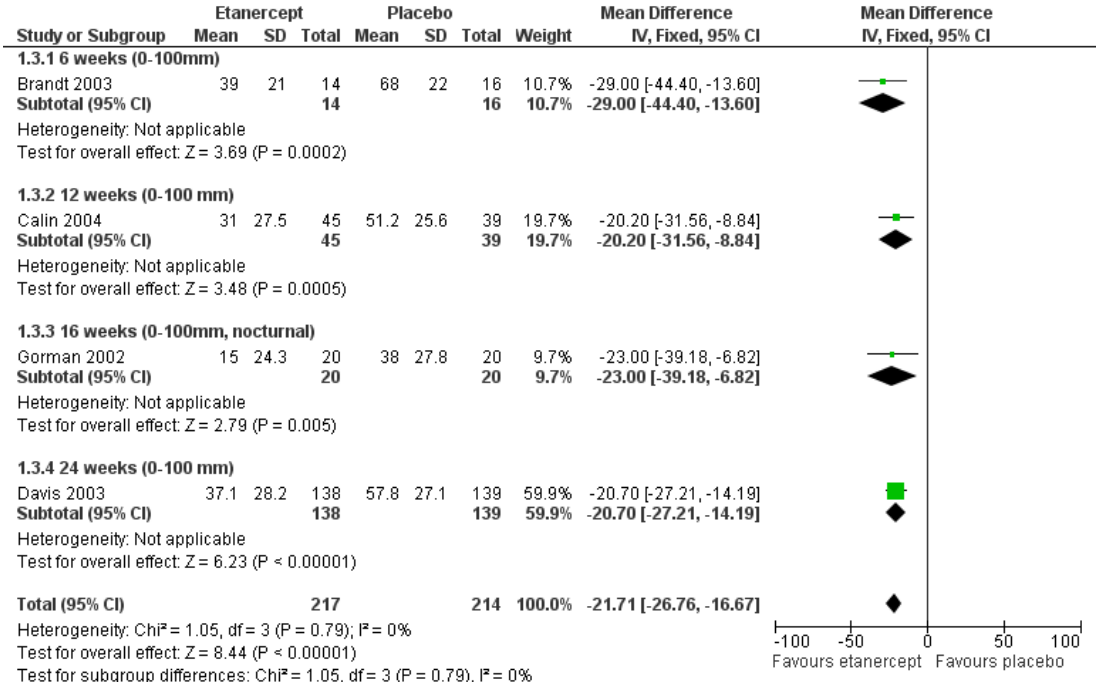
1.1 BASDAI (0-100 scale)



1.2 Fatigue (0-10 scale)

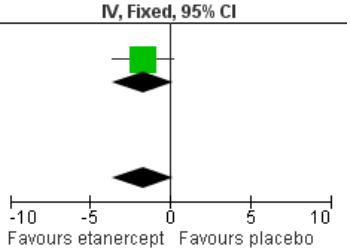


1.3 Spinal pain (0-100 scale)



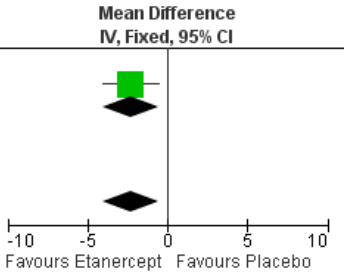
1.4 Peripheral joint pain (0-10 scale)

Study or Subgroup	Etanercept Mean	SD	Total	Placebo Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
1.4.1 6 weeks								
Brandt 2003	3.1	2.4	14	4.8	2.9	16	100.0%	-1.70 [-3.60, 0.20]
Subtotal (95% CI)			**14**			**16**	**100.0%**	**-1.70 [-3.60, 0.20]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 1.76 (P = 0.08)								
Total (95% CI)			**14**			**16**	**100.0%**	**-1.70 [-3.60, 0.20]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 1.76 (P = 0.08)								
Test for subgroup differences: Not applicable								



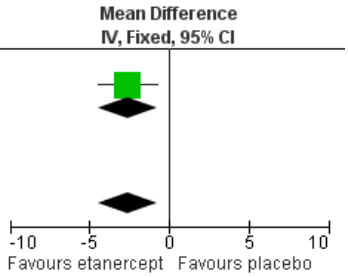
1.5 Entheseal pain (0-10 scale)

Study or Subgroup	Etanercept Mean	SD	Total	Placebo Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
1.5.1 6 weeks								
Brandt 2003	3.6	2.4	14	5.9	2.6	16	100.0%	-2.30 [-4.09, -0.51]
Subtotal (95% CI)			**14**			**16**	**100.0%**	**-2.30 [-4.09, -0.51]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 2.52 (P = 0.01)								
Total (95% CI)			**14**			**16**	**100.0%**	**-2.30 [-4.09, -0.51]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 2.52 (P = 0.01)								
Test for subgroup differences: Not applicable								



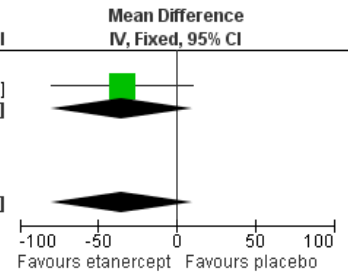
1.6 Morning stiffness - severity

Study or Subgroup	Etanercept Mean	SD	Total	Placebo Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
1.6.1 6 weeks								
Brandt 2003	3.29	2.3	14	5.9	3	16	100.0%	-2.61 [-4.51, -0.71]
Subtotal (95% CI)			**14**			**16**	**100.0%**	**-2.61 [-4.51, -0.71]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 2.69 (P = 0.007)								
Total (95% CI)			**14**			**16**	**100.0%**	**-2.61 [-4.51, -0.71]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 2.69 (P = 0.007)								
Test for subgroup differences: Not applicable								



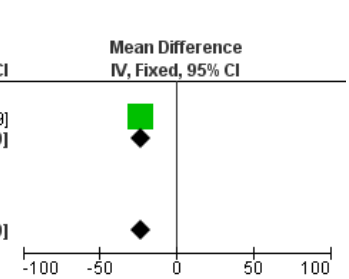
1.7 Morning stiffness - duration (minutes)

Study or Subgroup	Etanercept Mean	SD	Total	Placebo Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
1.7.1 16 weeks								
Gorman 2002	25	78.9	20	60	65	20	100.0%	-35.00 [-79.80, 9.80]
Subtotal (95% CI)			**20**			**20**	**100.0%**	**-35.00 [-79.80, 9.80]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 1.53 (P = 0.13)								
Total (95% CI)			**20**			**20**	**100.0%**	**-35.00 [-79.80, 9.80]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 1.53 (P = 0.13)								
Test for subgroup differences: Not applicable								



1.8 Morning stiffness - mean of severity and duration

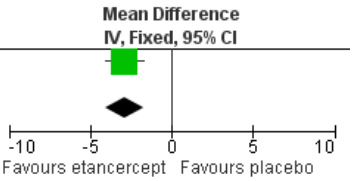
Study or Subgroup	Treatment Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
1.8.1 24 weeks								
Davis 2003	33.4	27	138	56.6	24.8	139	100.0%	-23.20 [-29.31, -17.09]
Subtotal (95% CI)			**138**			**139**	**100.0%**	**-23.20 [-29.31, -17.09]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 7.45 (P < 0.00001)								
Total (95% CI)			**138**			**139**	**100.0%**	**-23.20 [-29.31, -17.09]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 7.45 (P < 0.00001)								
Test for subgroup differences: Not applicable								



1.9 Morning stiffness - duration (VAS)

Study or Subgroup	Treatment Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Brandt 2003	1.4	0.8	14	4.3	2.3	16	100.0%	-2.90 [-4.10, -1.70]
Total (95% CI)			**14**			**16**	**100.0%**	**-2.90 [-4.10, -1.70]**

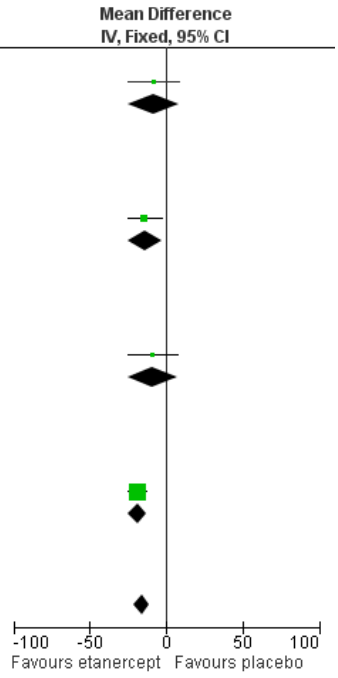
Heterogeneity: Not applicable
Test for overall effect: Z = 4.73 (P < 0.00001)



1.10 BASFI (0-100 scale, none to severe limitations)

Study or Subgroup	Etanercept Mean	SD	Total	Placebo Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
1.10.1 6 weeks								
Brandt 2003	43	23	14	51	24	16	8.3%	-8.00 [-24.84, 8.84]
Subtotal (95% CI)			**14**			**16**	**8.3%**	**-8.00 [-24.84, 8.84]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.93 (P = 0.35)								
1.10.2 12 weeks								
Calin 2004	39.6	28.3	45	53.9	24	39	18.9%	-14.30 [-25.48, -3.12]
Subtotal (95% CI)			**45**			**39**	**18.9%**	**-14.30 [-25.48, -3.12]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 2.51 (P = 0.01)								
1.10.3 16 weeks								
Gorman 2002	22	21	20	31	30	20	9.2%	-9.00 [-25.05, 7.05]
Subtotal (95% CI)			**20**			**20**	**9.2%**	**-9.00 [-25.05, 7.05]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 1.10 (P = 0.27)								
1.10.4 24 weeks								
Davis 2003	36	25.8	138	54.7	25.9	139	63.7%	-18.70 [-24.79, -12.61]
Subtotal (95% CI)			**138**			**139**	**63.7%**	**-18.70 [-24.79, -12.61]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 6.02 (P < 0.00001)								
Total (95% CI)			**217**			**214**	**100.0%**	**-16.09 [-20.95, -11.23]**

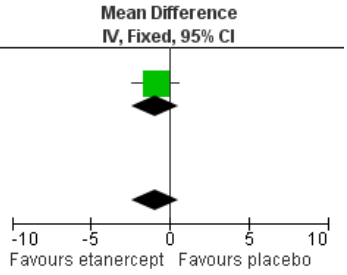
Heterogeneity: Chi² = 2.44, df = 3 (P = 0.49); I² = 0%
Test for overall effect: Z = 6.49 (P < 0.00001)
Test for subgroup differences: Chi² = 2.44, df = 3 (P = 0.49), I² = 0%



1.11 BASMI

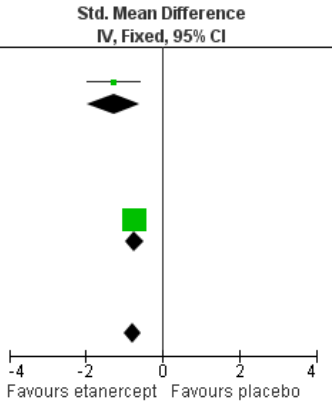
Study or Subgroup	Etanercept Mean	SD	Total	Placebo Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
1.11.1 6 weeks								
Brandt 2003	2.64	1.91	14	3.53	2.26	16	100.0%	-0.89 [-2.38, 0.60]
Subtotal (95% CI)			**14**			**16**	**100.0%**	**-0.89 [-2.38, 0.60]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 1.17 (P = 0.24)								
Total (95% CI)			**14**			**16**	**100.0%**	**-0.89 [-2.38, 0.60]**

Heterogeneity: Not applicable
Test for overall effect: Z = 1.17 (P = 0.24)
Test for subgroup differences: Not applicable



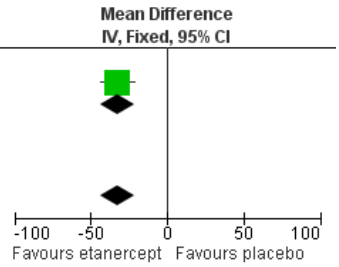
1.12 Patient global assessment

Study or Subgroup	Etanercept Mean	SD	Total	Placebo Mean	SD	Total	Weight	Std. Mean Difference IV, Fixed, 95% CI
1.12.1 16 weeks (disease activity 0-5 scale)								
Gorman 2002	2	0.6	20	3	0.9	20	11.2%	-1.28 [-1.97, -0.59]
Subtotal (95% CI)			**20**			**20**	**11.2%**	**-1.28 [-1.97, -0.59]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 3.66 (P = 0.0003)								
1.12.2 24 weeks (disease activity 0-100 VAS)								
Davis 2003	36	28.2	138	56.3	27.1	139	88.8%	-0.73 [-0.98, -0.49]
Subtotal (95% CI)			**138**			**139**	**88.8%**	**-0.73 [-0.98, -0.49]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 5.90 (P < 0.00001)								
Total (95% CI)			**158**			**159**	**100.0%**	**-0.79 [-1.02, -0.56]**
Heterogeneity: Chi² = 2.18, df = 1 (P = 0.14); I² = 54%								
Test for overall effect: Z = 6.78 (P < 0.00001)								
Test for subgroup differences: Chi² = 2.18, df = 1 (P = 0.14), I² = 54.2%								



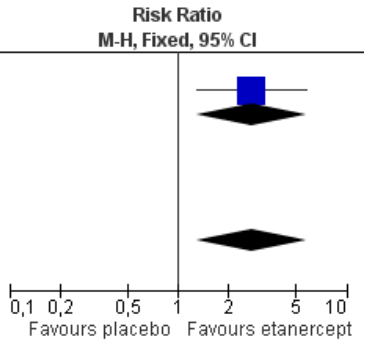
1.13 Physician global assessment

Study or Subgroup	Etanercept Mean	SD	Total	Placebo Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
1.13.1 16 weeks (0-100 mm)								
Gorman 2002	23	10.6	20	55.5	22.7	20	100.0%	-32.50 [-43.48, -21.52]
Subtotal (95% CI)			**20**			**20**	**100.0%**	**-32.50 [-43.48, -21.52]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 5.80 (P < 0.00001)								
Total (95% CI)			**20**			**20**	**100.0%**	**-32.50 [-43.48, -21.52]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 5.80 (P < 0.00001)								
Test for subgroup differences: Not applicable								

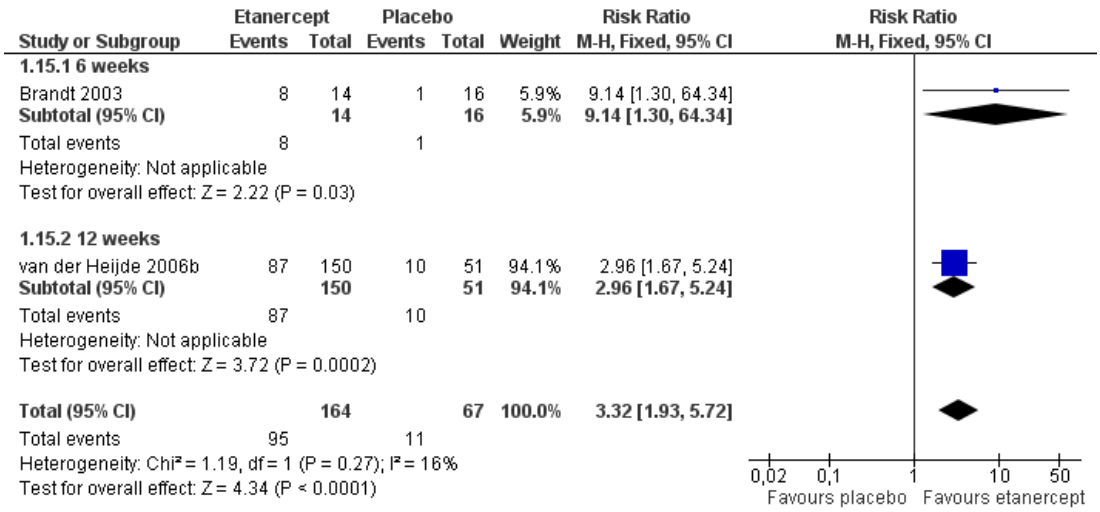


1.14 >20% improvement in BASDAI

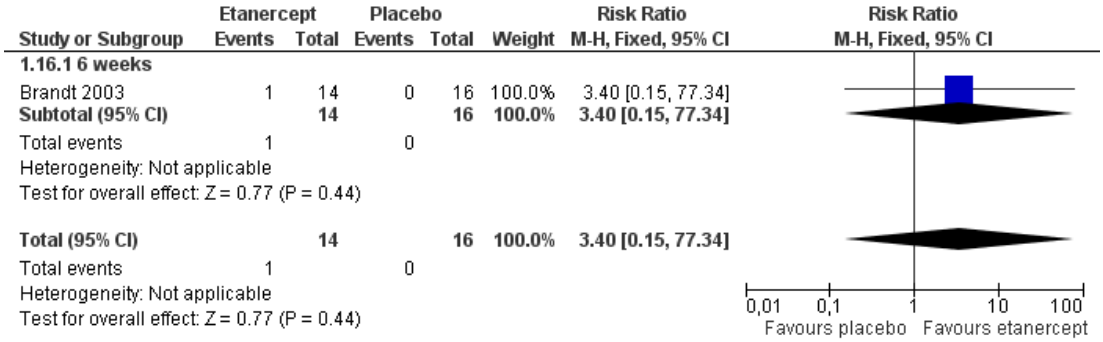
Study or Subgroup	Etanercept Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
1.14.1 6 weeks						
Brandt 2003	12	14	5	16	100.0%	2.74 [1.29, 5.85]
Subtotal (95% CI)		**14**		**16**	**100.0%**	**2.74 [1.29, 5.85]**
Total events	12		5			
Heterogeneity: Not applicable						
Test for overall effect: Z = 2.61 (P = 0.009)						
Total (95% CI)		**14**		**16**	**100.0%**	**2.74 [1.29, 5.85]**
Total events	12		5			
Heterogeneity: Not applicable						
Test for overall effect: Z = 2.61 (P = 0.009)						



1.15 >50% improvement in BASDAI

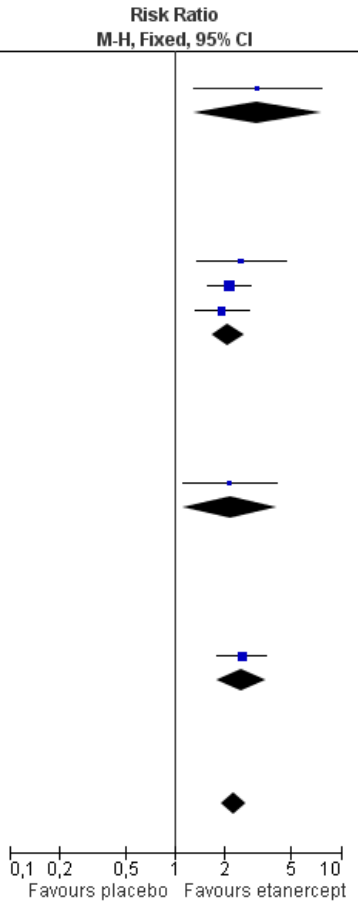


1.16 >70% improvement in BASDAI



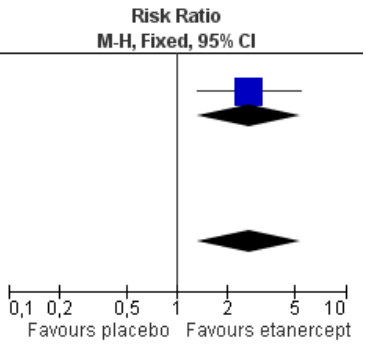
1.17 ASAS 20

Study or Subgroup	Etanercept Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
1.17.1 6 weeks						
Brandt 2003	11	14	4	16	3.2%	3.14 [1.29, 7.67]
Subtotal (95% CI)		**14**		**16**	**3.2%**	**3.14 [1.29, 7.67]**
Total events	11		4			
Heterogeneity: Not applicable						
Test for overall effect: Z = 2.52 (P = 0.01)						
1.17.2 12 weeks						
Calin 2004	26	45	9	39	8.1%	2.50 [1.34, 4.68]
Davis 2003	82	138	39	139	32.8%	2.12 [1.57, 2.86]
van der Heijde 2006b	107	150	19	51	23.9%	1.91 [1.32, 2.77]
Subtotal (95% CI)		**333**		**229**	**64.9%**	**2.09 [1.68, 2.60]**
Total events	215		67			
Heterogeneity: Chi² = 0.54, df = 2 (P = 0.76); I² = 0%						
Test for overall effect: Z = 6.62 (P < 0.00001)						
1.17.3 16 weeks						
Gorman 2002	15	20	7	20	5.9%	2.14 [1.12, 4.10]
Subtotal (95% CI)		**20**		**20**	**5.9%**	**2.14 [1.12, 4.10]**
Total events	15		7			
Heterogeneity: Not applicable						
Test for overall effect: Z = 2.30 (P = 0.02)						
1.17.4 24 weeks						
Davis 2003	78	138	31	139	26.1%	2.53 [1.80, 3.57]
Subtotal (95% CI)		**138**		**139**	**26.1%**	**2.53 [1.80, 3.57]**
Total events	78		31			
Heterogeneity: Not applicable						
Test for overall effect: Z = 5.31 (P < 0.00001)						
Total (95% CI)		**505**		**404**	**100.0%**	**2.24 [1.88, 2.67]**
Total events	319		109			
Heterogeneity: Chi² = 2.02, df = 5 (P = 0.85); I² = 0%						
Test for overall effect: Z = 9.10 (P < 0.00001)						



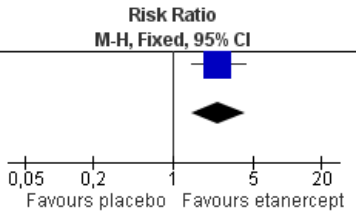
1.18 ASAS 20% 3/5

Study or Subgroup	Treatment Events	Total	Control Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
1.18.1 16 weeks						
Gorman 2002	16	20	6	20	100.0%	2.67 [1.32, 5.39]
Subtotal (95% CI)		**20**		**20**	**100.0%**	**2.67 [1.32, 5.39]**
Total events	16		6			
Heterogeneity: Not applicable						
Test for overall effect: Z = 2.73 (P = 0.006)						
Total (95% CI)		**20**		**20**	**100.0%**	**2.67 [1.32, 5.39]**
Total events	16		6			
Heterogeneity: Not applicable						
Test for overall effect: Z = 2.73 (P = 0.006)						



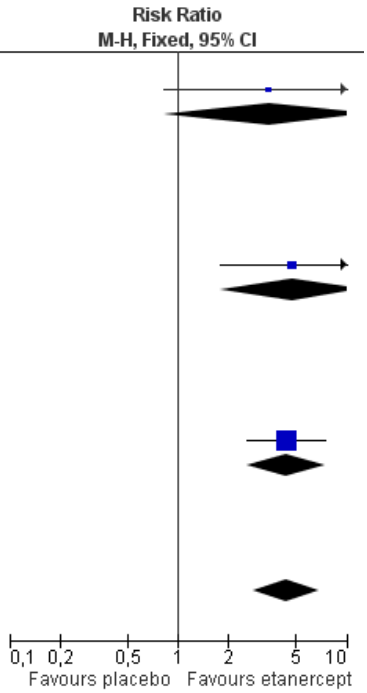
1.19 ASAS 40

Study or Subgroup	Etanercept Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
van der Heijde 2006b	80	150	11	51	100.0%	2.47 [1.43, 4.26]
Total (95% CI)		**150**		**51**	**100.0%**	**2.47 [1.43, 4.26]**
Total events	80		11			
Heterogeneity: Not applicable						
Test for overall effect: Z = 3.26 (P = 0.001)						



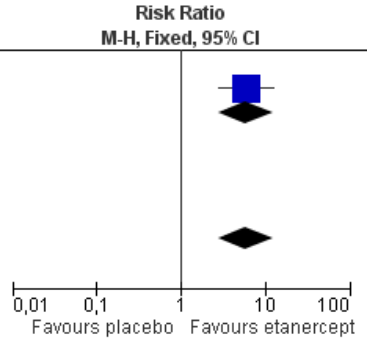
1.20 ASAS 50

Study or Subgroup	Etanercept Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
1.20.1 6 weeks						
Brandt 2003	6	14	2	16	9.3%	3.43 [0.82, 14.33]
Subtotal (95% CI)		**14**		**16**	**9.3%**	**3.43 [0.82, 14.33]**
Total events	6		2			
Heterogeneity: Not applicable						
Test for overall effect: Z = 1.69 (P = 0.09)						
1.20.2 12 weeks						
Calin 2004	22	45	4	39	21.3%	4.77 [1.80, 12.64]
Subtotal (95% CI)		**45**		**39**	**21.3%**	**4.77 [1.80, 12.64]**
Total events	22		4			
Heterogeneity: Not applicable						
Test for overall effect: Z = 3.14 (P = 0.002)						
1.20.3 24 weeks						
Davis 2003	61	138	14	139	69.4%	4.39 [2.58, 7.46]
Subtotal (95% CI)		**138**		**139**	**69.4%**	**4.39 [2.58, 7.46]**
Total events	61		14			
Heterogeneity: Not applicable						
Test for overall effect: Z = 5.46 (P < 0.00001)						
Total (95% CI)		**197**		**194**	**100.0%**	**4.38 [2.81, 6.83]**
Total events	89		20			
Heterogeneity: Chi² = 0.14, df = 2 (P = 0.93); I² = 0%						
Test for overall effect: Z = 6.52 (P < 0.00001)						



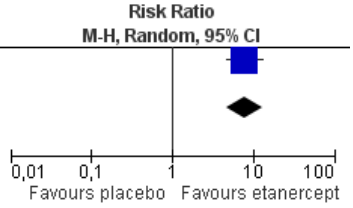
1.21 ASAS 70

Study or Subgroup	Etanercept Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
1.21.1 24 weeks						
Davis 2003	41	138	7	139	100.0%	5.90 [2.74, 12.69]
Subtotal (95% CI)		**138**		**139**	**100.0%**	**5.90 [2.74, 12.69]**
Total events	41		7			
Heterogeneity: Not applicable						
Test for overall effect: Z = 4.54 (P < 0.00001)						
Total (95% CI)		**138**		**139**	**100.0%**	**5.90 [2.74, 12.69]**
Total events	41		7			
Heterogeneity: Not applicable						
Test for overall effect: Z = 4.54 (P < 0.00001)						



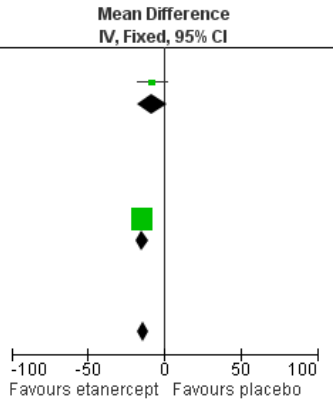
1.22 ASAS 5/6 - 24 weeks

Study or Subgroup	Etanercept Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Random, 95% CI
van der Heijde 2006b	108	150	14	150	100.0%	7.71 [4.64, 12.83]
Total (95% CI)		**150**		**150**	**100.0%**	**7.71 [4.64, 12.83]**
Total events	108		14			
Heterogeneity: Not applicable						
Test for overall effect: Z = 7.87 (P < 0.00001)						



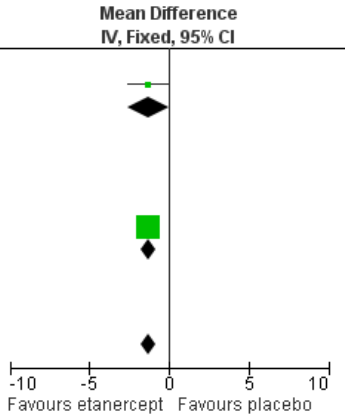
1.23 ESR

Study or Subgroup	Etanercept Mean	SD	Total	Placebo Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
1.23.1 16 weeks (mm/hr)								
Gorman 2002	8.5	12.8	20	16.5	18.7	20	15.2%	-8.00 [-17.93, 1.93]
Subtotal (95% CI)			**20**			**20**	**15.2%**	**-8.00 [-17.93, 1.93]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 1.58 (P = 0.11)								
1.23.2 24 weeks (mm/hr)								
Davis 2003	11.2	11.7	138	25.9	22.4	139	84.8%	-14.70 [-18.90, -10.50]
Subtotal (95% CI)			**138**			**139**	**84.8%**	**-14.70 [-18.90, -10.50]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 6.85 (P < 0.00001)								
Total (95% CI)			**158**			**159**	**100.0%**	**-13.68 [-17.55, -9.81]**
Heterogeneity: Chi² = 1.48, df = 1 (P = 0.22); I² = 33%								
Test for overall effect: Z = 6.93 (P < 0.00001)								
Test for subgroup differences: Chi² = 1.48, df = 1 (P = 0.22), I² = 32.6%								



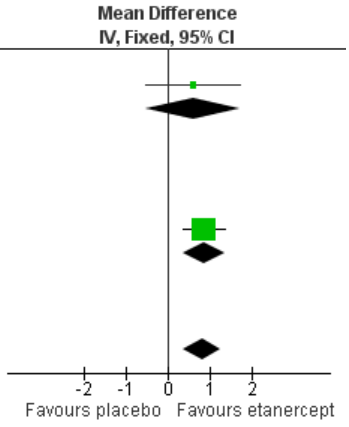
1.24 C-reactive protein

Study or Subgroup	Etanercept Mean	SD	Total	Placebo Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
1.24.1 16 weeks (mg/dl)								
Gorman 2002	0.7	1.1	20	2	2.8	20	10.3%	-1.30 [-2.62, 0.02]
Subtotal (95% CI)			**20**			**20**	**10.3%**	**-1.30 [-2.62, 0.02]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 1.93 (P = 0.05)								
1.24.2 24 weeks (mg/dl)								
Davis 2003	0.6	1.2	138	1.9	2.4	139	89.7%	-1.30 [-1.75, -0.85]
Subtotal (95% CI)			**138**			**139**	**89.7%**	**-1.30 [-1.75, -0.85]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 5.71 (P < 0.00001)								
Total (95% CI)			**158**			**159**	**100.0%**	**-1.30 [-1.72, -0.88]**
Heterogeneity: Chi² = 0.00, df = 1 (P = 1.00); I² = 0%								
Test for overall effect: Z = 6.03 (P < 0.00001)								
Test for subgroup differences: Chi² = 0.00, df = 1 (P = 1.00), I² = 0%								



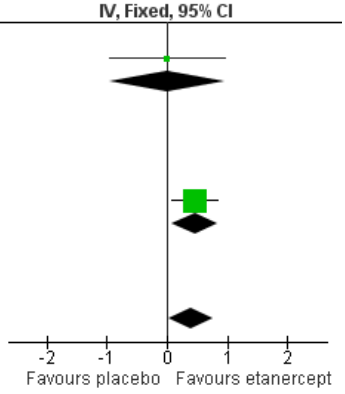
1.25 Chest expansion (cm)

Study or Subgroup	Etanercept Mean	SD	Total	Placebo Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
1.25.1 16 weeks								
Gorman 2002	3.5	1.9	20	2.9	1.7	20	16.5%	0.60 [-0.52, 1.72]
Subtotal (95% CI)			**20**			**20**	**16.5%**	**0.60 [-0.52, 1.72]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 1.05 (P = 0.29)								
1.25.2 24 weeks								
Davis 2003	3.85	2.3	138	3.01	1.9	139	83.5%	0.84 [0.34, 1.34]
Subtotal (95% CI)			**138**			**139**	**83.5%**	**0.84 [0.34, 1.34]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 3.31 (P = 0.0009)								
Total (95% CI)			**158**			**159**	**100.0%**	**0.80 [0.35, 1.25]**
Heterogeneity: Chi² = 0.15, df = 1 (P = 0.70); I² = 0%								
Test for overall effect: Z = 3.45 (P = 0.0006)								
Test for subgroup differences: Chi² = 0.15, df = 1 (P = 0.70), I² = 0%								



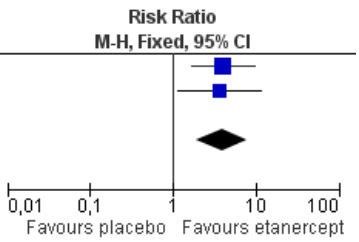
1.26 Modified Schober's lumbar flexion (cm)

Study or Subgroup	Etanercept Mean	SD	Total	Placebo Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
1.26.1 16 weeks								
Gorman 2002	13.4	1.6	20	13.4	1.5	20	14.0%	0.00 [-0.96, 0.96]
Subtotal (95% CI)			**20**			**20**	**14.0%**	**0.00 [-0.96, 0.96]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.00 (P = 1.00)								
1.26.2 24 weeks								
Davis 2003	3.34	1.76	138	2.88	1.53	139	86.0%	0.46 [0.07, 0.85]
Subtotal (95% CI)			**138**			**139**	**86.0%**	**0.46 [0.07, 0.85]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 2.32 (P = 0.02)								
Total (95% CI)			**158**			**159**	**100.0%**	**0.40 [0.04, 0.76]**
Heterogeneity: Chi² = 0.76, df = 1 (P = 0.38); I² = 0%								
Test for overall effect: Z = 2.15 (P = 0.03)								
Test for subgroup differences: Chi² = 0.76, df = 1 (P = 0.38), I² = 0%								



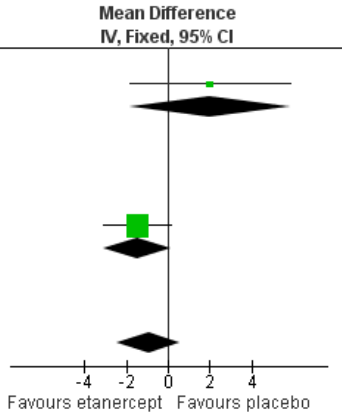
1.27 ASAS Partial remission

Study or Subgroup	Etanercept Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Davis 2003	24	138	6	139	57.2%	4.03 [1.70, 9.55]
van der Heijde 2006b	32	150	3	51	42.8%	3.63 [1.16, 11.34]
Total (95% CI)		**288**		**190**	**100.0%**	**3.86 [1.93, 7.70]**
Total events	56		9			
Heterogeneity: Chi² = 0.02, df = 1 (P = 0.88); I² = 0%						
Test for overall effect: Z = 3.82 (P = 0.0001)						



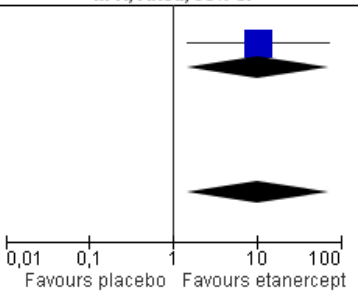
1.28 Occiput to wall (cm)

Study or Subgroup	Etanercept Mean	SD	Total	Placebo Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
1.28.1 16 weeks								
Gorman 2002	4.7	7.6	20	2.7	4.4	20	15.6%	2.00 [-1.85, 5.85]
Subtotal (95% CI)			**20**			**20**	**15.6%**	**2.00 [-1.85, 5.85]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 1.02 (P = 0.31)								
1.28.2 24 weeks								
Davis 2003	4.53	5.87	138	6.01	8.02	139	84.4%	-1.48 [-3.13, 0.17]
Subtotal (95% CI)			**138**			**139**	**84.4%**	**-1.48 [-3.13, 0.17]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 1.75 (P = 0.08)								
Total (95% CI)			**158**			**159**	**100.0%**	**-0.94 [-2.46, 0.58]**
Heterogeneity: Chi² = 2.65, df = 1 (P = 0.10); I² = 62%								
Test for overall effect: Z = 1.21 (P = 0.23)								
Test for subgroup differences: Chi² = 2.65, df = 1 (P = 0.10), I² = 62.3%								



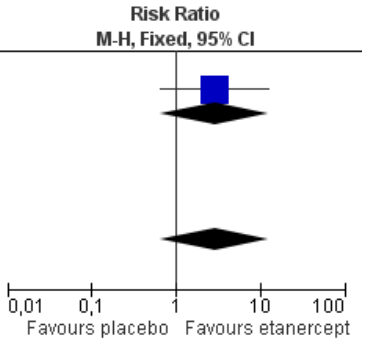
1.29 Reduction in NSAID use by 50%

Study or Subgroup	Etanercept Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
1.29.1 6 weeks						
Brandt 2003	9	14	1	16	100.0%	10.29 [1.48, 71.40]
Subtotal (95% CI)		**14**		**16**	**100.0%**	**10.29 [1.48, 71.40]**
Total events	9		1			
Heterogeneity: Not applicable						
Test for overall effect: Z = 2.36 (P = 0.02)						
Total (95% CI)		**14**		**16**	**100.0%**	**10.29 [1.48, 71.40]**
Total events	9		1			
Heterogeneity: Not applicable						
Test for overall effect: Z = 2.36 (P = 0.02)						



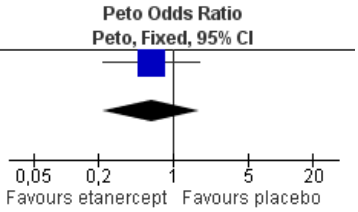
1.30 Stopped use of NSAIDs

Study or Subgroup	Etanercept Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
1.30.1 6 weeks						
Brandt 2003	5	14	2	16	100.0%	2.86 [0.65, 12.48]
Subtotal (95% CI)		**14**		**16**	**100.0%**	**2.86 [0.65, 12.48]**
Total events	5		2			
Heterogeneity: Not applicable						
Test for overall effect: Z = 1.40 (P = 0.16)						
Total (95% CI)		**14**		**16**	**100.0%**	**2.86 [0.65, 12.48]**
Total events	5		2			
Heterogeneity: Not applicable						
Test for overall effect: Z = 1.40 (P = 0.16)						



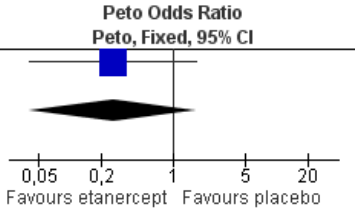
1.31 Total withdrawals

Study or Subgroup	Etanercept Events	Total	Placebo Events	Total	Weight	Peto Odds Ratio Peto, Fixed, 95% CI
van der Heijde 2006b	14	150	7	51	100.0%	0.63 [0.22, 1.77]
Total (95% CI)		**150**		**51**	**100.0%**	**0.63 [0.22, 1.77]**
Total events	14		7			
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.88 (P = 0.38)						



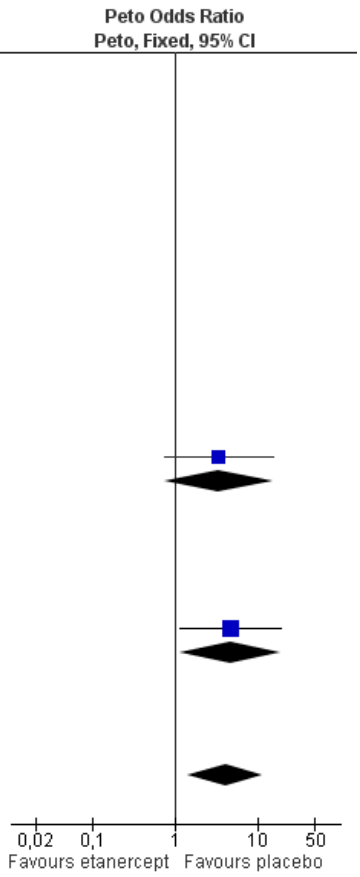
1.32 Withdrawals due to inefficacy

Study or Subgroup	Etanercept Events	Total	Placebo Events	Total	Weight	Peto Odds Ratio Peto, Fixed, 95% CI
van der Heijde 2006b	3	150	3	51	100.0%	0.26 [0.04, 1.70]
Total (95% CI)		**150**		**51**	**100.0%**	**0.26 [0.04, 1.70]**
Total events	3		3			
Heterogeneity: Not applicable						
Test for overall effect: Z = 1.40 (P = 0.16)						

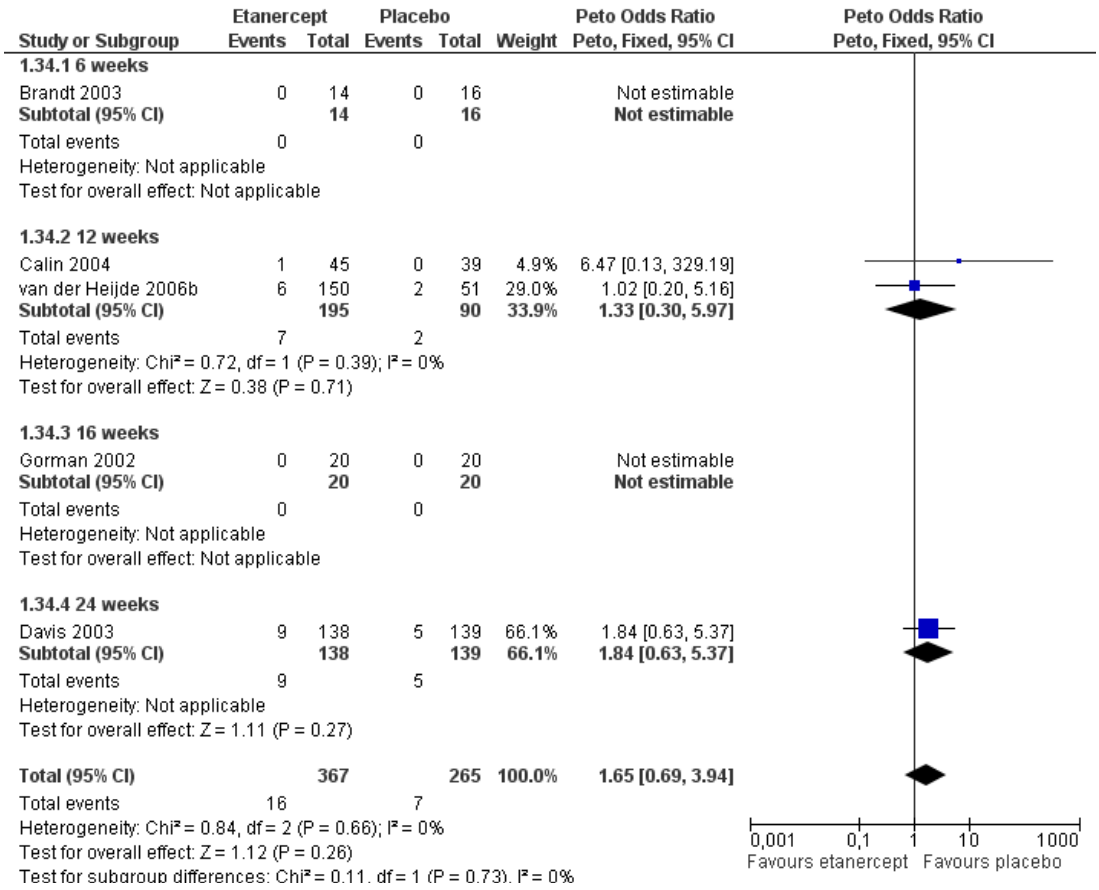


1.33 Withdrawals due to adverse events

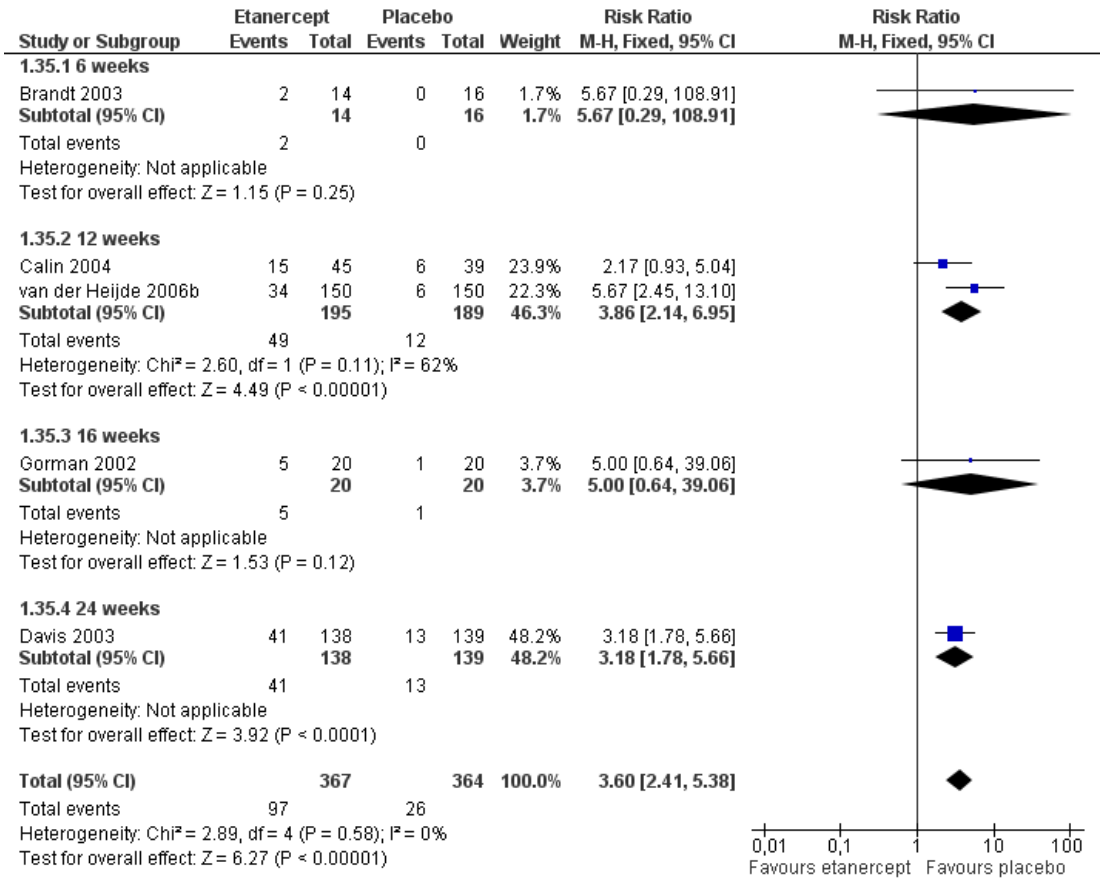
Study or Subgroup	Etanercept Events	Etanercept Total	Placebo Events	Placebo Total	Weight	Peto Odds Ratio Peto, Fixed, 95% CI
1.33.1 6 weeks						
Brandt 2003	0	14	0	16		Not estimable
Subtotal (95% CI)		**14**		**16**		**Not estimable**
Total events	0		0			
Heterogeneity: Not applicable						
Test for overall effect: Not applicable						
1.33.2 16 weeks						
Gorman 2002	0	20	0	20		Not estimable
Subtotal (95% CI)		**20**		**20**		**Not estimable**
Total events	0		0			
Heterogeneity: Not applicable						
Test for overall effect: Not applicable						
1.33.3 12 weeks						
Calin 2004	0	45	0	39		Not estimable
van der Heijde 2006b	14	305	0	51	45.9%	3.36 [0.73, 15.41]
Subtotal (95% CI)		**350**		**90**	**45.9%**	**3.36 [0.73, 15.41]**
Total events	14		0			
Heterogeneity: Not applicable						
Test for overall effect: Z = 1.56 (P = 0.12)						
1.33.4 24 weeks						
Davis 2003	7	138	1	139	54.1%	4.69 [1.15, 19.11]
Subtotal (95% CI)		**138**		**139**	**54.1%**	**4.69 [1.15, 19.11]**
Total events	7		1			
Heterogeneity: Not applicable						
Test for overall effect: Z = 2.16 (P = 0.03)						
Total (95% CI)		**522**		**265**	**100.0%**	**4.03 [1.43, 11.30]**
Total events	21		1			
Heterogeneity: Chi² = 0.10, df = 1 (P = 0.75); I² = 0%						
Test for overall effect: Z = 2.64 (P = 0.008)						
Test for subgroup differences: Chi² = 0.10, df = 1 (P = 0.75), I² = 0%						



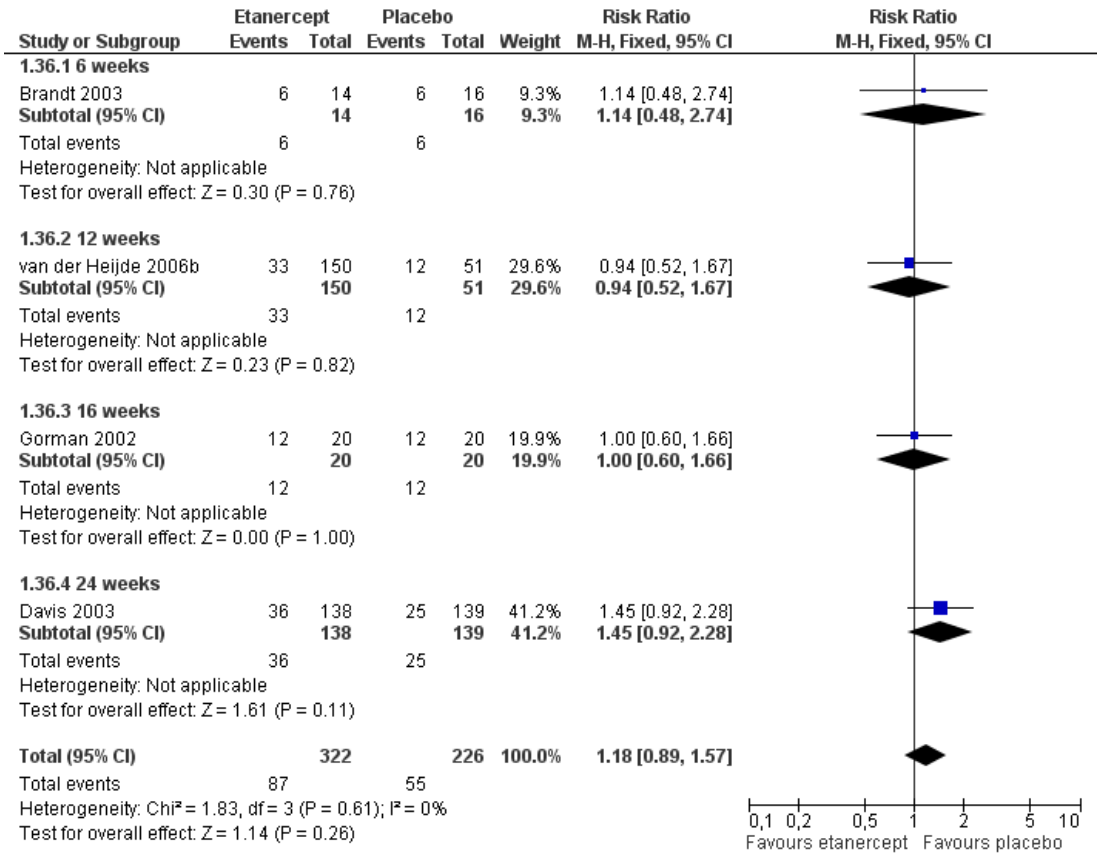
1.34 Serious adverse events



1.35 Injection site reaction

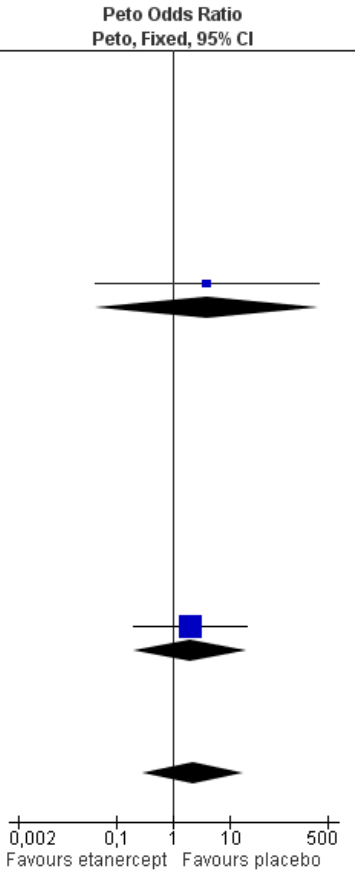


1.36 Infections



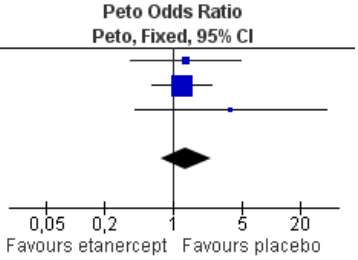
1.37 Serious infections

Study or Subgroup	Etanercept Events	Total	Placebo Events	Total	Weight	Peto Odds Ratio Peto, Fixed, 95% CI
1.37.1 6 weeks						
Brandt 2003	0	14	0	16		Not estimable
Subtotal (95% CI)		**14**		**16**		**Not estimable**
Total events	0		0			
Heterogeneity: Not applicable						
Test for overall effect: Not applicable						
1.37.2 12 weeks						
Calin 2004	0	45	0	39		Not estimable
van der Heijde 2006b	1	150	0	51	20.3%	3.82 [0.04, 345.21]
Subtotal (95% CI)		**195**		**90**	**20.3%**	**3.82 [0.04, 345.21]**
Total events	1		0			
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.58 (P = 0.56)						
1.37.3 16 weeks						
Gorman 2002	0	20	0	20		Not estimable
Subtotal (95% CI)		**20**		**20**		**Not estimable**
Total events	0		0			
Heterogeneity: Not applicable						
Test for overall effect: Not applicable						
1.37.4 24 weeks						
Davis 2003	2	138	1	139	79.7%	1.97 [0.20, 19.11]
Subtotal (95% CI)		**138**		**139**	**79.7%**	**1.97 [0.20, 19.11]**
Total events	2		1			
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.59 (P = 0.56)						
Total (95% CI)		**367**		**265**	**100.0%**	**2.25 [0.30, 17.13]**
Total events	3		1			
Heterogeneity: Chi² = 0.07, df = 1 (P = 0.80); I² = 0%						
Test for overall effect: Z = 0.79 (P = 0.43)						
Test for subgroup differences: Chi² = 0.07, df = 1 (P = 0.80), I² = 0%						



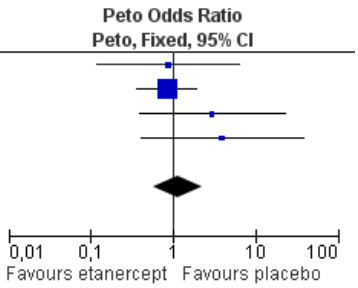
1.38 Headache

Study or Subgroup	Etanercept Events	Total	Placebo Events	Total	Weight	Peto Odds Ratio Peto, Fixed, 95% CI
Calin 2004	6	45	4	39	20.9%	1.34 [0.36, 4.98]
Davis 2003	19	138	16	139	72.1%	1.23 [0.60, 2.49]
van der Heijde 2006b	4	150	0	51	7.0%	3.90 [0.40, 37.70]
Total (95% CI)		**333**		**229**	**100.0%**	**1.35 [0.74, 2.47]**
Total events	29		20			
Heterogeneity: Chi² = 0.91, df = 2 (P = 0.63); I² = 0%						
Test for overall effect: Z = 0.99 (P = 0.32)						



1.39 Diarrhoea

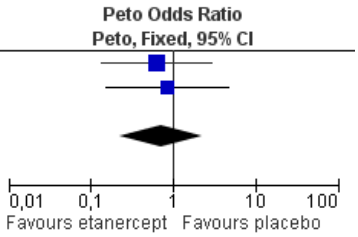
Study or Subgroup	Etanercept Events	Total	Placebo Events	Total	Weight	Peto Odds Ratio Peto, Fixed, 95% CI
Calin 2004	2	45	2	39	11.8%	0.86 [0.12, 6.38]
Davis 2003	11	138	13	139	67.7%	0.84 [0.36, 1.94]
Gorman 2002	3	20	1	20	11.4%	2.95 [0.38, 22.72]
van der Heijde 2006b	4	150	0	51	9.2%	3.90 [0.40, 37.70]
Total (95% CI)		**353**		**249**	**100.0%**	**1.12 [0.56, 2.23]**
Total events	20		16			
Heterogeneity: Chi² = 2.55, df = 3 (P = 0.47); I² = 0%						
Test for overall effect: Z = 0.32 (P = 0.75)						



1.40 Nausea

Study or Subgroup	Etanercept Events	Total	Placebo Events	Total	Weight	Peto Odds Ratio Peto, Fixed, 95% CI
Calin 2004	3	45	4	39	55.7%	0.63 [0.13, 2.94]
van der Heijde 2006b	5	150	2	51	44.3%	0.84 [0.15, 4.73]
Total (95% CI)		**195**		**90**	**100.0%**	**0.71 [0.23, 2.26]**
Total events	8		6			

Heterogeneity: Chi² = 0.06, df = 1 (P = 0.81); I² = 0%
Test for overall effect: Z = 0.57 (P = 0.57)

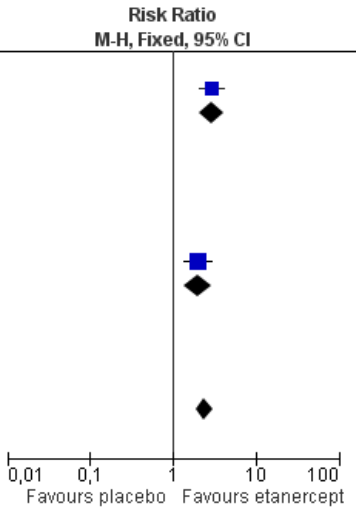


2 Etanercept (50mg once weekly versus placebo)

2.1 ASAS 20 responder

Study or Subgroup	Etanercept Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
2.1.1 ASAS 20 - 6 weeks						
Huang 2008	64	74	23	78	43.9%	2.93 [2.06, 4.18]
Subtotal (95% CI)		**74**		**78**	**43.9%**	**2.93 [2.06, 4.18]**
Total events	64		23			
Heterogeneity: Not applicable						
Test for overall effect: Z = 5.94 (P < 0.00001)						
2.1.2 ASAS 20 -12 weeks						
van der Heijde 2006b	115	155	19	51	56.1%	1.99 [1.38, 2.88]
Subtotal (95% CI)		**155**		**51**	**56.1%**	**1.99 [1.38, 2.88]**
Total events	115		19			
Heterogeneity: Not applicable						
Test for overall effect: Z = 3.67 (P = 0.0002)						
Total (95% CI)		**229**		**129**	**100.0%**	**2.41 [1.86, 3.10]**
Total events	179		42			

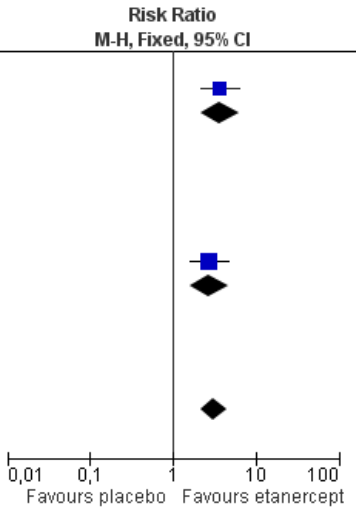
Heterogeneity: Chi² = 2.21, df = 1 (P = 0.14); I² = 55%
Test for overall effect: Z = 6.75 (P < 0.00001)
Test for subgroup differences: Not applicable



2.2 ASAS 40

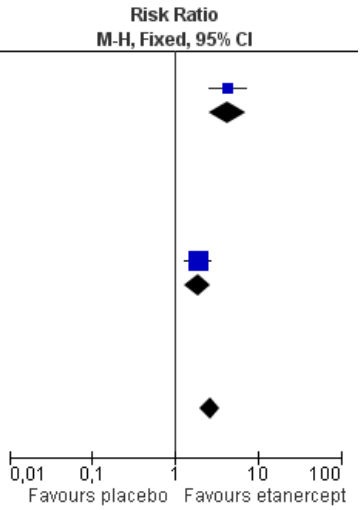
Study or Subgroup	Etanercept Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
2.2.1 ASAS 40 - 6 weeks						
Huang 2008	45	74	13	78	43.3%	3.65 [2.15, 6.19]
Subtotal (95% CI)		**74**		**78**	**43.3%**	**3.65 [2.15, 6.19]**
Total events	45		13			
Heterogeneity: Not applicable						
Test for overall effect: Z = 4.80 (P < 0.00001)						
2.2.2 ASAS 40 - 12 weeks						
van der Heijde 2006b	90	155	11	51	56.7%	2.69 [1.57, 4.62]
Subtotal (95% CI)		**155**		**51**	**56.7%**	**2.69 [1.57, 4.62]**
Total events	90		11			
Heterogeneity: Not applicable						
Test for overall effect: Z = 3.59 (P = 0.0003)						
Total (95% CI)		**229**		**129**	**100.0%**	**3.11 [2.13, 4.53]**
Total events	135		24			

Heterogeneity: Chi² = 0.63, df = 1 (P = 0.43); I² = 0%
Test for overall effect: Z = 5.88 (P < 0.00001)
Test for subgroup differences: Not applicable



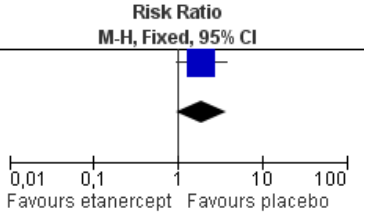
2.3 ASAS 5/6

Study or Subgroup	Etanercept Events	Etanercept Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
2.3.1 ASAS 5/6 - 6 weeks						
Huang 2008	53	74	13	78	30.7%	4.30 [2.56, 7.20]
Subtotal (95% CI)		**74**		**78**	**30.7%**	**4.30 [2.56, 7.20]**
Total events	53		13			
Heterogeneity: Not applicable						
Test for overall effect: Z = 5.53 (P < 0.00001)						
2.3.2 ASAS 5/6 - 12 weeks						
van der Heijde 2006b	109	155	19	51	69.3%	1.89 [1.30, 2.73]
Subtotal (95% CI)		**155**		**51**	**69.3%**	**1.89 [1.30, 2.73]**
Total events	109		19			
Heterogeneity: Not applicable						
Test for overall effect: Z = 3.36 (P = 0.0008)						
Total (95% CI)		**229**		**129**	**100.0%**	**2.63 [1.95, 3.54]**
Total events	162		32			
Heterogeneity: Chi² = 6.54, df = 1 (P = 0.01); I² = 85%						
Test for overall effect: Z = 6.36 (P < 0.00001)						
Test for subgroup differences: Not applicable						



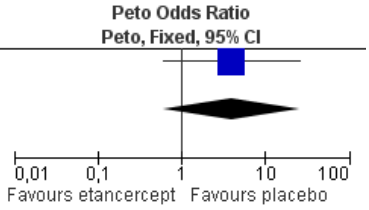
2.4 Total adverse events

Study or Subgroup	Etanercept Events	Etanercept Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Huang 2008	18	74	10	78	100.0%	1.90 [0.94, 3.84]
Total (95% CI)		**74**		**78**	**100.0%**	**1.90 [0.94, 3.84]**
Total events	18		10			
Heterogeneity: Not applicable						
Test for overall effect: Z = 1.78 (P = 0.07)						



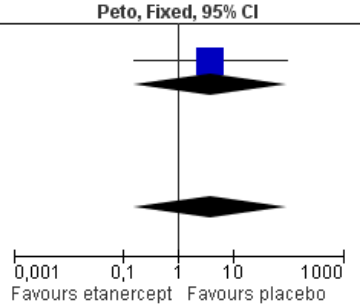
2.5 Withdrawals due to AE

Study or Subgroup	Etanercept Events	Etanercept Total	Placebo Events	Placebo Total	Weight	Peto Odds Ratio Peto, Fixed, 95% CI
van der Heijde 2006b	6	155	0	51	100.0%	3.91 [0.60, 25.51]
Total (95% CI)		**155**		**51**	**100.0%**	**3.91 [0.60, 25.51]**
Total events	6		0			
Heterogeneity: Not applicable						
Test for overall effect: Z = 1.42 (P = 0.15)						



2.6 Serious infections

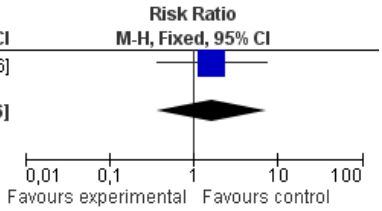
Study or Subgroup	Etanercept Events	Etanercept Total	Placebo Events	Placebo Total	Weight	Peto Odds Ratio Peto, Fixed, 95% CI
2.6.1 12 weeks						
van der Heijde 2006b	2	155	0	51	100.0%	3.80 [0.15, 95.06]
Subtotal (95% CI)		**155**		**51**	**100.0%**	**3.80 [0.15, 95.06]**
Total events	2		0			
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.81 (P = 0.42)						
Total (95% CI)		**155**		**51**	**100.0%**	**3.80 [0.15, 95.06]**
Total events	2		0			
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.81 (P = 0.42)						
Test for subgroup differences: Not applicable						



2.7 Serious adverse events

Study or Subgroup	Etanercept Events	Etanercept Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
van der Heijde 2006b	10	155	2	51	100.0%	1.65 [0.37, 7.26]
Total (95% CI)		**155**		**51**	**100.0%**	**1.65 [0.37, 7.26]**
Total events	10		2			

Heterogeneity: Not applicable
Test for overall effect: Z = 0.66 (P = 0.51)



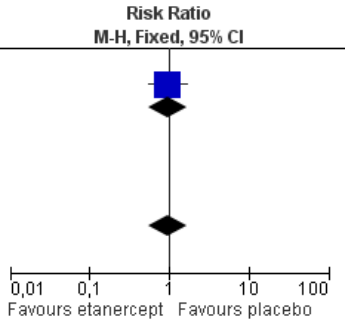
2.8 Infections

Study or Subgroup	Etanercept 50mg Events	Etanercept 50mg Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
2.8.1 12 weeks						
van der Heijde 2006b	35	155	12	51	100.0%	0.96 [0.54, 1.70]
Subtotal (95% CI)		**155**		**51**	**100.0%**	**0.96 [0.54, 1.70]**
Total events	35		12			

Heterogeneity: Not applicable
Test for overall effect: Z = 0.14 (P = 0.89)

Total (95% CI)		**155**		**51**	**100.0%**	**0.96 [0.54, 1.70]**
Total events	35		12			

Heterogeneity: Not applicable
Test for overall effect: Z = 0.14 (P = 0.89)
Test for subgroup differences: Not applicable



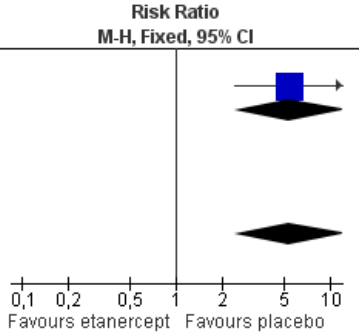
2.10 Injection site reaction

Study or Subgroup	Etanercept Events	Etanercept Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
2.10.1 12 weeks						
van der Heijde 2006b	66	305	6	150	100.0%	5.41 [2.40, 12.19]
Subtotal (95% CI)		**305**		**150**	**100.0%**	**5.41 [2.40, 12.19]**
Total events	66		6			

Heterogeneity: Not applicable
Test for overall effect: Z = 4.07 (P < 0.0001)

Total (95% CI)		**305**		**150**	**100.0%**	**5.41 [2.40, 12.19]**
Total events	66		6			

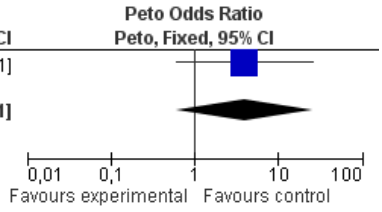
Heterogeneity: Not applicable
Test for overall effect: Z = 4.07 (P < 0.0001)



2.11 Headache

Study or Subgroup	Etanercept Events	Etanercept Total	Placebo Events	Placebo Total	Weight	Peto Odds Ratio Peto, Fixed, 95% CI
van der Heijde 2006b	6	155	0	51	100.0%	3.91 [0.60, 25.51]
Total (95% CI)		**155**		**51**	**100.0%**	**3.91 [0.60, 25.51]**
Total events	6		0			

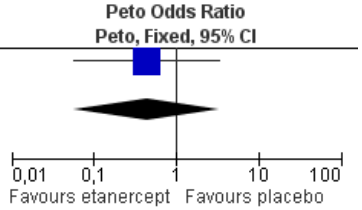
Heterogeneity: Not applicable
Test for overall effect: Z = 1.42 (P = 0.15)



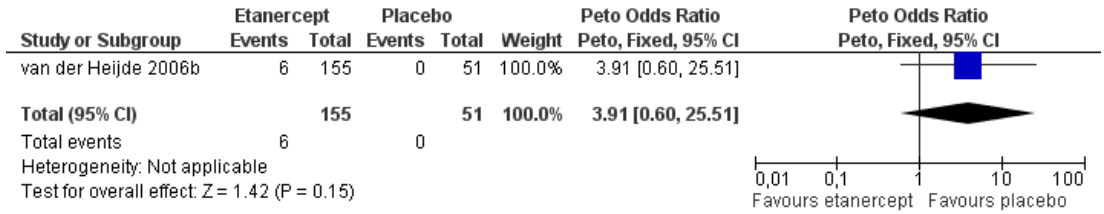
2.13 Nausea

Study or Subgroup	Etanercept Events	Etanercept Total	Placebo Events	Placebo Total	Weight	Peto Odds Ratio Peto, Fixed, 95% CI
van der Heijde 2006b	3	155	2	51	100.0%	0.43 [0.06, 3.38]
Total (95% CI)		**155**		**51**	**100.0%**	**0.43 [0.06, 3.38]**
Total events	3		2			

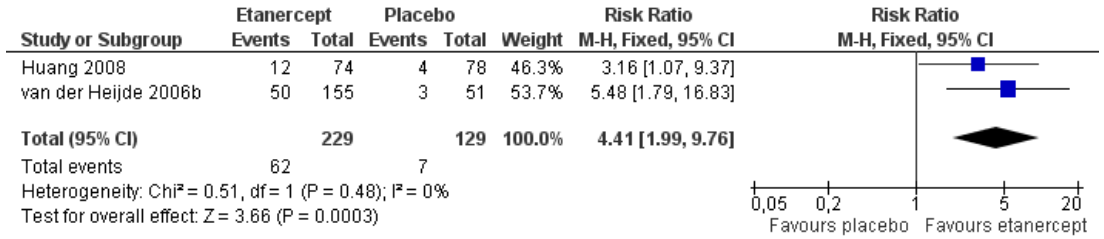
Heterogeneity: Not applicable
Test for overall effect: Z = 0.80 (P = 0.43)



2.12 Diarrhoea

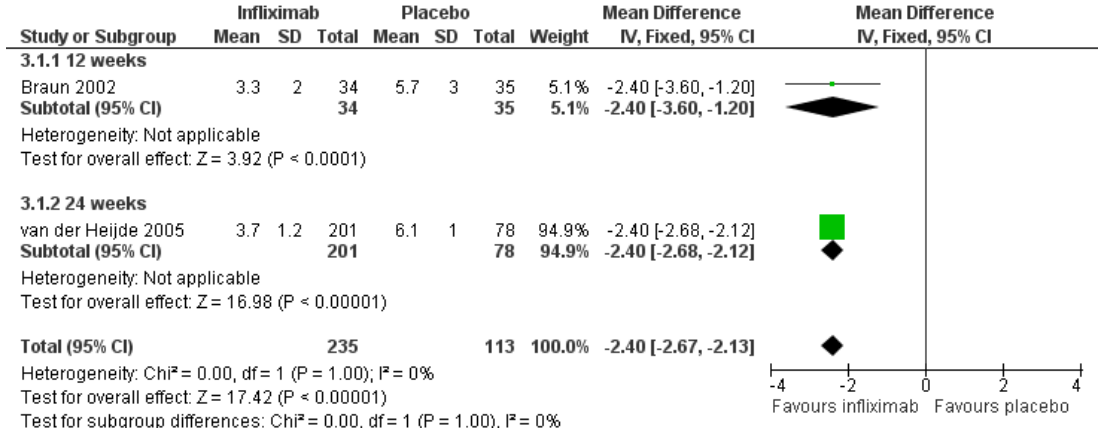


2.15 ASAS Partial remission

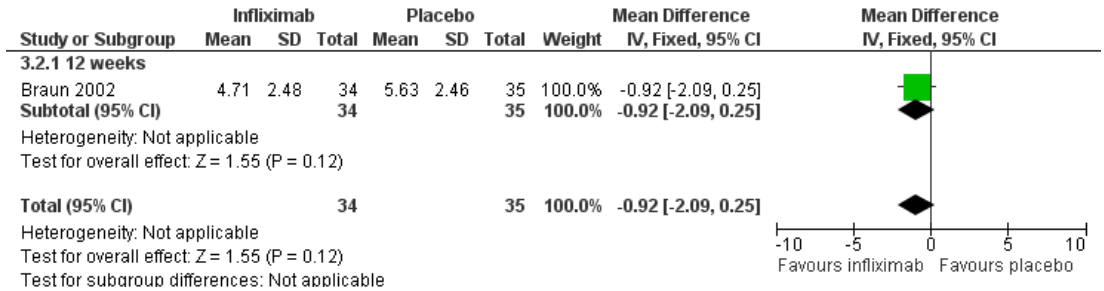


3 Infliximab versus placebo

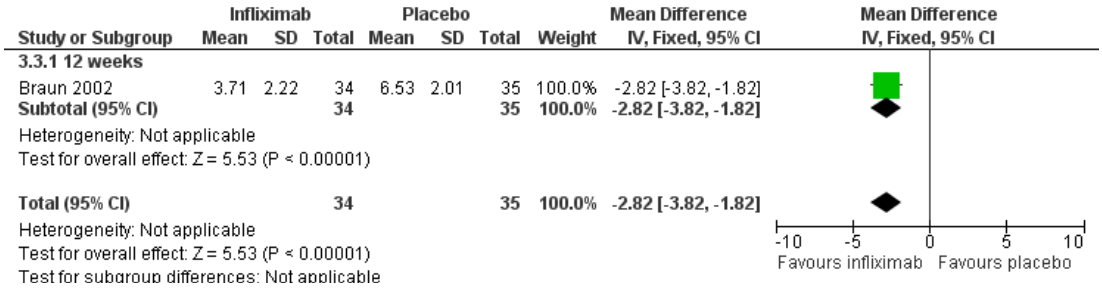
3.1 BASDAI (0-10 scale)



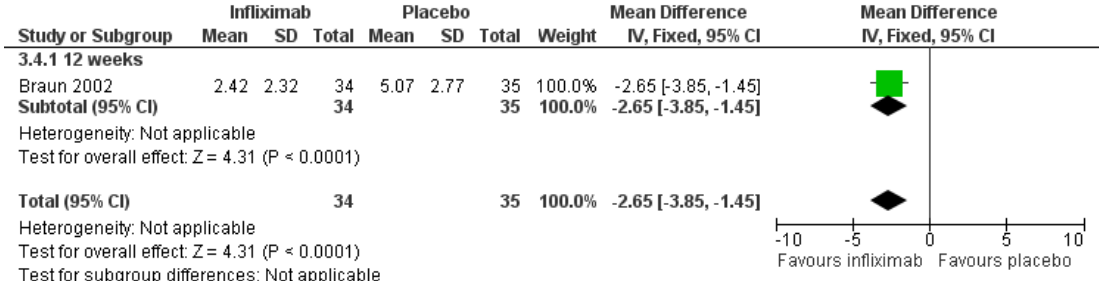
3.2 Fatigue



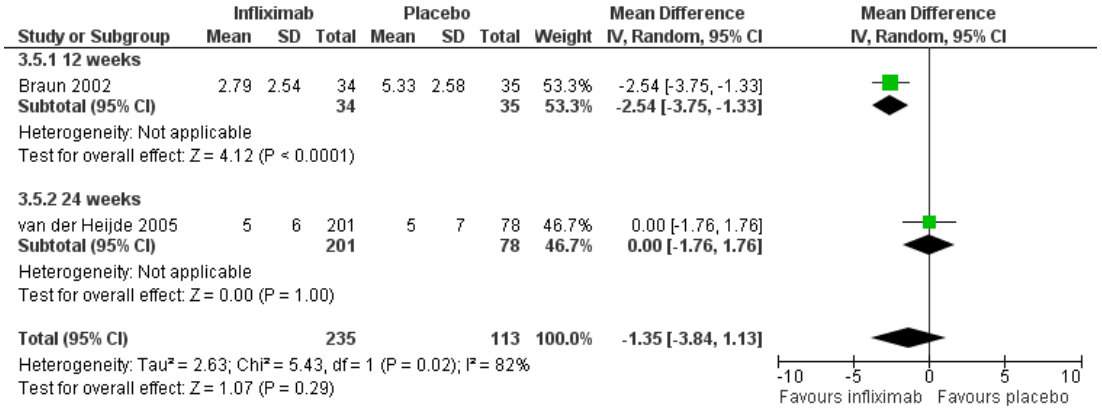
3.3 Spinal pain



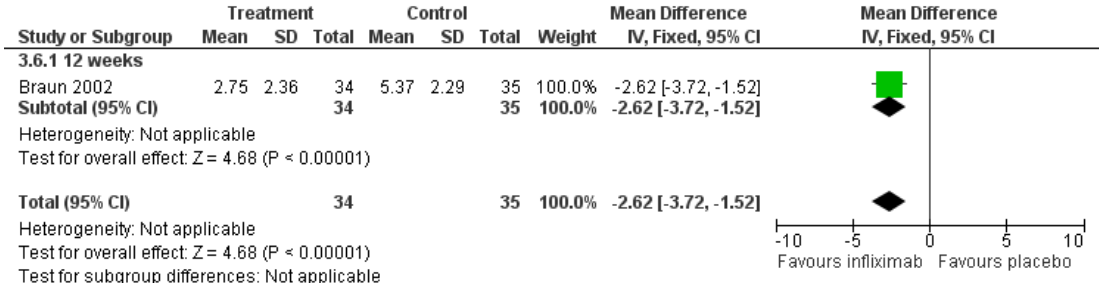
3.4 Peripheral joint pain



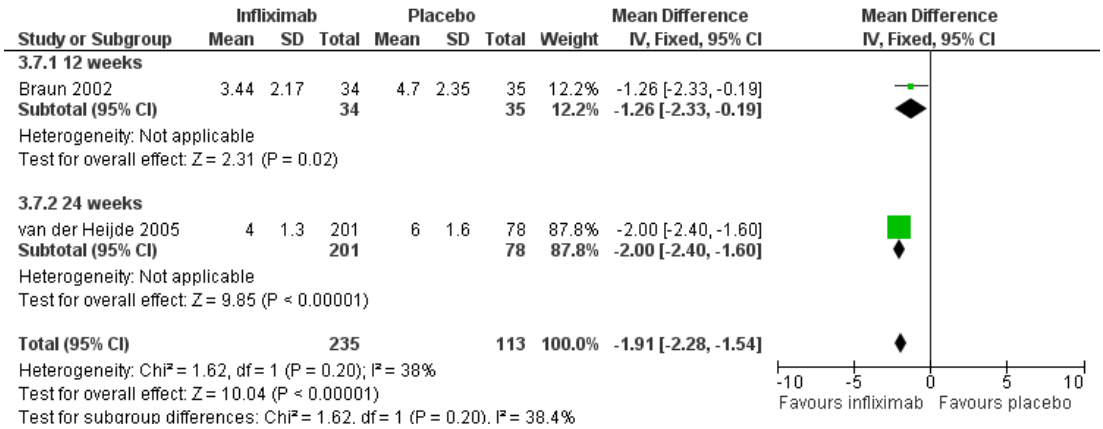
3.5 Entheseal pain



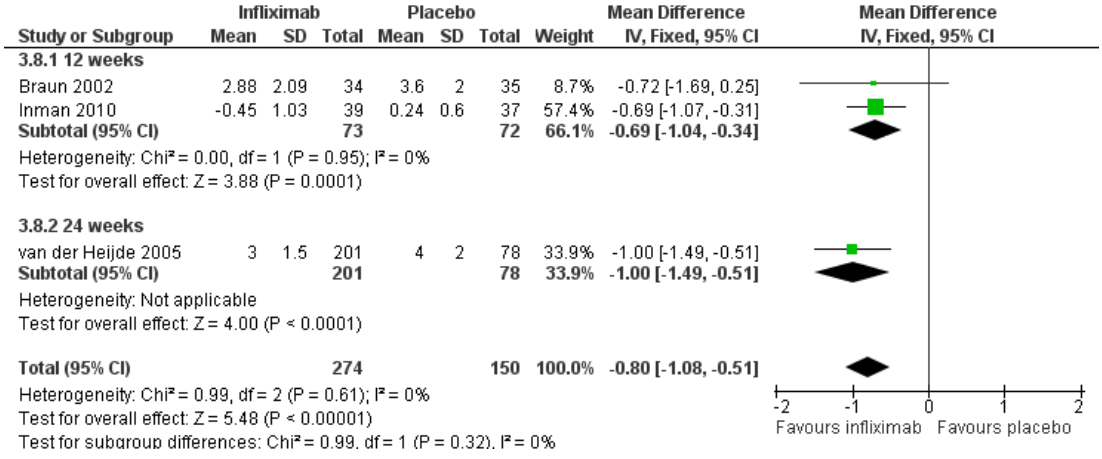
3.6 Morning stiffness - mean severity and duration



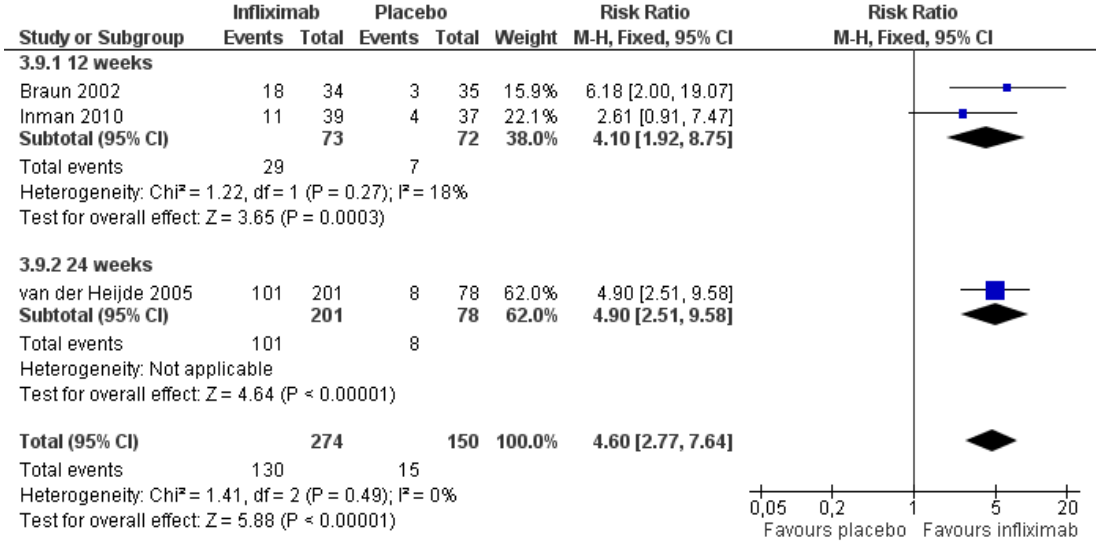
3.7 BASFI



3.8 BASMI

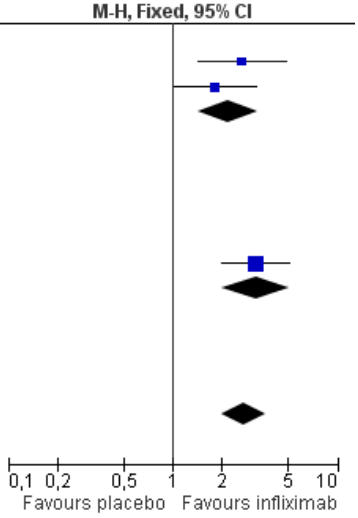


3.9 >50% improvement in BASDAI



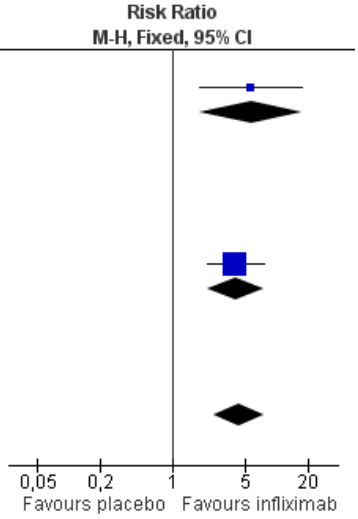
3.10 ASAS 20

Study or Subgroup	Infliximab Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
3.10.1 12 weeks						
Braun 2002	23	34	9	35	21.2%	2.63 [1.43, 4.84]
Inman 2010	21	39	11	37	27.0%	1.81 [1.02, 3.22]
Subtotal (95% CI)		**73**		**72**	**48.3%**	**2.17 [1.43, 3.29]**
Total events	44		20			
Heterogeneity: Chi² = 0.76, df = 1 (P = 0.38); I² = 0%						
Test for overall effect: Z = 3.65 (P = 0.0003)						
3.10.2 24 weeks						
van der Heijde 2005	123	201	15	78	51.7%	3.18 [1.99, 5.08]
Subtotal (95% CI)		**201**		**78**	**51.7%**	**3.18 [1.99, 5.08]**
Total events	123		15			
Heterogeneity: Not applicable						
Test for overall effect: Z = 4.85 (P < 0.00001)						
Total (95% CI)		**274**		**150**	**100.0%**	**2.69 [1.96, 3.71]**
Total events	167		35			
Heterogeneity: Chi² = 2.33, df = 2 (P = 0.31); I² = 14%						
Test for overall effect: Z = 6.09 (P < 0.00001)						



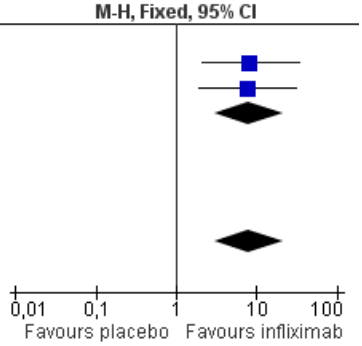
3.11 ASAS 40

Study or Subgroup	Infliximab Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
3.11.1 12 weeks						
Inman 2010	18	39	3	37	19.2%	5.69 [1.83, 17.74]
Subtotal (95% CI)		**39**		**37**	**19.2%**	**5.69 [1.83, 17.74]**
Total events	18		3			
Heterogeneity: Not applicable						
Test for overall effect: Z = 3.00 (P = 0.003)						
3.11.2 24 weeks						
van der Heijde 2005	93	201	9	78	80.8%	4.01 [2.13, 7.55]
Subtotal (95% CI)		**201**		**78**	**80.8%**	**4.01 [2.13, 7.55]**
Total events	93		9			
Heterogeneity: Not applicable						
Test for overall effect: Z = 4.31 (P < 0.0001)						
Total (95% CI)		**240**		**115**	**100.0%**	**4.33 [2.50, 7.52]**
Total events	111		12			
Heterogeneity: Chi² = 0.28, df = 1 (P = 0.60); I² = 0%						
Test for overall effect: Z = 5.21 (P < 0.00001)						



3.12 ASAS 50

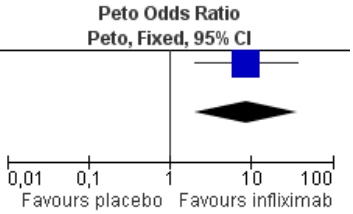
Study or Subgroup	Infliximab Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
3.12.1 12 weeks						
Braun 2002	16	34	2	35	49.0%	8.24 [2.05, 33.13]
Inman 2010	16	39	2	37	51.0%	7.59 [1.87, 30.76]
Subtotal (95% CI)		**73**		**72**	**100.0%**	**7.91 [2.95, 21.21]**
Total events	32		4			
Heterogeneity: Chi² = 0.01, df = 1 (P = 0.94); I² = 0%						
Test for overall effect: Z = 4.11 (P < 0.0001)						
Total (95% CI)		**73**		**72**	**100.0%**	**7.91 [2.95, 21.21]**
Total events	32		4			
Heterogeneity: Chi² = 0.01, df = 1 (P = 0.94); I² = 0%						
Test for overall effect: Z = 4.11 (P < 0.0001)						



3.13 ASAS 70

Study or Subgroup	Experimental Events	Experimental Total	Control Events	Control Total	Weight	Peto Odds Ratio Peto, Fixed, 95% CI
Inman 2010	8	39	0	37	100.0%	8.58 [2.00, 36.79]
Total (95% CI)		**39**		**37**	**100.0%**	**8.58 [2.00, 36.79]**
Total events	8		0			

Heterogeneity: Not applicable
Test for overall effect: Z = 2.89 (P = 0.004)



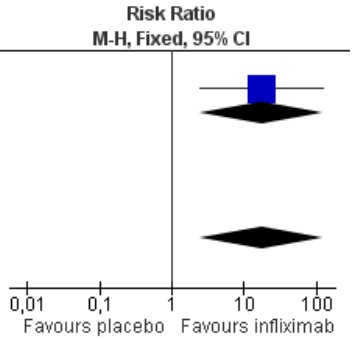
3.14 ASAS partial remission

Study or Subgroup	Infliximab Events	Infliximab Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
3.14.1 24 weeks						
van der Heijde 2005	45	201	1	78	100.0%	17.46 [2.45, 124.51]
Subtotal (95% CI)		**201**		**78**	**100.0%**	**17.46 [2.45, 124.51]**
Total events	45		1			

Heterogeneity: Not applicable
Test for overall effect: Z = 2.85 (P = 0.004)

Total (95% CI)		**201**		**78**	**100.0%**	**17.46 [2.45, 124.51]**
Total events	45		1			

Heterogeneity: Not applicable
Test for overall effect: Z = 2.85 (P = 0.004)



3.15 ASAS 5 out of 6

Study or Subgroup	Infliximab Events	Infliximab Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
3.15.1 12 weeks						
Inman 2010	20	39	1	37	10.6%	18.97 [2.68, 134.35]
Subtotal (95% CI)		**39**		**37**	**10.6%**	**18.97 [2.68, 134.35]**
Total events	20		1			

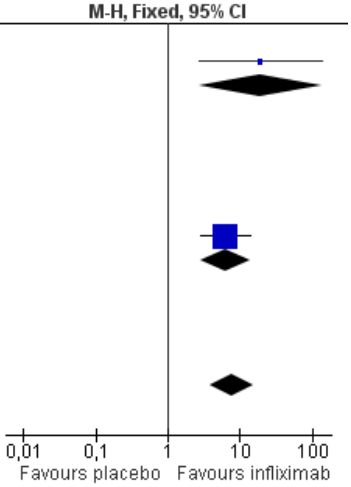
Heterogeneity: Not applicable
Test for overall effect: Z = 2.95 (P = 0.003)

3.15.2 24 weeks						
van der Heijde 2005	97	201	6	78	89.4%	6.27 [2.87, 13.71]
Subtotal (95% CI)		**201**		**78**	**89.4%**	**6.27 [2.87, 13.71]**
Total events	97		6			

Heterogeneity: Not applicable
Test for overall effect: Z = 4.60 (P < 0.00001)

Total (95% CI)		**240**		**115**	**100.0%**	**7.62 [3.73, 15.56]**
Total events	117		7			

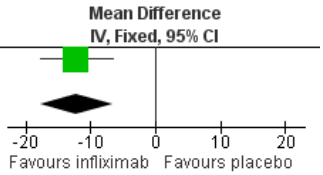
Heterogeneity: Chi² = 1.07, df = 1 (P = 0.30); I² = 7%
Test for overall effect: Z = 5.58 (P < 0.00001)



3.16 ESR

Study or Subgroup	Infliximab Mean	Infliximab SD	Infliximab Total	Placebo Mean	Placebo SD	Placebo Total	Weight	Mean Difference IV, Fixed, 95% CI
Inman 2010	-10.66	12.5	39	1.59	12.42	37	100.0%	-12.25 [-17.85, -6.65]
Total (95% CI)			**39**			**37**	**100.0%**	**-12.25 [-17.85, -6.65]**

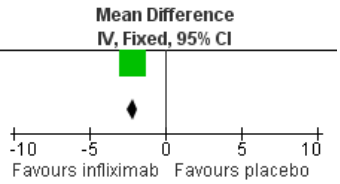
Heterogeneity: Not applicable
Test for overall effect: Z = 4.28 (P < 0.0001)



3.17 C-reactive protein (mg/dl)

Study or Subgroup	Infliximab Mean	SD	Total	Placebo Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
van der Heijde 2005	-0.5	1.3	201	1.7	1.3	78	100.0%	-2.20 [-2.54, -1.86]
Total (95% CI)			**201**			**78**	**100.0%**	**-2.20 [-2.54, -1.86]**

Heterogeneity: Not applicable
Test for overall effect: Z = 12.69 (P < 0.00001)



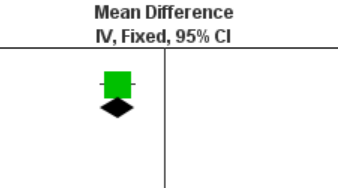
3.18 Patient global assessment

Study or Subgroup	Infliximab Mean	SD	Total	Placebo Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
3.18.1 12 weeks								
Braun 2002	3.3	2.6	34	6.3	2.1	35	100.0%	-3.00 [-4.12, -1.88]
Subtotal (95% CI)			**34**			**35**	**100.0%**	**-3.00 [-4.12, -1.88]**

Heterogeneity: Not applicable
Test for overall effect: Z = 5.26 (P < 0.00001)

Total (95% CI)			**34**			**35**	**100.0%**	**-3.00 [-4.12, -1.88]**

Heterogeneity: Not applicable
Test for overall effect: Z = 5.26 (P < 0.00001)
Test for subgroup differences: Not applicable



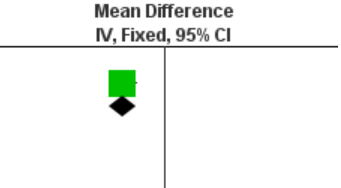
3.19 Physician global assessment

Study or Subgroup	Infliximab Mean	SD	Total	Placebo Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
3.19.1 12 weeks								
Braun 2002	3	2.1	34	5.7	1.6	35	100.0%	-2.70 [-3.58, -1.82]
Subtotal (95% CI)			**34**			**35**	**100.0%**	**-2.70 [-3.58, -1.82]**

Heterogeneity: Not applicable
Test for overall effect: Z = 5.99 (P < 0.00001)

Total (95% CI)			**34**			**35**	**100.0%**	**-2.70 [-3.58, -1.82]**

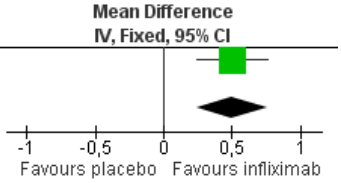
Heterogeneity: Not applicable
Test for overall effect: Z = 5.99 (P < 0.00001)
Test for subgroup differences: Not applicable



3.20 Chest expansion (cm)

Study or Subgroup	Infliximab Mean	SD	Total	Placebo Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
van der Heijde 2005	3.5	1	201	3	1	78	100.0%	0.50 [0.24, 0.76]
Total (95% CI)			**201**			**78**	**100.0%**	**0.50 [0.24, 0.76]**

Heterogeneity: Not applicable
Test for overall effect: Z = 3.75 (P = 0.0002)



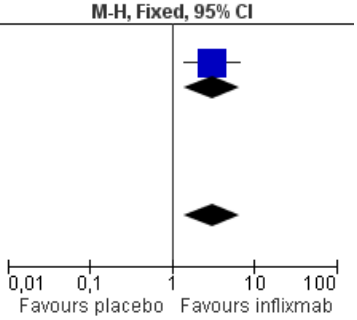
3.21 Reduction in NSAID use by 50%

Study or Subgroup	Infliximab Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
3.21.1 12 weeks						
Braun 2002	18	34	6	35	100.0%	3.09 [1.40, 6.83]
Subtotal (95% CI)		**34**		**35**	**100.0%**	**3.09 [1.40, 6.83]**
Total events	18		6			

Heterogeneity: Not applicable
Test for overall effect: Z = 2.78 (P = 0.005)

Total (95% CI)		**34**		**35**	**100.0%**	**3.09 [1.40, 6.83]**
Total events	18		6			

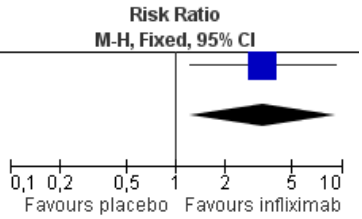
Heterogeneity: Not applicable
Test for overall effect: Z = 2.78 (P = 0.005)



3.22 Stopped use of NSAIDs

Study or Subgroup	Treatment Events	Total	Control Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Braun 2002	13	34	4	35	100.0%	3.35 [1.21, 9.24]
Total (95% CI)		**34**		**35**	**100.0%**	**3.35 [1.21, 9.24]**
Total events	13		4			

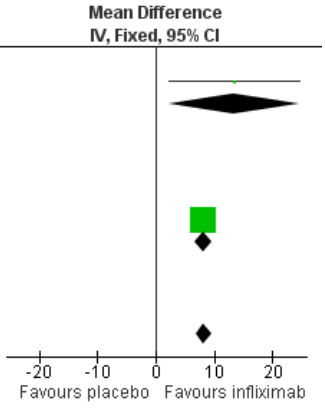
Heterogeneity: Not applicable
Test for overall effect: Z = 2.33 (P = 0.02)



3.23 SF-36 Physical functioning

Study or Subgroup	Infliximab Mean	SD	Total	Placebo Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
3.23.1 12 weeks								
Braun 2002	61.53	23.02	34	48.17	24.53	35	1.6%	13.36 [2.14, 24.58]
Subtotal (95% CI)			**34**			**35**	**1.6%**	**13.36 [2.14, 24.58]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 2.33 (P = 0.02)								
3.23.2 24 weeks								
van der Heijde 2005	39	5	201	30.9	5.7	78	98.4%	8.10 [6.66, 9.54]
Subtotal (95% CI)			**201**			**78**	**98.4%**	**8.10 [6.66, 9.54]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 11.01 (P < 0.00001)								
Total (95% CI)			**235**			**113**	**100.0%**	**8.19 [6.76, 9.62]**

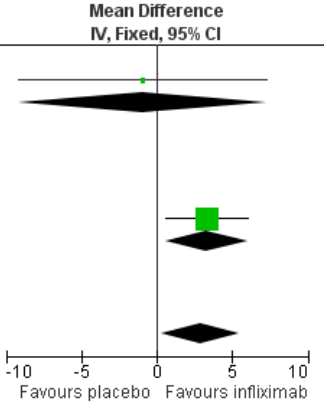
Heterogeneity: Chi² = 0.83, df = 1 (P = 0.36); I² = 0%
Test for overall effect: Z = 11.22 (P < 0.00001)
Test for subgroup differences: Chi² = 0.83, df = 1 (P = 0.36), I² = 0%



3.24 SF-36 Mental health

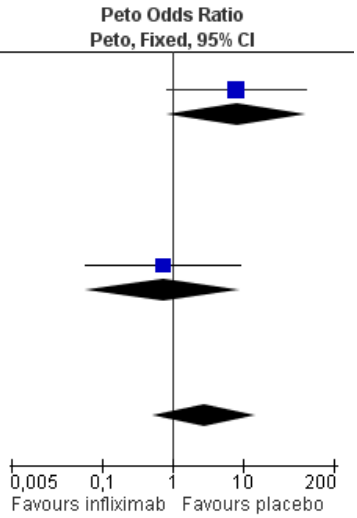
Study or Subgroup	Infliximab Mean	SD	Total	Placebo Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
3.24.1 12 weeks								
Braun 2002	62.38	17.04	34	63.33	17.86	35	9.7%	-0.95 [-9.19, 7.29]
Subtotal (95% CI)			**34**			**35**	**9.7%**	**-0.95 [-9.19, 7.29]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.23 (P = 0.82)								
3.24.2 24 weeks								
van der Heijde 2005	50.3	8.7	201	47	10.9	78	90.3%	3.30 [0.60, 6.00]
Subtotal (95% CI)			**201**			**78**	**90.3%**	**3.30 [0.60, 6.00]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 2.39 (P = 0.02)								
Total (95% CI)			**235**			**113**	**100.0%**	**2.89 [0.32, 5.45]**

Heterogeneity: Chi² = 0.92, df = 1 (P = 0.34); I² = 0%
Test for overall effect: Z = 2.20 (P = 0.03)
Test for subgroup differences: Chi² = 0.92, df = 1 (P = 0.34), I² = 0%



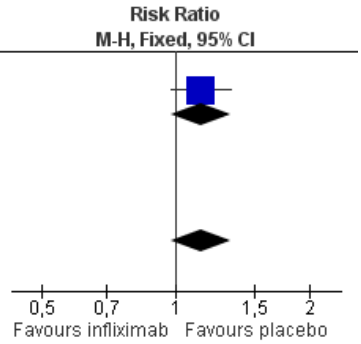
3.25 Withdrawals due to adverse events

Study or Subgroup	Infliximab Events	Total	Placebo Events	Total	Weight	Peto Odds Ratio Peto, Fixed, 95% CI
3.25.1 12 weeks						
Braun 2002	3	34	0	35	55.3%	8.09 [0.81, 80.51]
Subtotal (95% CI)		**34**		**35**	**55.3%**	**8.09 [0.81, 80.51]**
Total events	3		0			
Heterogeneity: Not applicable						
Test for overall effect: Z = 1.78 (P = 0.07)						
3.25.2 24 weeks						
van der Heijde 2005	2	202	1	75	44.7%	0.73 [0.06, 9.36]
Subtotal (95% CI)		**202**		**75**	**44.7%**	**0.73 [0.06, 9.36]**
Total events	2		1			
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.24 (P = 0.81)						
Total (95% CI)		**236**		**110**	**100.0%**	**2.76 [0.50, 15.22]**
Total events	5		1			
Heterogeneity: Chi² = 1.89, df = 1 (P = 0.17); I² = 47%						
Test for overall effect: Z = 1.16 (P = 0.24)						
Test for subgroup differences: Chi² = 1.89, df = 1 (P = 0.17), I² = 47.1%						



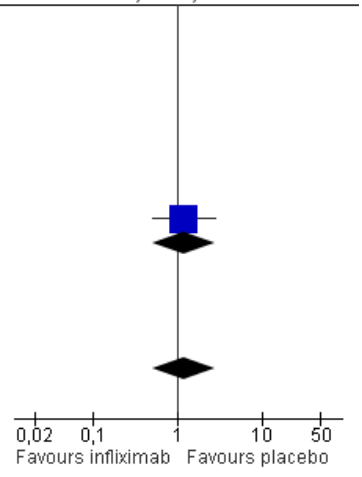
3.26 Total adverse events

Study or Subgroup	Infliximab Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
3.26.1 24 weeks						
van der Heijde 2005	166	202	54	75	100.0%	1.14 [0.98, 1.33]
Subtotal (95% CI)		**202**		**75**	**100.0%**	**1.14 [0.98, 1.33]**
Total events	166		54			
Heterogeneity: Not applicable						
Test for overall effect: Z = 1.67 (P = 0.09)						
Total (95% CI)		**202**		**75**	**100.0%**	**1.14 [0.98, 1.33]**
Total events	166		54			
Heterogeneity: Not applicable						
Test for overall effect: Z = 1.67 (P = 0.09)						



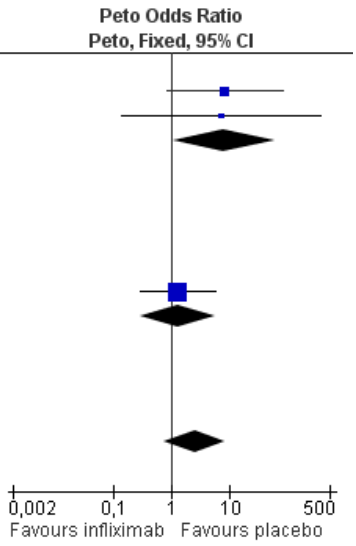
3.27 Infusion reaction

Study or Subgroup	Infliximab Events	Total	Placebo Events	Total	Weight	Peto Odds Ratio Peto, Fixed, 95% CI
3.27.1 12 weeks						
Braun 2002	0	34	0	35		Not estimable
Subtotal (95% CI)		**34**		**35**		**Not estimable**
Total events	0		0			
Heterogeneity: Not applicable						
Test for overall effect: Not applicable						
3.27.2 24 weeks						
van der Heijde 2005	22	202	7	75	100.0%	1.18 [0.50, 2.80]
Subtotal (95% CI)		**202**		**75**	**100.0%**	**1.18 [0.50, 2.80]**
Total events	22		7			
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.38 (P = 0.71)						
Total (95% CI)		**236**		**110**	**100.0%**	**1.18 [0.50, 2.80]**
Total events	22		7			
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.38 (P = 0.71)						
Test for subgroup differences: Not applicable						



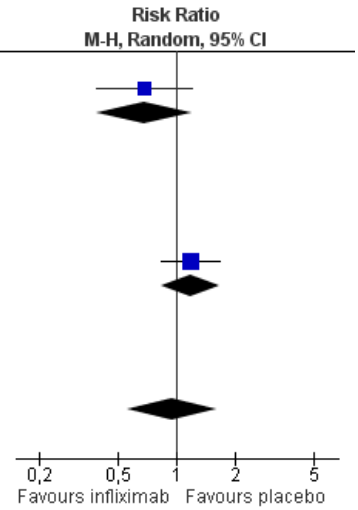
3.28 Serious adverse events

Study or Subgroup	Infliximab Events	Total	Placebo Events	Total	Weight	Peto Odds Ratio Peto, Fixed, 95% CI
3.28.1 12 weeks						
Braun 2002	3	34	0	35	26.9%	8.09 [0.81, 80.51]
Inman 2010	1	39	0	37	9.2%	7.02 [0.14, 354.25]
Subtotal (95% CI)		**73**		**72**	**36.2%**	**7.80 [1.07, 56.65]**
Total events	4		0			
Heterogeneity: Chi² = 0.00, df = 1 (P = 0.95); I² = 0%						
Test for overall effect: Z = 2.03 (P = 0.04)						
3.28.2 24 weeks						
van der Heijde 2005	7	202	2	75	63.8%	1.29 [0.29, 5.73]
Subtotal (95% CI)		**202**		**75**	**63.8%**	**1.29 [0.29, 5.73]**
Total events	7		2			
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.33 (P = 0.74)						
Total (95% CI)		**275**		**147**	**100.0%**	**2.47 [0.75, 8.14]**
Total events	11		2			
Heterogeneity: Chi² = 2.03, df = 2 (P = 0.36); I² = 1%						
Test for overall effect: Z = 1.49 (P = 0.14)						
Test for subgroup differences: Chi² = 2.03, df = 1 (P = 0.15), I² = 50.6%						



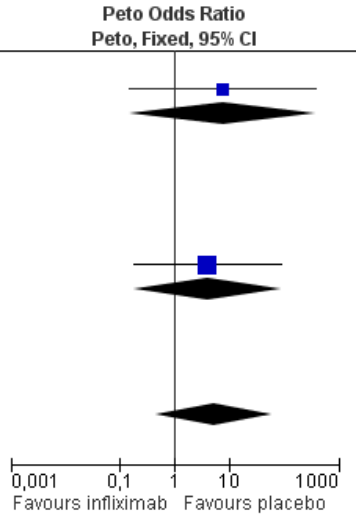
3.29 Any Infections

Study or Subgroup	Infliximab Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Random, 95% CI
3.29.1 12 weeks						
Braun 2002	12	34	18	35	41.5%	0.69 [0.39, 1.20]
Subtotal (95% CI)		**34**		**35**	**41.5%**	**0.69 [0.39, 1.20]**
Total events	12		18			
Heterogeneity: Not applicable						
Test for overall effect: Z = 1.32 (P = 0.19)						
3.29.2 24 weeks						
van der Heijde 2005	86	202	27	75	58.5%	1.18 [0.84, 1.66]
Subtotal (95% CI)		**202**		**75**	**58.5%**	**1.18 [0.84, 1.66]**
Total events	86		27			
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.96 (P = 0.34)						
Total (95% CI)		**236**		**110**	**100.0%**	**0.94 [0.56, 1.60]**
Total events	98		45			
Heterogeneity: Tau² = 0.09; Chi² = 2.66, df = 1 (P = 0.10); I² = 62%						
Test for overall effect: Z = 0.22 (P = 0.83)						



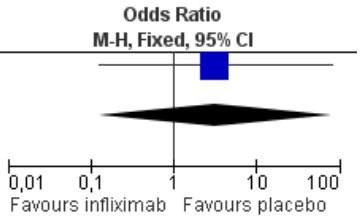
3.30 Serious infections

Study or Subgroup	Infliximab Events	Total	Placebo Events	Total	Weight	Peto Odds Ratio Peto, Fixed, 95% CI
3.30.1 12 weeks						
Braun 2002	1	34	0	35	38.8%	7.61 [0.15, 383.66]
Subtotal (95% CI)		**34**		**35**	**38.8%**	**7.61 [0.15, 383.66]**
Total events	1		0			
Heterogeneity: Not applicable						
Test for overall effect: Z = 1.01 (P = 0.31)						
3.30.2 24 weeks						
van der Heijde 2005	2	202	0	75	61.2%	3.96 [0.17, 90.10]
Subtotal (95% CI)		**202**		**75**	**61.2%**	**3.96 [0.17, 90.10]**
Total events	2		0			
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.86 (P = 0.39)						
Total (95% CI)		**236**		**110**	**100.0%**	**5.10 [0.44, 58.76]**
Total events	3		0			
Heterogeneity: Chi² = 0.07, df = 1 (P = 0.80); I² = 0%						
Test for overall effect: Z = 1.31 (P = 0.19)						
Test for subgroup differences: Chi² = 0.07, df = 1 (P = 0.80), I² = 0%						



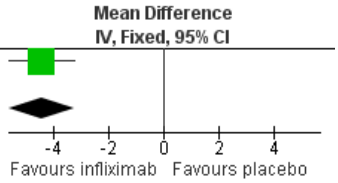
3.31 Tuberculosis

Study or Subgroup	Infliximab Events	Total	Placebo Events	Total	Weight	Odds Ratio M-H, Fixed, 95% CI
Braun 2002	1	34	0	35	100.0%	3.18 [0.13, 80.79]
Total (95% CI)		**34**		**35**	**100.0%**	**3.18 [0.13, 80.79]**
Total events	1		0			
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.70 (P = 0.48)						



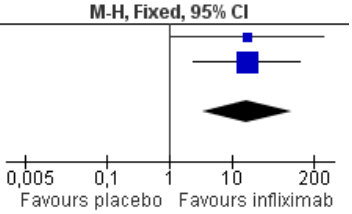
3.32 Spinal inflammation (MRI)

Study or Subgroup	Infliximab Mean	SD	Total	Placebo Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
van der Heijde 2005	-5.02	6.2	194	-0.6	3.4	72	100.0%	-4.42 [-5.59, -3.25]
Total (95% CI)			**194**			**72**	**100.0%**	**-4.42 [-5.59, -3.25]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 7.38 (P < 0.00001)								



3.33 Partial remission

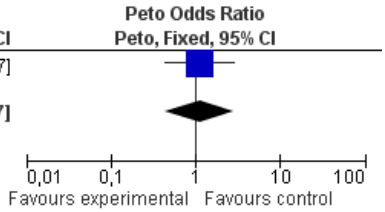
Study or Subgroup	Infliximab Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Braun 2002	8	34	0	35	25.5%	17.49 [1.05, 291.59]
van der Heijde 2005	45	201	1	78	74.5%	17.46 [2.45, 124.51]
Total (95% CI)		**235**		**113**	**100.0%**	**17.47 [3.42, 89.14]**
Total events	53		1			
Heterogeneity: Chi² = 0.00, df = 1 (P = 1.00); I² = 0%						
Test for overall effect: Z = 3.44 (P = 0.0006)						



3.38 Headache

Study or Subgroup	Etanercept Events	Total	Placebo Events	Total	Weight	Peto Odds Ratio Peto, Fixed, 95% CI
van der Heijde 2005	18	202	6	75	100.0%	1.12 [0.44, 2.87]
Total (95% CI)		**202**		**75**	**100.0%**	**1.12 [0.44, 2.87]**
Total events	18		6			

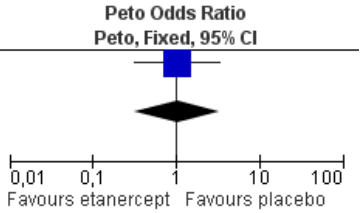
Heterogeneity: Not applicable
Test for overall effect: Z = 0.24 (P = 0.81)



3.39 Diarrhoea

Study or Subgroup	Etanercept Events	Total	Placebo Events	Total	Weight	Peto Odds Ratio Peto, Fixed, 95% CI
van der Heijde 2005	11	202	4	75	100.0%	1.02 [0.32, 3.29]
Total (95% CI)		**202**		**75**	**100.0%**	**1.02 [0.32, 3.29]**
Total events	11		4			

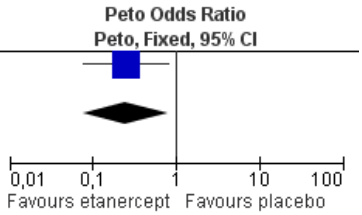
Heterogeneity: Not applicable
Test for overall effect: Z = 0.04 (P = 0.97)



3.40 Nausea

Study or Subgroup	Etanercept Events	Total	Placebo Events	Total	Weight	Peto Odds Ratio Peto, Fixed, 95% CI
van der Heijde 2005	7	202	8	75	100.0%	0.25 [0.08, 0.79]
Total (95% CI)		**202**		**75**	**100.0%**	**0.25 [0.08, 0.79]**
Total events	7		8			

Heterogeneity: Not applicable
Test for overall effect: Z = 2.35 (P = 0.02)

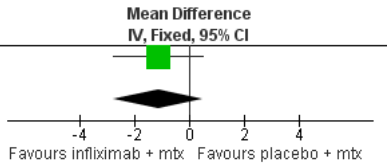


4 Infliximab + methotrexate versus placebo + methotrexate

4.1 BASDAI

Study or Subgroup	Infliximab + MTX Mean	SD	Total	Placebo + MTX Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Marzo-Ortega 2005	4.6	2.85	28	5.74	2.34	14	100.0%	-1.14 [-2.76, 0.48]
Total (95% CI)			**28**			**14**	**100.0%**	**-1.14 [-2.76, 0.48]**

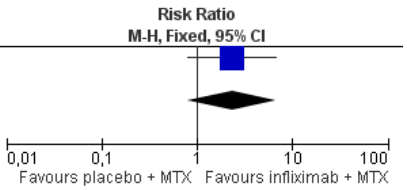
Heterogeneity: Not applicable
Test for overall effect: Z = 1.38 (P = 0.17)



4.2 ASAS20

Study or Subgroup	Infliximab + MTX Events	Total	Placebo + MTX Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Marzo-Ortega 2005	14	28	3	14	100.0%	2.33 [0.80, 6.80]
Total (95% CI)		**28**		**14**	**100.0%**	**2.33 [0.80, 6.80]**
Total events	14		3			

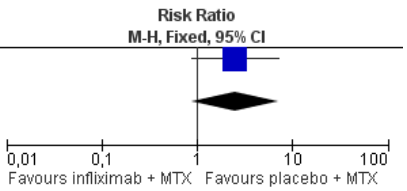
Heterogeneity: Not applicable
Test for overall effect: Z = 1.55 (P = 0.12)



4.3 >50% BASDAI

Study or Subgroup	Infliximab + MTX Events	Total	Placebo + MTX Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Marzo-Ortega 2005	15	28	3	14	100.0%	2.50 [0.87, 7.22]
Total (95% CI)		**28**		**14**	**100.0%**	**2.50 [0.87, 7.22]**
Total events	15		3			

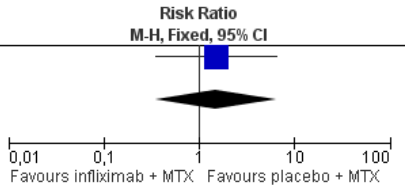
Heterogeneity: Not applicable
Test for overall effect: Z = 1.69 (P = 0.09)



4.4 Any infection

Study or Subgroup	Infliximab + MTX Events	Total	Placebo + MTX Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Marzo-Ortega 2005	6	28	2	14	100.0%	1.50 [0.35, 6.50]
Total (95% CI)		**28**		**14**	**100.0%**	**1.50 [0.35, 6.50]**
Total events	6		2			

Heterogeneity: Not applicable
Test for overall effect: Z = 0.54 (P = 0.59)

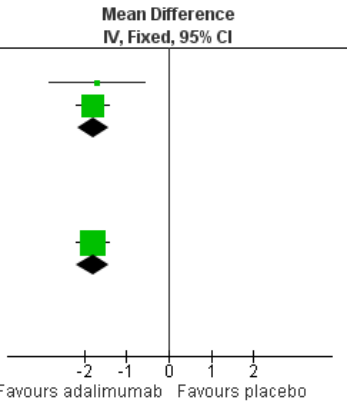


5 Adalimumab versus placebo

5.1 BASDAI (0-10 cm VAS)

Study or Subgroup	Adalimumab Mean	SD	Total	Placebo Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
5.1.1 12 weeks								
Lambert 2007	4.2	3	38	5.9	2.1	44	10.8%	-1.70 [-2.84, -0.56]
van der Heijde 2006a	3.7	1.7	208	5.5	1.7	107	89.2%	-1.80 [-2.20, -1.40]
Subtotal (95% CI)			**246**			**151**	**100.0%**	**-1.79 [-2.16, -1.41]**
Heterogeneity: Chi² = 0.03, df = 1 (P = 0.87); I² = 0%								
Test for overall effect: Z = 9.37 (P < 0.00001)								
5.1.2 24 weeks								
van der Heijde 2006a	3.7	1.7	208	5.5	1.7	107	100.0%	-1.80 [-2.20, -1.40]
Subtotal (95% CI)			**208**			**107**	**100.0%**	**-1.80 [-2.20, -1.40]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 8.90 (P < 0.00001)								

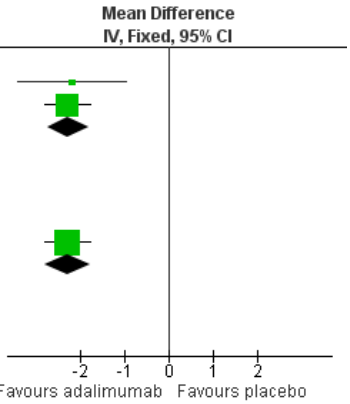
Test for subgroup differences: Chi² = 0.00, df = 1 (P = 0.97), I² = 0%



5.2 Total back pain (0-10cm)

Study or Subgroup	Adalimumab Mean	SD	Total	Placebo Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
5.2.1 12 weeks								
Lambert 2007	4.2	3.3	38	6.4	2.1	44	14.7%	-2.20 [-3.42, -0.98]
van der Heijde 2006a	3.8	2.1	208	6.1	2.2	107	85.3%	-2.30 [-2.81, -1.79]
Subtotal (95% CI)			**246**			**151**	**100.0%**	**-2.29 [-2.75, -1.82]**
Heterogeneity: Chi² = 0.02, df = 1 (P = 0.88); I² = 0%								
Test for overall effect: Z = 9.60 (P < 0.00001)								
5.2.2 24 weeks								
van der Heijde 2006a	3.7	2.1	208	6	2.2	107	100.0%	-2.30 [-2.81, -1.79]
Subtotal (95% CI)			**208**			**107**	**100.0%**	**-2.30 [-2.81, -1.79]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 8.92 (P < 0.00001)								

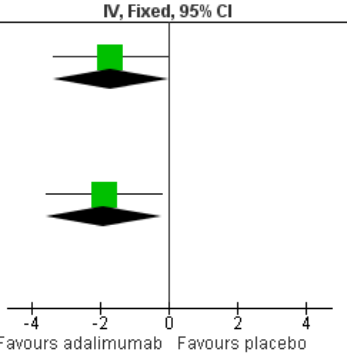
Test for subgroup differences: Chi² = 0.00, df = 1 (P = 0.97), I² = 0%



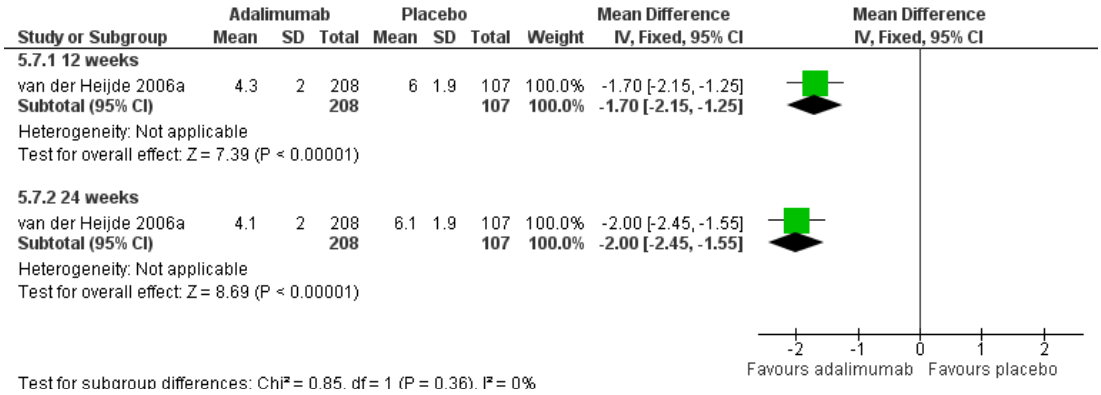
5.4 Entheseal pain MASES (range 0-13)

Study or Subgroup	Adalimumab Mean	SD	Total	Placebo Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
5.4.1 12 weeks								
van der Heijde 2006a	3.7	6.8	208	5.4	7.5	107	100.0%	-1.70 [-3.40, -0.00]
Subtotal (95% CI)			**208**			**107**	**100.0%**	**-1.70 [-3.40, -0.00]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 1.97 (P = 0.05)								
5.4.2 24weeks								
van der Heijde 2006a	3.2	6.8	208	5.1	7.5	107	100.0%	-1.90 [-3.60, -0.20]
Subtotal (95% CI)			**208**			**107**	**100.0%**	**-1.90 [-3.60, -0.20]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 2.20 (P = 0.03)								

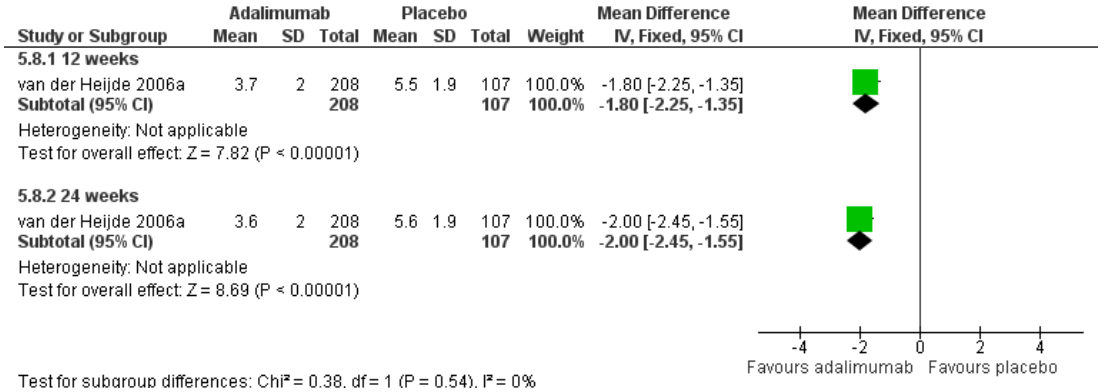
Test for subgroup differences: Chi² = 0.03, df = 1 (P = 0.87), I² = 0%



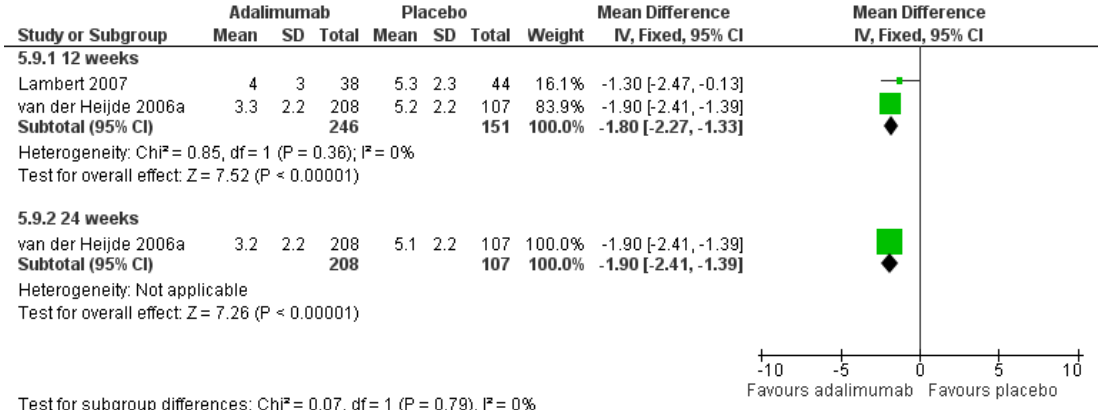
5.7 Fatigue (0-10cm VAS)



5.8 Morning stiffness - mean intensity and duration (0-10cm VAS)

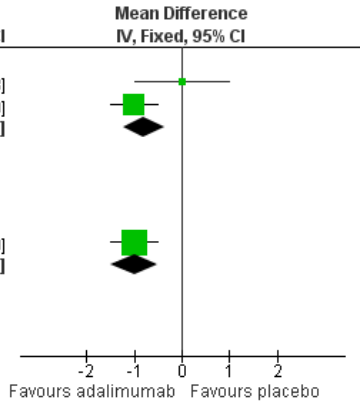


5.9 BASFI (0-10 VAS)



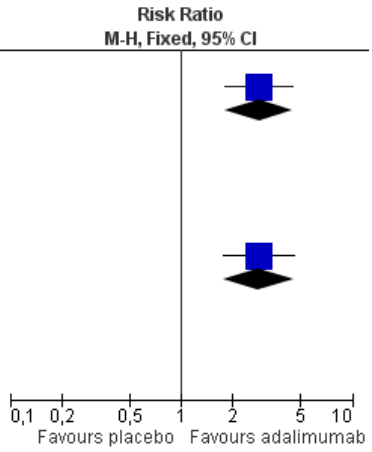
5.10 BASMI (0-10)

Study or Subgroup	Adalimumab Mean	SD	Total	Placebo Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
5.10.1 12 weeks								
Lambert 2007	3.7	2.3	38	3.7	2.2	44	20.6%	0.00 [-0.98, 0.98]
van der Heijde 2006a	3.3	2.2	208	4.3	2.1	107	79.4%	-1.00 [-1.50, -0.50]
Subtotal (95% CI)			**246**			**151**	**100.0%**	**-0.79 [-1.24, -0.35]**
Heterogeneity: Chi² = 3.19, df = 1 (P = 0.07); I² = 69%								
Test for overall effect: Z = 3.51 (P = 0.0004)								
5.10.2 24 weeks								
van der Heijde 2006a	3.2	2.2	208	4.2	2.1	107	100.0%	-1.00 [-1.50, -0.50]
Subtotal (95% CI)			**208**			**107**	**100.0%**	**-1.00 [-1.50, -0.50]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 3.94 (P < 0.0001)								
Test for subgroup differences: Chi² = 0.37, df = 1 (P = 0.55), I² = 0%								



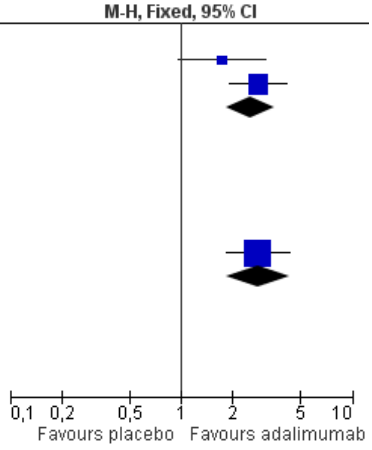
5.11 >50% improvement in BASDAI

Study or Subgroup	Adalimumab Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
5.11.1 12 weeks						
van der Heijde 2006a	94	208	17	107	100.0%	2.84 [1.79, 4.51]
Subtotal (95% CI)		**208**		**107**	**100.0%**	**2.84 [1.79, 4.51]**
Total events	94		17			
Heterogeneity: Not applicable						
Test for overall effect: Z = 4.45 (P < 0.00001)						
5.11.2 24 weeks						
van der Heijde 2006a	88	208	16	107	100.0%	2.83 [1.75, 4.57]
Subtotal (95% CI)		**208**		**107**	**100.0%**	**2.83 [1.75, 4.57]**
Total events	88		16			
Heterogeneity: Not applicable						
Test for overall effect: Z = 4.26 (P < 0.0001)						



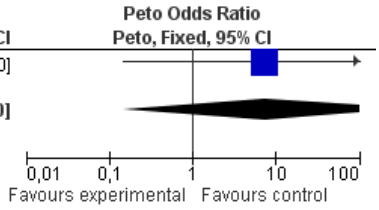
5.13 ASAS 20

Study or Subgroup	Adalimumab Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
5.13.1 12 weeks						
Lambert 2007	18	38	12	44	27.7%	1.74 [0.97, 3.13]
van der Heijde 2006a	121	208	22	107	72.3%	2.83 [1.92, 4.18]
Subtotal (95% CI)		**246**		**151**	**100.0%**	**2.53 [1.82, 3.51]**
Total events	139		34			
Heterogeneity: Chi² = 1.89, df = 1 (P = 0.17); I² = 47%						
Test for overall effect: Z = 5.55 (P < 0.00001)						
5.13.2 24 weeks						
van der Heijde 2006a	104	208	19	107	100.0%	2.82 [1.83, 4.33]
Subtotal (95% CI)		**208**		**107**	**100.0%**	**2.82 [1.83, 4.33]**
Total events	104		19			
Heterogeneity: Not applicable						
Test for overall effect: Z = 4.72 (P < 0.00001)						



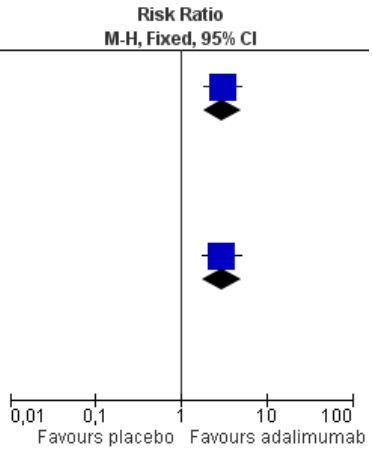
5.14 Hepatotoxicity (clinical symptoms)

Study or Subgroup	Adalimumab Events	Adalimumab Total	Placebo Events	Placebo Total	Weight	Peto Odds Ratio Peto, Fixed, 95% CI
van der Heijde 2006a	1	208	0	207	100.0%	7.35 [0.15, 370.60]
Total (95% CI)		**208**		**207**	**100.0%**	**7.35 [0.15, 370.60]**
Total events	1		0			
Heterogeneity: Not applicable						
Test for overall effect: Z = 1.00 (P = 0.32)						



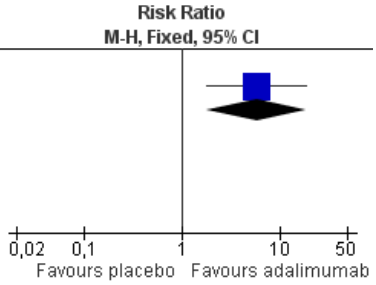
5.15 ASAS 40

Study or Subgroup	Adalimumab Events	Adalimumab Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
5.15.1 12 weeks						
van der Heijde 2006a	83	208	14	107	100.0%	3.05 [1.82, 5.11]
Subtotal (95% CI)		**208**		**107**	**100.0%**	**3.05 [1.82, 5.11]**
Total events	83		14			
Heterogeneity: Not applicable						
Test for overall effect: Z = 4.24 (P < 0.0001)						
5.15.2 24 weeks						
van der Heijde 2006a	82	208	14	107	100.0%	3.01 [1.80, 5.05]
Subtotal (95% CI)		**208**		**107**	**100.0%**	**3.01 [1.80, 5.05]**
Total events	82		14			
Heterogeneity: Not applicable						
Test for overall effect: Z = 4.18 (P < 0.0001)						



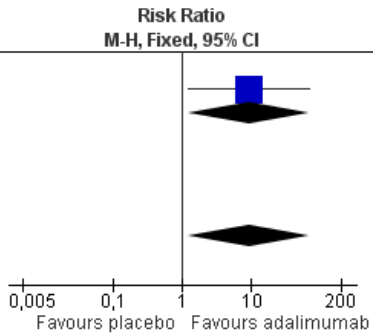
5.16 ASAS 50

Study or Subgroup	Adalimumab Events	Adalimumab Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
5.16.1 ASAS 50 -12 weeks						
Lambert 2007	15	38	3	44	100.0%	5.79 [1.81, 18.49]
Subtotal (95% CI)		**38**		**44**	**100.0%**	**5.79 [1.81, 18.49]**
Total events	15		3			
Heterogeneity: Not applicable						
Test for overall effect: Z = 2.96 (P = 0.003)						



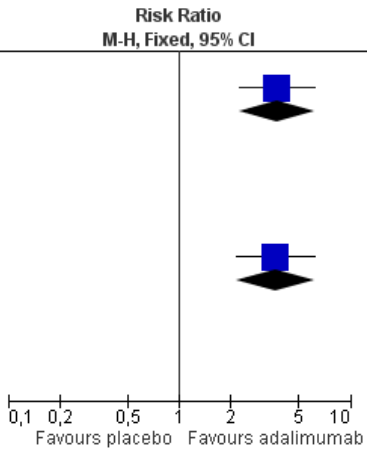
5.17 ASAS 70

Study or Subgroup	Adalimumab Events	Adalimumab Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
5.17.1 ASAS 70 -12 weeks						
Lambert 2007	8	38	1	44	100.0%	9.26 [1.21, 70.75]
Subtotal (95% CI)		**38**		**44**	**100.0%**	**9.26 [1.21, 70.75]**
Total events	8		1			
Heterogeneity: Not applicable						
Test for overall effect: Z = 2.15 (P = 0.03)						
Total (95% CI)		**38**		**44**	**100.0%**	**9.26 [1.21, 70.75]**
Total events	8		1			
Heterogeneity: Not applicable						
Test for overall effect: Z = 2.15 (P = 0.03)						



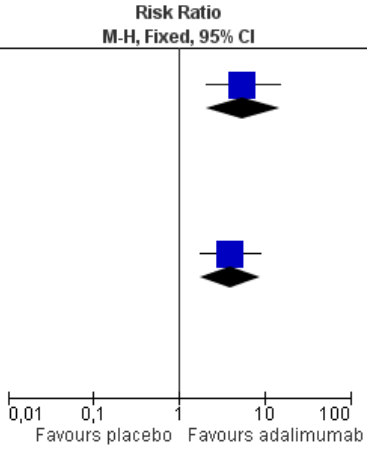
5.18 ASAS 5 out of 6

Study or Subgroup	Adalimumab Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
5.18.1 12 weeks						
van der Heijde 2006a	101	208	14	107	100.0%	3.71 [2.23, 6.17]
Subtotal (95% CI)		**208**		**107**	**100.0%**	**3.71 [2.23, 6.17]**
Total events	101		14			
Heterogeneity: Not applicable						
Test for overall effect: Z = 5.06 (P < 0.00001)						
5.18.2 24 weeks						
van der Heijde 2006a	93	208	13	107	100.0%	3.68 [2.16, 6.26]
Subtotal (95% CI)		**208**		**107**	**100.0%**	**3.68 [2.16, 6.26]**
Total events	93		13			
Heterogeneity: Not applicable						
Test for overall effect: Z = 4.81 (P < 0.00001)						



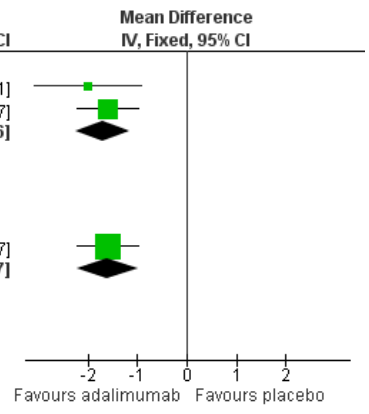
5.19 ASAS partial remission

Study or Subgroup	Adalimumab Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
5.19.1 12 weeks						
van der Heijde 2006a	43	208	4	107	100.0%	5.53 [2.04, 15.00]
Subtotal (95% CI)		**208**		**107**	**100.0%**	**5.53 [2.04, 15.00]**
Total events	43		4			
Heterogeneity: Not applicable						
Test for overall effect: Z = 3.36 (P = 0.0008)						
5.19.2 24 weeks						
van der Heijde 2006a	46	208	6	107	100.0%	3.94 [1.74, 8.94]
Subtotal (95% CI)		**208**		**107**	**100.0%**	**3.94 [1.74, 8.94]**
Total events	46		6			
Heterogeneity: Not applicable						
Test for overall effect: Z = 3.29 (P = 0.001)						

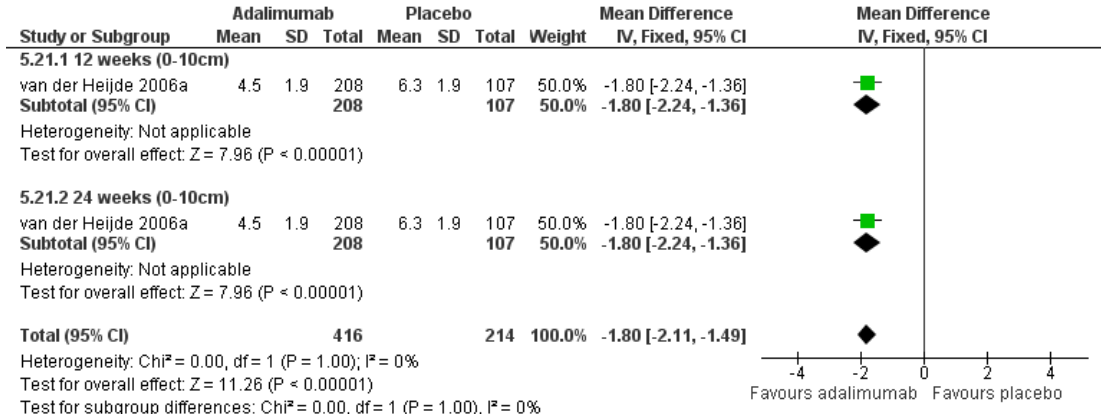


5.20 C-reactive protein (mg/dl)

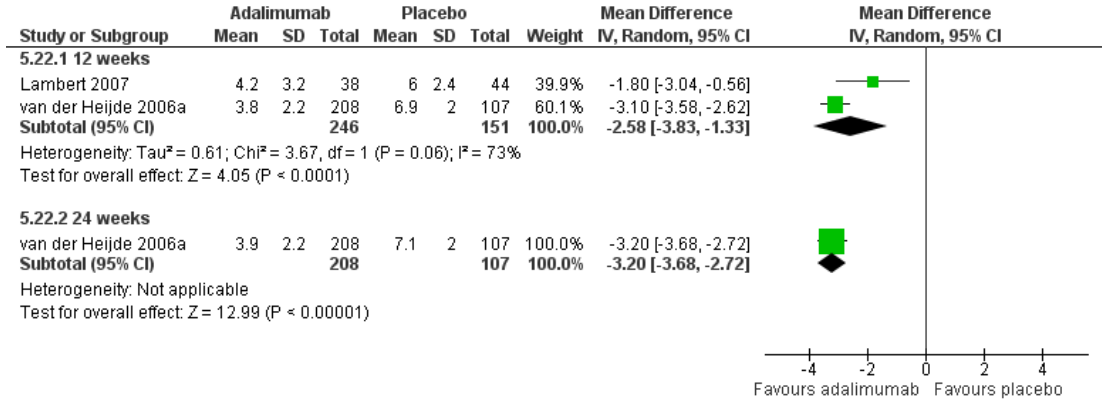
Study or Subgroup	Adalimumab Mean	SD	Total	Placebo Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
5.20.1 12 weeks								
Lambert 2007	0.4	0.7	38	2.4	3.6	44	24.9%	-2.00 [-3.09, -0.91]
van der Heijde 2006a	0.5	2.2	208	2.1	2.9	107	75.1%	-1.60 [-2.23, -0.97]
Subtotal (95% CI)			**246**			**151**	**100.0%**	**-1.70 [-2.24, -1.16]**
Heterogeneity: Chi² = 0.39, df = 1 (P = 0.53); I² = 0%								
Test for overall effect: Z = 6.14 (P < 0.00001)								
5.20.2 24 weeks								
van der Heijde 2006a	0.5	2.2	208	2.1	2.9	107	100.0%	-1.60 [-2.23, -0.97]
Subtotal (95% CI)			**208**			**107**	**100.0%**	**-1.60 [-2.23, -0.97]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 5.01 (P < 0.00001)								
Test for subgroup differences: Chi² = 0.06, df = 1 (P = 0.81), I² = 0%								



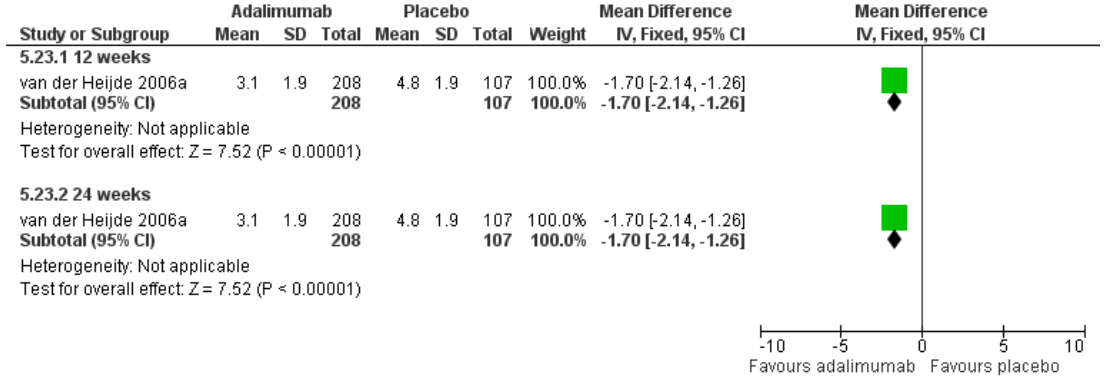
5.21 BASG



5.22 Patient global assessment (0-10cm VAS)

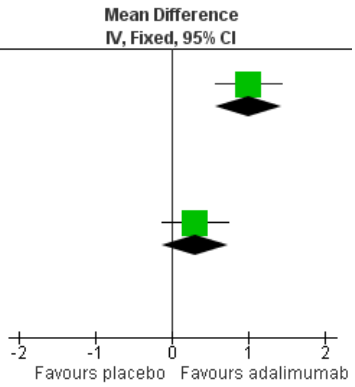


5.23 Physician global assessment (0-10cm VAS)



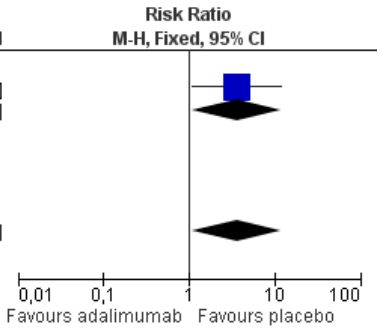
5.24 Chest expansion (cm)

Study or Subgroup	Adalimumab Mean	SD	Total	Placebo Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
5.24.1 12 weeks								
van der Heijde 2006a	4.3	1.8	208	3.3	1.9	107	100.0%	1.00 [0.56, 1.44]
Subtotal (95% CI)			**208**			**107**	**100.0%**	**1.00 [0.56, 1.44]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 4.50 (P < 0.00001)								
5.24.2 24 weeks								
van der Heijde 2006a	3.7	1.8	208	3.4	1.9	107	100.0%	0.30 [-0.14, 0.74]
Subtotal (95% CI)			**208**			**107**	**100.0%**	**0.30 [-0.14, 0.74]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 1.35 (P = 0.18)								



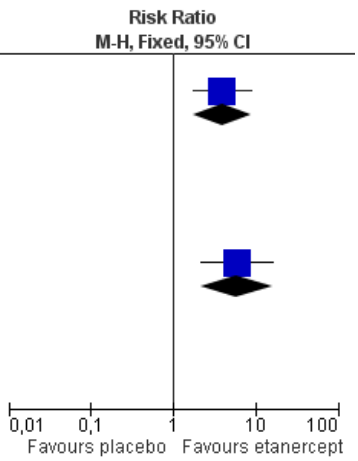
5.25 Injection site reactions

Study or Subgroup	Adalimumab Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
5.25.1 24 weeks						
van der Heijde 2006a	21	208	3	107	100.0%	3.60 [1.10, 11.80]
Subtotal (95% CI)		**208**		**107**	**100.0%**	**3.60 [1.10, 11.80]**
Total events	21		3			
Heterogeneity: Not applicable						
Test for overall effect: Z = 2.12 (P = 0.03)						
Total (95% CI)		**208**		**107**	**100.0%**	**3.60 [1.10, 11.80]**
Total events	21		3			
Heterogeneity: Not applicable						
Test for overall effect: Z = 2.12 (P = 0.03)						



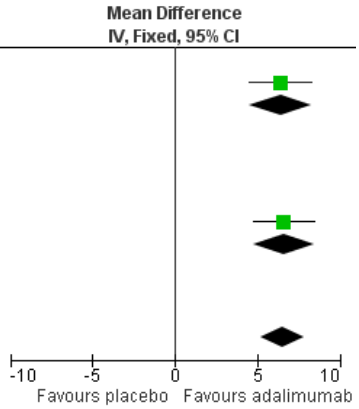
5.26 Partial remission

Study or Subgroup	Adalimumab Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
5.26.1 12 weeks						
van der Heijde 2006a	46	208	6	107	100.0%	3.94 [1.74, 8.94]
Subtotal (95% CI)		**208**		**107**	**100.0%**	**3.94 [1.74, 8.94]**
Total events	46		6			
Heterogeneity: Not applicable						
Test for overall effect: Z = 3.29 (P = 0.001)						
5.26.2 24 weeks						
van der Heijde 2006a	46	208	4	107	100.0%	5.92 [2.19, 16.00]
Subtotal (95% CI)		**208**		**107**	**100.0%**	**5.92 [2.19, 16.00]**
Total events	46		4			
Heterogeneity: Not applicable						
Test for overall effect: Z = 3.50 (P = 0.0005)						



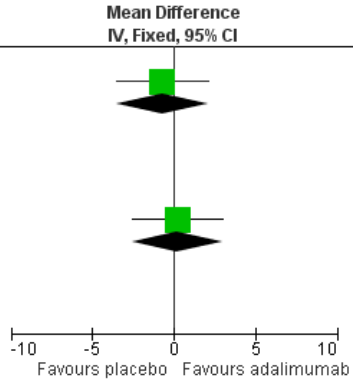
5.30 SF-36 Physical functioning

Study or Subgroup	Adalimumab Mean	SD	Total	Placebo Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
5.30.1 SF-36 PCS - 12 weeks								
van der Heijde 2006a	39.8	8	208	33.4	8	107	50.0%	6.40 [4.53, 8.27]
Subtotal (95% CI)			**208**			**107**	**50.0%**	**6.40 [4.53, 8.27]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 6.72 (P < 0.00001)								
5.30.2 SF-36 PCS - 24 weeks								
van der Heijde 2006a	40.3	8	208	33.7	8	107	50.0%	6.60 [4.73, 8.47]
Subtotal (95% CI)			**208**			**107**	**50.0%**	**6.60 [4.73, 8.47]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 6.93 (P < 0.00001)								
Total (95% CI)			**416**			**214**	**100.0%**	**6.50 [5.18, 7.82]**
Heterogeneity: Chi² = 0.02, df = 1 (P = 0.88); I² = 0%								
Test for overall effect: Z = 9.66 (P < 0.00001)								
Test for subgroup differences: Chi² = 0.02, df = 1 (P = 0.88), I² = 0%								



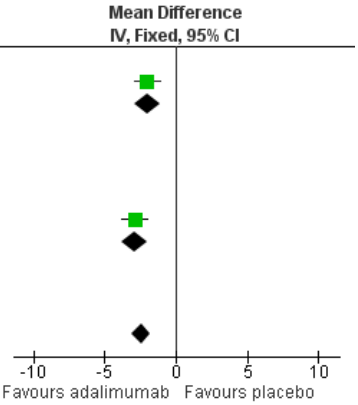
5.31 SF-36 Mental health

Study or Subgroup	Adalimumab Mean	SD	Total	Placebo Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
5.31.1 SF-36 MCS - 12 weeks								
van der Heijde 2006a	46.1	12	208	46.8	12	107	100.0%	-0.70 [-3.50, 2.10]
Subtotal (95% CI)			**208**			**107**	**100.0%**	**-0.70 [-3.50, 2.10]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.49 (P = 0.62)								
5.31.2 SF-36 MCS - 24 weeks								
van der Heijde 2006a	47	12	208	46.8	12	107	100.0%	0.20 [-2.60, 3.00]
Subtotal (95% CI)			**208**			**107**	**100.0%**	**0.20 [-2.60, 3.00]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.14 (P = 0.89)								



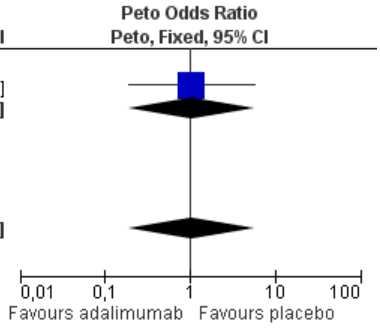
5.32 ASQoL

Study or Subgroup	Adalimumab Mean	SD	Total	Placebo Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
5.32.1 ASQoL - 12 weeks								
van der Heijde 2006a	7	4	208	9	4	107	50.0%	-2.00 [-2.93, -1.07]
Subtotal (95% CI)			**208**			**107**	**50.0%**	**-2.00 [-2.93, -1.07]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 4.20 (P < 0.0001)								
5.32.2 ASQoL - 24 weeks								
van der Heijde 2006a	6.6	4	208	9.5	4	107	50.0%	-2.90 [-3.83, -1.97]
Subtotal (95% CI)			**208**			**107**	**50.0%**	**-2.90 [-3.83, -1.97]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 6.09 (P < 0.00001)								
Total (95% CI)			**416**			**214**	**100.0%**	**-2.45 [-3.11, -1.79]**
Heterogeneity: Chi² = 1.79, df = 1 (P = 0.18); I² = 44%								
Test for overall effect: Z = 7.28 (P < 0.00001)								
Test for subgroup differences: Chi² = 1.79, df = 1 (P = 0.18), I² = 44.1%								



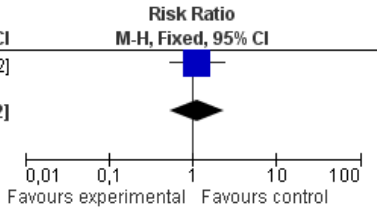
5.33 Withdrawals due to adverse events

Study or Subgroup	Adalimumab Events	Adalimumab Total	Placebo Events	Placebo Total	Weight	Peto Odds Ratio Peto, Fixed, 95% CI
5.33.1 24 weeks						
van der Heijde 2006a	4	208	2	107	100.0%	1.03 [0.19, 5.65]
Subtotal (95% CI)		**208**		**107**	**100.0%**	**1.03 [0.19, 5.65]**
Total events	4		2			
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.03 (P = 0.97)						
Total (95% CI)		**208**		**107**	**100.0%**	**1.03 [0.19, 5.65]**
Total events	4		2			
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.03 (P = 0.97)						
Test for subgroup differences: Not applicable						



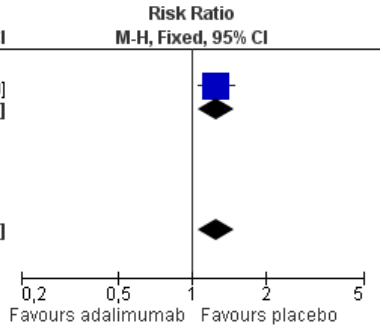
5.34 Headache

Study or Subgroup	Adalimumab Events	Adalimumab Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
van der Heijde 2006a	20	208	9	107	100.0%	1.14 [0.54, 2.42]
Total (95% CI)		**208**		**107**	**100.0%**	**1.14 [0.54, 2.42]**
Total events	20		9			
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.35 (P = 0.73)						



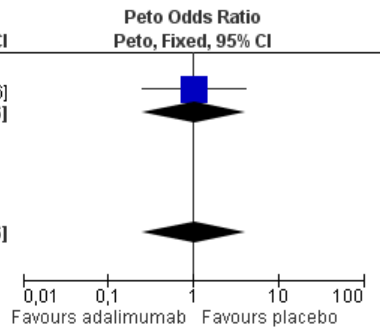
5.35 Total adverse events

Study or Subgroup	Adalimumab Events	Adalimumab Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Fixed, 95% CI
5.35.1 24 weeks						
van der Heijde 2006a	156	208	64	107	100.0%	1.25 [1.05, 1.49]
Subtotal (95% CI)		**208**		**107**	**100.0%**	**1.25 [1.05, 1.49]**
Total events	156		64			
Heterogeneity: Not applicable						
Test for overall effect: Z = 2.55 (P = 0.01)						
Total (95% CI)		**208**		**107**	**100.0%**	**1.25 [1.05, 1.49]**
Total events	156		64			
Heterogeneity: Not applicable						
Test for overall effect: Z = 2.55 (P = 0.01)						

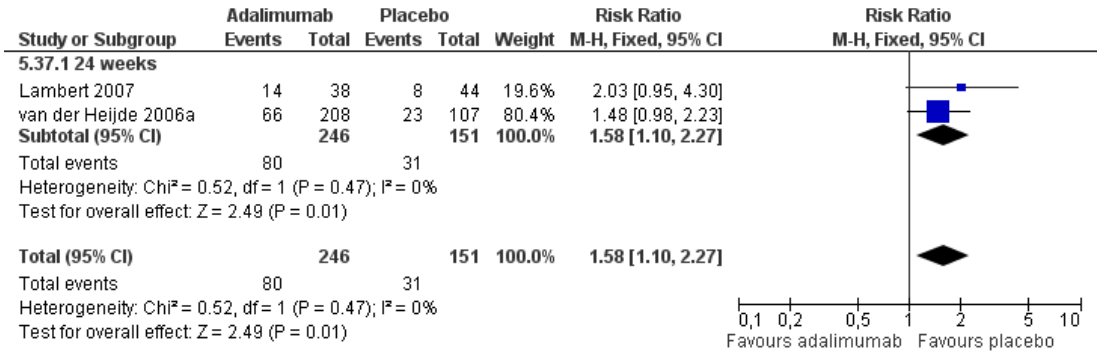


5.36 Serious adverse events

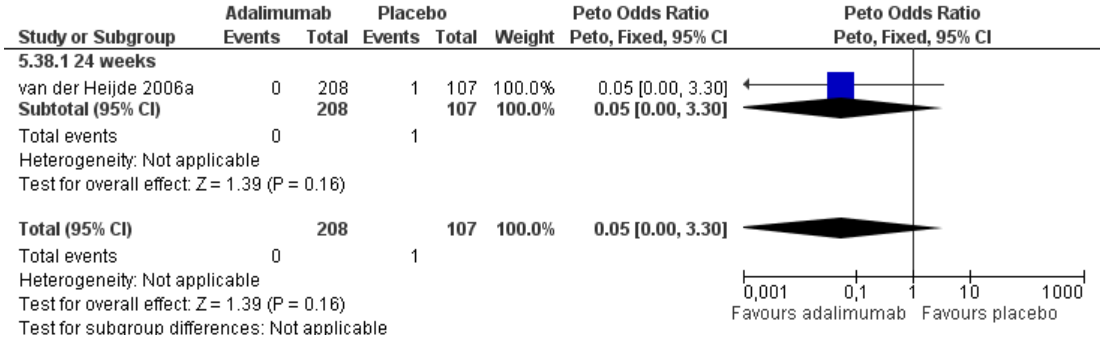
Study or Subgroup	Adalimumab Events	Adalimumab Total	Placebo Events	Placebo Total	Weight	Peto Odds Ratio Peto, Fixed, 95% CI
5.36.1 24 weeks						
van der Heijde 2006a	6	208	3	107	100.0%	1.03 [0.25, 4.16]
Subtotal (95% CI)		**208**		**107**	**100.0%**	**1.03 [0.25, 4.16]**
Total events	6		3			
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.04 (P = 0.97)						
Total (95% CI)		**208**		**107**	**100.0%**	**1.03 [0.25, 4.16]**
Total events	6		3			
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.04 (P = 0.97)						
Test for subgroup differences: Not applicable						



5.37 Any infection

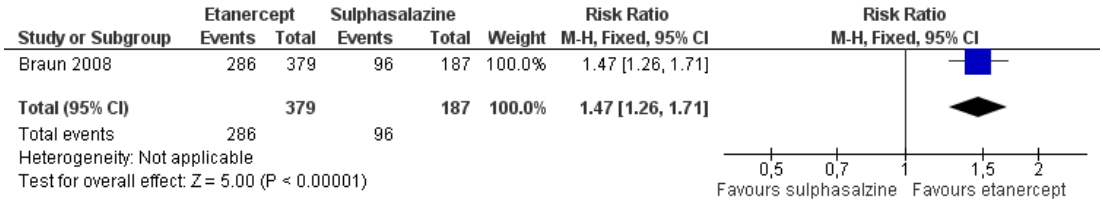


5.38 Serious infections

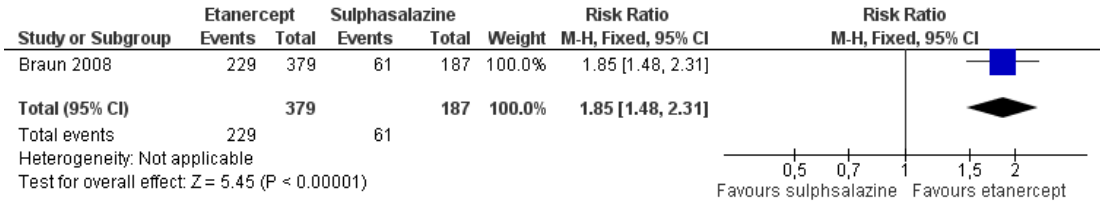


6 Etanercept versus sulphasalazine

6.1 ASAS20 - 16wk



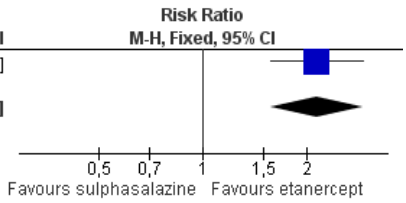
6.2 ASAS40 - 16 wk



6.3 ASAS5/6 - 16 wk

Study or Subgroup	Etanercept Events	Total	Sulphasalazine Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Braun 2008	164	375	38	185	100.0%	2.13 [1.57, 2.89]
Total (95% CI)		**375**		**185**	**100.0%**	**2.13 [1.57, 2.89]**
Total events	164		38			

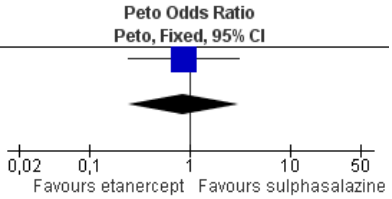
Heterogeneity: Not applicable
Test for overall effect: Z = 4.84 (P < 0.00001)



6.4 Serious adverse events - 16 wk

Study or Subgroup	Etanercept Events	Total	Sulphasalazine Events	Total	Weight	Peto Odds Ratio Peto, Fixed, 95% CI
Braun 2008	7	379	4	187	100.0%	0.86 [0.24, 3.05]
Total (95% CI)		**379**		**187**	**100.0%**	**0.86 [0.24, 3.05]**
Total events	7		4			

Heterogeneity: Not applicable
Test for overall effect: Z = 0.24 (P = 0.81)

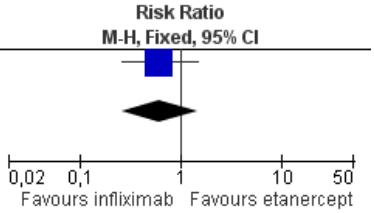


7 Etanercept versus infliximab

7.1 ASAS 50-6 weeks

Study or Subgroup	Etanercept Events	Total	Infliximab Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Giardina 2009	6	26	9	24	100.0%	0.62 [0.26, 1.47]
Total (95% CI)		**26**		**24**	**100.0%**	**0.62 [0.26, 1.47]**
Total events	6		9			

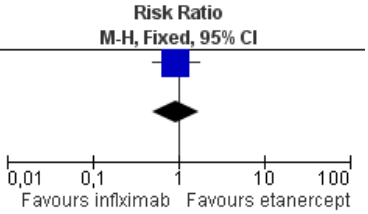
Heterogeneity: Not applicable
Test for overall effect: Z = 1.09 (P = 0.27)



7.2 ASAS 50-104 weeks

Study or Subgroup	Etanercept Events	Total	Infliximab Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Giardina 2009	11	26	11	24	100.0%	0.92 [0.49, 1.72]
Total (95% CI)		**26**		**24**	**100.0%**	**0.92 [0.49, 1.72]**
Total events	11		11			

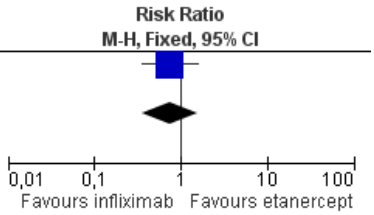
Heterogeneity: Not applicable
Test for overall effect: Z = 0.25 (P = 0.80)



7.3 BASDAI 50-6 weeks

Study or Subgroup	Etanercept Events	Total	Infliximab Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Giardina 2009	8	26	10	24	100.0%	0.74 [0.35, 1.56]
Total (95% CI)		**26**		**24**	**100.0%**	**0.74 [0.35, 1.56]**
Total events	8		10			

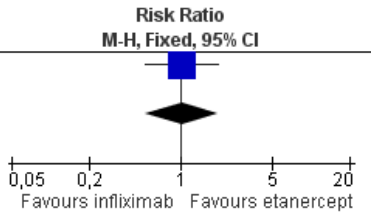
Heterogeneity: Not applicable
Test for overall effect: Z = 0.80 (P = 0.43)



7.4 BASDAI 50-104 weeks

Study or Subgroup	Etanercept Events	Total	Infliximab Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Giardina 2009	11	26	10	24	100.0%	1.02 [0.53, 1.95]
Total (95% CI)		**26**		**24**	**100.0%**	**1.02 [0.53, 1.95]**
Total events	11		10			

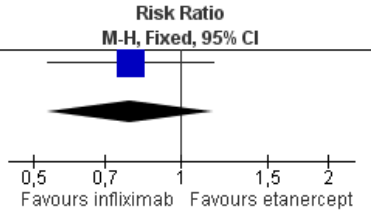
Heterogeneity: Not applicable
Test for overall effect: Z = 0.05 (P = 0.96)



7.5 ASAS 20 - 12 weeks

Study or Subgroup	Etanercept Events	Total	Infliximab Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Giardina 2009	15	25	19	25	100.0%	0.79 [0.54, 1.16]
Total (95% CI)		**25**		**25**	**100.0%**	**0.79 [0.54, 1.16]**
Total events	15		19			

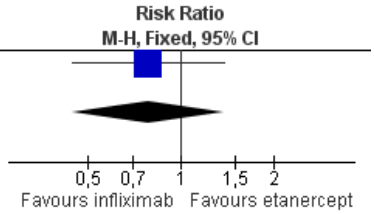
Heterogeneity: Not applicable
Test for overall effect: Z = 1.19 (P = 0.23)



7.6 ASAS40

Study or Subgroup	Etanercept Events	Total	Infliximab Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Giardina 2009	11	25	14	25	100.0%	0.79 [0.45, 1.38]
Total (95% CI)		**25**		**25**	**100.0%**	**0.79 [0.45, 1.38]**
Total events	11		14			

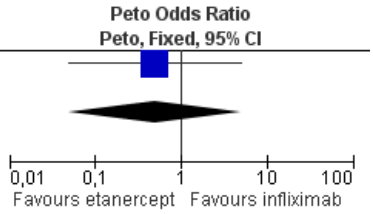
Heterogeneity: Not applicable
Test for overall effect: Z = 0.84 (P = 0.40)



7.7 Serious infections

Study or Subgroup	Etanercept Events	Total	Infliximab Events	Total	Weight	Peto Odds Ratio Peto, Fixed, 95% CI
Giardina 2009	1	25	2	25	100.0%	0.50 [0.05, 5.03]
Total (95% CI)		**25**		**25**	**100.0%**	**0.50 [0.05, 5.03]**
Total events	1		2			

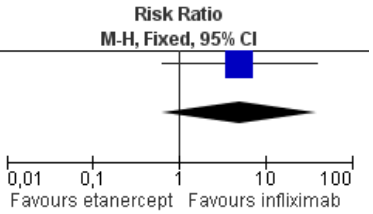
Heterogeneity: Not applicable
Test for overall effect: Z = 0.59 (P = 0.56)



7.8 Injection site reaction

Study or Subgroup	Etancerept Events	Total	Infliximab Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Giardina 2009	5	25	1	25	100.0%	5.00 [0.63, 39.79]
Total (95% CI)		**25**		**25**	**100.0%**	**5.00 [0.63, 39.79]**
Total events	5		1			

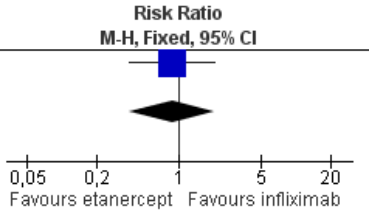
Heterogeneity: Not applicable
Test for overall effect: Z = 1.52 (P = 0.13)



7.9 Headache

Study or Subgroup	Etanercept Events	Total	Infliximab Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Giardina 2009	7	25	8	25	100.0%	0.88 [0.37, 2.05]
Total (95% CI)		**25**		**25**	**100.0%**	**0.88 [0.37, 2.05]**
Total events	7		8			

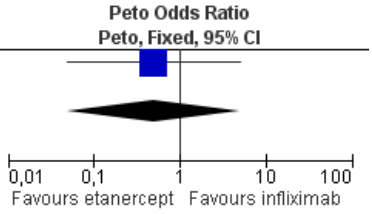
Heterogeneity: Not applicable
Test for overall effect: Z = 0.31 (P = 0.76)



7.10 Diarrhoea

Study or Subgroup	Etanercept Events	Total	Infliximab Events	Total	Weight	Peto Odds Ratio Peto, Fixed, 95% CI
Giardina 2009	1	25	2	25	100.0%	0.50 [0.05, 5.03]
Total (95% CI)		**25**		**25**	**100.0%**	**0.50 [0.05, 5.03]**
Total events	1		2			

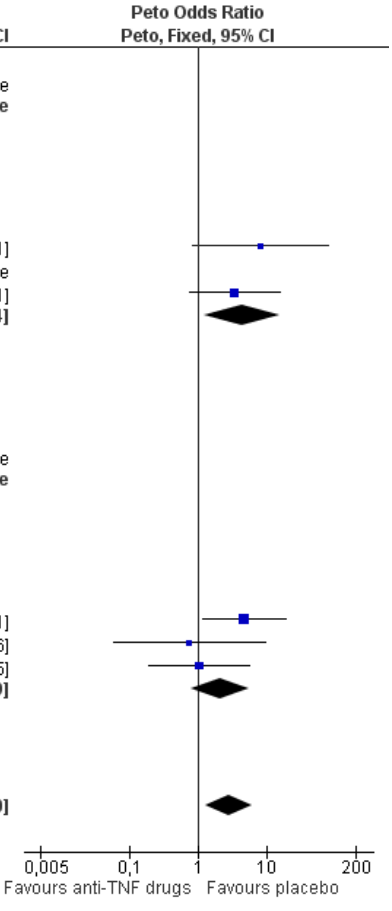
Heterogeneity: Not applicable
Test for overall effect: Z = 0.59 (P = 0.56)



8 TNF-inhibitors versus placebo

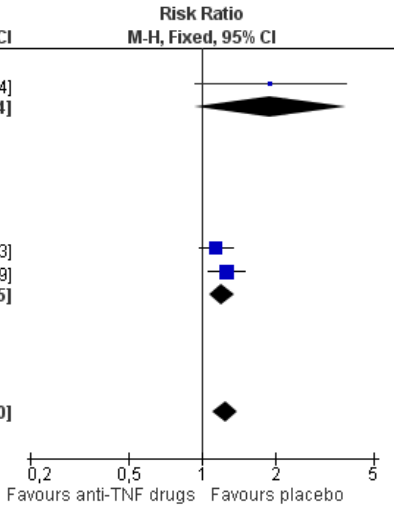
8.1 Withdrawals due to AE

Study or Subgroup	Anti-TNF drugs Events	Total	Placebo Events	Total	Weight	Peto Odds Ratio Peto, Fixed, 95% CI
8.1.1 6 weeks						
Brandt 2003	0	14	0	16		Not estimable
Subtotal (95% CI)		**14**		**16**		**Not estimable**
Total events	0		0			
Heterogeneity: Not applicable						
Test for overall effect: Not applicable						
8.1.2 12 weeks						
Braun 2002	3	34	0	35	11.7%	8.09 [0.81, 80.51]
Calin 2004	0	45	0	39		Not estimable
van der Heijde 2006b	14	305	0	51	26.5%	3.36 [0.73, 15.41]
Subtotal (95% CI)		**384**		**125**	**38.2%**	**4.39 [1.23, 15.64]**
Total events	17		0			
Heterogeneity: Chi² = 0.39, df = 1 (P = 0.53); I² = 0%						
Test for overall effect: Z = 2.28 (P = 0.02)						
8.1.3 16 weeks						
Gorman 2002	0	20	0	20		Not estimable
Subtotal (95% CI)		**20**		**20**		**Not estimable**
Total events	0		0			
Heterogeneity: Not applicable						
Test for overall effect: Not applicable						
8.1.4 24 weeks						
Davis 2003	7	138	1	139	31.2%	4.69 [1.15, 19.11]
van der Heijde 2005	2	202	1	75	9.4%	0.73 [0.06, 9.36]
van der Heijde 2006a	4	208	2	107	21.2%	1.03 [0.19, 5.65]
Subtotal (95% CI)		**548**		**321**	**61.8%**	**2.10 [0.77, 5.69]**
Total events	13		4			
Heterogeneity: Chi² = 2.60, df = 2 (P = 0.27); I² = 23%						
Test for overall effect: Z = 1.46 (P = 0.14)						
Total (95% CI)		**966**		**482**	**100.0%**	**2.78 [1.27, 6.10]**
Total events	30		4			
Heterogeneity: Chi² = 3.79, df = 4 (P = 0.43); I² = 0%						
Test for overall effect: Z = 2.56 (P = 0.01)						
Test for subgroup differences: Chi² = 0.80, df = 1 (P = 0.37), I² = 0%						



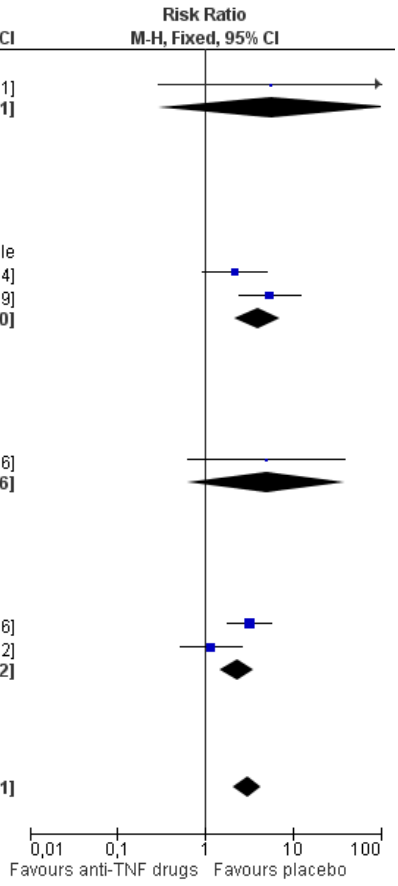
8.2 Total adverse events

Study or Subgroup	Anti-TNF drugs Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
8.2.1 6 weeks						
Huang 2008	18	74	10	78	5.6%	1.90 [0.94, 3.84]
Subtotal (95% CI)		**74**		**78**	**5.6%**	**1.90 [0.94, 3.84]**
Total events	18		10			
Heterogeneity: Not applicable						
Test for overall effect: Z = 1.78 (P = 0.07)						
8.2.2 24 weeks						
van der Heijde 2005	166	202	54	75	45.5%	1.14 [0.98, 1.33]
van der Heijde 2006a	156	208	64	107	48.9%	1.25 [1.05, 1.49]
Subtotal (95% CI)		**410**		**182**	**94.4%**	**1.20 [1.07, 1.35]**
Total events	322		118			
Heterogeneity: Chi² = 0.64, df = 1 (P = 0.42); I² = 0%						
Test for overall effect: Z = 3.04 (P = 0.002)						
Total (95% CI)		**484**		**260**	**100.0%**	**1.24 [1.10, 1.40]**
Total events	340		128			
Heterogeneity: Chi² = 2.50, df = 2 (P = 0.29); I² = 20%						
Test for overall effect: Z = 3.51 (P = 0.0004)						
Test for subgroup differences: Not applicable						

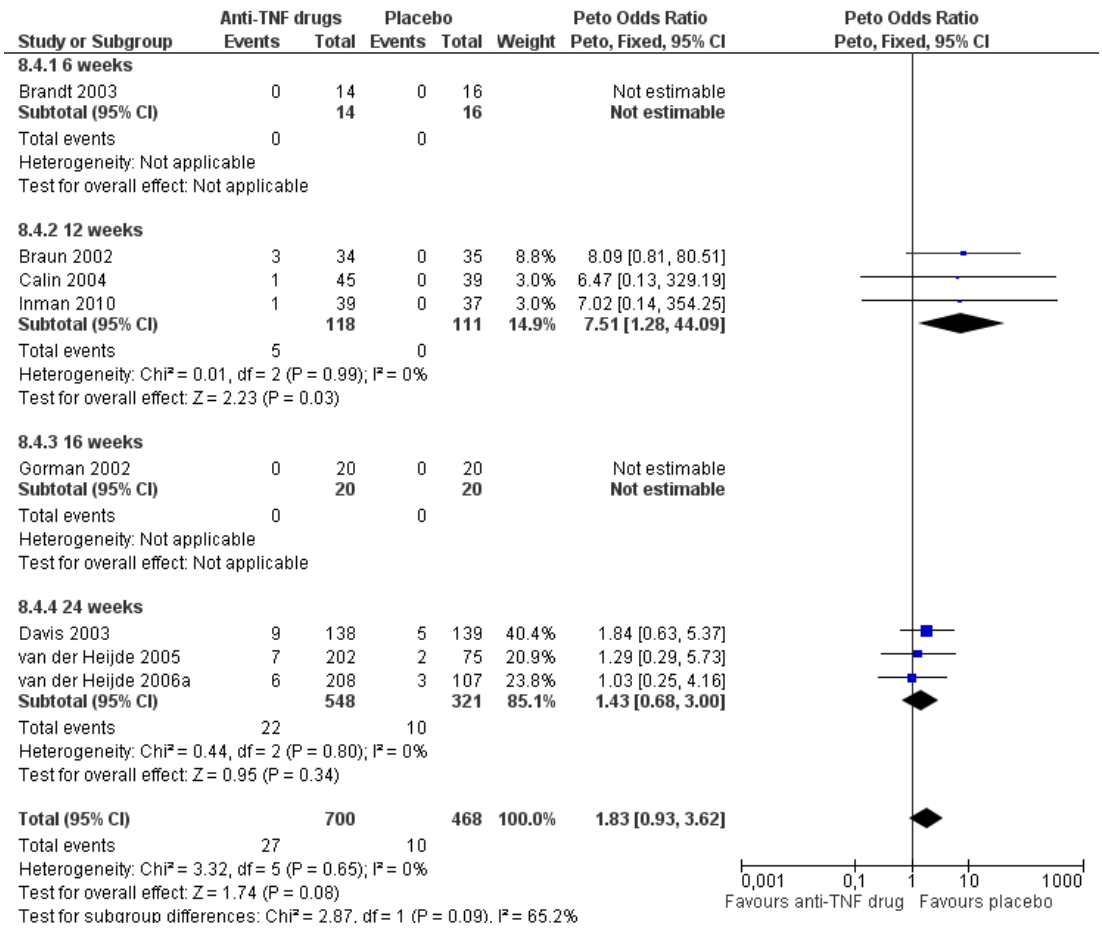


8.3 Injection/Infusion site reaction

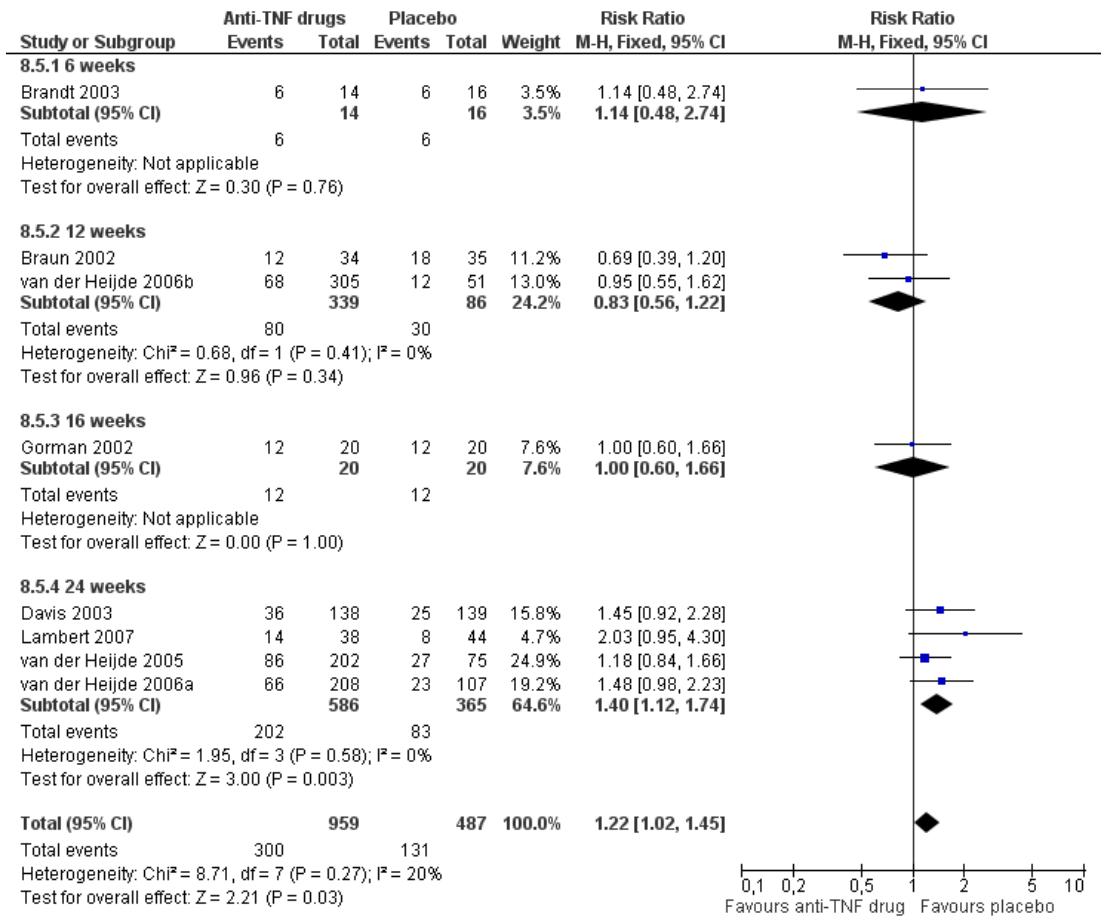
Study or Subgroup	Anti-TNF drugs Events	Total	Placebo Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
8.3.1 6 weeks						
Brandt 2003	2	14	0	16	1.2%	5.67 [0.29, 108.91]
Subtotal (95% CI)		**14**		**16**	**1.2%**	**5.67 [0.29, 108.91]**
Total events	2		0			
Heterogeneity: Not applicable						
Test for overall effect: Z = 1.15 (P = 0.25)						
8.3.2 12 weeks						
Braun 2002	0	34	0	35		Not estimable
Calin 2004	15	45	6	39	16.4%	2.17 [0.93, 5.04]
van der Heijde 2006b	66	305	6	150	20.6%	5.41 [2.40, 12.19]
Subtotal (95% CI)		**384**		**224**	**37.0%**	**3.97 [2.19, 7.20]**
Total events	81		12			
Heterogeneity: Chi² = 2.53, df = 1 (P = 0.11); I² = 61%						
Test for overall effect: Z = 4.54 (P < 0.00001)						
8.3.3 16 weeks						
Gorman 2002	5	20	1	20	2.6%	5.00 [0.64, 39.06]
Subtotal (95% CI)		**20**		**20**	**2.6%**	**5.00 [0.64, 39.06]**
Total events	5		1			
Heterogeneity: Not applicable						
Test for overall effect: Z = 1.53 (P = 0.12)						
8.3.4 24 weeks						
Davis 2003	41	138	13	139	33.1%	3.18 [1.78, 5.66]
van der Heijde 2005	22	202	7	75	26.1%	1.17 [0.52, 2.62]
Subtotal (95% CI)		**340**		**214**	**59.2%**	**2.29 [1.45, 3.62]**
Total events	63		20			
Heterogeneity: Chi² = 3.91, df = 1 (P = 0.05); I² = 74%						
Test for overall effect: Z = 3.55 (P = 0.0004)						
Total (95% CI)		**758**		**474**	**100.0%**	**3.02 [2.12, 4.31]**
Total events	151		33			
Heterogeneity: Chi² = 8.33, df = 5 (P = 0.14); I² = 40%						
Test for overall effect: Z = 6.10 (P < 0.00001)						



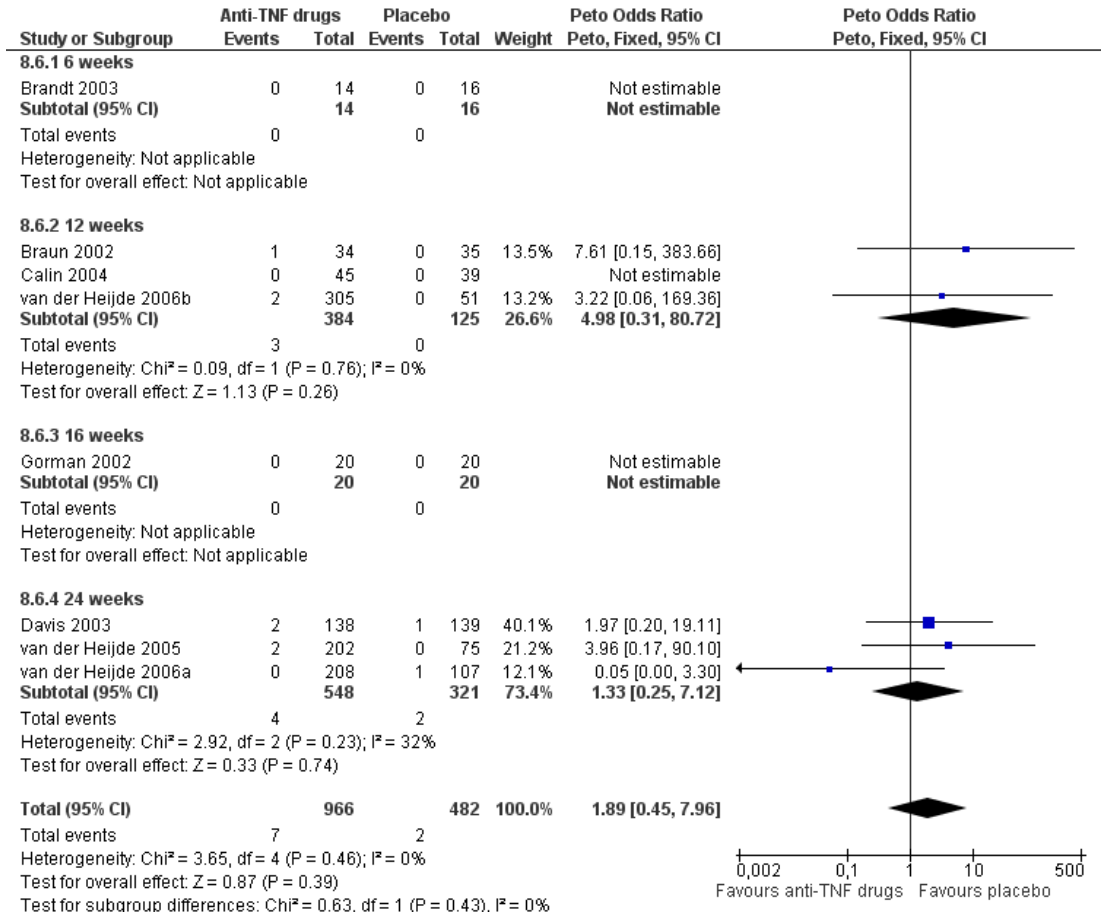
8.4 Serious adverse events



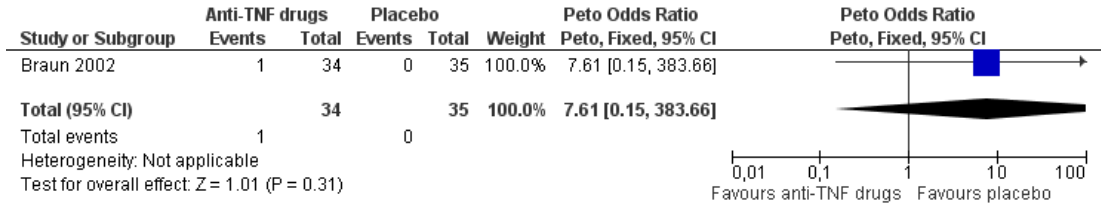
8.5 Infections



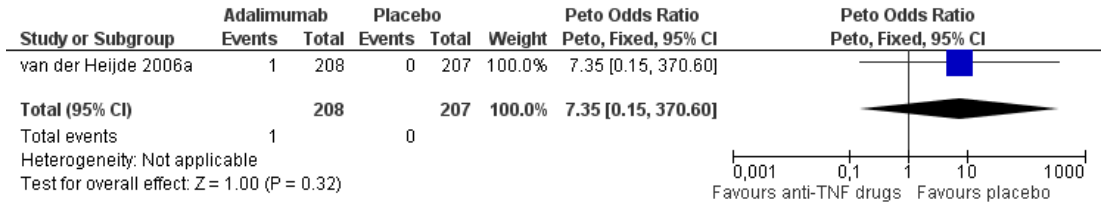
8.6 Serious infections



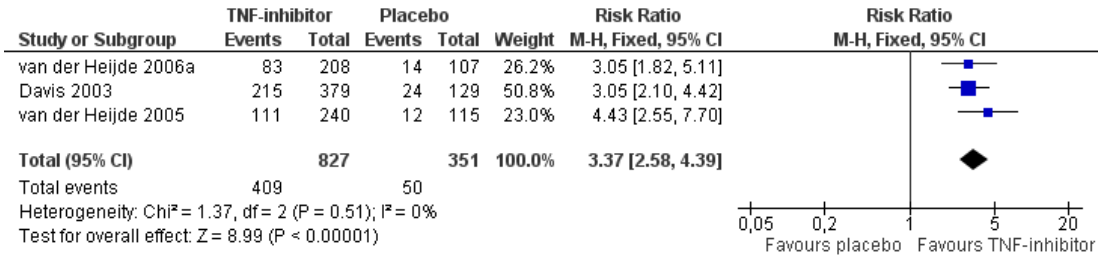
8.7 Tuberculosis



8.10 Hepatotoxicity (clinical symptoms)



8.12 ASAS 40


TNF-inhibitor
Placebo
Risk Ratio
Risk Ratio
Study or Subgroup
Events
Total
Events
Total
Weight
M-H, Fixed, 95% CI
M-H, Fixed, 95% CI
van der Heijde 2006a 83 208 14 107 26.2% 3.05 [1.82, 5.11]
Davis 2003 215 379 24 129 50.8% 3.05 [2.10, 4.42]
van der Heijde 2005 111 240 12 115 23.0% 4.43 [2.55, 7.70]
Total (95% CI) 827 351 100.0% 3.37 [2.58, 4.39]
Total events 409 50
Heterogeneity: Chi² = 1.37, df = 2 (P = 0.51); I² = 0%
Test for overall effect: Z = 8.99 (P < 0.00001)
0,05
0,2
1
5
20
Favours placebo
Favours TNF-inhibitor

CHAPTER FIVE: A SYSTEMATIC ASSESSMENT OF ADVERSE EFFECTS OF TNF-INHIBITORS FOR ANKYLOSING SPONDLYLITIS

The following is a manuscript prepared for publication, based on a systematic review of the literature, focusing on the potential adverse effects of TNF-inhibitors for treating AS. The previous manuscript was a Cochrane systematic review of the efficacy and harms of anti-TNF therapy. This manuscript focuses on this systematic assessment of the adverse effect data and aims to reach a wider clinical audience by being submitted for publication in the Journal of Rheumatology.

This manuscript was co-authored by the student (LJM) and her co-supervisors, Dr. Peter Tugwell and Dr. George Wells, as well as advisor to the thesis, Dr. Annelies Boonen. Dr Zochling screened search results, performed data extraction, and provided comments on drafts of the review. The student is the first author on this manuscript as she had the primary responsibility for data collection, analysis, interpretation, and writing. Drs Tugwell, Wells, and Boonen provided valuable feedback throughout the process

Assessment of adverse effects in the use of TNF-inhibitors (adalimumab, etanercept, and infliximab) for ankylosing spondylitis: A Cochrane systematic review

Lara Maxwell, Jane Zochling, Annelies Boonen, Jasvinder Singh, Peter Tugwell, George Wells

Abstract

Objective: To perform a systematic review of adverse effects of TNF- inhibitors (adalimumab, etanercept, and infliximab) for ankylosing spondylitis (AS).

Methods: We searched MEDLINE, EMBASE, The Cochrane Library, ACP Journal Club, CINAHL, and ISI Web of Knowledge to March 2010 for randomized controlled trials (RCTs) and non-randomized studies (NRS). Two reviewers independently assessed search results, risk of bias, and extracted data. We performed network meta-analysis to obtain indirect estimates given the lack of head-to-head studies.

Results: Fifteen RCTs and 4 NRS with a total of 2,459 participants were included. Up to 24 weeks, anti-TNF treatment compared to placebo was associated with a statistically significant higher rate of infections (relative risk (RR) 1.22, 95% CI 1.02 to 1.45; absolute increase in harm 6% (95% CI 1% to 11%)), total number of adverse events (RR 1.24, 95% CI 1.10 to 1.40; absolute increase 13% (95% CI 6% to 19%)), withdrawals due to adverse events (Peto OR 2.78, 95% CI 1.27 to 6.10; absolute increase 2% (95% CI 1% to 4%)), and injection or infusion site reactions (RR 3.02, 95% CI 2.12 to 4.31; absolute increase 14% (95% CI 10% to 18%)). Serious adverse events (Peto OR, 1.83 (95% CI 0.93 to 3.6); absolute increase 2% (95% CI 0% to 4%) and serious infections (Peto OR 1.89 (95% CI 0.45, 7.96); absolute increase 0% (95% CI -1% to 2%)) were not statistically different between the anti-TNF and placebo groups. Few cases were seen of tuberculosis, lymphoma, demyelinating disease, and congestive heart failure. Data from NRS at 2 and 3 years were at high risk of bias and difficult to interpret, but there was no large signal of key safety outcomes such as serious infections and malignancies.

Conclusion: Though the total number of infections was increased, we did not find evidence of an increase in serious adverse events or serious infections from short-term RCTs,. Data from NRS were at high risk of bias and difficult to interpret, but there was no large signal of key safety outcomes such as serious infections and malignancies over a three-year period. Based on indirect comparison methodology, there did not appear to differences between anti-TNF agents in terms of the key safety outcomes. Biologic registries with appropriate control patients as well as data from post-marketing surveillance are required to assess for longer-term harms.

Key indexing terms: systematic review, safety, ankylosing spondylitis

Word count: 4048
Abstract: 350

Introduction

TNF (tumour necrosis factor)-inhibitors have revolutionized (1) the treatment of ankylosing spondylitis (AS) , a sero-negative inflammatory disease characterized by back pain due to sacroiliitis and spondylitis which leads to ankylosis and functional impairment. Prior to their introduction, the mainstay of treatment was physical therapy (2) and non-steroidal anti-inflammatory drugs (NSAIDs) (3;4). However, at least one- third of patients respond insufficiently to NSAID therapy or experience serious side effects from NSAIDs and thus require disease-controlling drugs in addition to symptom modifying treatment.

TNF-inhibitors work by reducing the amount of TNF in the joints which in turn reduces pain and inflammation. The efficacy of three of the biologics approved for use in treating AS, adalimumab, etanercept, and infliximab, has recently been systematically assessed. They appear to significantly reduce disease activity and improve function, fatigue, and quality of life in studies up to 24 weeks. Reduction in radiographic progression with the use of TNF-inhibitors has not been established (5).

However, the use of these biologics has raised concerns about potential harms, especially in terms of serious infections, including tuberculosis, heart and liver problems, central nervous system disorders, cancer, and autoimmune disorders. Randomized controlled trials (RCTs) in rheumatoid arthritis patients have suggested anti-TNF use increases the risk of serious infections and malignancies (6) as has recent registry data (7;8).

While RCTs may provide the least biased estimate of effect of an intervention, their strict inclusion criteria and commonly short duration make it challenging to use the results of RCTs for the assessment of adverse events. This is especially true for rare or latent effects.

We conducted a systematic review of both randomized and non-randomized evidence to assess the adverse events associated with the use of anti-TNF agents adalimumab, etanercept, and infliximab for the treatment of AS. Given the lack of head-to-head studies, we also performed a network meta-analysis to estimate the risk of adverse events of each of the three anti-TNF agents compared to each other.

Materials and Methods

Inclusion criteria

Types of studies, participants, and interventions: We included randomized controlled trials (RCTs), controlled clinical trials (CCTs), and non-randomized studies consisting of the following designs: cohort studies (prospective (e.g. long-term extension of RCT) or retrospective), case-control studies, case-series, and published registry data on adalimumab, etanercept, or infliximab. The non-randomized studies had to follow patients on anti-TNF agents for at least one year, have a sample size > 100 patients (9;10), and report an included outcome. Patients had to meet the 1961 Rome, 1966 New York, or modified 1984 New York AS classification criteria.

Outcomes: We collected data on the following adverse effect outcome measures: total number of adverse events, serious adverse events, withdrawals, and withdrawals due to adverse events as well as inefficacy. Specific adverse events of interest were: headache, diarrhea, nausea, infections, serious infections, tuberculosis, lymphoma or malignancy, hepatotoxicity, chronic or congestive heart failure, injection- or infusion-site reactions, autoimmunity symptoms and disorders, demyelinating disease, and mortality (11).

Search methods for identification of studies and additional data: A Cochrane-trained information specialist developed and ran the search strategies in the following electronic databases to January 2009 for the RCT search and May 2009 for a specific adverse effects search: MEDLINE, EMBASE, The Cochrane Library, ACP Journal Club, CINAHL, ISI Web of Knowledge. An updated search from January 2008 to

March 2010 was run in Ovid MEDLINE, Pre-MEDLINE and EMBASE. We also searched specifically for full-text articles for any abstracts that were included from the earlier search.

Two search strategies were developed; one for RCTs and one to specifically search for adverse effects (Appendix 1). The references of retrieved articles were reviewed. Conference proceedings from the American College of Rheumatology and European Congress of Rheumatology meetings were hand searched (by a single author, LM) from 2005 to 2009. We contacted the authors of included studies for clarifications. There was no language, year or type of publication restriction. We also searched the websites of the regulatory agencies (US Food and Drug Administration, European Medicines Evaluation Agency, Australian Adverse Drug Reactions Bulletin, and UK pharmacovigilance and drug safety updates) using the terms "ankylosing spondylitis," "remicade," "infliximab", "enbrel", "etanercept", "humira", "adalimumab" on April 1, 2010.

Data collection and analysis. Analyses were determined a priori and published as a Cochrane Library protocol. The more detailed adverse event search was pre-specified in a MSc thesis proposal. Search results and data extraction was conducted independently by two people (LM, JZ, or BD), disagreement was resolved by consensus, and a structured extraction form was used. For missing data, we contacted trial authors. Data was entered into RevMan 5 for analysis.

Assessment of risk of bias. For RCTs, we used the recommended domains in the Cochrane Handbook of: randomization sequence generation, allocation sequence concealment, blinding of participants, personnel and outcome assessors, incomplete outcome data (dropout rates and reasons for withdrawal, appropriate imputation of missing data, an overall completion rate $\geq 80\%$), and selective outcome reporting. We added two other criteria, specific to the assessment of adverse effects: ascertainment of outcome - did the researchers actively monitor for AEs (low risk of bias) or did they simply provide spontaneous reporting of AEs that arise (high risk of

bias) and definition of adverse outcomes - were definitions provided for 'serious adverse event'? Each criterion was explicitly judged as follows: Yes = low risk of bias; No = high risk of bias; Unclear = either lack of information or uncertainty about potential for bias.

For the non-RCTs, we adapted the Newcastle-Ottawa Scale (NOS) (12) for single-group cohort studies. We decided that the NOS criteria assessed the important potential risk of bias criteria concerning the included studies with the exception of two criteria for single-group cohorts which were not applicable to a single-group cohort; namely 'selection of the non-exposed cohort' and 'comparability of cohorts on the basis of the design or analysis'. We did not judge these criteria in our assessment.

Statistical analysis: Results from the three biologics were analyzed separately. One further analysis pooled adverse effects data from all three anti-TNF agents to assess for a class-effect. The results of the studies were analyzed using RevMan 5.0. Data was summarized using meta-analysis when it was sufficiently homogeneous, both clinically and statistically. Continuous data was expressed as weighted mean difference (WMD) or standardized mean difference (SMD) (depending on similarity of scales). Dichotomous data was expressed as relative risk (RR), unless the data consisted of rare events (<10%) and then the Peto OR was used. The Peto OR is not recommended when there is a severe imbalance in the number of people in the treatment groups (13), and we used the rule of thumb that this is a concern when there are at least four times as many participants in one group as another as suggested in Systematic Reviews in Health Care (p.295,(14)). We made a post-hoc decision to check the robustness of estimates obtained by Peto OR by also using the Mantel-Haenzsal OR with the standard continuity correction of 0.5.
Heterogeneity of the data was tested for using the I-squared statistic (15). A value greater than 50% may be considered substantial heterogeneity. In the absence of significant heterogeneity, a fixed effects model was used, otherwise a random effects model was used for analysis. Where available, the analyses were based on intention-to-treat data from the individual studies. We planned to undertake visual inspection

of funnel plots to assess publication bias; however, the few studies combined for individual outcomes made this difficult.

A priori defined sub-group analyses were: 1. Intervention: different dosage, trial duration 2. characteristics of participants: different AS classification criteria, severity of baseline disease (based on BASDAI , BASFI), age, disease duration, sex. A sensitivity analysis was pre-specified to assess the effect of study quality (proper generation of randomization sequence, adequate allocation concealment and blinding) on the overall estimates of effect.

Given the paucity of head-to-head studies, we undertook an indirect comparison analysis of the randomized controlled trials using three different methods of indirect comparisons to obtain summary estimates for comparisons of each of the three anti-TNF agents versus each other, and to compare and contrast the results from these different methods: i. the Bucher Adjusted Indirect Treatment Comparison (ITC) approach (16) using the ITC software designed for the Wells 2009 report (17); ii. the Generalized Linear Mixed Models (GLIMMIX) approach; iii. the Mixed Treatment Comparisons (MTC) approach (also referred to as "network meta-analysis"). We performed indirect comparisons on the following outcomes: withdrawals due to AE, all infections, and serious infections. The outcomes of withdrawals due to AE and serious infections had rare events, so in the MTC analysis, a random effects model was used and the zero events were adjusted for by adding 0.5. Assumptions assessed were: i. homogeneity; ii. trial similarity; iii. consistency of evidence. Inconsistency was assessed using the back-calculation method (18).

Results

Figures 1 and 2 show the results of the searches. For the RCT search, fifteen studies met our inclusion criteria with a total of 2,459 participants. 303 people received infliximab (19-21;28) in combination with methotrexate (22); 995 received etanercept (23-30) and 246 received adalimumab (31;32). One study was an open-label head to head study of etanercept (N=25) and infliximab (N=25) (33).

For the non-RCT search, only four extension studies from included RCTs met our inclusion criteria, with six publications from those four studies. For etanercept, there were two published open-label extension (OLE) studies of the Davis 2003 trial, with 96-and 192-week follow-up. For infliximab, two-year results from the follow-up of the van der Heijde 2005 RCT on 277 people were published in a full-text article (34). Patients remained blinded during the extension phase of this trial. One abstract was found on the two-year follow up of 89 people in the EASIC cohort (35), which was an extension of the ASSERT trial. For adalimumab, two and three-year OLE data from the van der Heijde 2006a trial on 311 people were reported in a full-text article (36) and an abstract (37). Appendix 2 provides the RCT and non-RCT study characteristics.

The results of the search for warnings on pharmacovigilance websites is summarized in Table 1**.**

The results of the risk of bias assessment of RCTs and non-RCTs are summarized in Appendix 3. Overall, most of the adverse event outcome data from the RCTs were judged to be at moderate risk of bias and were downgraded due to imprecise estimates given the low event rates and that the two criteria specific to bias assessment of adverse events were mostly judged as 'unclear'. The adverse event outcomes from the OLE studies were judged to be at a high risk of bias. There is significant selection bias since participants entering the RCT phase of the study were already a highly select group of participants, with various co-morbidities excluded. Those who then continued in the extension phase would have been participants who did not have a significant adverse event during the RCT-phase. Therefore, the participants exposed to anti-TNF agents for a follow-up period are not representative of those in the community who may be taking this treatment. Ascertainment of exposure to the anti-TNF therapy during the extension study was not provided. Most studies had less than 80% follow-up.

Due to lack of studies, publication bias could only be assessed for injection/infusion reactions in the pooled anti-TNF agents against placebo estimate. The funnel plot of

six studies had an empty spot in the bottom left of the funnel plot, indicating that there may be bias against small studies with small estimates of harm being published. However, this interpretation must be viewed with caution given the small number of studies included in the plot.

Pooled RCT results from all three anti-TNF agents versus placebo

We pooled the results from the RCTs of etanercept, infliximab, and adalimumab to assess for a class-effect of short-term (up to 24 weeks) adverse effects of TNF-inhibitors. Table 2 provides a summary of these results.

There was a statistically significantly greater total number of adverse events and withdrawals due to adverse events in the anti-TNF versus placebo group (RR 1.24, 95% CI 1.10 to 1.40, absolute increase 13%, 95% CI 6% to 19%) and Peto OR 2.78, 95% CI 1.27 to 6.10, absolute increase 2%, 95% CI 1% to 4%) respectively). The risk of injection or infusion site reactions was significantly greater in the anti-TNF group (RR 3.02, 95% CI 2.12 to 4.31, absolute increase 14%, 95% CI 10% to 18%). The risk of any infection was also statistically increased in the anti-TNF group (RR 1.22, 95% CI 1.02 to 1.45, absolute increase 6% (95% CI 1% to 11%).

Serious adverse events were not statistically different between the anti-TNF and placebo groups and neither were serious infections (Peto OR 1.89, 95% CI 0.45 to 7.96, absolute increase 0%, 95% CI -1% to 2%). There was one case of tuberculosis reported in (19) in the infliximab-treated group compared with none in the placebo group (Peto OR 7.61, 95% CI 0.15, 383.66, absolute increase 3%, 95% CI -5% to 11%). No cases of tuberculosis were reported in four other RCTs (21;27;32;33).

Four trials reported the absence of malignancies or lymphoma (21;30;32;33) these were not reported as an outcome in the other trials. Similarly, three studies (30;32;33) stated they looked for demyelinating or autoimmune disease or congestive heart failure, but no events occurred. Five studies reported on liver

function tests, but only one person required hospitalization for hepatotoxicity (on adalimumab treatment (32).

Adverse effects reported in RCTs, by biologic: We report here on those adverse effects which were found to be statistically significantly increased in the biologic group compared to the placebo group.

Adalimumab

There was a significant increase between adalimumab and placebo in terms of total adverse events (RR 1.25, 95% CI 1.05 to 1.49, absolute increase 15%, 95% CI 4% to 26%) and any infection (RR 1.58, 95% CI 1.10 to 2.27, absolute increase 12%, 95% CI 3% to 21%).

Etanercept

There were significantly more withdrawals due to adverse events (Peto OR 4.03, 95% CI 1.43 to 11.30, absolute increase 3%, 95% CI 1% to 6%) in the treatment group compared to placebo. As well, the treatment group had statistically significantly more injection site reactions than placebo in both the 25mg twice weekly and 50 mg once weekly groups (RR 3.60, 95% CI 2.41 to 5.38, absolute increase 19%, 95% CI 14% to 24% and RR 5.41, 95% CI 2.40 to 12.19, absolute increase 18%, 95% CI 12% to 23%, respectively).

Infliximab

At 12 weeks, there was a significant increase in serious adverse events (Peto OR 7.80, 95% CI 1.07 to 56.65), but when 12 and 24 week data was pooled, the result was no longer significant (Peto OR 2. 47, 95 % CI 0.75 to 8.14, absolute increase 3%, 95% -1% to 6%).

Adverse effects reported in one head-to-head RCT:

The incidence of adverse events in terms of serious infections, headaches and diarrhea were similar among etanercept and infliximab groups (33). There were no withdrawals due to adverse events in either group. As well, there were "no cases of cases of opportunistic infections, tuberculosis, congestive heart disease, demyelinating disorders, lupus-like syndromes, and malignancy."

Adverse effects reported in studies with comparators different from placebo:

The effect of etanercept (50mg weekly) versus sulphasalazine (3g) daily (25) and infliximab + methotrexate (MTX) versus placebo + MTX (22) were assessed in two RCTs. In the Braun 2008 study, there was no difference in the risk of serious adverse events between etanercept and placebo (RR 1.00 (95% CI 0.98 to 1.03). This was the only type of adverse event reported in the abstract.

There was no significant difference between the infliximab + MTX and placebo+ MTX groups in terms of the relative risk of infections (RR 1.50, 95% CI 0.35, 6.50, absolute increase 7%, 95% CI -17% to 31%). Two mild infusion reactions occurred in the combination group and none in the MTX group. No severe adverse events occurred in either group.

Adverse effects from open-label extensions

The results from the open-label extension studies were reported in different manners; in adalimumab studies, adverse effects data were reported separately by those who had received weekly exposure to adalimumab and those who had received at least one dose of adalimumab. For infliximab, separate results of the extension study were reported for the original placebo group in the RCT, the treatment group in the RCT, and combined. The etanercept extension study reported those who had received at least one dose of etanercept.

Table 3 summarizes the results for the key harm outcomes across the NRS studies. While bearing in mind that these studies are judged to be at a high risk of bias, there is no large signal of an increase in risk over the 2 to 3 year follow up period.

Indirect comparisons

Using RCT data, we performed indirect comparisons on withdrawals due to AE, all infections, and serious infections, and found that against placebo, results for all outcomes were not statistically significant and almost all favoured the placebo group (Appendix 4). For the head-to-head analysis (Table 4), there were no statistically significant results and no consistency with respect as to which anti-TNF drug was favoured. Necessary assumptions were met and results were robust across three different methodologies.

Subgroup/sensitivity analyses

The studies were quite similar in terms of patient characteristics and interventions so we did not undertake subgroup analyses. Sensitivity analyses of assessing only those studies at high risk of bias were not significantly different from the pooled estimates of the adverse event outcomes. Rare event data was robust to different methods of estimating the effect.

Discussion

Although fairly broad in scope, we focused on adverse effects which had previously been raised as concerns in post-marketing surveillance of these three TNF-inhibitors. Fifteen RCTs with a total of 2,459 participants and four extension studies from those RCTs were included in this systematic review.

When data for all three anti-TNF agents was pooled together, the total number of adverse events, withdrawals due to adverse events, injection or infusion site reactions, and the risk of any infection outcomes over a 24-week period were statistically increased in the anti-TNF group. We did not find statistically significant increases for serious adverse events or for specific adverse events of interest such as serious infections, tuberculosis, or lymphoma, but the data on these outcomes in the RCTs was quite sparse, leading to wide confidence intervals. Results from these RCTs were judged to be at moderate risk of bias, mainly due to the imprecision of the estimates or concerns of selective outcome reporting of the adverse event. There is evidence that adverse events are selectively reported in RCTs (38-40) and the majority of studies in this review were judged to be at a high or unclear risk of bias for this domain. As well, the two criteria specific to AE assessment, method of monitoring for AE and definitions of SAE, were mainly unclear. It is difficult to assess these criteria without obtaining all the AE data from the full clinical trial reports.

Another systematic review and meta-analysis was recently performed which focused on assessing the risk of serious infections (41) in AS patients taking anti-TNF agents

and compared them to estimates of risk in groups of patients taking placebo and NSAIDs. All the included studies in this review were also included in the Fouque-Aubert review, though our systematic review, which was based on Cochrane methodology and looked at a broader range of adverse events, included non-RCTs and also conducted network meta-analysis to provide head-to-head estimates. Based on RCT data, the two reviews reached the same conclusion in that there does not appear to be a statistically significant increase in serious infections in AS patients exposed to anti-TNF agents. This is contrary to the serious infection risk that was evident in a systematic review of anti-TNF use in rheumatoid arthritis patients (6). The Fouque-Aubert review authors hypothesized that, compared to RA patients, AS patients may have fewer co-morbidities due to their younger average age and the differences in pathophysiology of the two diseases and concomitant treatments (NSAIDs in AS patients versus steroids with RA patients) may make RA patients more susceptible to infections (41).

A fair amount of literature has been written on the challenges of assessing adverse effects and the limitations associated with using RCTs to assess rare or delayed adverse events have been well established (42-45). Reporting of adverse effects is often poorly indexed, making the results difficult to find. We conducted an extensive electronic literature search, supplemented with handsearching of abstracts from two key conferences and review of major regulatory agency websites but it is possible that we missed some data, particularly unpublished data from pharmaceutical companies.

The RCTs included in this review focused on short- term results and were likely underpowered to find a significant difference between groups, especially for delayed events such as malignancies. Yazici et al highlights other important limitations in using RCTs for adverse effects assessment, including lack of reporting of timing of events; using patient-years inappropriately when events are not randomly distributed over time; and inadequate sample size to detect harms (46).

Given the limitations associated with using RCTs for AE assessment, we broadened our inclusion criteria to include NRS but restricted our NRS inclusion criteria to those studies with more than 100 people with follow-up for at least one year, as recommended by an EMEA 2009 guidance document on the treatment of AS (10). However, the choice of '100' is arbitrary, and a sample size calculation in the Fouque-Aubert review noted that, based on the risk difference estimate found in their review, a study with at least 5500 participants would be needed to find a difference in risk of serious infections between groups.

Applying these strict inclusion criteria to our NRS search meant that we only included four extension studies from RCTs from the results of the adverse events search. A major problem with RCT-extension studies is that the participants who enter them are highly selected in that they originally met strict RCT inclusion criteria and then did not experience adverse events that were serious enough to make them withdraw early from the trial before the extension phase began. Therefore, the results from these OLE studies are not generalizable to the population who are eligible for anti-TNF therapy. As well, since there is no control group, it is difficult to assess whether the results of important adverse events of interest such as serious infections, malignancies or death are much increased compared to what happens in a similar population. The lack of blinding places these studies at a higher risk of bias, as does the withdrawal rates and lack of ascertainment of exposure to the drug during the extension phase. In addition, the small numbers of participants entering an OLE mean they are underpowered to detect rare AE (47-50).

We did not find evidence of an increase in serious adverse events or serious infections from RCTs, though total number of infections was increased, and other adverse events had low event rates. Data from NRS were at high risk of bias and difficult to interpret, but there was no large signal of key safety outcomes such as serious infections and malignancies. Based on indirect comparison methodology, there did not appear to differences between anti-TNF agents in terms of the key safety outcomes.

It is important to use the right study design to answer the adverse event of interest (51). Studies based on large numbers of participants in biologic registry databases, such as the French registry RATIO (52) and the UK British Society for Rheumatology Biologics Register (53), with a well-matched control group, will be useful in clarifying the evidence base regarding rare or delayed adverse effects of anti-TNF agents.

References

(1) Braun J, Baraliakos X. Treatment of ankylosing spondylitis and other spondyloarthritides. Curr Opin Rheumatol 2009 Jul;21(4):324-34.

(2) Dagfinrud H, Kvien TK, Hagen KB. Physiotherapy interventions for ankylosing spondylitis. Cochrane Database Syst Rev 2004;(4):CD002822.

(3) Wanders A, Heijde D, Landewe R, Behier JM, Calin A, Olivieri I, et al. Nonsteroidal antiinflammatory drugs reduce radiographic progression in patients with ankylosing spondylitis: a randomized clinical trial. Arthritis Rheum 2005 Jun;52(6):1756-65.

(4) Song I.H., Poddubnyy D.A, Rudwaleit M, Sieper J. Benefits and risks of ankylosing spondylitis treatment with nonsteroidal antiinflammatory drugs. Arthritis Rheum 2008;58(4):929-38.

(5) Schett G, Landewe R, Van Der Heijde D. Tumour necrosis factor blockers and structural remodelling in ankylosing spondylitis: what is reality and what is fiction? Ann Rheum Dis 2007 Jun;66(6):709-11.

(6) Bongartz T, Sutton AJ, Sweeting MJ, Buchan I, Matteson EL, Montori V. Anti-TNF antibody therapy in rheumatoid arthritis and the risk of serious infections and malignancies: systematic review and meta-analysis of rare harmful effects in randomized controlled trials. JAMA 2006 May 17;295(19):2275-85.

(7) Galloway JB, Hyrich KL, Mercer LK, Dixon WG, Fu B, Ustianowski AP, et al. Anti-TNF therapy is associated with an increased risk of serious infections in patients with rheumatoid arthritis especially in the first 6 months of treatment: updated results from the British Society for Rheumatology Biologics Register with special emphasis on risks in the elderly. Rheumatology (Oxford) 2011 Jan;50(1):124-31.

(8) Mariette X, Tubach F, Bagheri H, Bardet M, Berthelot JM, Gaudin P, et al. Lymphoma in patients treated with anti-TNF: results of the 3-year prospective French RATIO registry. Ann Rheum Dis 2010 Feb;69(2):400-8.

(9) Gartlehner G, Hansen RA, Jonas BL, Thieda P, Lohr KN. The comparative efficacy and safety of biologics for the treatment of rheumatoid arthritis: a systematic review and metaanalysis. J Rheumatol 2006 Dec;33(12):2398-408.

(10) European Medicines Agency. Guideline on clinical investigation of medicinal products for the treatment of ankylosing spondylitis. European Medicines Agency; 2005.

(11) Lee S.J. KA. Biologic Agents in Rheumatology: Safety Considerations. Rheumatic Disease Clinics of North America Suppl 1, 3-10. 2005.

(12) Wells G, Shea B, O'Donnell D, Peterson J, Welch V, Losos J, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp

(13) Bradburn MJ, Deeks JJ, Berlin JA, Russell LA. Much ado about nothing: a comparison of the performance of meta-analytical methods with rare events. Stat Med 2007 Jan 15;26(1):53-77.

(14) Systematic Reviews in Health Care. BMJ Books; 2001.

(15) Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003 Sep 6;327(7414):557-60.

(16) Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. J Clin Epidemiol 1997 Jun;50(6):683-91.

(17) Wells G, Sultan S, Chen L, Khan M, Coyle D. Indirect evidence: indirect treatment comparisons in meta-analysis. Ottawa: Canadian Agency for Drugs and Technologies in Health; 2009.

(18) Dias S, Welton NJ, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. Stat Med 2010 Mar 30;29(7-8):932-44.

(19) Braun J, Brandt J, Listing J, Zink A, Alten R, Golder W, et al. Treatment of active ankylosing spondylitis with infliximab: a randomised controlled multicentre trial. Lancet 2002 Apr 6;359(9313):1187-93.

(20) Inman R, Maksymowych W, for the CANDLE group. A double-blind, placebo-controlled trial of low dose infliximab in ankylosing spondylitis. J Rheumatol 2010;37(6):1-8.

(21) Van Der Heijde D, Dijkmans B, Geusens P, Sieper J, DeWoody K, Williamson P, et al. Efficacy and safety of infliximab in patients with ankylosing spondylitis: results of a randomized, placebo-controlled trial (ASSERT). Arthritis Rheum 2005 Feb;52(2):582-91.

(22) Marzo-Ortega H, McGonagle D, Jarrett S, Haugeberg G, Hensor E, O'connor P, et al. Infliximab in combination with methotrexate in active ankylosing spondylitis: a clinical and imaging study. Ann Rheum Dis 2005 Nov;64(11):1568-75.

(23) Barkham N, Coates L, Keen H, Hensor E, Fraser A, Redmond A, et al. A double blind placebo controlled trial of etanercept in the prevention of work disability in ankylosing spondylitis. Rheumatology (Oxford) 47, ii72. 2008.

(24) Brandt J, Khariouzov A, Listing J, Haibel H, Sorensen H, Grassnickel L, et al. Six-month results of a double-blind, placebo-controlled trial of etanercept treatment in patients with active ankylosing spondylitis. Arthritis Rheum 2003 Jun;48(6):1667-75.

(25) Braun J, Burgos-Vargas R, van der Horst-Bruinsma I.E, Freundlich B, Vlahos B, Koenig A.S., et al. Assessment of clinical efficacy in a randomized, double-blind study of etanercept and sulphasalazine in patients with ankylosing spondylitis. Arthritis Rheum 2008;58:S415.

(26) Calin A, Dijkmans BA, Emery P, Hakala M, Kalden J, Leirisalo-Repo M, et al. Outcomes of a multicentre randomised clinical trial of etanercept to treat ankylosing spondylitis. Ann Rheum Dis 2004 Dec;63(12):1594-600.

(27) Davis JC, Jr., Van Der Heijde D, Braun J, Dougados M, Cush J, Clegg DO, et al. Recombinant human tumor necrosis factor receptor (etanercept) for treating ankylosing spondylitis: a randomized, controlled trial. Arthritis Rheum 2003 Nov;48(11):3230-6.

(28) Gorman JD, Sack KE, Davis JC, Jr. Treatment of ankylosing spondylitis by inhibition of tumor necrosis factor alpha. N Engl J Med 2002 May 2;346(18):1349-56.

(29) Huang F, MacPeek D, Shunle C, Li Z, Gu J, Wu D, et al. Etanercept is effective and well tolerated in Chinese patients with ankylosing spondylitis: 6-week placebo-

controlled trial followed by 6-week open label treatment. Ann.Rheum.Dis. 67[Suppl II], 521. 2008.

(30) Van Der Heijde D, Da Silva JC, Dougados M, Geher P, van dH-B, I, Juanola X, et al. Etanercept 50 mg once weekly is as effective as 25 mg twice weekly in patients with ankylosing spondylitis. Ann Rheum Dis 2006 Dec;65(12):1572-7.

(31) Lambert RG, Salonen D, Rahman P, Inman RD, Wong RL, Einstein SG, et al. Adalimumab significantly reduces both spinal and sacroiliac joint inflammation in patients with ankylosing spondylitis: a multicenter, randomized, double-blind, placebo-controlled study. Arthritis Rheum 2007 Dec;56(12):4005-14.

(32) Van Der Heijde D, Kivitz A, Schiff MH, Sieper J, Dijkmans BA, Braun J, et al. Efficacy and safety of adalimumab in patients with ankylosing spondylitis: results of a multicenter, randomized, double-blind, placebo-controlled trial. Arthritis Rheum 2006 Jul;54(7):2136-46.

(33) Giardina AR, Ferrante A, Ciccia F, Impastato R, Miceli MC, Principato A, et al. A 2-year comparative open label randomized study of efficacy and safety of etanercept and infliximab in patients with ankylosing spondylitis. Rheumatol Int 2009;30(11):1437-40.

(34) Braun J, Deodhar A, Dijkmans B, Geusens P, Sieper J, Williamson P, et al. Efficacy and safety of infliximab in patients with ankylosing spondylitis over a two-year period. Arthritis Rheum 2008 Sep 15;59(9):1270-8.

(35) Heldmann F, Brandt J, van der Horst-Bruinsma I.E, Landewe R, Sieper J, Burmester G, et al. European Ankylosing Spondylitis Infliximab Cohort (EASIC): Safety of longterm therapy in patients with ankylosing spondylitis. Ann.Rheum.Dis. 68[Suppl 3], 626. 2009.

(36) van Der Heijde D, Schiff MH, Sieper J, Kivitz AJ, Wong RL, Kupper H, et al. Adalimumab effectiveness for the treatment of ankylosing spondylitis is maintained for up to 2 years: long-term results from the ATLAS trial. Ann Rheum Dis 2009 Jun;68(6):922-9.

(37) van der Heijde DM, Dijkmans B, Schiff MH, Kivitz A, de Vlam K, Wordsworth P. Maintenance of reduction in disease activity and partial remission in patients with ankylosing spondylitis (AS) - 3 year results from the adalimumab

(Humira) trial evaluating long-term efficacy and safety in AS (ATLAS). Arthritis Rheum. S579. 2008.

(38) Ioannidis JP, Lau J. Completeness of safety reporting in randomized trials: an evaluation of 7 medical areas. JAMA 2001;285:437-43.

(39) Loke YK, Derry S. Reporting of adverse drug reactions in randomised controlled trials: a systematic survey. BMC Clin Pharmacol 2001;1(3).

(40) Hazell L, Shakir SA. Under-reporting of adverse drug reactions: a systematic review. Drug Saf 2006;29:385-96.

(41) Fouque-Aubert A, Jette-Paulin L, Combescure C, Basch A, Tebib J, Gossec L. Serious infections in patients with ankylosing spondylitis with and without TNF blockers: a systematic review and meta-analysis of randomised placebo-controlled trials. Ann Rheum Dis 2010 Oct;69(10):1756-61.

(42) Seminara N.M. Assessing long-term drug safety: Lessons (Re) learned from Raptiva. Seminars in Cutaneous Medicine and Surgery 2010;29:16-9.

(43) Kaufman D.W, Shapiro S. Epidemiological assessment of drug-induced disease. Lancet 2000;356:1339-43.

(44) Chou R, Helfand M. Challenges in systematic reviews that assess treatment harms. Ann Intern Med 2005;142:1090-9.

(45) Cochrane Handbook for Systematic Reviews of Interventions version 5.0.0. 2008.

(46) Yazici Y, Adler NM, Yazici H. Most tumour necrosis factor inhibitor trials in rheumatology are undeservedly called 'efficacy and safety' trials: a survey of power considerations. Rheumatology (Oxford) 2008 Jul;47(7):1054-7.

(47) Day RO, Williams KM. Open-label extension studies: do they provide meaningful information on the safety of new drugs? Drug Saf 2007;30(2):93-105.

(48) Hemming K, Hutton J, Maguire M, Marson A. Open label extension studies and patient selection biases. Journal of evaluation in clinical science 2008;14:141-4.

(49) Taylor GJ, Wainwright P. Open label extension studies: research or marketing? BMJ 2005 Sep 10;331(7516):572-4.

(50) Taylor WJ, Weatherall M. What are open-label extension studies for? J Rheumatol 2006;33(4):642-3.

(51) Vandenbroucke JP, Psaty BM. Benefits and risks of drug treatments: how to combine the best evidence on benefits with the best data about adverse effects. JAMA 2008 Nov 26;300(20):2417-9.

(52) Tubach F, Salmon-Ceron D, Ravaud P, Mariette X, for the RATIO study group. The RATIO observatory: French registry of opportunistic infections,severe bacterial infections, and lymphomas complicating anti-TnF alpha therapy. Joint Bone Spine 2005;72:456-60.

(53) Lord PA, Farragher TM, Lunt M, Watson KD, Symmons DP, Hyrich KL. Predictors of response to anti-TNF therapy in ankylosing spondylitis: results from the British Society for Rheumatology Biologics Register. Rheumatology (Oxford) 2010 Mar;49(3):563-70.

Tables and Figures

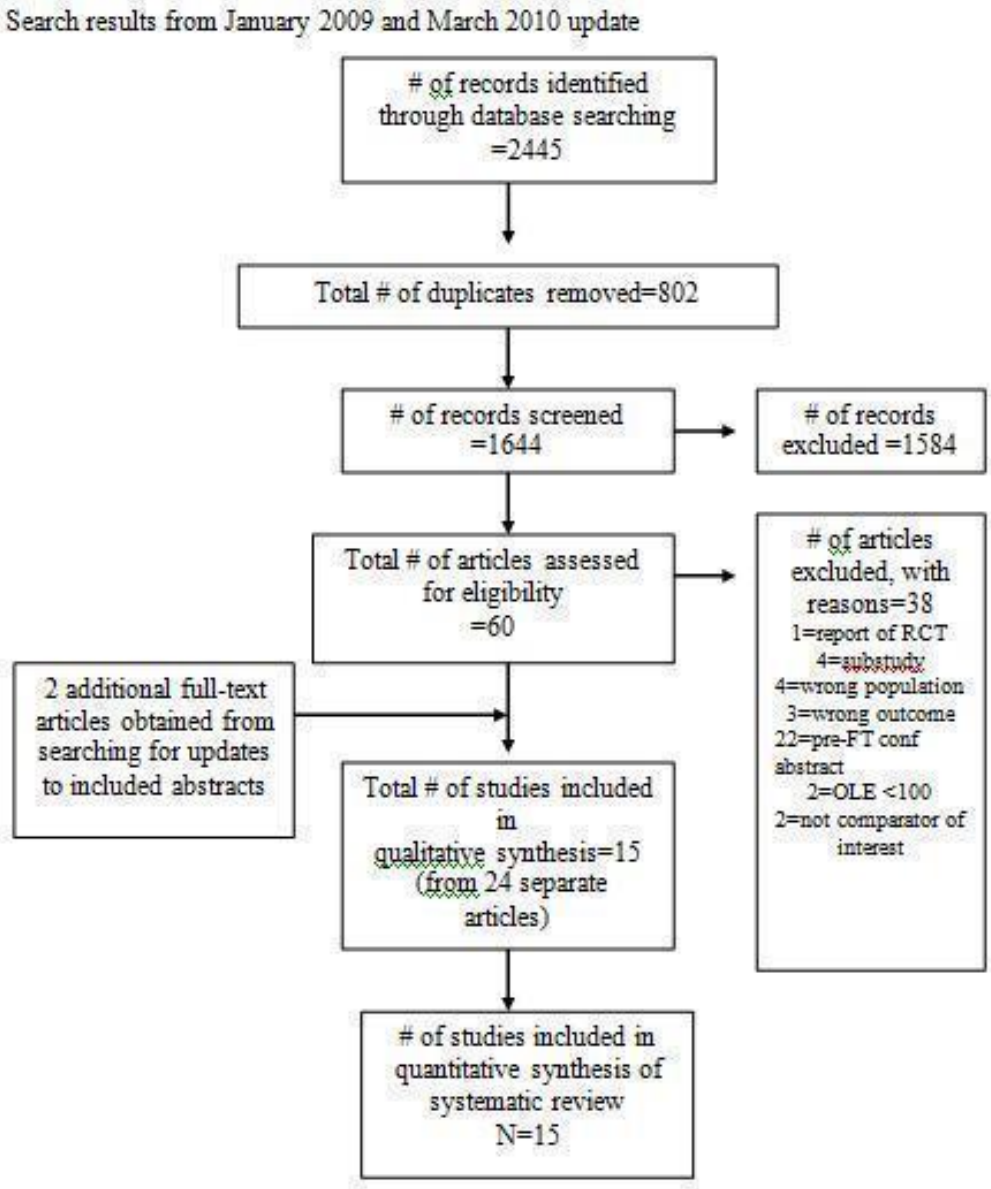


Figure 1: Flowchart of search results for randomized controlled trials

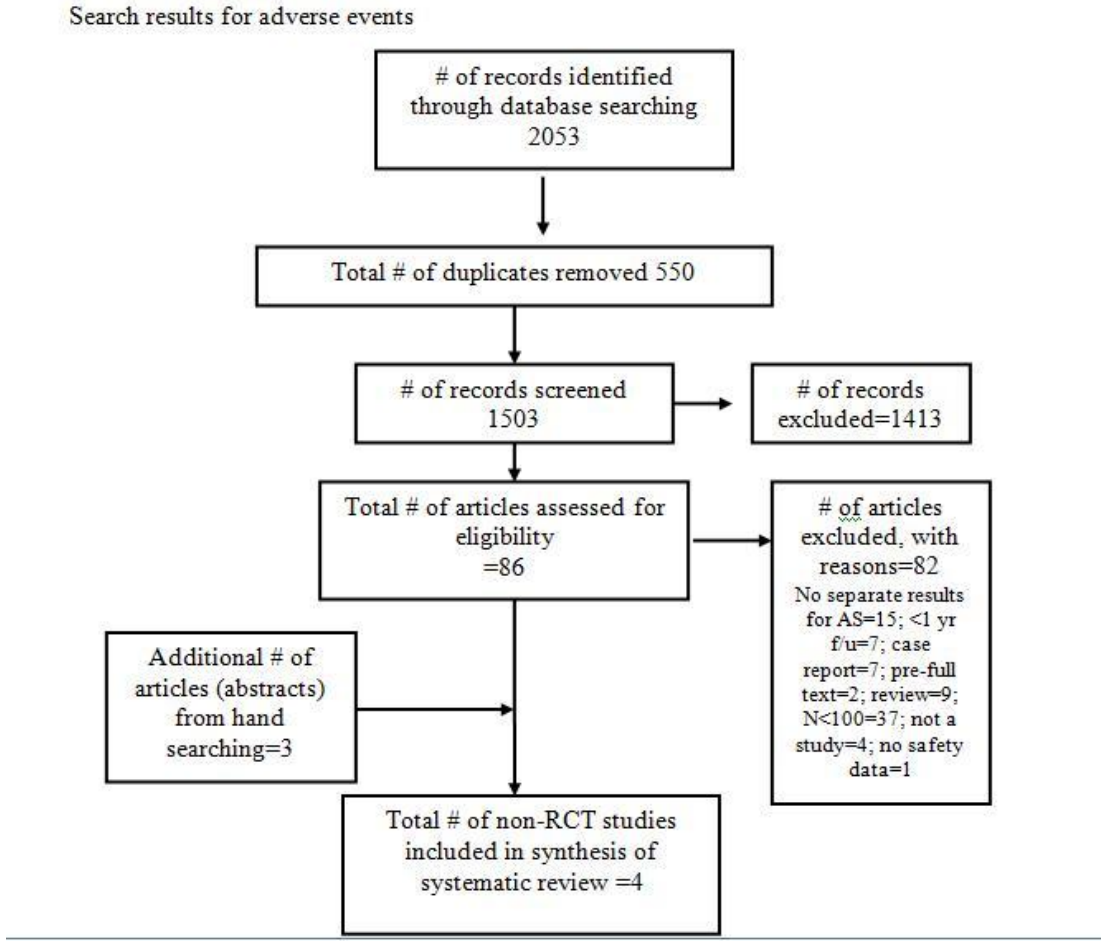


Figure 2: Flowchart of search results for non-randomized controlled trials

Table 1: Summary of warnings on the TNF-inhibitors from regulatory agency websites (search date: April 2010)

Regulatory agency/source	# relevant/# hits	Summary of Warnings and Conclusions
MedWatch: The FDA Safety Information and Adverse Event Reporting Program	11/16	- Recent warning of increased risk of lymphoma and other cancers associated with the use of TNF blockers in children and adolescents; as well specific warning of leukemia (malignancies already exists). - Label warnings added since 2000 for infliximab: hepatotoxicity; infections (pneumonia specifically added), lymphoma, tuberculosis, and other serious opportunistic infections including histoplasmosis, listeriosis, and pneumocystosis, malignancies. - Label warnings added since 2000 for etanercept: serious infections leading to hospitalization or death, including bacterial sepsis and tuberculosis; recommendation to screen for latent tuberculosis infection before beginning Enbrel; lymphoma and other malignancies, including acute and chronic leukemia - Label warnings added since 2000 for adalimumab: lymphoma and other malignancies; skin reactions: new or worsening psoriasis (all sub-types including pustular and palmoplantar) ; serious infections with the combined use of Humira (adalimumab) and anakinra, hypersensitivity reactions, including anaphylaxis, and hematologic events, including pancytopenia and aplastic anemia.
European Medicines Agency (EMEA)	5/73	Since 2000: - Public statements on Infliximab (Remicade) on the increased incidence of mortality and hospitalisation for worsening Congestive Heart Failure and increased tuberculosis Infections - Humira should not be used in people

		who may be hypersensitive (allergic) to adalimumab or any of the other ingredients. Humira must not be used in patients with tuberculosis, other severe infections, or moderate to severe heart failure (an inability of the heart to pump enough blood around the body). -Summary for public: Enbrel: risk of serious infections and should not be used in people who may be hypersensitive (allergic) to etanercept or any of the other ingredients. Enbrel must not be used in patients who have or are at risk of sepsis (when bacteria and toxins circulate in the blood and start to damage the organs), or in patients with infections. Before using Enbrel, doctors must check that the patient is free of infections including tuberculosis.
UK MHRA – Medicines and Healthcare products Regulatory Agency: Drug Safety Updates (formerly Current Problems in Pharmacovigilance)	4/292	- 2001: states infliximab patients should be screened for TB - July 2008: letter to health care providers re: reports of hepatosplenic T-cell lymphoma in patients treated with HUMIRA® (adalimumab) - congestive cardiac failure, cardiomyopathy, the frequency of blood dyscrasias, demyelination, infections, adult respiratory distress syndrome and TB should be kept under close monitoring by the MA (marketing authorization) Holder
Australian Adverse Drug Reactions Bulletin	2/72	- Drug-induced lupus erythematosus (June 2009): An emerging association with TNF inhibitors - TNA-alpha inhibitors (Dec. 2006): While extremely effective, TNFA inhibitors are associated with several serious reactions. These include: • Hypersensitivity reactions - immediately post-injection or delayed; • Serious and life-threatening infection and sepsis; • Recrudescence of tuberculosis and other granulomatous diseases;

		• Reactivation of hepatitis B; • Malignancy, including lymphoma; • Haematological reactions such as pancytopenia and aplastic anaemia; • Autoimmunity - eg, drug-induced lupus; • CNS reactions, including demyelinating disorders and seizures; • New-onset heart failure or worsening of advanced heart failure

Table 2: Summary of adverse event data from the RCTs included in this review

TNF-inhibitors versus placebo for ankylosing spondylitis

Patient or population: patients with ankylosing spondylitis
Settings: international hospital and clinic settings
Intervention: TNF-inhibitors (pooled results for adalimumab, etanercept, infliximab) versus placebo

Outcomes	Illustrative comparative risks* (95% CI)		Relative effect (95% CI)	No of Participants (studies)	Quality of the evidence (GRADE)	Comments
	Assumed risk **Control**	Corresponding risk **TNF-inhibitors versus placebo**				
Withdrawals due to AE Follow-up: 6-24 weeks	**8 per 1000**	**22 per 1000** (10 to 47)	**OR 2.78** (1.27 to 6.1)	1448 (8 studies[1])	⊕⊕⊕⊖ **moderate**	Absolute increase in harm: 2% (1% to 4%)
Total infections Follow-up: 6-24 weeks	**269 per 1000**	**328 per 1000** (274 to 390)	**RR 1.22** (1.05 to 1.45)	1446 (8 studies)	⊕⊕⊕⊖ **moderate**[2]	Absolute increase in harm: 6% (1% to 11%)
Injection/Infusion site reaction Follow-up: 6-24 weeks	**70 per 1000**	**211 per 1000** (148 to 302)	**RR 3.02** (2.12 to 4.31)	1232 (7 studies[3])	⊕⊕⊕⊖ **moderate**[2]	Absolute increase in harm: 14% (10% to 18%)
Serious adverse events Follow-up: 6-24 weeks	**21 per 1000**	**38 per 1000** (20 to 72)	**OR 1.83** (0.93 to 3.62)	1168 (8 studies[5])	⊕⊕⊕⊖ **moderate**[4]	Absolute increase in harm: 2% (0% to 4%)
Serious infections Follow-up: 6-24 weeks	**4 per 1000**	**8 per 1000** (2 to 31)	**OR 1.89** (0.45 to 7.96)	1448 (8 studies[1])	⊕⊕⊕⊖ **moderate**[4]	Absolute increase in harm: 0% (-1% to 2%)
Tuberculosis Follow-up: 12-24 weeks	**Medium risk population**[6]		**OR 7.61** (0.15 to 383.66)	69 (4 studies[7])	⊕⊕⊖⊖ **low**[2,4]	Absolute increase in harm: 3% (-5% to 11%)
	0 per 1000	**3 per 1000** (0 to 133)				
Lymphoma Follow-up: 6-24 weeks	See comment	See comment	Not estimable	478 (2 studies[8])	⊕⊕⊖⊖ **low**[2,4]	No events in van der Heijde 2005 or 2006b

*The basis for the **assumed risk** (e.g. the median control group risk across studies) is provided in footnotes. The **corresponding risk** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

CI: Confidence interval; **RR:** Risk ratio; **OR:** Odds ratio;

GRADE Working Group grades of evidence
High quality: Further research is very unlikely to change our confidence in the estimate of effect.
Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Very low quality: We are very uncertain about the estimate.

[1] Brandt 2003, Calin 2004, Gorman 2002 did not have any events in either the treatment or control groups
[2] Outcome not reported in all trials
[3] Braun 2002 did not have events in either the treatment or control group
[4] Few events and wide confidence interval
[5] Brandt 2003 and Gorman 2002 did not have any events in either the treatment or control groups
[6] Used control event rate from recent Cochrane overview of adverse events of biologics (in press, Singh et al. 2010)
[7] Davis 2003, van der Heijde 2005, van der Heijde 2006a did not have any events in either treatment or control groups
[8] van der Heijde 2005 and 2006a reported no events in either group

Table 3: Summary of adverse events reported in extension studies

Study	Total W/D	W/D due to AE	Total AE	Nausea	Total SAE	Any infection	Serious infection	TB	Malignancy or lymphoma	Abnormal (clinical) liver function	Injection site/ infusion reaction	CHF	Demye-lination	Auto-immune	Death
ETN: Davis 2005-Results reported for 72 week OLE and all entering OLE	57/257 (22.2%)	12/257 (5%)	NR	17/257 (6.6%)	17/257 (6.6%)	NR	5/257 (2%)	1/257 (0.4%) (PPD+ but no symptoms)	0/257 (0%)	NR	53/257 (20.6%)	NR	0/257 (0%)	NR	0/257 (0%)
ETN: Davis 2008-Results reported for 168 week OLE and all entering OLE	131/257 (51%)	21/257 (8.2%)	234/257 (91.1%)	NR	33/257 (12.8%)	187/257 (72.8%)	6/257 (2.3%)	1/257 (0.4%)	0	NR	57/257 (22.2%)	NR	0	0	0
INF: Braun08-Results reported for cumulative RCT and OLE (102 weeks) and only those originally in tx grp who entered OLE (N=201)	47/201 (23.4%)	NR	196/201 (97.5%)	NR	34/201 (16.9%)	158/201 (78.6%)	8/201 (4%)	2/201 (1%) (PPD+ no clinical symptoms)	2/201 (1%) (0 lymphoma)	0	42/201 (21.4%)	NR	NR	2/201 (1%)	0
INF: Heldmann09:Results for EASIC 2 yr OLE for those who completed ASSERT +1.3 yrs on INF before starting EASIC	NR	NR	75/89 (84.3%)	NR	NR	230**	3**	0	0	NR	4/89 (4.5%)	NR	NR	NR	NR

Study	Total W/D	W/D due to AE	Total AE	Nausea	Total SAE	Any infection	Serious infection	TB	Malignancy or lymphoma	Abnormal (clinical) liver function	Injection site/ infusion reaction	CHF	Demye-lination	Auto-immune	Death
ADA:vdh09:80 wk OLE; results reported for cumulative 24wk RCT +OLE(total ~2yr ADA exposure)	56/311 (18%)	24/311 (7.7%)	293/311 (94.2%)	27/311 (8.7%)	48/311 (15.4%)	213/311 (68.5%)	6/311 (1.9%)	0	4/311 (1.3%) (1 non-Hodgkin's lymphoma)	0	42/311 (13.5%)	0	0	0	0
ADA vdH08 OLE; results reported for cumulative 24wk RCT +OLE(total~3yr tx). AE rate/100 pt-yrs during 24 week double-blind period vs 3 yrs ADA	88/311 (28%)	3.6 vs 3.8/100 pt-yrs	NR	NR	10.2 vs 11.1 /100 pt-yrs	NR	0.0 vs 1.4 /100 pt-yrs	NR	0.0 vs 0.7 /100 pt-yrs	NR	NR	NR	NR	NR	1/311 (0.3%)

Footnotes: all events reported as # of patients with an event except those with **=number of events; ADA=adalimumab; INF=infliximab; ETN=etanercept; AE=adverse events; SAE=serious adverse events; W/D=withdrawals; CHF=chronic or congestive heart failure; TB=tuberculosis; pt-yrs=patient-years; OLE=open-label extension; NR=not reported

Table 4: Network meta-analysis results for head-to-head estimates

Outcomes	DRUG1	DRUG2	GLIMMIX			MTC			ITC			Favours
			OR	LCL	UCL	OR	LCL	UCL	OR	LCL	UCL	
Withdrawals due to AE	ADALIMUM	ETANERCE	0.381	0.008	18.315	0.013	0.002	4.484	0.25	0.03	1.87	ADA (NS)
	ADALIMUM	INFLIXIMAB	0.618	0.024	15.641	0.024	0.009	14.465	0.37	0.03	4.14	ADA (NS)
	ETANERCE	INFLIXIMAB	1.621	0.086	30.506	0.597	0.110	101.143	1.46	0.20	10.75	NC*
All Infections	ADALIMUM	ETANERCE	1.406	0.533	3.712	1.369	0.575	4.822	1.16	0.55	2.46	ETN (NS)
	ADALIMUM	INFLIXIMAB	1.534	0.487	4.835	1.759	0.702	7.886	2.08	0.74	5.81	INF (NS)
	ETANERCE	INFLIXIMAB	1.091	0.413	2.877	1.151	0.469	4.119	1.79	0.61	5.25	INF (NS)
Serious Infections	ADALIMUM	ETANERCE	0.173	0.005	6.359	0.062	0.000	7.849	0.02	0.00	0.76	ADA (NS)
	ADALIMUM	INFLIXIMAB	0.080	0.002	2.932	0.032	0.000	3.843	0.01	0.00	0.44	ADA (NS)
	ETANERCE	INFLIXIMAB	0.463	0.083	2.574	0.180	0.030	5.305	0.44	0.02	10.55	ETN (NS)
	DIRECT ETN	*INF*				*0.50*	*0.05*	*5.03*				ETN (NS)

ADA=adalimumab; ETN=etanercept; INF=infliximab; OR=odds ratio; LCL=lower confidence limit; UCL=upper confidence limit; NC*=not consistent; NS=not significant

Appendix 1 Search strategies

For RCTs: The MEDLINE search strategy is listed below. Database: Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R) <1950 to Present>

1 exp spondylitis, ankylosing/
2 exp spondylarthropathies/
3 (ankylosing or spondyl$).tw.
4 (bekhterev or bechterew).tw.
5 4 or 1 or 3 or 2
6 exp Receptors, Tumor Necrosis Factor/
7 exp tumor necrosis factor/
8 exp Antibodies, Monoclonal/
9 anti-tumo?r necrosis factor$.sh,rn,tw.
10 antitumo?r necrosis factor$.sh,rn,tw.
11 anti-tnf.sh,rn,tw.
12 antitnf.sh,rn,tw.
13 etanercept.sh,rn,tw.
14 enbrel.sh,rn,tw.
15 infliximab.sh,rn,tw.
16 remicade.sh,rn,tw.
17 adalimumab.sh,rn,tw.
18 humira.sh,rn,tw.
19 or/6-18
20 5 and 19
21 randomized controlled trial.pt.
22 controlled clinical trial.pt.
23 randomized.ab.
24 placebo.ab.
25 drug therapy.fs.
26 randomly.ab.
27 trial.ab.
28 groups.ab.
29 27 or 25 or 28 or 21 or 26 or 22 or 24 or 23
30 (animals not (humans and animals)).sh.
31 29 not 30
32 20 and 31
33 from 32 keep 1-10

Appendix 2 Characteristics of included studies

Table 2a: Summary of characteristics of included randomized controlled trials studies

Study	Duration (weeks)	Intervention and comparator groups	Participant characteristics		AE outcomes reported	Notes
			Treatment group	Comparator group		
Barkham 2008	12	ETN 25 mg twice weekly vs placebo	N=40; mean age =40.1; % male=80; disease duration (yrs)=17 (note, combined group characteristics reported)		W/D due to AE	Abstract
Brandt 2003	6	ETN 25 mg twice weekly vs placebo	N=14; mean age =40; % male=71; disease duration (yrs)=15	N=16; mean age =32; % male=75; disease duration (yrs)=11	W/D due to AE; injection reaction; any infection; serious infection; SAE;	"Supported by a grant (Kompentenzne tz Rheuma) from the German Ministry of Research and by Wyeth Pharma who provided the study drug"
Braun 2002	12	INF 5mg/kg vs placebo	N=34; age (mean years): 41; % male: 68; Disease duration (years): 16;	N=35; age (mean years):39; % male: 63; disease duration (years): 15	W/D due to AE; infusion reaction; any infection; serious infection; SAE;	"Funded by a grant (Kompentenzne tz Rheuma) from the German Ministry of Research and by Essex Pharma, Munich, who provided the study drug"

Braun 2008	16	ETN 50mg once weekly vs SSZ 3 g daily	N=379; mean age=41 years; 74% male; average disease duration=7.5 years. (note, combined group characteristics reported)	N=187	SAE	Abstract. Funding source not reported
Calin 2004	12	ETN 25 mg twice weekly vs placebo	N=45; placebo; age (mean, years): 45% %male: 80%; disease duration (years): 15	N=39; age (mean, years): 41; %male: 77%; disease duration (years): 10	W/D due to AE; injection reaction; serious infection; SAE; headache; nausea; diarrhea; MI; hepatotoxicity; table reported AE occurring in >5% of patients	"Trial was funded by Wyeth Research"
Davis 2003	24	ETN 25 mg twice weekly vs placebo	N=138; age (mean, years):42; % male: 76 ; Disease duration (years): 10	N=139; age (mean, years):42; % male: 76; Disease duration (years):10	W/D due to AE; injection reaction; opportunistic infections; TB; SAE; headache; diarrhea; table reported AE occurring in >5% of patients	"Supported by Immunex Corporation"
Giardina 2009	102	ETN 25mg twice weekly or 5mg/kg infliximab	N=25; age (mean, years): 32.6 SD 6.8; % male: 80; Infliximab group - 76% Disease duration (years): 15.7, SD 6.5	N=25; age (mean, years): 31.9 SD 9.2; % male: 76; Disease duration (years): 15.4, SD 10.6	W/D due to AE; injection/infusion reaction; serious infection; SAE; TB, malignancy, demyelinating disease, autoimmune disorders, congestive heart disease, headache; diarrhea;	Reported as a full-text and an abstract from a conference. Funding source not reported.
Gorman 2002	16	ETN 25 mg twice weekly versus placebo	N=20; age (median, years): 38; % male: 65; disease duration (years): 15	N=20; age (median, years): 39; % male: 90; disease duration (years): 12	W/D due to AE; injection reaction; upper respiratory infections, serious infection; SAE, diarrhea, hepatotoxicity;	"The majority of funding for the study was provided by the National Institute of Arthritis and Musculoskeletal and Skin Diseases. Immunex, the pharmaceutical

						funding source, supplied etanercept and placebo and provided partial funding."
Huang 2008	6	ETN 50mg once weekly vs placebo	N=74; no further details	N=78	Total AE, W/D due to AE, SAE, common AEs(≥3%) reported; hepatotoxicity	Abstract from conference proceeding. Funding source not reported.
Inman 2010	12	INF 3mg/kg or placebo	N=39; age (mean, years): 42.9 SD=10.4; % male:82; disease duration (years): 11.7, SD=10.6; Control group -11.1 (10.3)	N=37 age (mean, years): 39.3 SD=10.4; % male:78; disease duration (years): 11.1, SD=10.3	AE not clearly reported between RCT and OLE period; SAE	Reported in one full-text article and 3 abstracts from conferences. "Supported by Schering-Plough, Canada"
Lambert 2007	24	40mg ADA or placebo every other week	N=38; age (mean (SD), years): 41.9(11.1); % male: 76.3; disease duration (years (SD): 14.5 (9) 12.1 (8.7)	N=44; age (mean (SD), years): 40 (10.9); % male: 81.8; disease duration (years (SD): 12.1 (8.7)	AE not reported in full text article; abstract reports some AE (any infection and death), but timing of AE is unclear	Sponsored by Abbott Laboratories. Advisory committee, including authors from academic institutions and Abbott Laboratories. All authors were involved in preparation of manuscript.
Marzo-Ortega 2005	30	INF(5 mg/kg + MTX vs placebo +	N=28; age (mean , years): 41; % male: 82; disease duration	N=14; age (mean , years): 39; % male: 79; disease duration (median	"Drug-related AE" reported; total AE; infusion reaction; infections; serious infections;	This study was supported by a grant in aid

		MTX (both groups 7.5 mg – 10mg folic acid per week)	(median years): 8	years): 10	SAE	from Schering-Plough, UK.
Van der Heijde 2005	24	5 mg/kg INF vs placebo	N=201; age (mean, years): 40; % male: 78; disease duration (years): 8;	N=78; age (mean, years): 41; % male: 87; disease duration (years): 13	W/D due to AE; total AE; infusion reaction; any infection; serious infection; SAE; TB; lymphoma; death; headache; nausea; diarrhea; hepatotoxicity	"Supported by Centocor Inc"
van der Heijde 2006a	24	ADA 40 mg every other week vs placebo	N=208; age (mean, years): 42; % male: 76 %; disease duration (years): 11	N=107; age (mean, years): 43; % male: 74; disease duration (years): 10	W/D due to AE; total AE; injection reaction; any infection; serious infection; SAE; TB; lymphoma; demyelinating disease; autoimmune disease; congestive heart failure; death; headache	"Supported by Abbott Laboratories"
Van der Heijde 2006b	12	ETN 50mg once weekly vs ETN 25mg twice weekly vs placebo	ETN 50mg once weekly:N=155; mean age (SD)=41.5 (11.0); 69.7% male; disease duration, years (SD)=9.0 (8.7). Etanercept 25mg twice weekly:N=150; mean age (SD)=39.8 (10.7); 76% male; disease duration, years (SD)=10.0 (9.1).	Placebo:N=51; mean age (SD)=40.1 (10.9); 78.4% male; disease duration, years (SD)=8.5 (6.8).	W/D due to AE; injection reaction; any infection; serious infection; lymphoma; demyelinating disease; autoimmune disease; death; headache ; nausea; diarrhea	"Study was supported by Wyeth Pharmaceuticals , Collegeville, Pennsylvania, USA (study drug and grants to investigational sites).

Table 2b: Summary of characteristics of studies included for long-term adverse event assessment of adalimumab, etanercept, and infliximab

Study	Drug	Country	Study type	Funding source	Population at risk/Number completed study	Duration of observation & Reporting details	Outcomes
Davis 2005 (FT)	Etanercept	America, Canada, Netherlands, Germany, France	Open label extension (from Davis 2003 RCT)	Study funded by Immunex Corp, a wholly owned subsidiary of Amgen Inc, and by Wyeth Research	257 enrolled in OLE (out of 277 in RCT)	96 weeks (24 wk RCT + 72 wk OLE) Reported OLE data only. Reported AE combined ETN + PL patients	AE, SAE, infections, serious infections, lupus, demyelinating disorders, lymphoma, death
Davis 2008 (FT)	Etanercept	America, Canada, Netherlands, Germany, France	Open label extension (from Davis 2003 RCT)	Study funded by Immunex Corp, a wholly owned subsidiary of Amgen Inc, and by Wyeth Research	257 enrolled in OLE (out of 277 in RCT)/126 completed	192 weeks (24 wk RCT + 168 wk OLE) Reported OLE data only. Reported AE combined ETN + PL pts	AE, SAE, infections, serious infections, lupus, demyelinating disorders, lymphoma, death
van der Heijde 2009 (FT)	Adalimumab	43 centers in the US and Europe	Open label extension (from van der Heijde 2006a)	Funded by Abbott Laboratories	311 entered the OLE; 296/315 completed the 24 week RCT; 218/311 were exposed for at least 1.75 years	~2 years. (24wk RCT + 80 wk OLE). Reported cumulative AE data only. Reported AE combined ETN + PL pts	AE, SAE, infections, serious infections, injection-site reactions, congestive heart failure, lupus, demyelinating disorders, lymphoma, death
van der Heijde 2008 (AB)	Adalimumab	43 centers in the US and Europe	Open label extension (from van der Heijde 2006a)	Funded by Abbott Laboratories	311 entered the OLE; 296/315 completed the 24 week RCT; 227 received adalimumab at 3 years	~3 years. (24 RCT wk + OLE). Reported cumulative AE data only. Reported AE combined ETN + PL pts	AE, SAE, serious infections, malignancies, death
Braun 2008 (FT)	Infliximab	33 centers in the US, Canada, and Europe	Extension (blinding maintained; from van der Heijde 2005)	Funding not clearly stated; but RCT report states 'Supported by Centocor, Inc'	279 randomized; 61 (78.2%) randomized to placebo and 166 randomized to infliximab (82.6%) completed 102 weeks	~2 years (24 wk RCT + 72 wk blinded extension). Reported cumulative AE data. Reported separate results for patients originally in INF gp who entered OLE	AE, SAE, infection, serious infections, infusion reaction, lupus, malignancies, death

Heldmann 2009 (AB)	Infliximab	6 European countries	Open label extension; started 1.3+/-0.9 yrs after the van der Heijde 2005 RCT	Not disclosed	89/149 from van der Heijde 2005 RCT entered EASIC after infliximab treatment during lag phase	~ 5 years. Results presented only for the 2 year EASIC cohort study	AE, SAE, infections, serious infections, infusion reactions, malignancies
AB=abstract; AE=adverse event; FT=full-text; OLE=open-label extension; SAE=serious adverse event							

Appendix 3 Methodological quality of included studies

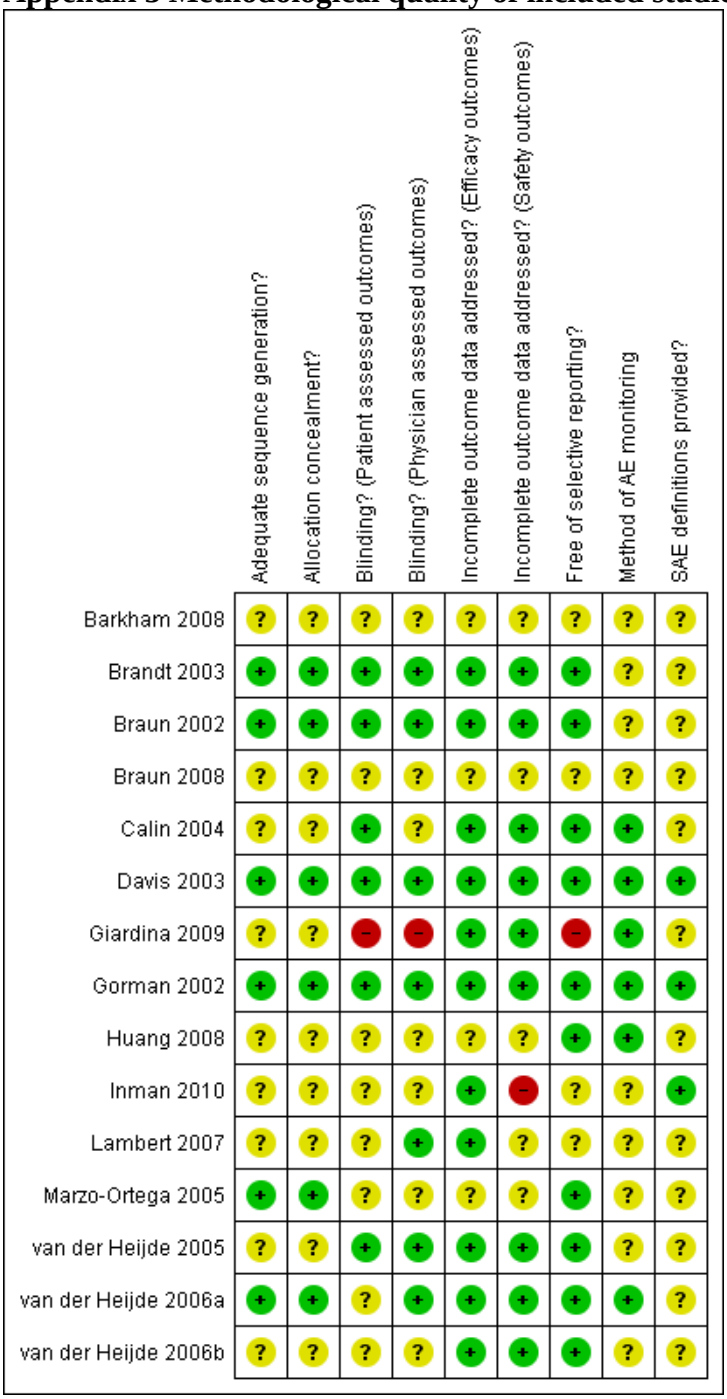


Figure 3a Summary of the risk of bias of included randomized controlled trials

Table: Assessment of risk of bias in non-RCTs using the Newcastle-Ottawa Scale for cohort studies

NOS Criteria	Study	Judgment:	Study	Judgment	Study	Judgment	Study	Judgment	Study	Judgment	Study	Judgment
Selection	vdH09		vdH08		Braun 08		Held-mann09		Davis 05		Davis08	
Representativeness of the exposed cohort	0	Highly selected for RCT; if a significant AE in RCT phase would not continue	0	Highly selected for RCT; if a significant AE in RCT phase would not continue	0	Highly selected for RCT; if a significant AE in RCT phase would not continue	0	Highly selected for RCT; if a significant AE in RCT phase would not continue	0	Highly selected for RCT; if a significant AE in RCT phase would not continue	0	Highly selected for RCT; if a significant AE in RCT phase would not continue
Selection of the non exposed cohort	N/A	Single cohort; all exposed	N/A	Single cohort; all exposed	N/A	Single cohort; all exposed	NA	Single cohort; all exposed	NA	Single cohort; all exposed	NA	Single cohort; all exposed
Ascertainment of exposure	0	none of the published articles stated how they ascertained that participants in the extension	0	none of the published articles stated how they ascertained that participants in the extension study were taking the anti-TNF therapy as	0	none of the published articles stated how they ascertained that participants in the extension study were taking the anti-TNF therapy as	0	none of the published articles stated how they ascertained that participants in the extension	0	none of the published articles stated how they ascertained that participants in the extension	0	none of the published articles stated how they ascertained that participants in the extension
Demonstration that outcome of interest was not present at start of study	*	Strict exclusion criteria for RCT (e.g TB, history of cardiac, renal, neurologic, psychiatric, endocrinologic metabolic, or hepatic disease; and a history of demyelinating disease, multiple sclerosis, or malignancy)	*	Strict exclusion criteria for RCT (e.g TB, history of cardiac, renal, neurologic, psychiatric, endocrinologic metabolic, or hepatic disease; and a history of demyelinating disease, multiple sclerosis, or malignancy)	*	Strict exclusion criteria for RCT (e.g TB, infection, CHF, multiple sclerosis, malignancy)	0	There was a 1.3 year lag after ASSERT and before starting EASIC so unclear if the outcome wss present at start of EASIC	*	Exclusion criteria for RCT (e.g had a serious infection (requiring hospitalization or IV antibiotics))	*	Exclusion criteria for RCT (e.g had a serious infection (requiring hospitalization or IV antibiotics))
Comparability												
Comparability of cohorts on the basis of the design or analysis	N/A	Nothing controlled for; single cohort	N/A	Nothing controlled for; single cohort	N/A	Nothing controlled for; single cohort	NA	Nothing controlled for; single cohort	NA	Nothing controlled for; single cohort	NA	Nothing controlled for; single cohort
Outcome												
Assessment of outcome	*	No blinding, but AE of interest had objective diagnostics	*	No blinding, but AE of interest had objective diagnostics	*	Only 1 patient unblinded over the 2 years (Fig1) and AE of interest had objective diagnostics	*	No blinding, but AE of interest had objective diagnostics	*	No blinding, but AE of interest had objective diagnostics	*	No blinding, but AE of interest had objective diagnostics
Was follow-up long enough for outcomes to occur	*	Yes, 2 yrs infections and malignancies	*	Yes, 2 yrs infections and malignancies	*	Yes, 2 yrs infections and malignancies	*	Yes, 5 years for infections and malignancies	*	Yes, 1.5 years for infections and malignancies	*	Yes, 3.5 years for infections and malignancies
Adequacy of follow up of cohorts	*	261/311 83.9%received adalimumab treatment to 2 years	0	227/311 (73%) received adalimumab for at least 3 years	*	61/76 (78.2%) and 166/201 (82.6%) completed 102 weeks	0	Not reported	0	200/257 (78%) completed 72 weeks of OLE	0	126/257 (49%) completed 168 weeks of OLE

Figure 3b: Summary of the risk of bias of included extension studies

Appendix 4 Network meta-analysis results

Table 4 Network meta-analysis results for placebo estimates

Outcomes	DRUG1	DRUG2	GLIMMIX			MTC			ORIGINAL (OR, RE)			Favours
			OR	LCL	UCL	OR	LCL	UCL	OR	LCL	UCL	
Withdrawals due to AE	ADALIMUM	Placebo	1.573	0.109	22.715	3.358	0.066	19.510	1.03	0.19	5.65	PL (NS)
	ETANERCE	Placebo	4.125	0.507	33.576	7.516	0.914	154.1	4.03	1.43	11.30	PL (NS)
	INFLIXIMAB	Placebo	2.545	0.348	18.629	8.225	0.305	39.120	2.76	0.50	15.22	PL(NS)
All Infections	ADALIMUM	Placebo	1.708	0.731	3.992	2.176	0.884	4.749	1.87	1.16	3.03	PL (NS)
	ETANERCE	Placebo	1.215	0.666	2.215	1.320	0.638	2.320	1.27	0.84	1.91	PL (NS)
	INFLIXIMAB	Placebo	1.114	0.478	2.597	1.028	0.371	2.081	0.90	0.36	2.22	PL (NS)
Serious Infections	ADALIMUM	Placebo	0.216	0.006	7.512	0.759	0.000	5.346	0.05	0.00	3.30	ADA (NS)
	ETANERCE	Placebo	1.247	0.246	6.320	2.663	0.194	12.800	2.25	0.30	17.13	PL (NS)
	INFLIXIMAB	Placebo	2.693	0.517	14.022	7.368	0.339	39.170	5.10	0.44	58.76	PL (NS)

ADA=adalimumab; ETN=etanercept; INF=infliximab; PL=placebo; NS=not significant; OR=odds ratio; LCL=lower confidence limit; UCL=upper confidence limit; RE=random effects

CHAPTER SIX: DISCUSSION AND CONCLUSIONS

6.1 Discussion

The purpose of this thesis was to examine the variability of three important outcome measures, namely, disease activity, function, and well-being, over the course of a year in a pre-biologic era cohort of AS patients and to conduct a systematic review on the efficacy and safety of using TNF-blockers in the treatment AS using both direct and indirect evidence. Decisions by patients and clinicians as to what course of treatment should be undertaken requires both an understanding of the natural history of the disease as well as the trade-off between benefit and harm of possible interventions.

This thesis is composed of three manuscripts. The exploration of the variability of disease activity, function and well-being in the first manuscript (chapter three) provided a useful characterization of the natural fluctuations of these outcomes in a cohort of AS patients, unexposed to anti-TNF therapy, and demonstrated that, in this particular cohort, the current guidelines regarding when TNF-blocker therapy should be initiated or maintained fit well with the natural course of disease activity in these patients. The introduction of anti-TNF agents to the list of therapeutic options for AS patients was heralded as a major breakthrough; however, no systematic review of the literature had been undertaken to establish and quantify the efficacy and harms of this treatment. The second manuscript (chapter four) is a systematic review of the literature following Cochrane methodology and formatting. As mentioned above, a key component of decision-making for therapeutic options requires balancing the potential for improvement against the potential for adverse effects. Assessment of harms often poses greater challenges to systematic reviewers yet is of great interest to clinicians and patients. The third manuscript (chapter five) focuses on the results of the systematic review of adverse events associated with TNF-blocker therapy in AS and highlights the difficulties in assessment of these effects. This manuscript was written for a specialty journal with the aim of reaching a wider clinical audience.

In chapter three, our analysis of frequently-measured key outcomes in a year-long cohort of AS patients demonstrated that there was large variation in these outcomes, both within and between-individuals, though the variability was greater between-individuals. As well, we found evidence of complex patterns (cubic and quadratic) in changes of these outcome measures over the year. This result reinforces the awareness of the nature of the flare and remission course of this disease and provides some characterization as to the patterns. Recent work on the patterns of flares by researchers in the UK identified four possible patterns of disease activity. The most frequently identified flare pattern in both our study cohort and the UK group was that of ongoing flares with disease activity in between the flares, thus demonstrating consistency in the patterns in different cohorts of AS patients. We also assessed the variability in disease activity in this cohort against current guidelines on the use of anti-TNF agents using different scenarios of timing of assessments. The majority of patients showed stability over the scenarios, thus indicating that the current guidelines of two assessments of disease activity four weeks apart fits well with the natural course of disease activity seen in this cohort.

In chapter four, a systematic review of three TNF-blockers concluded that there is mainly high quality evidence from RCTs that, when compared to placebo, these agents reduce disease activity and improve function in patients with AS. In addition to establishing efficacy against placebo, clinicians and patients are interested in knowing how each anti-TNF agent performs against others. One small head-to-head RCT was found during the literature search. To supplement this direct knowledge, we used indirect comparison methodology to obtain head-to-head estimates and, in keeping with the direct RCT results, found there was no significant difference between the efficacy of adalimumab, etanercept or infliximab against each other. Therefore, there was no evidence to recommend one TNF-blocker over another.

Assessment of adverse effects is often more challenging than establishing efficacy; specifically in terms of searching due to poor indexing, a lack of standardized reporting, and establishing causality. For rare or delayed adverse effects, RCTs are

often underpowered and/or too short in duration to provide enough information about these types of effects. Chapter five focused on the systematic assessment of harms associated with anti-TNF treatment in AS. A meta-analysis of the RCT data for all three anti-TNF agents did not find a statistically significant relative increase between anti-TNF agents and serious adverse events or serious infections in the short-term. For both of these outcomes, the point estimate favoured the placebo group, though the absolute increase in harm was small. There were statistically significant increases in total adverse events, total infections, and injection/infusion site reactions. Few cases of tuberculosis, lymphoma, demyelinating disease, and congestive heart failure were observed in the RCTs. Given the problems associated with adverse effect assessment using RCTs, we performed a systematic search and assessment of observational data. The only non-RCT studies that met our inclusion criteria were from four extension studies (single-cohort) following on from the included RCTs. While this data was judged to be at a high risk of bias, there was no large signal in safety concerns over the two- and three-year follow-up periods.

6.2 Future research

Future research should focus on the efficacy of these TNF-blockers in head-to-head studies, and to investigate risk factors for predicting which patients may benefit more from a specific anti-TNF agent as well as risk factors for different patterns of disease activity. The assessment of adverse effects can be much enhanced by the collaboration of both systematic review methodologists and pharmacoepidemiologists as proposed by Vandenbroucke (1). Data from biologic registries with a matched control group will be useful in providing estimates for rare and delayed adverse effects.

6.3 References

(1) Vandenbroucke JP, Psaty BM. Benefits and risks of drug treatments: how to combine the best evidence on benefits with the best data about adverse effects. JAMA 2008 Nov 26:300(20):2417-9

APPENDICES

Appendix A: BASDAI, BASFI, AND BAS-G questionnaires

NRS BASDAI

Please tick the box which represents your answer.
All questions refer to **last week**. (i.e. ☒)

41. How would you describe the overall level of fatigue/tiredness you have experienced?

0	1	2	3	4	5	6	7	8	9	10

none very severe

42. How would you describe the overall level of AS **neck, back or hip** pain you have had?

0	1	2	3	4	5	6	7	8	9	10

none very severe

43. How would you describe the overall level of pain/swelling in joints **other than** neck, back or hips you have had?

0	1	2	3	4	5	6	7	8	9	10

none very severe

44. How would you describe the overall level of discomfort you have had from any areas tender to touch or pressure?

0	1	2	3	4	5	6	7	8	9	10

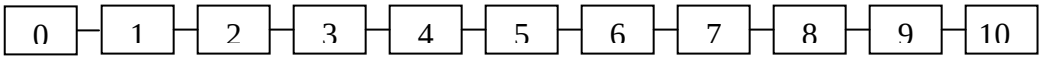
none very severe

45. How would you describe the overall level of morning stiffness you have had from the time you wake up?

0	1	2	3	4	5	6	7	8	9	10

none very severe

46. How long does your morning stiffness last from the time you wake up?



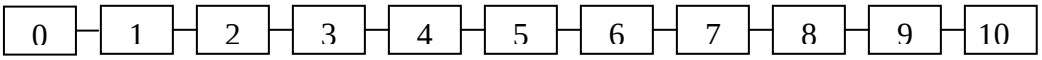
0 hr 1 hr 2 or more hrs

NRS BASFI

Please indicate your level of ability with each of the following activities during **the last week**.
(i.e. 10)

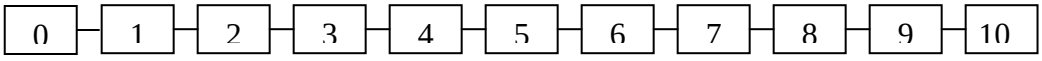
(An aid is a piece of equipment which helps you to perform an action or movement)

47. Putting on your socks or tights without help or aids (e.g. sock aid).



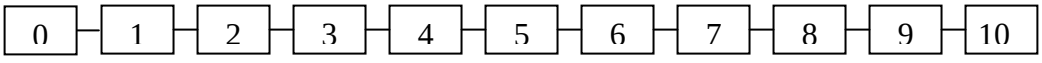
easy impossible

48. Bending forward from the waist to pick up a pen from the floor without an aid.



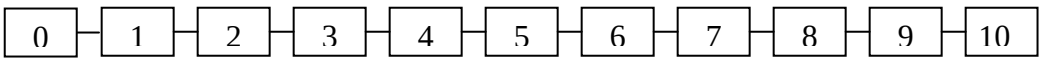
easy impossible

49. Reaching up to a high shelf without help or aids (e.g. helping hand).



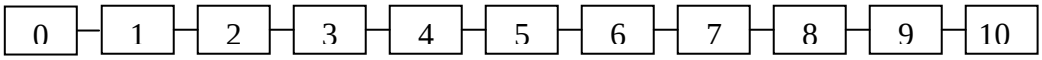
easy impossible

50. Getting up out of an armless dining room chair without using your hands or any other help.



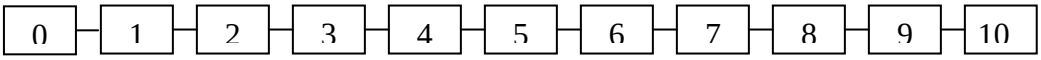
easy impossible

51. Getting up off the floor without help from lying on your back.



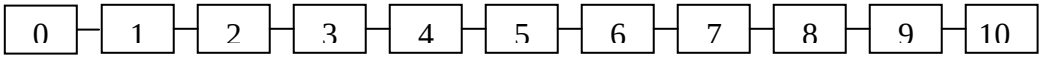
easy impossible

52. Standing unsupported for 10 minutes without discomfort.



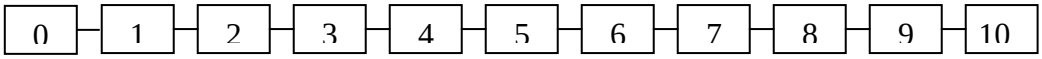
easy impossible

53. Climbing 12-15 steps without using a handrail or walking aid. **One foot at each step**.



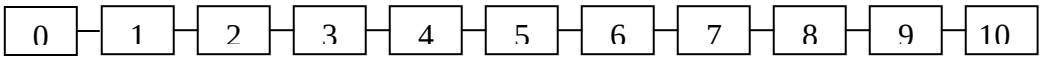
easy impossible

54. Looking over your shoulder without turning your body.



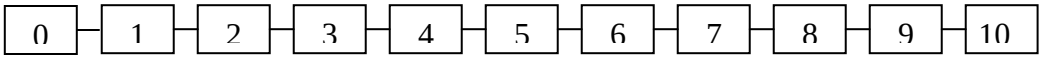
easy impossible

55. Doing physically demanding activities (e.g. physiotherapy exercises, gardening or sports).



easy impossible

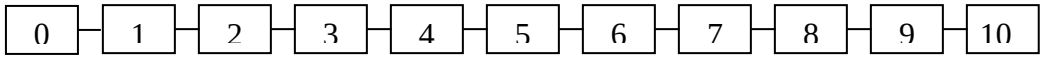
56. Doing a full days activities, whether it be at home or at work.



easy impossible

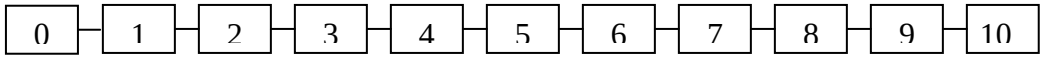
NRS BAS-G

39. Please tick a box to indicate the effect your disease has had on your well-being over the **last week**.



none very severe

40. Please indicate the effect your disease has had on your well-being over **the last six months**.



none very severe

Appendix B: Letter of ethics approval from the University of Ottawa

File Number: H01-09-05 **Date (mm/dd/yyyy):** 02/25/2009

Université d'Ottawa University of Ottawa
Service de subventions de recherche et déontologie Research Grants and Ethics Services

Ethics Approval Notice

Health Sciences and Science REB

Principal Investigator / Supervisor / Co-investigator(s) / Student(s)

First Name	Last Name	Affiliation	Role
George	Wells	Medicine / Medicine	Supervisor
Peter	Tugwell	Medicine / Medicine	Co-Supervisor
Lara	Maxwell	Medicine / Medicine	Student Researcher

File Number: H01-09-05

Type of Project: Secondary use of data

Title: Assessment of the Efficacy and Safety of Anti-TNFa Therapy and Intra-Individual Variability of Outcome Measures in Ankylosing Spondylitis

Approval Date (mm/dd/yyyy)	Expiry Date (mm/dd/yyyy)	Approval Type
02/25/2009	02/24/2010	Ia

(Ia: Approval, Ib: Approval for initial stage only)

Special Conditions / Comments:
N/A

File Number: H01-09-05 **Date (mm/dd/yyyy):** 02/25/2009

Université d'Ottawa University of Ottawa
Service de subventions de recherche et déontologie Research Grants and Ethics Services

This is to confirm that the University of Ottawa Research Ethics Board identified above, which operates in accordance with the Tri-Council Policy Statement and other applicable laws and regulations in Ontario, has examined and approved the application for ethical approval for the above named research project as of the Ethics Approval Date indicated for the period above and subject to the conditions listed the section above entitled "Special Conditions / Comments".

During the course of the study the protocol may not be modified without prior written approval from the REB except when necessary to remove subjects from immediate endangerment or when the modification(s) pertain to only administrative or logistical components of the study (e.g. change of telephone number). Investigators must also promptly alert the REB of any changes which increase the risk to participant(s), any changes which considerably affect the conduct of the project, all unanticipated and harmful events that occur, and new information that may negatively affect the conduct of the project and safety of the participant(s). Modifications to the project, information/consent documentation, and/or recruitment documentation, should be submitted to this office for approval using the "Modification to research project" form available at: http://www.rges.uottawa.ca/ethics/application_dwn.asp

Please submit an annual status report to the Protocol Officer 4 weeks before the above-referenced expiry date to either close the file or request a renewal of ethics approval. This document can be found at: http://www.rges.uottawa.ca/ethics/application_dwn.asp

If you have any questions, please do not hesitate to contact the Ethics Office at extension [redacted] or by e-mail at: [redacted]

[redacted]
Germain Zongo
Assistant Director, Ethics (Interim)
For Dr. Daniel Lagarec, Chair of the Health Sciences and Science REB

Appendix C: Copyright permission for use of Figure 1 in chapter 3

This is a License Agreement between Lara J Maxwell ("You") and Oxford University Press ("Oxford University Press"). The license consists of your order details, the terms and conditions provided by Oxford University Press, and the payment terms and conditions.

Get the printable license.

License Number	2562651239558
License date	Dec 05, 2010
Licensed content publisher	Oxford University Press
Licensed content publication	Rheumatology
Licensed content title	Assessment of the impact of flares in ankylosing spondylitis disease activity using the Flare Illustration:
Licensed content author	M. A. Stone, E. Pomeroy, A. Keat, R. Sengupta, S. Hickey, P. Dieppe, R. Gooberman-Hill, R. Mogg, J. Richardson, R. D. Inman
Licensed content date	08/01/2008
Volume number	47
Issue number	8
Type of Use	Thesis/Dissertation
Requestor type	Academic/Educational institute
Format	Print and electronic
Portion	Figure/table
Number of figures/tables	5
Will you be translating?	No
Author of this OUP article	No
Order reference number	
Title of your thesis / dissertation	Assessment of the efficacy and safety of anti-TNFα therapy and variability of outcome measures in ankylosing spondylitis
Expected completion date	Jun 2011
Estimated size(pages)	200
Total	0.00 USD

Appendix D: Detailed description of methodology used in the analysis of the manuscript in chapter 3

The methods for each objective are expanded upon below in greater detail than is allowed in a publication-quality manuscript.

Objective 1. Describe the variability in the main outcomes, BASDAI, BASFI, and BAS-G, in of this cohort of patients:
1.1. We used box plots and time plots to describe the variability in each patient and overall over the 52 weeks.

1.2 Overall inter (between)- versus intra (within)-patient variability was assessed using a variance component model which estimated the effect of the subject (a random variable) to the variance of disease activity (Proc Varcomp in SAS), and intracluster correlation coefficients.

1.3 We used mixed effects modeling to investigate the general patterns of change over time for BASDAI, BASFI, and BAS-G. Objective 4 below provides greater detail on mixed modeling methods.

1.4 The minimally clinical important difference (MCID) of disease activity, function, and global well-being have been identified as BASDAI = 1, BASFI = 0.7, and BAS-G = 1.5, all on a 0 to 10 scale (1). We investigated the number of patients who exceeded the MCID from their median score of the outcome for each of these three outcomes. For example, for patient 1, their median BASDAI was 4.9. We counted the number of times during the 52-week study that their BASDAI score exceeded 4.9 +/- 1 (since the MCID for BASDAI is 1). We did this for all patients for the three outcomes of interest. We displayed this as a histogram for each outcome, showing a count of the number of weeks each patient exceeded the MCID.

1.5 Recent work by Stone et al (2) identifies four different graphical illustrations of flare and remission patterns of disease activity (Figure 1, page 32). The patterns are:

1. Flares and remissions with periods of symptom-free remission; 2. Flares with disease activity in between flares; 3. Severe flare of long duration followed by long symptom-free remission; and 4. Severe flare of long duration followed by continuing, constant symptoms. In this cohort, we also had some patients who did not have a flare over the course of the 52 weeks and so created a fifth category to describe this lack of a flare.

By visually inspecting graphs of disease activity over time generated for each patient, we categorized each patient in this cohort into one of these five disease patterns. In the study by Stone, a flare was defined as an "exacerbation of the disease that may have required additional treatment or necessitated a visit to a health care professional". Given the retrospective nature of our data, we used a change in 1 point or more in the BASDAI (since the BASDAI MCID =1) to define a flare in our cohort. We used the chi-square procedure to test for an association between the disease pattern groupings and sex. We used ANOVA and Kruskal-Wallis to test for differences between the disease pattern groupings and age and disease duration for normally distributed variables and non-normally distributed variables, respectively. Statistical significance was defined as $p<0.05$.

Objective 2 To determine the number of consecutive assessments that need to be averaged to provide a stable estimate of disease activity, specifically in the context of establishing baseline disease activity for a randomized controlled trial, we used a moving average approach. That is, consecutive observations were averaged to reduce variability within our observations. In an RCT setting the specific question of interest is: should we take one baseline measurement and then randomize the participant, or should we take an average of several weekly scores to obtain an estimate that reduces the variability, and provides a more stable estimate of disease activity before randomizing the participant? By averaging the weekly observations, variability will naturally be reduced, but the important question is whether it is actually worth the effort to obtain these additional observations to be averaged? To answer this question we focused on the first four weekly observations in the dataset.

We chose a maximum of four weeks as we felt that this was most feasible in terms of time and cost in the context of entry criteria for a RCT. For each patient, we determined the mean BASDAI for the first week, and then for the next three weeks, averaging the scores together after each additional week. We calculated the cohort mean and standard deviation with their respective 95% confidence intervals, as well as the coefficient of variation (CV, SD/mean) for each of the four categories. We hypothesized that the mean scores for each of the moving averages would not change significantly, but that a large reduction in the variability (as measured by the SD and CV) would indicate that the repeated measures are useful in providing a more stable estimate of disease activity.

Objective 3 We used the Pearson correlation statistic to investigate the relationship between the three main outcomes of interest: disease activity (BASDAI) and function (BASFI) disease activity (BASDAI) and well-being (BAS-G), and function (BASFI) and well-being (BAS-G) for each patient and presented descriptive statistics. We also calculated the correlation for the entire sample for each of the above three pairs of outcomes. To assess the robustness of our results we repeated the analysis using the Spearman correlation statistic which is a non-parametric test and less sensitive to outliers. To determine whether disease activity or well-being contribute independently to physical function, we used mixed effects modeling with BASDAI or BAS-G as the dependent variable and adjusted for age, sex, and disease duration.

Objective 4 We used mixed effects modeling to investigate the general patterns of change over time in disease activity, function, and well-being and to explore the association of age, sex, and disease duration with these outcomes. Mixed modeling is the most appropriate type of analysis for this longitudinal study. It allows for flexible management of the correlation of the observations due to the repeated measures nature of the data. Since we were interested in the pattern of change over time, rather than the measurement of outcomes at specific times, we modeled time as a random factor. We expected that we would find highly variable slopes and

intercepts so we planned to use a random coefficient analysis which included both a random intercept and a random slope with time. Therefore, the most appropriate covariance structure for this analysis is modeled through the random effects (Proc Mixed in SAS with the Random statement) (3). As well, we included linear, quadratic, and cubic functions of time given the various types of curves visible in the individual profiles over time. Since similar results were seen in the time plots of BASFI and BAS-G, we modeled them in the same way. The following equation describes this relationship:

$$Y_{it} = \beta 0_i + \beta 1_{it} + \varepsilon_{it}$$

where Y_{it} are observations for subject i at time t, $\beta 0_i$ is the random intercept for subject i, t is time, $\beta 1_i$ is the regression coefficient for time, and ε_{it} is the error of subject i at time t. (pg.78 (4)). We used $P<0.05$ as the cut-point for keeping variables in the model and used backward elimination to remove non-significant variables. In selecting the best fit model we used Akaike's and Schawarz's information criteria (AIC and BIC) where the model with the smaller information criteria is preferred. We entered the following variables in the maximum model: age, sex, disease duration, daily dose of NSAID, time modeled as linear, quadratic, and cubic functions, and interaction terms of time with the patient characteristic variables mentioned above. We undertook model diagnostics, specifically assessing the normality and distribution of residuals and the distribution of the random effects (intercept and the random effects of time) to check the fit of the final model. Any statistically significant interaction terms were further explored graphically in an attempt to provide an explanation for the interaction.

Objective 5 As this cohort of patients were assessed in 2000, before anti-TNF therapy was available as a treatment option they can be considered a historical control group that can be used to make decisions on when biologics are indicated or should be switched. Based on the ASAS recommendation of patient selection for initiation of anti-TNF therapy (5) of active disease for >4 weeks and BASDAI > 4 (0–10 scale), we assumed that two consecutive BASDAI scores greater than four and measured four weeks apart would be sufficient to initiate biologic therapy. We then

assessed whether the BASDAI score every four weeks, from the point of initiation, either stayed above 4 or declined below 4 for twelve consecutive weeks. The choice of twelve weeks was made based on the ASAS guidelines which state that the assessment for a reduction in active disease should be made between six and twelve weeks; otherwise it is recommended that anti-TNFα therapy is either switched to another anti-TNFα agent or withdrawn. Week one was taken as the arbitrary starting point of the assessment.

We categorized patients into one of three categories:

Category 1: patient achieved a level of disease activity that would have qualified them for biologic treatment (defined as a BASDAI >=4 for 2 consecutive visits 4 weeks apart), and the natural course of their BASDAI suggests that initiation and continuation of biologic therapy would have been warranted (ie, BASDAI was maintained >=4 at 12 weeks follow-up);

Category 2: while two consecutive >=4 BASDAI scores qualified him for biologic treatment, twelve weeks after this point there was a natural decline in disease activity and biologic therapy may not have been necessary;

Category 3: there was no point at which biologic treatment would have been warranted because there were no two consecutive BASDAI scores >=4, measured four weeks apart.

The percentage of patients in each category was determined. We then repeated the above analysis using different assessment scenarios of three assessments, four weeks apart and four assessments, four weeks apart, while keeping the same follow-up time of twelve weeks, to determine if the number of patients in each category changed significantly depending on the assessment scenario. We hypothesized that if many patients were switching categories in the different scenarios, then more than the two currently recommended assessments may be required.

References

(1) Pavy S, Brophy S, Calin A. Establishment of the minimum clinically important difference for the bath ankylosing spondylitis indices: a prospective study. J Rheumatol 2005 Jan;32(1):80-5.

(2) Stone MA, Pomeroy E, Keat A, Sengupta R, Hickey S, Dieppe P, et al. Assessment of the impact of flares in ankylosing spondylitis disease activity using the Flare Illustration. Rheumatology (Oxford) 2008 Aug;47(8):1213-8.

(3) Cheng J, Edwards LJ, Maldonado-Molina MM, Komro KA, Muller KE. Real longitudinal data analysis for real people: Building a good enough mixed model. Statist Med 2010;29:504-20.

(4) Twisk J.W.R. Applied longitudinal data analysis for epidemiology: a practical guide. Cambridge: Cambridge University Press; 2003.

(5) Braun J, Davis J, Dougados M, Sieper J, van der Linden S, Van Der Heijde D. First update of the international ASAS consensus statement for the use of anti-TNF agents in patients with ankylosing spondylitis. Ann Rheum Dis 2006 Mar;65(3):316-20.