Making best use of evidence for explicit decisions in health care

Volume 2 of 2

Marta Soares

PhD

University of York

Health Sciences

April 2017

List of Contents

Volume 2

Appendices (continued)

Paper 4.

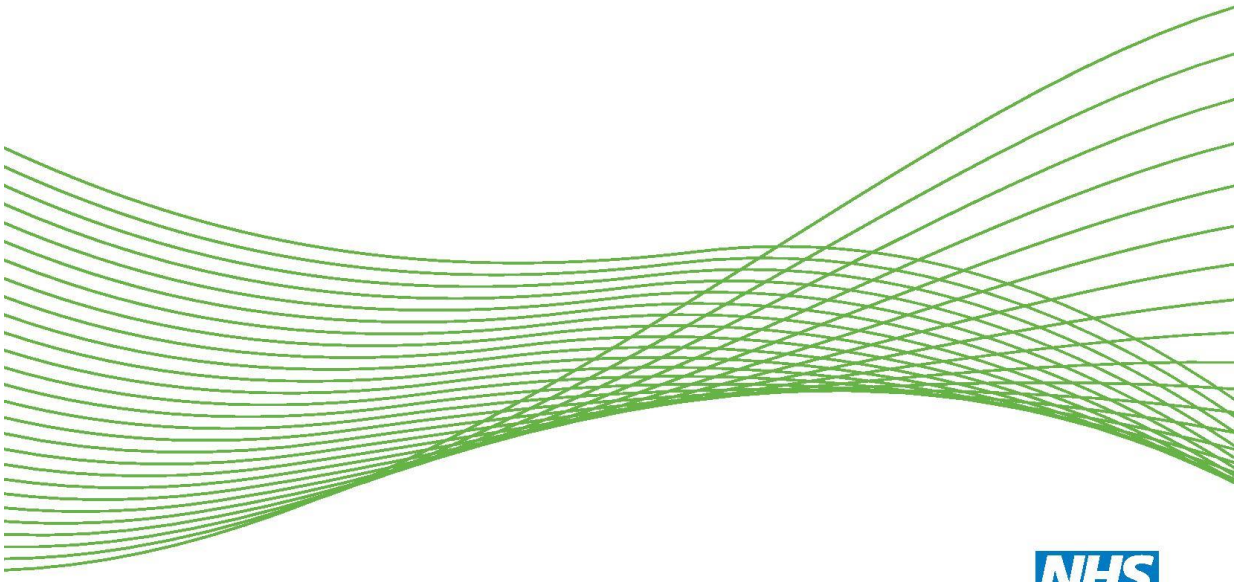
Corbett M, Soares M, Jhuti G, Rice S, Spackman E, Sideris E et al. Tumour necrosis factor-α inhibitors for ankylosing spondylitis and non-radiographic axial spondyloarthritis: a systematic review and economic evaluation. Health Technology Assessment. 2016 Feb;20(9):1-334.

HEALTH TECHNOLOGY ASSESSMENT

VOLUME 20 ISSUE 9 FEBRUARY 2016
ISSN 1366-5278

Tumour necrosis factor-α inhibitors for ankylosing spondylitis and non-radiographic axial spondyloarthritis: a systematic review and economic evaluation

Mark Corbett, Marta Soares, Gurleen Jhuti, Stephen Rice, Eldon Spackman, Eleftherios Sideris, Thirimon Moe-Byrne, Dave Fox, Helena Marzo-Ortega, Lesley Kay, Nerys Woolacott and Stephen Palmer



DOI 10.3310/hta20090

Tumour necrosis factor-α inhibitors for ankylosing spondylitis and non-radiographic axial spondyloarthritis: a systematic review and economic evaluation

Mark Corbett,[1] Marta Soares,[2] Gurleen Jhuti,[2] Stephen Rice,[1] Eldon Spackman,[2] Eleftherios Sideris,[2] Thirimon Moe-Byrne,[1] Dave Fox,[1] Helena Marzo-Ortega,[3] Lesley Kay,[3] Nerys Woolacott[1]* and Stephen Palmer[2]

[1]Centre for Reviews and Dissemination, University of York, York, UK
[2]Centre for Health Economics, University of York, York, UK
[3]Division of Rheumatic and Musculoskeletal Disease, Chapel Allerton Hospital, Leeds Teaching Hospitals NHS Trust and University of Leeds, Leeds, UK
[4]Department of Rheumatology, Freeman Hospital, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK

*Corresponding author

Declared competing interests of authors: Lesley Kay has received sponsorship to attend meetings by AbbVie and Merck Sharp & Dohme Limited in 2014. Helena Marzo-Ortega has received grants, sponsorship and/or honoraria from AbbVie, Janssen, Merck Sharp & Dohme Limited, Pfizer and UCB.

Published February 2016
DOI: 10.3310/hta20090

This report should be referenced as follows:

Corbett M, Soares M, Jhuti G, Rice S, Spackman E, Sideris E, *et al*. Tumour necrosis factor-α inhibitors for ankylosing spondylitis and non-radiographic axial spondyloarthritis: a systematic review and economic evaluation. *Health Technol Assess* 2016;**20**(9).

Health Technology Assessment is indexed and abstracted in *Index Medicus*/MEDLINE, *Excerpta Medica*/EMBASE, *Science Citation Index Expanded* (SciSearch®) and *Current Contents*®/Clinical Medicine.

Health Technology Assessment

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 5.027

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, The Cochrane Library and the ISI Science Citation Index.

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: nihredit@southampton.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

The HTA programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

For more information about the HTA programme please visit the website: http://www.nets.nihr.ac.uk/programmes/hta

This report

The research reported in this issue of the journal was commissioned and funded by the HTA programme on behalf of NICE as project number 13/46/01. The protocol was agreed in June 2014. The assessment report began editorial review in January 2015 and was accepted for publication in April 2015. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research funded by the National Institute for Health Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).

DOI: 10.3310/hta20090 HEALTH TECHNOLOGY ASSESSMENT 2016 VOL. 20 NO. 9

Abstract

Tumour necrosis factor-α inhibitors for ankylosing spondylitis and non-radiographic axial spondyloarthritis: a systematic review and economic evaluation

Mark Corbett,[1] Marta Soares,[2] Gurleen Jhuti,[2] Stephen Rice,[1] Eldon Spackman,[2] Eleftherios Sideris,[2] Thirimon Moe-Byrne,[1] Dave Fox,[1] Helena Marzo-Ortega,[3] Lesley Kay,[3] Nerys Woolacott[1]* and Stephen Palmer[2]

[1]Centre for Reviews and Dissemination, University of York, York, UK
[2]Centre for Health Economics, University of York, York, UK
[3]Division of Rheumatic and Musculoskeletal Disease, Chapel Allerton Hospital, Leeds Teaching Hospitals NHS Trust and University of Leeds, Leeds, UK
[4]Department of Rheumatology, Freeman Hospital, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK

*Corresponding author nerys.woolacott@york.ac.uk

Background: Tumour necrosis factor (TNF)-α inhibitors (anti-TNFs) are typically used when the inflammatory rheumatologic diseases ankylosing spondylitis (AS) and non-radiographic axial spondyloarthritis (nr-AxSpA) have not responded adequately to conventional therapy. Current National Institute for Health and Care Excellence (NICE) guidance recommends treatment with adalimumab, etanercept and golimumab in adults with active (severe) AS only if certain criteria are fulfilled but it does not recommend infliximab for AS. Anti-TNFs for patients with nr-AxSpA have not previously been appraised by NICE.

Objective: To determine the clinical effectiveness, safety and cost-effectiveness within the NHS of adalimumab, certolizumab pegol, etanercept, golimumab and infliximab, within their licensed indications, for the treatment of severe active AS or severe nr-AxSpA (but with objective signs of inflammation).

Design: Systematic review and economic model.

Data sources: Fifteen databases were searched for relevant studies in July 2014.

Review methods: Clinical effectiveness data from randomised controlled trials (RCTs) were synthesised using Bayesian network meta-analysis methods. Results from other studies were summarised narratively. Only full economic evaluations that compared two or more options and considered both costs and consequences were included in the systematic review of cost-effectiveness studies. The differences in the approaches and assumptions used across the studies, and also those in the manufacturer's submissions, were examined in order to explain any discrepancies in the findings and to identify key areas of uncertainty. A de novo decision model was developed with a generalised framework for evidence synthesis that pooled change in disease activity (BASDAI and BASDAI 50) and simultaneously synthesised information on function (BASFI) to determine the long-term quality-adjusted life-year and cost burden of the disease in the economic model. The decision model was developed in accordance with the NICE reference case. The model has a lifetime horizon (60 years) and considers costs from the perspective of the NHS and personal social services. Health effects were expressed in terms of quality-adjusted life-years.

ABSTRACT

Results: In total, 28 eligible RCTs were identified and 26 were placebo controlled (mostly up to 12 weeks); 17 extended into open-label active treatment-only phases. Most RCTs were judged to have a low risk of bias overall. In both AS and nr-AxSpA populations, anti-TNFs produced clinically important benefits to patients in terms of improving function and reducing disease activity; for AS, the relative risks for ASAS 40 ranged from 2.53 to 3.42. The efficacy estimates were consistently slightly smaller for nr-AxSpA than for AS. Statistical (and clinical) heterogeneity was more apparent in the nr-AxSpA analyses than in the AS analyses; both the reliability of the nr-AxSpA meta-analysis results and their true relevance to patients seen in clinical practice are questionable. In AS, anti-TNFs are approximately equally effective. Effectiveness appears to be maintained over time, with around 50% of patients still responding at 2 years. Evidence for an effect of anti-TNFs delaying disease progression was limited; results from ongoing long-term studies should help to clarify this issue. Sequential treatment with anti-TNFs can be worthwhile but the drug survival response rates and benefits are reduced with second and third anti-TNFs. The de novo model, which addressed many of the issues of earlier evaluations, generated incremental cost-effectiveness ratios ranging from £19,240 to £66,529 depending on anti-TNF and modelling assumptions.

Conclusions: In both AS and nr-AxSpA populations anti-TNFs are clinically effective, although more so in AS than in nr-AxSpA. Anti-TNFs may be an effective use of NHS resources depending on which assumptions are considered appropriate.

Future work recommendations: Randomised trials are needed to identify the nr-AxSpA population who will benefit the most from anti-TNFs.

Study registration: This study is registered as PROSPERO CRD42014010182.

Funding: The National Institute for Health Research Health Technology Assessment programme.

Contents

List of tables

List of figures

List of boxes

Glossary

Adverse effect An abnormal or harmful effect caused by, and attributable to, exposure to a chemical (e.g. a drug), which is indicated by some result such as death, a physical symptom or visible illness. An effect may be classed as adverse if it causes functional or anatomical damage, causes irreversible change in the homeostasis of the organism or increases the susceptibility of the organism to other chemical or biological stress.

Ankylosing spondylitis A rheumatic disease that affects the spine and may lead to some degree of stiffness in the back. As the inflammation reduces and healing takes place, bone grows out from both sides of the vertebrae and may join the two together; this stiffening is called ankylosis. If definite, changes to spinal and/or pelvic joints are present on plain radiographs.

Articular Of or relating to the joints.

Axial spondyloarthritis Refers to a form of arthritis in which the predominant symptom is back pain due to inflammation of spinal and/or pelvic joints. If definite changes on plain radiographs are present, the disease is classified as ankylosing spondylitis, but if they are absent the disease is classified as non-radiographic axial spondyloarthritis. Further tests may indicate that in some patients it is very likely that non-radiographic axial spondyloarthritis is ankylosing spondylitis, only at an earlier stage of disease.

Between-study variance Between-study variance is a measure of statistical heterogeneity that depends on the scale of the outcome measured. It represents the variation in reported study effects over and above the variation expected given the within-study variation.

Biologic therapies (synonym: biological) Medical preparations derived from living organisms. Includes anti-tumour necrosis factor drugs and other new drugs which target pathologically active T cells.

Biosimilar An imitation biological medical product (such as an anti-tumour necrosis factor) usually marketed by a different manufacturer to the original biological product, once a patent has expired. The biosimilar should be similar to the original licensed product in terms of safety and efficacy.

C-reactive protein Concentrations of this protein in the blood can be measured as a test of inflammation or disease activity, for example in ankylosing spondylitis and non-radiographic axial spondyloarthritis.

Corticosteroid A synthetic hormone, similar to that produced naturally by the adrenal glands, which is available in pill, topical and injectable forms.

Cost–benefit analysis An economic analysis that converts the effects or consequences of interventions into the same monetary terms as the costs and compares them using a measure of net benefit or a cost–benefit ratio.

Cost-effectiveness analysis An economic analysis that expresses the effects or consequences of interventions on a single dimension. This would normally be expressed in 'natural' units (e.g. cases cured, life-years gained or additional strokes prevented). The difference between interventions in terms of costs and effects is typically expressed as an incremental cost-effectiveness ratio (e.g. the incremental cost per life-year gained).

Cost–utility analysis The same as a cost-effectiveness analysis but the effects or consequences of interventions are expressed in generic units of health gain, usually quality-adjusted life-years.

Credible interval In Bayesian statistics, a credible interval is a posterior probability interval estimation which incorporates problem-specific contextual information from the prior distribution. Credible intervals are used for the purposes similar to those of confidence intervals in frequentist statistics.

Disease-modifying antirheumatic drugs Disease-modifying antirheumatic drugs are drugs capable of modifying the progression of rheumatic disease. The term is, however, applied to what are now considered to be traditional disease-modifying drugs, in particular sulphasalazine, methotrexate and ciclosporin, as well as azathioprine, cyclophosphamide, antimalarials, penicillamine and gold. The newer agent leflunomide may be included as a disease-modifying antirheumatic drug. The biologics such as etanercept and infliximab are not generally referred to as disease-modifying antirheumatic drugs.

Erythrocyte sedimentation rate One of the tests designed to measure the degree of inflammation.

Fixed-effect model A statistical model that stipulates that the units under analysis (e.g. people in a trial or study in a meta-analysis) are the ones of interest, and thus constitute the entire population of units. Only within-study variation is taken to influence the uncertainty of results (as reflected in the confidence interval) of a meta-analysis using a fixed-effect model.

Heterogeneity In systematic reviews heterogeneity refers to variability or differences between studies in the estimates of effects. A distinction is sometimes made between 'statistical heterogeneity' (differences in the reported effects), 'methodological heterogeneity' (differences in study design) and 'clinical heterogeneity' (differences between studies in key characteristics of the participants, interventions or outcome measures).

***I*-squared (I^2)** A measure of 'statistical heterogeneity' (differences in the reported effects). It varies between 0 and 1, for which 0 indicates that the differences in reported effects are entirely consistent with the within-study uncertainty and 1 indicates that the differences in reported effects are entirely explained by study characteristics that vary across studies.

Intention to treat An intention-to-treat analysis is one in which all the participants in a trial are analysed according to the intervention to which they were allocated, whether or not they received it.

Monoclonal antibody An antibody produced in a laboratory from a single clone that recognises only one antigen.

Non-radiographic axial spondyloarthritis Axial spondyloarthritis for which definite changes to spinal and/or pelvic joints on plain radiographs are not present. Further tests may indicate that in some patients it is very likely that non-radiographic axial spondyloarthritis ankylosing spondylitis, only at an earlier stage of disease.

Non-steroidal anti-inflammatory drugs Consists of a large range of drugs of the aspirin family, prescribed for different kinds of arthritis which reduce inflammation and control pain, swelling and stiffness.

Open-label study A type of study in which both participants and researchers know which treatment is being administered.

Placebo An inactive substance or procedure administered to a patient, usually to compare its effects with those of a real drug or other intervention but sometimes for the psychological benefit to the patient through a belief that s/he is receiving treatment.

Quality-adjusted life-year An index of health gain in which survival duration is weighted or adjusted by the patient's quality of life during the survival period. Quality-adjusted life-years have the advantage of incorporating changes in both quantity (mortality) and quality (morbidity) of life.

Quality of life A concept incorporating all the factors that might impact on an individual's life, including factors such as the absence of disease or infirmity as well as other factors which might affect their physical, mental and social well-being.

Random-effects model A statistical model sometimes used in meta-analysis in which both within-study sampling error (variance) and between-studies variation are included in the assessment of the uncertainty (confidence interval) of the results of a meta-analysis.

Randomised controlled trial (synonym: randomised clinical trial) An experiment in which investigators randomly allocate eligible people into intervention groups to receive, or not to receive, one or more interventions that are being compared.

Relative risk (synonym: risk ratio) The ratio of risk in the intervention group to the risk in the control group. The risk (proportion, probability or rate) is the ratio of people with an event in a group to the total in the group. A relative risk of one indicates no difference between comparison groups. For undesirable outcomes, a relative risk that is less than one indicates that the intervention was effective in reducing the risk of that outcome.

Sensitivity analysis An analysis used to determine how sensitive the results of a study or systematic review are to changes in how it was done. Sensitivity analyses are used to assess how robust the results are to uncertain decisions or assumptions about the data and the methods that were used.

Tumour necrosis factor One of the cytokines, or messengers, known to be involved in the process of systemic inflammation.

Weighted mean difference (in meta-analysis) A method of meta-analysis used to combine measures on continuous scales, where the mean, standard deviation and sample size in each group are known. The weight given to each study is determined by the precision of its estimate of effect and is equal to the inverse of the variance. This method assumes that all of the trials have measured the outcome on the same scale.

List of abbreviations

AE	adverse event
AiC	academic in confidence
AS	ankylosing spondylitis
ASAS	Assessment in Ankylosing Spondylitis
ASDAS	Ankylosing Spondylitis Disease Activity Score
ASQoL	Ankylosing Spondylitis Quality of Life
ASSERT	Ankylosing Spondylitis Study for the Evaluation of Recombinant Infliximab Therapy
ATLAS	Adalimumab Trial Evaluating Long-term Efficacy and Safety for Ankylosing Spondylitis
ATP	adalimumab target population
axSpA	axial spondyloarthritis
BASDAI	Bath Ankylosing Spondylitis Disease Activity Index
BASFI	Bath Ankylosing Spondylitis Functional Index
BASMI	Bath Ankylosing Spondylitis Metrology Index
BSR	British Society for Rheumatology
BSRBR	British Society for Rheumatology Biologics Register
CC	conventional care
CI	confidence interval
CiC	commercial in confidence
CrI	credible interval
CRP	C-reactive protein
DES	discrete event simulation
DIC	deviance information criterion
DMARD	disease-modifying antirheumatic drug
DSU	Decision Support Unit
EQ-5D	European Quality of Life-5 Dimensions
ERG	Evidence Review Group
ESR	erythrocyte sedimentation rate
FDA	Food and Drug Administration
GESPIC	German Spondyloarthritis Inception Cohort
HLA	human leucocyte antigen
HRG	Healthcare Resource Group
HRQoL	health-related quality of life
HUI	Health Utilities Index
ICER	incremental cost-effectiveness ratio
ITT	intention to treat
LOCF	last observation carried forward
LRiG	Liverpool Reviews and Implementation Group
MAIC	match-adjusted indirect comparison
MASES	Maastricht Ankylosing Spondylitis Enthesitis Score
MCID	minimum clinically important difference
mCMC	Markov chain Monte Carlo
MCS	mental component summary
mITT	modified intention to treat
MRI	magnetic resonance imaging
mSASSS	modified Stoke Ankylosing Spondylitis Spinal Score
MTC	mixed treatment comparison
NHS EED	NHS Economic Evaluation Database
NICE	National Institute for Health and Care Excellence
NNH	number needed to harm
nr-AxSpA	non-radiographic axial spondyloarthritis

NRI	non-responder imputation
NSAID	non-steroidal anti-inflammatory drug
OASIS	Outcomes in Ankylosing Spondylitis International Study
OMERACT	Outcome Measures in Rheumatology
OR	odds ratio
PAS	patient access scheme
PCS	physical component summary
PSA	probabilistic sensitivity analysis
PSS	personal social services
PSSRU	Personal Social Services Research Unit
QALY	quality-adjusted life-year
RCT	randomised controlled trial
RLDQ	Revised Leeds Disability Questionnaire
SAE	serious adverse event
SD	standard deviation
SE	standard error
SF-36	Short Form questionnaire-36 items
SIRAS	Scotland and Ireland Registry for Ankylosing Spondylitis
SMART	St Mary RheumaToid Arthritis
SMR	standardised mortality ratio
SpA	spondyloarthritis
TA	technology appraisal
TNF	tumour necrosis factor
VAS	visual analogue scale

Note

This monograph is based on the Technology Assessment Report produced for NICE. The full report contained a considerable number of data that were deemed academic-in-confidence and commercial-in-confidence. The full report was used by the Appraisal Committee at NICE in their deliberations. The full report with each piece of academic-in-confidence data and commercial-in-confidence removed and replaced by the statement 'academic-in-confidence and/or commercial-in-confidence information (or data) removed' is available on the NICE website: www.nice.org.uk.

The present monograph presents as full a version of the report as is possible while retaining readability, but some sections, sentences, tables and figures have been removed. Readers should bear in mind that the discussion, conclusions and implications for practice and research are based on all the data considered in the original full NICE report.

Plain English summary

Axial spondyloarthritis is a progressive form of arthritis which causes severe back pain because of inflammation of spinal and/or pelvic joints. If definite changes on plain radiographs are present, the disease is classified as ankylosing spondylitis (AS), but if they are absent the disease is classified as non-radiographic axial spondyloarthritis (nr-AxSpA). Usual therapy includes anti-inflammatory drugs, exercise and physiotherapy. Tumour necrosis factor inhibitors (also known as anti-TNFs) are typically used when the disease has not responded adequately to this.

This project systematically reviewed the evidence on five anti-TNF treatments (adalimumab, certolizumab pegol, etanercept, golimumab and infliximab), for treating severe active AS or nr-AxSpA. The objective of this project was to assess the benefits and adverse effects of these anti-TNFs and to run an economic model using both response to treatment and the impact of disease progression, to evaluate if their use to treat these patients is a cost-effective use of NHS resources.

In total, 28 eligible randomised controlled trials were identified and 26 were placebo controlled (most of the trials which used a placebo did so for no more than 12 weeks); the majority were good quality and 17 were extended into active treatment-only phases. In both AS and nr-AxSpA populations, anti-TNFs produced clinically important benefits to patients in terms of improving function and reducing disease activity. The benefit of treatment was consistently slightly smaller for nr-AxSpA than for AS. In AS the different anti-TNFs are approximately equally effective and effectiveness appears to be maintained over time. The results of the economic model indicated that anti-TNFs may be an effective use of NHS resources depending on which assumptions are considered appropriate.

Scientific summary

Background

Spondyloarthritis encompasses a heterogeneous group of inflammatory rheumatologic diseases. Spondyloarthritis can be categorised as having predominantly axial or peripheral involvement. In people with axial spondyloarthritis (axSpA), the predominant symptoms are back pain and stiffness developed before age 45 years. For axSpA patients to be classified as having ankylosing spondylitis (AS), imaging evidence of joint damage using radiography is required. Patients with non-radiographic axial spondyloarthritis (nr-AxSpA) may, or may not, have signs of sacroiliac joint inflammation on a magnetic resonance image. The use of magnetic resonance imaging allows for earlier detection of axSpA, as joint damage may not become evident on radiographs for many years. Progression of axSpA is difficult to predict.

Tumour necrosis factor (TNF)-α inhibitors (anti-TNFs) are typically used when the disease has not responded adequately to conventional therapy. Current National Institute for Health and Care Excellence (NICE) guidance recommends treatment with adalimumab, etanercept and golimumab in adults with active (severe) AS only if certain criteria are fulfilled, but it does not recommend infliximab for AS. Anti-TNFs for patients with nr-AxSpA have not previously been appraised by NICE.

Objectives

To determine the clinical effectiveness, safety and cost-effectiveness within the NHS of adalimumab, certolizumab pegol, etanercept, golimumab and infliximab, within their respective licensed indications, for the treatment of severe active AS or severe nr-AxSpA (but with objective signs of inflammation).

Methods

For the systematic review of clinical efficacy, randomised controlled trials (RCTs) were eligible, including any open-label extensions. Adverse events data were sought from existing reviews of anti-TNFs used in any disease and from other appropriately large studies. For studies of natural history, long-term effectiveness, adherence and sequential use, published analyses based on large and long-term data sets (registry data) were eligible. Eligible studies were of adults with either severe active AS or severe nr-AxSpA but with objective signs of inflammation. The treatments of interest were adalimumab, certolizumab pegol, etanercept, golimumab, infliximab or any of their biosimilars. The relevant comparators were conventional management strategies (either with or without placebo) and alternative anti-TNFs. Key outcomes included multiple domain response criteria [such as Assessment in Ankylosing Spondylitis (ASAS) 40] and measures of disease activity [Bath Ankylosing Spondylitis Disease Activity Index (BASDAI)] and function [Bath Ankylosing Spondylitis Functional Index (BASFI)].

Fifteen databases were searched for relevant studies in July 2014. Clinical effectiveness data from RCTs were synthesised using Bayesian network meta-analysis methods. Sensitivity analyses were performed in which trials at risk of bias were excluded. Results from other studies were summarised narratively.

A systematic review of cost-effectiveness studies was undertaken to assess the relevance of existing data from the perspective of the NHS. Three databases were searched. Only full economic evaluations that compared two or more options and considered both costs and consequences were included. The

differences in the approaches and assumptions used across the studies were examined in order to explain any discrepancies in the findings and to identify key areas of uncertainty. A separate review of the manufacturer's submissions was also undertaken and the findings were compared with those found in the review of previously published studies.

The findings from the clinical effectiveness and cost-effectiveness reviews were used to inform the development of a de-novo decision model to assess the cost-effectiveness of the alternative anti-TNFs in accordance with their licences for the separate indications. We developed a generalised framework for evidence synthesis that pools evidence on the change in BASDAI by considering those studies that report this measure directly and also those that report the proportion of patients achieving a BASDAI 50 response (a $\geq$ 50% improvement in BASDAI score). We expressed BASDAI 50 as a function of the absolute change in BASDAI and we used this relationship in the extended synthesis. We also aimed to simultaneously synthesise information on BASFI (function) score, a measure that is used together with the BASDAI score to determine the long-term quality-adjusted life-year (QALY) and cost burden of the disease in the economic model. The decision model was a cohort model structured as a modified decision tree tracking response at 12 weeks and treatment failure at subsequent time points within the time horizon. These events determine changes in BASDAI and BASFI scores, which are further used to define costs and utilities. The model considers the independent effects on BASFI as a result of disease activity (BASDAI) and the extent and progression of radiographic disease (as measured by the modified Stoke Ankylosing Spondylitis Spinal Score). The model was developed in accordance with the NICE reference case. The model has a lifetime horizon (60 years) and considers costs from the perspective of the NHS and personal social services. Health effects were expressed in terms of QALYs.

Results

Clinical efficacy from randomised controlled trials

Twenty-eight eligible RCTs were identified, with 24 being suitable for data synthesis. All but seven of the trials were extended into open-label active treatment-only phases. Most RCTs were judged to have a low risk of bias overall.

For the AS population, the 10- to 16-week data showed consistent effects across the different anti-TNFs when compared with placebo: for ASAS 20 the pooled relative risks ranged from 1.80 (certolizumab pegol) to 2.45 (infliximab); for the ASAS 40 data the relative risks ranged from 2.53 (certolizumab pegol) to 3.42 (adalimumab) and for BASDAI 50 the relative risks ranged from 3.16 (adalimumab) to 4.86 (infliximab). Adalimumab, certolizumab pegol, etanercept and infliximab produced statistically significant and clinically important reductions in disease activity, with BASDAI reductions ranging from 1.46 units (certolizumab pegol) to 2.28 units (infliximab), and function, with BASFI reductions ranging from 1.1 units (certolizumab pegol) to 2.16 units (infliximab).

When analysed as a class, anti-TNFs were statistically significantly more likely than placebo to result in patients with AS achieving an ASAS 20 response (relative risk 2.21), an ASAS 40 response (relative risk 3.06), and a BASDAI 50 response (relative risk 3.37). They also produced statistically significant improvements (calculated using mean difference in change from baseline) in disease activity (BASDAI mean difference –1.66 units) and in function (BASFI mean difference –1.38 units). There was little evidence of statistical heterogeneity for the key outcomes (ASAS outcomes, BASFI, BASDAI and BASDAI 50) but substantial heterogeneity was seen for other outcomes. Results of the sensitivity analyses performed for the AS studies were very similar to the main analyses.

For the nr-AxSpA population, five RCTs were included. When anti-TNFs were considered as a class, statistically significant improvements were found for ASAS 20 (relative risk 1.65); ASAS 40 (relative risk 2.74); BASDAI 50 (relative risk 2.31); BASDAI (mean difference –1.32 units); and BASFI (mean difference

–0.99 units). For the disease activity, function and responder outcomes, these common class efficacy estimates were consistently slightly smaller for nr-AxSpA than for AS, most noticeably for BASFI and BASDAI 50. Statistical heterogeneity (when such estimates could be calculated) was apparent in the nr-AxSpA analyses.

Long-term efficacy

For AS, across all the anti-TNFs, after around 2 years and 5 years of treatment, roughly half of patients were still achieving a good level of response to therapy. However, the long-term studies produced less reliable data than the RCTs. Fewer studies were available of nr-AxSpA patients, although the results were broadly similar to those of the AS studies.

Evidence for an effect of anti-TNFs on radiographic disease progression was limited; the relatively short-term follow-up available to date and the insensitivity of radiography as an imaging tool precluded the drawing of firm conclusions regarding the role of anti-TNFs in preventing or delaying the progression of AS. There are some data to suggest an identifiable benefit from around 4 years, but results from ongoing long-term studies should help to clarify this issue.

Registry data demonstrate that around 60% of patients with AS treated with a first anti-TNF will still be on treatment at 2 years. Sequential treatment with anti-TNFs can be worthwhile but the drug survival response rates and benefits are reduced with second and third anti-TNFs, with the proportion of BASDAI 50 responders falling approximately 10% with each subsequent anti-TNF and the median BASDAI and BASFIs achieved increasing (worsening).

Adverse effects

Data from large systematic reviews, which included patients with a wide range of diseases, suggest that, in the short term, anti-TNFs as a group are associated with significantly higher rates of serious infections, tuberculosis reactivation, non-melanoma skin cancer, total adverse events (AEs) and withdrawals because of AEs than control treatments. Specifically, infliximab is associated with significantly higher rates of total AEs and withdrawals because of AEs and certolizumab pegol is associated with significantly higher rates of serious infections and serious AEs. The available open-label data on AEs were limited by the small sample sizes and non-randomised study designs.

Cost-effectiveness reported in existing published studies and manufacturer's submissions

A total of six UK studies reporting on the cost-effectiveness of anti-TNFs were identified, all for the treatment of AS. There appear marked differences between the results of the previously published industry-funded assessments in AS and the results reported in a previous independent assessment. Although all models reviewed used changes in BASDAI and/or BASFI to quantitatively model the short- and longer-term costs and quality-of-life effects, there appeared significant variation in the assumptions employed. We identified important conceptual issues with all existing models relating to the subsequent projection of BASDAI and BASFI scores over a longer time horizon.

Manufacturers submitted de novo analyses for both AS (AbbVie, UCB, Pfizer, Merck Sharp & Dohme) and nr-AxSpA (AbbVie, UCB, Pfizer) populations. Despite the different model structures and assumptions applied across the various manufacturer's submissions, the incremental cost-effectiveness ratios (ICERs) reported for the anti-TNFs versus conventional care (CC) appeared consistent in AS. Across the separate base-case analyses, the ICERs ranged from £16,391 to £44,448 for the alternative anti-TNFs compared with CC alone. Infliximab was routinely reported to have the highest ICER. When infliximab was excluded from consideration, the ICERs ranged from £16,391 to £21,972 for the other anti-TNFs.

The differences in structural and parameter assumptions appear more evident in the cost-effectiveness results for the nr-AxSpA population. The ICERs for adalimumab, certolizumab and etanercept ranged between £12,866 and £50,692 per QALY. Importantly, when the results in the separate populations were compared, no consistent relationship appeared to emerge across the manufacturer's submissions regarding the cost-effectiveness on anti-TNFs in AS compared with the nr-AxSpA population. In addition, many of the same conceptual concerns identified from the review of published cost-effectiveness studies were also still evident.

An independent model was developed to address the conceptual concerns and areas of remaining uncertainty. Although it shared several of the assumptions and parameter estimates from the manufacturer models, it has a different conceptual structure (linking BASFI progression to evidence from radiographic assessments) and applies a more generalised framework for the synthesis of clinical-effectiveness data. The extended synthesis approach showed the effectiveness of the different anti-TNFs to be similar. Consequently, the treatment effects for the anti-TNFs were assumed to come from a 'common' distribution, that is a 'class effect'. We developed a simulation model that allowed prediction of the conditional change scores for responders/non-responders to BASDAI 50 at 12 weeks and to explore differences in the baseline BASDAI/BASFI scores according to response status.

Base-case cost-effectiveness results were presented for two alternative 'rebound' assumptions. In the rebound equal to gain scenario, the ICER of the alternative anti-TNFs varied between £19,240 [certolizumab with the proposed patient access scheme (PAS)] to £40,467 per additional QALY (infliximab) in AS patients. In the rebound to CC scenario, the ICER of the alternative anti-TNFs varied between £33,762 (certolizumab with the proposed PAS) to £66,529 per additional QALY (infliximab) in AS patients.

In the rebound equal to gain scenario, the ICER of the alternative anti-TNFs for nr-AxSpA patients varied between £28,247 (certolizumab with the proposed PAS) to £29,784 per additional QALY (etanercept) in AS patients. In the rebound to CC scenario, the ICER of the alternative anti-TNFs for nr-AxSpA patients varied between £32,528 (certolizumab with the proposed PAS) to £34,232 (etanercept) per additional QALY.

Discussion

The key strengths of the systematic review are the rigorous methods used and the extensive breadth of the types of study included. The York model confers several advantages over current cost-effectiveness studies by linking changes in function to a more explicit clinical/biological process and facilitating a more formal consideration of the potential impact of anti-TNFs on function, via the specific effects these drugs have on the different processes which independently relate to this parameter.

The meta-analysis results derived from a substantial and generally high-quality evidence base demonstrated that anti-TNFs produce clinically important benefits to AS patients in terms of improved function and reduced disease activity following around 3 months of treatment with an anti-TNF. Smaller benefits were seen across outcomes in patients with nr-AxSpA, which was a more heterogeneous population. Less reliable data were available on long-term efficacy, although it appears that around half of patients still achieve a good level of response after around 2 years of treatment.

Although there are a number of important differences in approaches both among the different manufacturer models and compared with the York model, the comparison of ICERs based on the York rebound equal to gain scenario appears broadly consistent with that reported by the manufacturers in both populations.

Conclusions

- In both AS and nr-AxSpA populations anti-TNFs produce clinically important benefits to patients in terms of improving function and reducing disease activity. The efficacy estimates were consistently slightly smaller for nr-AxSpA than for AS.
- Statistical (and clinical) heterogeneity was more apparent in the nr-AxSpA analyses than in the AS analyses; both the reliability of the nr-AxSpA meta-analysis results and their true relevance to patients seen in clinical practice are questionable.
- In AS anti-TNFs can be assumed to have a class effect, with the treatments being equally effective.
- Effectiveness appears to be maintained over time in about 50% of patients at 2 years.
- Evidence for an effect of anti-TNFs delaying disease progression was limited; results from ongoing long-term studies should help to clarify this issue.
- Sequential treatment with anti-TNFs can be worthwhile but the drug survival response rates and benefits are reduced with second and third anti-TNFs.
- The de novo model, which had addressed many of the issues of earlier evaluations, generated ICERs ranging from £19,240 to £66,529 depending on anti-TNF and modelling assumptions.

Suggested research priorities

Randomised trials are needed to identify the nr-AxSpA population that will benefit the most from anti-TNFs. Long-term studies are needed to clarify the effect of anti-TNFs on the progression of structural damage in AS and to help clarify the characteristics of nr-AxSpA patients who go on to develop AS. Studies are also needed to better inform the efficacy estimates relating to sequential use of anti-TNFs.

Study registration

This study is registered as PROSPERO CRD42014010182.

Funding

Funding for this study was provided by the Health Technology Assessment programme of the National Institute for Health Research.

Chapter 1 Background

Description of health problem

Spondyloarthritis (SpA) encompasses a heterogeneous group of inflammatory rheumatologic diseases including ankylosing spondylitis (AS), reactive arthritis, psoriatic arthritis, inflammatory bowel disease-related arthritis and undifferentiated SpA.[1] SpA can be categorised as having predominantly axial (sacroiliac joints or spine) or peripheral involvement. In people with axial spondyloarthritis (axSpA), the predominant symptom is back pain (due to inflammation of the sacroiliac joints, the spine, or both) but there may also be extra-articular and peripheral joint manifestations.

In practice, and in clinical trials, AS is commonly diagnosed using the modified New York criteria (*Box 1*); sometimes in practice radiography may not be performed routinely (because of the radiation doses involved) or magnetic resonance imaging (MRI) may be preferred as a diagnostic tool. The recently developed Assessment of SpondyloArthritis International (ASAS) Society classification criteria encompass a broad range of patients with axSpA, including patients with AS and patients with non-radiographic axial spondyloarthritis (nr-AxSpA).[3] All axSpA patients will have developed chronic back pain ($\geq$ 3 months) before age 45 years. Classifications can be made using the imaging or clinical arms of the criteria. The imaging arm requires evidence of joint damage (erosions or fusion) due to sacroiliitis, using either radiography (when the disease is classified as AS) or MRI (when the disease is classified as nr-AxSpA);[4] additionally, at least one of the following SpA features is also required: inflammatory back pain, arthritis, enthesitis (heel), uveitis, dactylitis, psoriasis, Crohn's/colitis, good response to non-steroidal anti-inflammatory drugs (NSAIDs), family history of SpA, human leucocyte antigen (HLA)-B27 genetic marker, and elevated C-reactive protein (CRP). People with axSpA often have the genetic marker HLA-B27. To be classified as having axSpA via the clinical arm of the criteria, patients must be HLA-B27 positive and also have at least three of the aforementioned SpA features.

The use of MRI allows for earlier detection of axSpA, as joint damage may not become evident on radiography for many years. Patients with nr-AxSpA may, or may not, have signs of sacroiliac joint inflammation on a magnetic resonance image. There may be other objective signs of inflammation such as an abnormally raised erythrocyte sedimentation rate (ESR) or CRP level, although these are less sensitive and specific for AS. A MRI diagnosis may therefore provide the opportunity for treatment to reduce the possibility of long-term structural damage (and associated burden of symptoms).[5] However, there is some concern that the diagnostic criteria for nr-AxSpA may be too liberal and may include patients who do not have axSpA and will never progress to AS, particularly with respect to patients who are diagnosed without evidence of imaging (MRI) changes.[6–8] The differences between AS and nr-AxSpA are explored further in *Chapter 3*.

BOX 1 Modified New York criteria for AS (1984)[2]

Clinical criteria

- Low back pain and stiffness for more than 3 months that improves with exercise but is not relieved by rest.
- Limitation of motion of the lumbar spine in the sagittal and frontal planes.
- Limitation of chest expansion relative to normal values correlated for age and sex.

Radiological criterion

- Sacroiliitis grade > 2 bilaterally or grade 3–4 unilaterally.
- Definite AS if the radiological criterion is associated with at least one clinical criterion.

Prognosis

Axial spondyloarthritis is a painful, progressive form of inflammatory arthritis. It mainly affects the spine but can also affect other joints, tendons and ligaments. Other areas such as the eyes and bowel can also sometimes be involved in non-radiographic and radiographic (AS) forms of axSpA.[9] The pain and stiffness of axSpA adversely affects optimal daily functioning. These symptoms are a result of a combination of reversible components of the disease, such as inflammation and flares, and irreversible components, such as syndesmophytes and vertebral bridging (bony deposition).[10] Most patients with AS develop the first symptoms at 25–45 years of age.[11] Progression of the disease is variable and difficult to predict.[12] There is often a delay of many years between patients first noticing symptoms and the diagnosis of axSpA being received. Many people with axSpA have AS, with evidence of bony deposition as well as inflammation. In later-stage AS, joints and bones may fuse together, a process that can occur over a long period of time and cause restricted movement. The functional impairment because of inflammation and/or bony deposition can have a profound effect on health and quality of life, and lead to withdrawal from active employment, with resultant adverse financial consequences; the burden of disease is greater in more socially deprived patients.[13] The prognosis is poor, although there is some evidence that deterioration plateaus in well-established AS.[14] Paradoxically, early disease (nr-AxSpA) may be less readily diagnosed and patients offered fewer treatment options even though it can be as, or even more, debilitating that established AS.[15]

Ankylosing spondylitis is associated with an increased risk of death; it is estimated that patients have a standardised mortality ratio (SMR) of ≥ 1.5. The increased risk appears to be greater in men, with one study reporting a statistically significant increase in SMR of 1.63 in men but no significant increase in women (SMR 1.38) with AS.[16] This study found that, after correcting for age, sex, disease duration and pre-existing cardiovascular disease, independent predictors of increased mortality were elevated CRP level, diagnostic delay, not using NSAIDs and work disability. According to British Society for Rheumatology (BSR) guidelines, the excess mortality is mainly accounted for by cardiac valvular disease, amyloidosis and fractures.[17] nr-AxSpA affects approximately equal numbers of men and women, but it is more likely that men will develop AS.[18]

Epidemiology

Currently, only limited epidemiological data are available for axSpA defined according to ASAS criteria. For AS, the prevalence is thought to be around 0.25% in European populations.[19] It is around three times more common in men than in women.[20] A recent study published in the USA reported an estimated AS prevalence of 0.52–0.55%, and the prevalence of axSpA as approximately 1.0–1.4%.[21] The proportion of nr-AxSpA among patients with axSpA is estimated to be between 20% and 80%.[22] Each year in the UK an estimated 2% of patients in a general practice will present with back pain and up to 5% of these will show features of AS.[23]

Measurement of disease

There are a number of components and measures of disease activity in axSpA;[24] a patient's health-related quality of life (HRQoL) is determined by both by physical functioning and by disease activity. In turn, physical function is determined by spinal mobility and disease activity, and spinal mobility is determined by structural damage and inflammation of the spine.[24] In nr-AxSpA, a patient may have significant inflammation but no detectable structural damage; in AS, a patient may have both significant inflammation and structural damage; and in late AS, there may be less inflammation but extensive structural damage.

The main tools used for the assessment of various components of the disease are listed in *Table 1*.

Placebo response

The term 'placebo effect' can be used to describe different types of 'effect' but it generally encompasses one or more of three different meanings. First, there is the temporal (before–after) change after placebo medication, in which the effects of a placebo intervention cannot be distinguished from the natural course of the disease or regression to the mean. Second, there is the causal effect of placebo intervention associated with the treatment ritual, and, finally, there is the effect of all the psychological processes

TABLE 1 Disease assessment tools

Assessment measures		
Tool	Disease component	Description
BASDAI	Disease activity	Consists of a 1–10 scale (1 being no problem and 10 being the worst problem) which is used to answer six questions pertaining to the five major symptoms of AS: • fatigue • spinal pain • joint pain/swelling • areas of localised tenderness (also called enthesitis or inflammation of tendons and ligaments) • morning stiffness duration • morning stiffness severity
BASFI	Functional ability	Patient assesses difficulty on a 10-point scale (1 is easy and 10 is impossible) for each of 10 items: • putting on your socks or tights without help or aids (e.g. sock aid) • bending from the waist to pick up a pen from the floor without aid • reaching up to a high shelf without help or aids (e.g. helping hand) • getting up from an armless chair without your hands or any other help • getting up off the floor without help from lying on your back • standing unsupported for 10 minutes without discomfort • climbing 12–15 steps without using a handrail or walking aid • looking over your shoulder without turning your body • doing physically demanding activities (e.g. physiotherapy exercises, gardening or sports) • doing a full day's activities whether it be at home or at work
BASMI	Disease activity, spinal mobility	Clinician assessment of cervical rotation, tragus-to-wall distance, lumbar side flexion, modified Schober's, intermalleolar distance
ASDAS	Disease activity	Calculated from BASDAI questions on spinal pain, peripheral arthritis and duration of morning stiffness, patients global assessment of disease activity and CRP level (or ESR if CRP level not available)
mSASSS	Structural damage	In the mSASSS the anterior vertebral corners of the cervical (lower border of C2 to upper border of T1) and lumbar (lower border of T12 to upper border of S1) segments (a total of 24 VCs) are scored at a lateral view, for the presence of erosion and/or sclerosis and/or squaring (1 point), syndesmophyte (2 points) and bridging syndesmophyte (3 points). The total score ranges from 0 to 72. The mSASSS has shown better reliability and sensitivity to change than other radiographic scoring methods[25]

continued

TABLE 1 Disease assessment tools (*continued*)

MRI assessments		
Measures of response		
BASDAI 50	Response criterion	≥ 50% improvement in BASDAI
ASAS 20	Response criterion	≥ 20% improvement and ≥ 1 unit absolute improvement (range 1–10) in three of four domains with no worsening of ≥ 20% improvement and ≥ 1 unit absolute in the fourth domain: BASFI, spinal pain, patient GDA and inflammation (BASDAI Q5 and 6)
ASAS 40	Response criterion	≥ 40% improvement and ≥ 2 units absolute improvement (range 1–10) in three of four domains with no worsening at all in the fourth domain: BASFI, spinal pain, patient GDA and inflammation (BASDAI Q5 and 6)
ASAS partial remission	Response criterion	A value of ≥ 2 units absolute improvement (range 1–10) in each of four domains: ASFI, spinal pain, patient GDA and inflammation (BASDAI Q5 and 6)
ASAS 5/6	Response criterion	Improvement in five out of six domains (using pre-defined % improvements) without deterioration in the sixth domain: pain, patient global assessment, function, inflammation, spinal mobility, CRP level
ASDAS major improvement	Response criterion	≥ 2 units improvement in ASDAS

ASDAS, Ankylosing Spondylitis Disease Activity Score; BASDAI, Bath Ankylosing Spondylitis Disease Activity Index; BASFI, Bath Ankylosing Spondylitis Functional Index; BASMI, Bath Ankylosing Spondylitis Metrology Index; GDA, Global Disease Activity; mSASSS, modified Stroke Ankylosing Spondylitis Spinal Score; VC, vertebral corner.

involved in the interaction between doctor and patient.[26] For the placebo-controlled trials in AS and nr-AxSpA these non-pharmacological components can be assumed to act equally in the anti-tumour necrosis factor (TNF) and placebo arms. Results from the placebo arms measure the non-pharmacological effects and the difference between the anti-TNF and placebo arms measures the pharmacological effect. All three components of the placebo effect could be important to consider when evaluating trials in this assessment, although once the trial treatment periods have ended, it is likely that the effect of the natural course of the disease becomes the most important factor of any 'placebo' effect. Estimated cost-effectiveness ratios and associated policy decisions may be sensitive to assumptions regarding the mechanism underlying placebo responses.[27]

The natural course of disease activity in AS is known to vary over time with exacerbations, or flares, being common. In a study of flares in patients with AS, clinically relevant changes in Bath Ankylosing Spondylitis Disease Activity Index (BASDAI; but not in function) were noted during minor/localised flares (which occurred in 59% of patients in any given week). Although major/generalised flares were less common (reported in 12% of patients in any given week) they were associated with clinically relevant changes in both disease activity and function.[28] Pain is a key component of BASDAI and the ASAS responder outcomes; a Cochrane systematic review of placebos for all clinical conditions found that placebo interventions can influence patient-reported outcomes, especially pain (and nausea).[29] The authors also concluded that it was difficult to distinguish patient-reported effects of placebo from biased reporting, and that the effect on pain varied from negligible to clinically important, even among trials with low risk of bias.

Current service provision

Management of disease

Short- and long-term treatment goals for axSpA include minimising pain and stiffness, maintaining function and posture, arresting disease progression and maintaining quality of life and ability to work. Current conventional therapy for axSpA includes acute anti-inflammatory treatment with NSAIDs and physiotherapy and exercise.

Conventional therapy for AS is limited to NSAIDs (despite very limited supporting clinical trial evidence)[30] and recommendations regarding appropriate physical activity. Other statements in the ASAS/EULAR (European League Against Rheumatism) recommendations for the management of AS include analgesics such as paracetamol and opioid-like drugs that may be considered for residual pain. Glucocorticoid injections into the direct site of inflammation (but not systemic) may be of benefit. The use of disease-modifying antirheumatic drugs (DMARDs, such as methotrexate and sulfasalazine) has been all but abandoned after evidence of lack of benefit. The cornerstone of non-pharmacological treatment of patients with AS is patient education and regular exercise; home exercises are effective. Physical therapy with supervised exercises, land- or water-based, individually or in a group, should be preferred, as these are more effective than home exercises. Patient associations and self-help groups may be useful. A Cochrane review of 11 trials concluded that the current best available evidence suggests that physiotherapy is beneficial for people with AS, but that it is still not clear which treatment protocol, duration and intensity should be recommended in the management of AS.[31] Physiotherapy is universally recommended[32] but variable in practice.

Biologic drugs are the only treatment shown to be efficacious in the treatment of symptoms and signs of disease activity in axSpA and AS. Current National Institute for Health and Care Excellence (NICE) and BSR guidance recommends treatment with the anti-TNFs adalimumab, etanercept and golimumab in adults with active (severe) AS only if certain criteria are fulfilled, but it does not recommend infliximab for AS.[17,33]

Description of technology under assessment

Tumour necrosis factor-α inhibitors (adalimumab, certolizumab pegol, etanercept, golimumab and infliximab), also referred to as anti-TNFs, are typically used when the disease has not responded adequately to conventional therapy. They target the activation of TNF-α and its subsequent activation of downstream inflammatory processes, and as such have the potential to offer symptom control as well as alter disease progression. Adalimumab, certolizumab pegol, golimumab and infliximab are monoclonal antibodies, whereas etanercept is a recombinant human TNF receptor fusion protein.

Adalimumab, etanercept, golimumab and infliximab are licensed in the UK for the treatment of adults with severe active AS that has responded inadequately to conventional therapy. Certolizumab pegol is licensed for the treatment of adults with severe active AS whose disease has responded inadequately to, or who are intolerant of, NSAIDs.

Adalimumab, etanercept and certolizumab pegol are also licensed for the treatment of adults with severe nr-AxSpA with objective signs of inflammation (including elevated CRP level and/or positive MRI), whose disease has responded inadequately to, or who are intolerant of, NSAIDs. Golimumab and infliximab do not currently have a UK marketing authorisation for nr-AxSpA. Current NICE guidance recommends treatment with adalimumab, etanercept or golimumab in adults with active (severe) AS only if certain criteria are fulfilled (including a stipulation that patients must have tried at least two different NSAIDs, which have failed to control symptoms), but it does not recommend infliximab for AS.[17,33] Anti-TNFs for patients with nr-AxSpA have not previously been appraised by NICE.

Chapter 2 Definition of decision problem

Decision problem in terms of Population, Intervention, Comparator, Outcome, Study design and other key issues

The decision problem relates to the optimal use of adalimumab, certolizumab pegol, etanercept, golimumab and infliximab, within their licensed indications, for the treatment of severe/active AS or severe axSpA without radiographic evidence of AS (but with objective signs of inflammation).

Previous National Institute for Health and Care Excellence appraisals

In the previous NICE technology appraisal (TA) 143, adalimumab, etanercept and infliximab were evaluated for AS, while in TA233[33] golimumab was evaluated for AS. A number of key areas of uncertainty and potential limitations of the evidence base were identified from these appraisals. These include:

1. a lack of direct head-to-head trial evidence evaluating the relative efficacy and safety of the TNF-α inhibitors
2. a lack of evidence on the efficacy and safety of the sequential use of TNF-α inhibitors
3. the long-term effectiveness of TNF-α inhibitors in controlling disease activity
4. the rate of disease progression in responders and non-responders to treatment, and in those on placebo
5. the proportion of patients who may experience a significant improvement in their condition without TNF-α inhibitor treatment
6. the rate of treatment withdrawal on TNF-α inhibitors and the degree to which a patient's condition might be expected to rebound if therapy is withdrawn
7. the adverse effects associated with the long-term use of TNF-α inhibitors
8. the impact of TNF-α inhibitors on the progression of structural damage in the spine and functional disability associated with ankylosis
9. the time horizon appropriate for considering the cost-effectiveness of TNF-α inhibitors
10. a lack of registry data of patients receiving TNF-α inhibitors for severe active AS.

This assessment would consider each of these areas of uncertainty and identify the relevant evidence available to inform the limitations of the previous appraisals.

Overall aims and objectives of assessment

The aim of the study is to determine the clinical effectiveness, safety and cost-effectiveness within the NHS of adalimumab, certolizumab pegol, etanercept, golimumab and infliximab, within their licensed indications, for the treatment of severe active AS or severe axSpA without radiographic evidence of AS (but with objective signs of inflammation). If evidence allows, the clinical effectiveness and cost-effectiveness of sequential use of these treatments will also be evaluated.

Chapter 3 Assessment of clinical effectiveness

Methods for reviewing effectiveness

Inclusion criteria

Two reviewers independently screened all titles and abstracts. Full manuscripts of any titles/abstracts that were relevant were obtained where possible and the relevance of each study assessed by two reviewers according to the criteria below. Any discrepancies were resolved by consensus and, when necessary, a third reviewer was consulted. Studies available only as abstracts were included.

Study design

For the review of clinical efficacy randomised controlled trials (RCTs) were eligible, including any open-label extensions of RCTs. Adverse event (AE) data were sought from existing reviews and other appropriately large studies. For studies of natural history, long-term effectiveness, adherence and sequential use, published analyses based on large and long-term data sets (including studies of registry data) were eligible.

Interventions

Adalimumab, certolizumab pegol, etanercept, golimumab, infliximab or any of their biosimilars were eligible.

Comparators

Relevant comparators were conventional management strategies (either with or without placebo) and also the different TNF-α inhibitors listed above (i.e. head-to-head trials).

Participants

Studies of adults with either severe active AS or severe axSpA without radiographic evidence of AS but with objective signs of inflammation (such as elevated CRP levels or a positive MRI) were eligible. Patients with predominantly peripheral spondyloarthritis were excluded. Data relating to serious adverse effects associated with anti-TNF agents used in other indications were also considered.

Outcomes

Studies reporting the following outcomes were eligible:

- multiple domain response criteria: (e.g. ASAS 20, ASAS 40, ASAS 5/6 and ASAS partial remission)
- disease activity (e.g. BASDAI)
- functional capacity [e.g. Bath Ankylosing Spondylitis Functional Index (BASFI)]
- disease progression [e.g. modified Stoke Ankylosing Spondylitis Spinal Score (mSASSS)]
- pain [e.g. visual analogue scale (VAS) scores]
- peripheral symptoms (including enthesitis, peripheral arthritis and dactylitis)
- symptoms of extra-articular manifestations (including anterior uveitis, inflammatory bowel disease and psoriasis)
- HRQoL [e.g. European Quality of Life-5 Dimensions (EQ-5D)]
- rates of treatment discontinuation and withdrawal
- AEs.

For AEs the evaluation specifically focused on known possible AEs of anti-TNFs, such as reactivation of latent tuberculosis, malignancies, non-melanoma skin cancer, severe infections, congestive heart failure, and injection site reactions. Withdrawals because of AEs and events categorised as serious adverse events (SAEs) were also evaluated.

Searches

The following databases were searched for relevant clinical effectiveness and cost-effectiveness research:

- MEDLINE
- EMBASE
- Cumulative Index to Nursing and Allied Health Literature Plus
- Science Citation Index
- ClinicalTrials.gov
- Cochrane Central Register of Controlled Trials
- Cochrane Database of Systematic Reviews
- Database of Abstracts of Reviews of Effects
- International Prospective Register of Systematic Reviews (PROSPERO)
- Health Technology Assessment Database
- Conference Proceedings Citation Index – Science
- National Guidelines Clearinghouse
- NHS Evidence
- NHS Clinical Knowledge Summaries
- NHS Economic Evaluation Database (NHS EED).

The terms for search strategies were identified through discussion within the research team, by scanning the background literature and browsing the MEDLINE medical subject headings. No date or language limits were applied. As several databases were searched, some degree of duplication resulted. To manage this issue, the titles and abstracts of bibliographic records were imported into EndNote bibliographic management software (version X7, Thomson Reuters, CA, USA) to remove duplicate records. Databases were searched from inception, with most of the searches being performed in June or July 2014. The full search strategies used in each database, together with the search dates, are listed in *Appendix 1*.

Data extraction

Data relating to study design, outcome results and quality were extracted by one reviewer using a standardised data extraction form and independently checked for accuracy by a second reviewer. Disagreements were resolved through consensus, and, when necessary, a third reviewer was consulted. Data from studies with multiple publications were extracted and reported as a single study. Data were also extracted from the manufacturer's submissions when they were not available from other sources.[34–37] Clinicaltrials.gov records and relevant US Food and Drug Administration (FDA) or European Medicines Agency reports were also used to extract any missing data. When data could only be estimated from graphs, the estimates used in the previous assessment report[38] were used when available. In the light of the multidomain outcomes which incorporated pain scores (the ASAS and BASDAI outcomes), it was decided that pain scores on their own would not be extracted.

This report contains reference to confidential information provided as part of the NICE appraisal process. This information has been removed from the report and the results, discussions and conclusions of the report do not include the confidential information. These sections are clearly marked in the report.

Critical appraisal

The quality of RCTs was assessed using the Cochrane risk of bias tool,[39] with additional assessments made for baseline imbalance of important prognostic indicators.[40] The relevant prognostic and treatment response indicators were identified from both published research and clinical advice. The risk of bias assessments were performed by one reviewer, and independently checked by a second. Disagreements were resolved through consensus, and, when necessary, a third reviewer was consulted. Open-label extension studies were evaluated based on the imputation methods and patient withdrawal criteria used.

Methods of data synthesis

This section describes the data set construction and meta-analyses conducted for the different outcomes individually. *Chapter 5* provides detailed evidence synthesis methods that incorporate different outcomes within one analysis and presents clinical outcome estimates appropriate for the economic model.

Results of the data extraction in terms of study characteristics and quality assessment are presented in tables and summarised narratively. Results of open-label studies, drug survival and switching studies and natural history studies were also summarised narratively. As several of the RCTs were placebo-controlled up to 24 weeks, only time points beyond 24 weeks were evaluated in the open-label studies. AE data from the RCTs were pooled when enough data were identified; otherwise, the AE data and the other studies relating specifically to AEs were summarised narratively.

Clinical effectiveness data were synthesised using Bayesian meta-analysis methods. The main analysis was of outcomes reported from 10 to 16 weeks. A sensitivity analysis was done of outcomes reported from 24 to 30 weeks.

Dosage and pooling of trial arms

The doses included in the analyses were:

- adalimumab – 40 mg every other week
- certolizumab pegol – 200 mg every 2 weeks, 400 mg every 4 weeks
- etanercept – 25 mg twice weekly, 50 mg weekly
- golimumab – 50 mg every month
- infliximab – 5 mg/kg at 0, 2 or $\geq$ 6 weeks.

Golimumab of 100 mg every 4 weeks was excluded when it was not used according to its licence.

Data from active treatment arms were pooled in trials which studied different doses. This occurred for certolizumab pegol 200 mg every 2 weeks and 400 mg every 4 weeks and for etanercept 25 mg twice weekly and 50 mg weekly.

Data imputation and assumptions

Medians were treated as means. Although the median may not be exactly the same as the mean, the median was considered to give sufficiently accurate information. Standard deviations (SDs) were estimated from interquartile ranges, the method of which is described in *Appendix 2*. Where no SD was reported, the highest SD from the other trials was used as a conservative estimate.

In the meta-analyses, 'change from baseline' outcomes were used in the analysis for continuous outcomes. Where these were not reported, but adequate baseline and final value outcomes were reported, the change from baseline and its SD were derived from the baseline and final values and their SDs. The detailed methods are described in *Appendix 2*.

The imputation of change from baseline or final values required a within-trial correlation estimate, and trials that reported the SDs of baseline, change from baseline and final values were used to estimate the within-trial correlation. For BASDAI the within-study correlation varied from 0.33 to 0.67 across four trials. Given the small samples of some trials, the within-study correlation can vary significantly from trial to trial. For the base-case analysis, a correlation estimate of 0.3 was used and an estimate of 0.7 was tested in sensitivity analysis. For the calculation of final values, the lowest possible correlation was used when 0.3 or 0.7 were not feasible solutions (see *Appendix 2*).

Change from baseline was imputed for three trials for BASDAI, five trials for BASFI, one trial for Bath Ankylosing Spondylitis Metrology Index (BASMI), two trials for Short Form questionnaire-36 items (SF-36) physical component summary (PCS), and one trial for SF-36 mental component summary (MCS). For each of these outcomes, one of the imputations was for a trial with a non-radiographic population.

Binary event outcomes

Odds ratios (ORs) were derived for binary event outcomes. Relative risks were also derived from the ORs using the placebo absolute risks estimated from all the trials measuring the relevant outcome within weeks 10 to 16. The relative risk estimates are therefore based on the population distribution of the trials across the interventions. As the placebo absolute risk was based on more trials than those informing the ORs for some outcomes, the 95% credible interval (CrI) estimates of the relative risk were narrower than the CrI estimates of the OR. The placebo absolute risk was estimated using both fixed- and random-effect models within WinBUGS (Medical Research Council Biostatistics Unit, Cambridge, UK). As the random-effect model for the placebo absolute risk was a better fit than the fixed effect model according to the deviance information criterion (DIC) statistic, the placebo absolute risks from the random-effect models were used. For the ASAS outcomes, fewer trials reported the greater response outcomes, so a prior distribution was used for the between-study SD based on the closest ASAS outcome (see *Appendix 2*).

Analyses

Analyses were conducted in WinBUGS version 1.4.3. See *Chapter 5* for more details on the models. For each outcome, multiple-treatment meta-analyses were conducted assuming that the treatments had independent effects [related to models A1 (fixed effect) and A2 (random effects) in *Chapter 5*]. They were also run assuming that they had a common class effect [related to models A3 (fixed effect) and A4 (random effects) in *Chapter 5*] and that the DIC statistic was used to determine the model that best fitted the data. The random-effect models with independent treatment effects were assumed to have a common between-study variance across the comparisons in the network.

The sensitivity of random-effect models to the between-study SD priors was tested. I^2 statistics for heterogeneity were calculated for random-effect models that were insensitive to change in the prior distribution for the between-study SD. Results were only presented for random-effect models.

Clinical effectiveness results

Quantity and quality of research available

The electronic database searches identified 2284 references. After screening titles and abstracts, full copies of 198 papers were assessed for inclusion in the review. Three trials of axSpA populations were excluded because results were not available separately for the AS and nr-AxSpA populations.[41–43] One study of adalimumab appeared likely to be eligible but was excluded as it was only available as a ClinicalTrials.gov record, without any results or further study details.[44] One excluded study was an ongoing trial of golimumab (called GO-AHEAD).[45]

Twenty-eight eligible RCTs were identified, with 24 being suitable for data synthesis. Three etanercept trials were not suitable for data synthesis because the study durations were only 6 weeks,[46–48] and one infliximab trial was unsuitable because a (currently) unlicensed dose (3 mg/kg) had been studied.[49] The Barkham 2009 trial[50,51] of infliximab in nr-AxSpA patients (see *Table 2*) was included in the clinical efficacy section because, even though infliximab is not currently licensed for patients with nr-AxSpA, the dose used in this trial was the same as that licensed for AS. Furthermore, there was no reason to think it could not be considered in the same class as the other anti-TNFs when treating a nr-AxSpA population. The results of the trial therefore had the potential to be useful to help inform the relative efficacy of anti-TNFs for nr-AxSpA.

Of the 17 RCTs in which participants were studied beyond the randomised phase (i.e. in open-label studies), 71 additional full publications or conference abstracts were identified. *Figure 1* illustrates the flow of studies through the review process.

Study characteristics

Table 2 lists the 24 eligible RCTs (and all the RCT-related references) which were eligible for inclusion in the network meta-analysis. Six trials compared adalimumab versus placebo, one compared certolizumab pegol versus placebo, seven compared etanercept versus placebo, three compared golimumab versus placebo, five compared infliximab versus placebo, one compared etanercept with infliximab and one compared infliximab with an infliximab biosimilar (CT-P13). Most placebo-controlled phases lasted for 12 weeks. All but seven of the trials were extended into open-label (unblinded) phases, with 11 studies having a total duration of at least 1 year.

Of the trials suitable for analysis, most were conducted in Europe and/or North America; four were conducted in China. Four studies recruited a nr-AxSpA population, 19 an AS population and one recruited both populations.[64] *Table 3* details the baseline characteristics of the populations studied. In the nr-AxSpA studies around half of the participants were male, whereas in the AS studies around three-quarters were male. All trials recruited participants with active disease; half the trials specified that participants had to have failed one or more NSAID, and a BASDAI score of ≥ 4 was used as an entry requirement in most, with the exception of six early trials in which a BASDAI criterion was not stated.[72,79,83,86,100] Notwithstanding these entry criteria, the recruited participants mostly still took a NSAID (around between 80% and 90% of participants, although reported in only 12 trials) and had quite high mean (or median) BASDAI scores: most were between 5.5 and 6.5 (the range across all trial arms was 5.3–7.0). BASFI scores varied more widely, ranging between 3.2 and 6.7. Variation in CRP levels was also apparent, with lower values in the nr-AxSpA trials being evident. Trials which reported both mean and median CRP showed skewed distributions, with means being higher than medians.[58,64,95] The upper limits of normal used for defining elevated CRP level in the nr-AxSpA trials were either unclear[58] or varied, being 3 mg/l,[76] 6 mg/l[51] or 7.9 mg/l.[64] One nr-AxSpA study recruited only MRI-positive patients.[50] In the remaining nr-AxSpA trials the proportion of MRI-positive patients ranged from 51%[58] to 81%.[76]

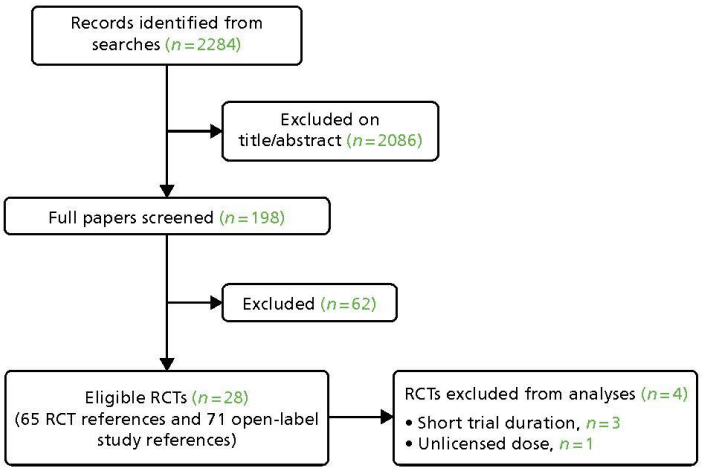


FIGURE 1 Flow chart showing the number of studies identified and included.

TABLE 2 General trial characteristics

Study	Interventions	Anti-TNF dose	Country/continent	Population	Duration of placebo-controlled phase (weeks)	Total duration of study, including any open-label extension phase
Haibel 2008[52–54]	Adalimumab; placebo	40 mg every other week	Germany	nr-AxSpA with inflammation, inadequate response/intolerance to NSAIDs	12	1 year
Hu 2012[55]	Adalimumab; placebo	40 mg every other week	China	AS, inadequate response/ intolerance to NSAIDs	12	24 weeks
Huang 2014[56]	Adalimumab; placebo	40 mg every other week	China	AS, inadequate response/ intolerance to NSAIDs	12	24 weeks
Lambert 2007[57]	Adalimumab; placebo	40 mg every other week	Canada	AS, inadequate response to a NSAID or DMARD	12	1 year
ABILITY-1 2013[58–60]	Adalimumab; placebo	40 mg every other week	Australia, Europe, North America	nr-AxSpA with inflammation, inadequate response/ contraindication to NSAIDs	12	3 years
ATLAS 2006[61–63]	Adalimumab; placebo	40 mg every other week	USA, Europe	AS, inadequate response to a NSAID or DMARD	12	5 years
RAPID-axSpA 2014[64–70]	Certolizumab pegol; placebo	200 mg every 2 weeks or 400 mg every 4 weeks	Europe, North America, Latin America	AS, nr-AxSpA with inflammation, inadequate response/intolerance to NSAIDs	12	96 weeks
Barkham 2010[71]	Etanercept; placebo	25 mg twice weekly	UK	AS	12	12 weeks
Davis 2003[72,73]	Etanercept; placebo	25 mg twice weekly	North America, Europe	AS	24	168 weeks
Dougados 2011[74,75]	Etanercept; placebo	50 mg weekly	Europe	AS, inadequate response to NSAIDs	12	24 weeks
Dougados 2014[76–78]	Etanercept; placebo	50 mg weekly	Europe, Asia, South America	nr-AxSpA, inadequate response to NSAIDs	12	48 weeks
Gorman 2002[79–82]	Etanercept; placebo	25 mg twice weekly	USA	AS	16	40 weeks
Calin 2004[83–85]	Etanercept; placebo	25 mg twice weekly	Europe	AS	12	5 years
van der Heijde 2006[86,87]	Etanercept; placebo	25 mg twice weekly or 50 mg weekly	Europe	AS	12	12 weeks

Study	Interventions	Anti-TNF dose	Country/continent	Population	Duration of placebo-controlled phase (weeks)	Total duration of study, including any open-label extension phase
Giardina 2010[88,89]	Etanercept; infliximab	50 mg weekly; 5 mg/kg (at week 0, 2, 6 and every 6 weeks)	Italy	AS, inadequate response to NSAIDs	N/A	12 weeks
GO-RAISE 2008[90–94]	Golimumab; placebo	50 mg or 100 mg every 4 weeks	North America, Europe, Asia	AS	16	4 years
Bao 2014[95,96]	Golimumab; placebo	50 mg every 4 weeks	China	AS	14	1 year
Tam 2014[97]	Golimumab; placebo	50 mg every 4 weeks	China (Hong Kong)	AS, inadequate response to NSAIDs	24	1 year
Barkham 2009[50,51]	Infliximab; placebo	5 mg/kg (at 0, 2, 6 and 12 weeks)	UK	nr-AxSpA with inflammation	16	16 weeks
Braun 2002[98,99]	Infliximab; placebo	5 mg/kg (at weeks 0, 2 and 6)	Germany	AS	12	8 years
Marzo-Ortega 2005[100]	Infliximab + methotrexate; placebo + methotrexate	5 mg/kg (at weeks 0, 2, 6, 14 and 22)	UK	AS	30	30 weeks
Van den Bosch 2002[101]	Infliximab; placebo	5 mg/kg (at weeks 0, 2 and 6)	Belgium	AS	12	12 weeks
ASSERT[102–109]	Infliximab; placebo	5 mg/kg (at weeks 0, 2, 6, 12 and 18)	North America, Europe	AS, inadequate response/ intolerance to NSAIDs	24	2 years
PLANETAS 2013[110,111]	CT-P13 biosimilar; infliximab	Both 5 mg/kg	Europe, Asia, Latin America	AS	N/A	2 years (using randomised interventions up to 54 weeks)

ASSERT, Ankylosing Spondylitis Study for the Evaluation of Recombinant Infliximab Therapy; ATLAS, Adalimumab Trial Evaluating Long-term Efficacy and Safety for Ankylosing Spondylitis; DMARD, disease-modifying antirheumatic drugs; PLANETAS, Programme evaluating the Autoimmune disease iNvEstigational drug CT-p13 in Ankylosing Spondylitis; N/A, not applicable.
All RCT-related references have been included for each study.

TABLE 3 Baseline characteristics of trial populations

Trial	Patient group	Trial arm	*n*	% male	Age (years)	% on a NSAID	Symptom duration (years)	BASDAI score	BASFI score	BASMI score	CRP level mg/l, (SD)	% HLA-B27 positive	SF-36 MCS score	SF-36 PCS score	ASQoL score
Haibel 2008[52]	nr-AxSpA	Adalimumab	22	41	Mean 38	NR	Mean 7	Mean 6.5 (SD 1.2)	Mean 5.4 (SD 2)	Mean 1.3 (SD 1.2)	Mean 6.2 (SD 5.8)	59	Mean 41.3 (SD 12.5)	Mean 28.8 (SD 7.6)	Mean 10.8 (SD 3.7)
	nr-AxSpA	Placebo	24	50	Mean 37	NR	Mean 8	Mean 6.2 (SD 1.3)	Mean 4.9 (SD 1.6)	Mean 1.3 (SD 1.6)	Mean 7.8 (SD 7.0)	75	Mean 43.6 (SD 11.1)	Mean 30.7 (SD 6)	Mean 9.5 (SD 3)
Hu 2012[55]	AS	Adalimumab	26	92	Mean 28.2 (SD 6.9)	NR	Mean 7.4	Mean 5.9 (SD 1.4)	Mean 3.7 (SD 2.1)	–	Mean 24.6	96	–	–	–
	AS	Placebo	20	100	Mean 27.4 (SD 7.2)	NR	Mean 7.6	Mean 6.2 (SD 1.1)	Mean 3.9 (SD 2)	–	Mean 32.1	95	–	–	–
Huang 2014[56]	AS	Adalimumab	229	81	Mean 30.1 (SD 8.7)	80	Mean 8.1	Mean 6.0 (SD 1.4)	Mean 4.3 (SD 2.3)	Mean 3.4 (SD 1.4)	Mean 22.4 (SD 24)	96	Mean 36.2 (SD 10.7)	Mean 33.8 (SD 7)	–
	AS	Placebo	115	83	Mean 29.6 (SD 7.5)	78	Mean 7.7	Mean 6.2 (SD 1.4)	Mean 4.4 (SD 2.3)	Mean 3.4 (SD 1.5)	Mean 23 (SD 30)	95	Mean 35 (SD 10.6)	Mean 32.2 (SD 6.7)	–
Lambert 2007[57]	AS	Adalimumab	38	76	Mean 41.9 (SD 11.1)	NR	Mean 14.5	Mean 6.2 (SD 1.7)	Mean 5.3 (SD 2)	–	Mean 18	87	–	–	–
	AS	Placebo	44	82	Mean 40 (SD 10.9)	NR	Mean 12.1	Mean 6.5 (SD 1.6)	Mean 5.6 (SD 2.2)	–	Mean 23	82	–	–	–
[a]ABILITY-1 2013[58]	nr-AxSpA	Adalimumab	69	46	Mean 38.3 (SD 11.7)	NR	Mean 10.7	Mean 6.4 (SD 1.6)	Mean 4.5 (SD 2.1)	Mean 2.7	Mean 8.6 (SD 13.1)	NR	–	Mean 33.3 (SD 7.8)	–
	nr-AxSpA	Placebo	73	45	Mean 38.3 (SD 10.5)	NR	Mean 10.5	Mean 6.4 (SD 1.5)	Mean 4.8 (SD 2.3)	Mean 2.7	Mean 9.3 (SD 10.9)	NR	–	Mean 33.2 (SD 8.2)	–
ATLAS 2006[61]	AS	Adalimumab	208	76	Mean 41.7 (SD 11.7)	80	Mean 11.3	Mean 6.3 (SD 1.7)	Mean 5.2 (SD 2.2)	Mean 3.8 (SD 2.2)	Mean 18	78	Mean 43.4 (SD 12)	Mean 32.9 (SD 8)	Mean 10.2 (SD 4)
	AS	Placebo	107	74	Mean 43.4 (SD 11.3)	79	Mean 10	Mean 6.3 (SD 1.7)	Mean 5.6 (SD 2.2)	Mean 4.2 (SD 2.1)	Mean 22	79	Mean 44.4 (SD 12)	Mean 31.8 (SD 8)	Mean 10.6 (SD 4)
RAPID-axSpA 2014[64]	AS	Certolizumab pegol 200 mg	65	72	Mean 41 (SD 10.8)	91	Median 8.8	Mean 6.5 (SD 1.7)	Mean 5.6 (SD 2.3)	Mean 4.2 (SD 1.6)	Median 14	82	–	–	–
	AS	Certolizumab pegol 400 mg	56	73	Mean 41.9 (SD 11.5)	91	Median 8.8	Mean 6.2 (SD 1.3)	Mean 5.7 (SD 2.3)	Mean 4.3 (SD 1.8)	Median 12.9	79	–	–	–
	AS	Placebo	57	72	Mean 41.6 (SD 12.8)	90	Median 10.2	Mean 6.4 (SD 1.9)	Mean 6.0 (SD 2)	Mean 4.7 (SD 1.6)	Median 16.6	84	–	–	–

Trial	Patient group	Trial arm	*n*	% male	Age (years)	% on a NSAID	Symptom duration (years)	BASDAI score	BASFI score	BASMI score	CRP level mg/l, (SD)	% HLA-B27 positive	SF-36 MCS score	SF-36 PCS score	ASQoL score
RAPID-axSpA 2014[64]	nr-AxSpA	Certolizumab pegol 200 mg	46	44	Mean 36.6 (SD 13)	83	Median 4.8	Mean 6.5 (SD 1.4)	Mean 4.8 (SD 2.2)	Mean 3.1 (SD 1.4)	Median 10	74	–	–	–
	nr-AxSpA	Certolizumab pegol 400 mg	51	53	Mean 37.5 (SD 10.8)	86	Median 7.3	Mean 6.6 (SD 1.6)	Mean 5.1 (SD 2.4)	Mean 3.3 (SD 1.5)	Median 12.1	73	–	–	–
	nr-AxSpA	Placebo	50	48	Mean 38 (SD 11.8)	82	Median 4.5	Mean 6.4 (SD 1.5)	Mean 4.9 (SD 2.2)	Mean 3.1 (SD 1.6)	Median 13.5	78	–	–	–
Barkham 2010[71]	AS	Etanercept	20	75	Mean 40.8 (SD 9.7)	NR	Median 11	Mean 6.1 (SD 1.7)	Mean 5.6 (SD 2.0)	–	–	NR	–	–	–
	AS	Placebo	20	85	Mean 39.4 (SD 10.1)	NR	Median 20	Mean 5.5 (SD 1.7)	Mean 5.3 (SD 1.8)	–	–	NR	–	–	–
Davis 2003[72]	AS	Etanercept	138	76	Mean 42.1	91	Mean 10.1	Mean 5.8 (SE 0.15)	Mean 5.2	–	Mean 19	84	–	–	–
	AS	Placebo	139	76	Mean 41.9	92	Mean 10.5	Mean 6.0 (SE 0.14)	Mean 5.6	–	Mean 20	84	–	–	–
Dougados 2011[74]	AS	Etanercept	39	95	Mean 46 (SD 11)	NR	Mean 19	Mean 6.4 (SD 1.2)	Mean 6.3 (SD 2.0)	Mean 5.7 (SD 1.4)	Mean 25 (31)	79	–	–	–
	AS	Placebo	43	91	Mean 48 (SD 10)	NR	Mean 23	Mean 5.8 (SD 1.5)	Mean 5.7 (SD 1.9)	Mean 5.8 (SD 1.3)	Mean 17 (19)	86	–	–	–
[b]Dougados 2014[16]	nr-AxSpA	Etanercept	106	64	Mean 31.9 (SD 7.8)	CiC information has been removed	Mean 2.4	CiC information has been removed	CiC information has been removed	Mean 1.4 (SD 1.3)	Mean 6.8	67	–	CiC information has been removed	CiC information has been removed
	nr-AxSpA	Placebo	109	57	Mean 32 (SD 7.8)	CiC information has been removed	Mean 2.5	CiC information has been removed	CiC information has been removed	Mean 1.2 (SD 1.3)	Mean 6.4	76	–	CiC information has been removed	CiC information has been removed
Gorman 2002[79]	AS	Etanercept	20	65	CiC information has been removed	80	CiC information has been removed	–	Mean 4.5 (SD 2.1)	–	Mean 20	95	–	–	–
	AS	Placebo	20	90	CiC information has been removed	95	CiC information has been removed	–	Mean 3.2 (SD 2.5)	–	Mean 15	90	–	–	–

continued

TABLE 3 Baseline characteristics of trial populations (*continued*)

Trial	Patient group	Trial arm	*n*	% male	Age (years)	% on a NSAID	Symptom duration (years)	BASDAI score	BASFI score	BASMI score	CRP level mg/l, (SD)	% HLA-B27 positive	SF-36 MCS score	SF-36 PCS score	ASQoL score
Calin 2004[83]	AS	Etanercept	45	80	Mean 45.3 (SD 9.5)	89	Mean 15.0	Mean 6.1	Mean 6.0	–	Median 154	NR	–	–	–
	AS	Placebo	39	77	Mean 40.7 (SD 11.4)	85	Mean 9.7	Mean 5.9	Mean 5.7	–	Median 97	NR	–	–	–
van der Heijde 2006[86]	AS	Etanercept 25 mg	150	76	Mean 39.8 (SD 10.7)	85	Mean 10.0	Mean 5.9 (SD 1.7)	Mean 5.8 (SD 2.0)	–	Mean 19.8 (SD 20.8)	NR	–	–	–
	AS	Etanercept 50 mg	155	70	Mean 41.5 (SD 11)	80	Mean 9.0	Mean 6.2 (SD 1.7)	Mean 6.1 (SD 2.0)	–	Mean 21.7 (SD 24.6)	NR	–	–	–
	AS	Placebo	51	78	Mean 40.1 (SD 10.9)	78	Mean 8.5	Mean 6.1 (SD 1.4)	Mean 6.0 (SD 1.9)	–	Mean 22 (SD 22.9)	NR	–	–	–
Giardina 2010[88]	AS	Etanercept	25	80	Mean 32.6 (SD 6.8)	NR	Mean 15.7	Mean 6.6 (SD 1.1)	Mean 6.5 (SD 1.1)	Mean 3.9 (SD 1.7)	Mean 22.9	96	–	–	–
	AS	Infliximab	25	76	Mean 31.9 (SD 9.2)	NR	Mean 15.4	Mean 6.5 (SD 1.2)	Mean 6.1 (SD 0.9)	Mean 3.7 (SD 1.6)	Mean 25	92	–	–	–
GO-RAISE 2008[90]	AS	Golimumab 50 mg	138	74	Median 38	90	Median 11	Median 6.6 (IQR 5.6–7.6)	Median 5 (IQR 3.2–6.7)	Median 3 (IQR 2–4)	Mean 11	82	Median 46.5 (IQR 36.8–54.1)	Median 29.7 (IQR 22.5–35.3)	–
	AS	Golimumab 100 mg	140	70	Median 38	88	Median 9.5	Median 7 (IQR 6.0–7.9)	Median 5.4 (IQR 3.4 to 7.3)	Median 3 (IQR 2–5)	Mean 9	84	Median 43.1 (IQR 33.5–53.5)	Median 29.8 (IQR 25.2–35.5)	–
	AS	Placebo	78	71	Median 41	92	Median 16.0	Median 6.6 (IQR 5.7–7.7)	Median 4.9 (IQR 3.5–6.8)	Median 4 (IQR 2–5)	Mean 11.5	85	Median 46.2 (IQR 37.1–54.8)	Median 28.3 (IQR 23.8–34.1)	–
Bao 2014[95]	AS	Golimumab	108	83	Mean 30.5 (SD 10.3)	67	Mean 6.8	Mean 6.6 (1.3)	Mean 5 (SD 2.4)	Mean 4 (SD 1.9)	Mean 20.6	–	Mean 36.5 (SD 10.5)	Mean 33.2 (SD 7.8)	–
	AS	Placebo	105	83	Mean 30.6 (SD 8.6)	72	Mean 7.5	Mean 6.5 (1.5)	Mean 5 (SD 2.4)	Mean 3.8 (SD 1.6)	Mean 18.6	–	Mean 36.2 (SD 11.5)	Mean 33.9 (SD 7.7)	–
Tam 2014[97]	AS	Golimumab	20	90	Mean 35.6 (SD 9.9)	85	Mean 8.0	Mean 6.2 (1.0)	Mean 4.6 (SD 1.9)	Median 5.0 (IQR 4.0–7.0)	Mean 23.9 (SD 18.6)	–	–	–	–
	AS	Placebo	21	90	Mean 34.2 (SD 10)	100	Mean 11.0	Mean 6.2 (1.5)	Mean 4.1 (SD 2.3)	Median 3 (IQR 2.0–5.5)	Mean 19.9 (SD 14.0)	–	–	–	–

Trial	Patient group	Trial arm	*n*	% male	Age (years)	% on a NSAID	Symptom duration (years)	BASDAI score	BASFI score	BASMI score	CRP level mg/l, (SD)	% HLA-B27 positive	SF-36 MCS score	SF-36 PCS score	ASQoL score
Barkham 2009[50]	nr-AxSpA	Infliximab	20	75	Mean 29.5	90	Mean 13.4	Mean 5.9	Mean 4.4	–	Median 5	100	–	–	Mean 10
	nr-AxSpA	Placebo	20	75	Mean 28.2	90	Mean 17.2	Mean 5.8	Mean 4.1	–	Median 11.5	100	–	–	Mean 11
Braun 2002[88]	AS	Infliximab	34	68	Mean 40.6 (SD 8)	NR	Mean 16.4	Mean 6.5 (1.2)	Mean 5.4 (SD 1.8)	Mean 3.7 (SD 2.0)	Mean 24	91	Mean 51.5 (SD 22.6)	Mean 46.5 (SD 22.6)	-
	AS	Placebo	35	63	Mean 39 (SD 9.1)	NR	Mean 14.9	Mean 6.3 (1.4)	Mean 5.1 (SD 2.2)	Mean 3.7 (SD 2.2)	Mean 18	88	Mean 65.4 (SD 18.4)	Mean 47.6 (SD 23.4)	–
Marzo-Ortega 2005[100]	AS	Infliximab	28	82	Mean 41	89	Median 8	Mean 6.5 (1.9)	Median 6.7	–	Median 30.5	96	–	–	Median 14
	AS	Placebo	14	79	Mean 39	86	Median 10	Mean 6.6 (2.1)	Median 6	–	Median 30	86	–	–	Median 13.5
Van den Bosch 2002[101]	AS	Infliximab	9	78	Mean 44.3	NR	Mean 10	Median 5.9	Median 4.7	Median 5	Mean 41.0	89	–	–	–
	AS	Placebo	12	83	Mean 46.4	NR	Mean 17	Median 5.3	Median 5.9	Median 4	Mean 25.7	75	–	–	–
ASSERT 2005[102]	AS	Infliximab	201	78	Median 40	NR	Mean 7.7	Median 6.6 (IQR 5.2–7.1)	Median 5.7 (IQR 4.5–7.1)	–	Mean 15	87	Median 47.6 (IQR 37.6–54.9)	Median 28.8 (IQR 23.8–33.7)	–
Park 2013[110]	AS	Placebo	78	87	Median 41	NR	Mean 13.2	Median 6.5 (IQR 5.3–7.6)	Median 6 (IQR 4.1–7.2)	–	Mean 17	89	Median 45 (IQR 33.7–55.5)	Median 30.1 (IQR 24.9–36.2)	–
PLANETAS 2013[110]	AS	CT-P13	125	79	Median 38	NR	–	Mean 6.7 (SD 1.4)	Mean 6.2 (SD 1.9)	Mean 4 (SD 2.1)	Median 11	–	–	–	–
	AS	Infliximab	125	82	Median 38	NR	–	Mean 6.6 (SD 1.6)	Mean 6.2 (SD 2.2)	Mean 4.1 (SD 2.1)	Median 14	–	–	–	–

ASQoL, Ankylosing Spondylitis Quality of Life; ASSERT, Ankylosing Spondylitis Study for the Evaluation of Recombinant Infliximab Therapy; ATLAS, Adalimumab Trial Evaluating Long-term Efficacy and Safety for Ankylosing Spondylitis; CER, certolizumab pegol; CiC, commercial-in-confidence; IQR, interquartile range; NR not reported; SE, standard error.

a Licensed population.

b Includes a small proportion (12%) of unlicensed patients.

Risk of bias

Results of the risk of bias judgements are presented in *Table 4*. Further details, including judgement reasons and the prognostic indicators of important baseline imbalance, are available in *Appendix 3*. Most trials were judged to have a low risk of bias overall; when possible bias was detected, there was little indication to suggest that this varied across the different anti-TNF trials.

Over half the trials did not report adequate details about methods of randomisation and allocation concealment, although in the majority of those trials (8 out of 14) an assessment could be made of whether or not groups were balanced in all five of the important prognostic indicators of treatment response. Using both randomisation method details and a baseline assessment to judge the risk of selection bias, 15 trials were judged as having a low risk of selection bias, five trials were judged as having an unclear risk[51,71,86,96,112] and four as having a high risk;[50,79,83,101] in one of these four trials the risk was deemed likely to be a result of a chance effect.[79]

The risk of performance bias arising from lack of blinding of participants and personnel was low in 20 trials, unclear in three trials[55,57,64] and high in the one head-to-head trial, in which blinding would have been difficult to achieve because of the different modes and timings of delivery (weekly injection for etanercept vs. 6-weekly infusion for infliximab).[88] All except one of the trials were at low risk of detection bias, as they were all adequately placebo controlled (except the head-to-head trial), with nearly all the key outcomes being self-reported by patients (a notable exception being BASMI). The blinded patients were the outcome assessors, and the effect of any unblinded study personnel on patient questionnaire responses was likely to be minimal at most. The proportion of patients withdrawing or dropping out of trials was generally low; most trials received low risk judgements for attrition bias. In two of the trials with unclear risk judgements, there were nevertheless reasons to suspect the possibility of important bias (see *Appendix 3*).[55,71] Of the studies with missing data which also reported details on the populations and imputations used in analyses, 'last observation carried forward' (LOCF) was used; this was done using a modified intention-to-treat (mITT) approach in just over half the trials (in which patients had to have received at least one dose of treatment) and an intention-to-treat (ITT) approach in the remaining trials (see *Appendix 3*). There was no evidence of reporting bias in any of the trials with all being judged as low risk, except for one trial with an unclear risk of bias.[55]

Clinical effectiveness results: efficacy results from randomised controlled trials

Individual results for all 24 trials are presented in *Appendix 4*.

Exclusions from the meta-analyses

Of the trials with results at between 10 and 16 weeks, one small head-to-head trial ($n = 50$) comparing etanercept with infliximab was excluded, as it was redundant in a class-effect model (in addition, blinding was not feasible in this trial).[88] One trial[110] was excluded because it compared infliximab and CT-P13, and therefore did not include any of the relevant comparators needed for meta-analysis. The maximum number of studies included for any one outcome was 16.

Exclusions from the sensitivity analyses

Five studies were excluded in the sensitivity analyses because of risk of bias judgements.[55,71,79,83,101] Further details can be found in *Appendix 3*. A sensitivity analysis of the nr-AxSpA trials was not performed, as the one trial judged to have a high risk of bias had only 40 patients;[50] any effect arising from the removal of such a small study would have been likely to have been minimal.

TABLE 4 Risk of bias assessment results

Trial	Bias domain							
	1. Sequence generation	2. Allocation concealment	3. Important baseline imbalance	Selection bias based on 1, 2, and 3	4. Blinding of participants and personnel	5. Blinding of outcome assessment	6. Incomplete outcome data	7. Selective reporting
	Risk of bias judgement							
Adalimumab vs. placebo								
Haibel 2008[52]	Unclear	Unclear	Unclear	Unclear	Low	Low	Low	Low
Hu 2012[55]	Unclear	Unclear	Low	Low	Unclear	Low	Unclear	Unclear
Huang 2014[56]	Low	Low	Low	Low	Low	Low	Low	Low
Lambert 2007[57]	Unclear	Unclear	Low	Low	Unclear	Low	Low	Low
ABILITY-1 2013[58]	Low	Low	Low	Low	Low	Low	Low	Low
ATLAS 2006[61]	Unclear	Unclear	Low	Low	Low	Low	Low	Low
Certolizumab pegol vs. placebo								
RAPID-axSpA 2014[64]	Low	Low	Low	Low	Unclear	Low	Low	Low
Etanercept vs. placebo								
Barkham 2010[71]	Unclear	Unclear	Unclear	Unclear	Low	Low	Unclear	Low
Davis 2003[72]	Unclear	Unclear	Low	Low	Low	Low	Low	Low
Dougados 2011[74]	Unclear	Unclear	Low	Low	Low	Low	Low	Low
Dougados 2014[76]	Low	Low	Low	Low	Low	Low	Low	Low
Gorman 2002[79]	Low	Low	High[a]	High[a]	Low	Low	Low	Low
Calin 2004[83]	Unclear	Unclear	High	High	Low	Low	Low	Low
van der Heijde 2006[86]	Unclear	Unclear	Unclear	Unclear	Low	Low	Low	Low

continued

TABLE 4 Risk of bias assessment results (*continued*)

	Bias domain							
	1. Sequence generation	2. Allocation concealment	3. Important baseline imbalance	Selection bias based on 1, 2, and 3	4. Blinding of participants and personnel	5. Blinding of outcome assessment	6. Incomplete outcome data	7. Selective reporting
Trial	Risk of bias judgement							
Etanercept vs. infliximab								
Giardina 2010[88]	High	High	Low	Low	High	High	Low	Low
Golimumab vs. placebo								
GO-RAISE 2008[90]	Low	Low	Low	Low	Low	Low	Low	Low
Bao 2014[95]	Unclear	Unclear	Unclear	Unclear	Low	Low	Unclear	Low
Tam 2014[97]	Low	Unclear	Unclear	Unclear	Low	Low	Low	Low
Infliximab vs. placebo								
Barkham 2009[50]	Unclear	Unclear	High	High	Low	Low	Low	Low
Braun 2002[98]	Low	Low	Low	Low	Low	Low	Low	Low
Marzo-Ortega 2005[100]	Low	Low	Unclear	Low	Low	Low	Unclear	Low
Van den Bosch 2002[101]	Unclear	Unclear	High	High	Low	Low	Low	Low
ASSERT[102]	Unclear	Unclear	Low	Low	Low	Low	Low	Low
Infliximab vs. biosimilar infliximab (InflectraCT-P13)								
PLANETAS 2013[110]	Low	Low	Unclear	Low	Low	Low	Low	Low

ASSERT, Ankylosing Spondylitis Study for the Evaluation of Recombinant Infliximab Therapy; ATLAS, Adalimumab Trial Evaluating Long-term Efficacy and Safety for Ankylosing Spondylitis.
a Judged to be likely to be because of chance.

The best model

Models were run when it was assumed that:

- there were different independent treatment effects
- there was just one treatment class effect.

In addition, fixed-effect and random-effects models were run when there were sufficient data. These models relate to models A1, A2, A3 and A4 in *Chapter 5*. For the non-radiographic population, there were very few studies and therefore only fixed-effect analyses were conducted.

The DIC and I^2 results for each outcome (other than injection site reactions) are shown for the AS population in *Table 5* and for the nr-AxSpA population in *Table 6*. The lower the DIC for a given outcome, the better the model fit. I^2 varies between 0% and 100%, with 0% representing no heterogeneity in the results and 100% indicating that all of the variation in the results can be explained by heterogeneity. The greater the value of I^2, the more likely it is that a random-effects model would be a better fit. But this is not always the case, as if there are few studies then there will be significant uncertainty around the between-study variance and therefore the I^2 also. Random-effect model results and I^2 results are not presented for some outcomes because of sensitivity to prior distributions in the model.

Overall, assuming a class effect for the treatments produced a better-fitting model than assuming independent treatment effects. In addition, a fixed-effect analysis was more often than not appropriate. The mean and median effects of the two analyses were also similar. Hence the fixed-effect results are reported in this chapter; these represent a common class effect.

For AS, the common class-effect model was found to be a much better fit than the independent treatment effect model. As described in *Chapter 5* the exchangeable class-effect model, not explored here, also fitted the data well, although not so well as the common class-effect model. It should be noted here that the common class-effect model may possibly underestimate the uncertainty around the treatment effect estimate. As explained in *Chapter 5*, if the differences between treatments are a result of systematic

TABLE 5 The AS population model DIC statistics

	Independent effects		Class effect		
Outcome	Fixed effect	Random effects	Fixed effect	Random effects	I^2 (%)
Related model in *Chapter 5*	A1	A2	A3	A4	–
BASDAI 50	16.82	–	10.86	12.71	21
BASDAI	16.76	18.22	13.53	15.12	21
BASFI	18.96	20.87	14.79	16.80	10
ASAS 20	10.68	17.05	9.98	8.73	16
ASAS 40	10.36	14.07	8.50	10.29	27
ASAS 50	8.38	–	6.68	8.11	52
ASAS 70	2.92	–	–	–	–
BASMI	–0.87	–	0.12	–3.01	77
SF-36 PCS	19.64	–	20.20	17.71	76
MASES	5.99	–	4.17	–	–
SF-36 MCS	19.20	–	16.67	18.26	47

MASES, Maastricht Ankylosing Spondylitis Enthesitis Score.

TABLE 6 The nr-AxSpA population model DIC statistics

	Independent effects		Class effect		
Outcome	Fixed effect	Random effects	Fixed effect	Random effects	I^2 (%)
Related model in *Chapter 5*	A1	A2	A3	A4	–
BASDAI 50	6.74	–	4.85	–	–
BASDAI	10.80	–	11.07	11.51	69
BASFI	11.45	–	13.74	10.70	83
ASAS 20	6.72	–	5.23	–	–
ASAS 40	11.17	–	7.96	9.30	49
ASAS 50	–	–	–	–	–
ASAS 70	–	–	–	–	–
BASMI	1.80	–	4.74	2.42	89
SF-36 PCS	16.67	–	20.18	–	–
MASES	–	–	–	–	–
SF-36 MCS	14.61	–	14.08	–	–

MASES, Maastricht Ankylosing Spondylitis Enthesitis Score.

differences in study design between treatments, then an exchangeable class-effect model may be appropriate. However, if in fact there is a true difference between treatments, such as between infliximab and the other TNF-inhibitors, then an exchangeable class-effect model may overestimate the uncertainty around the effect estimates. As the common class-effect model had a lower DIC than the exchangeable class-effect model, this is the model evaluated in this chapter. The economic model explores the assumption that treatment effect differences are in fact because of systematic differences in study design between treatments.

As there was very little difference between the results in which change from baseline was imputed assuming a within-study correlation of 0.3 or 0.7, only the results assuming a within-study correlation of 0.3 are reported here. A comparison of the results assuming different within-study correlations is presented in *Appendix 5*.

Individual anti-tumour necrosis factors compared with placebo

Binary responder outcomes at between 10 and 16 weeks

The results of the analyses of the responder outcomes between 10 and 16 weeks for patients with AS are presented in *Table 7*.

Assessment in Ankylosing Spondylitis improvement criteria: Assessment in Ankylosing Spondylitis 20, Assessment in Ankylosing Spondylitis 40, Assessment in Ankylosing Spondylitis 50 and Assessment in Ankylosing Spondylitis 70

For the AS population ASAS 20 data were available for all five anti-TNFs, although the number of participants studied varied considerably, ranging from 839 patients in five etanercept trials to 111 patients in two infliximab trials. A consistent effect was evident across the treatments with the pooled relative risks ranging from 1.80 (certolizumab pegol) to 2.45 (infliximab). ASAS 40 data were available for four anti-TNFs (no data were available for infliximab); the number of data available ranged from 178 patients in one certolizumab trial to 659 patients in two adalimumab trials. Again, a consistent effect was found, with relative risks ranging from 2.53 (certolizumab pegol) to 3.42 (adalimumab); all the relative risks were

TABLE 7 Results versus placebo for AS population: response outcomes at between 10 and 16 weeks

Intervention	Type of analysis	ASAS 20			ASAS 40			ASAS 50			BASDAI 50		
		Number of trials (number of patients)	Relative risk (95% CrI)	OR (95% CrI)	Number of trials (number of patients)	Relative risk (95% CrI)	OR (95% CrI)	Number of trials (number of patients)	Relative risk (95% CrI)	OR (95% CrI)	Number of trials (number of patients)	Relative risk (95% CrI)	OR (95% CrI)
Adalimumab	Main	3 (741)	2.28 (1.98 to 2.62)	4.52 (3.23 to 6.33)	2 (659)	3.42 (2.57 to 4.55)	5.67 (3.56 to 8.97)	1 (82)	2.75 (1.11 to 5.45)	3.58 (1.12 to 11.17)	2 (659)	3.16 (2.40 to 4.16)	4.68 (3.14 to 7.03)
	Sensitivity	3 (741)	2.27 (1.97 to 2.62)	4.52 (3.23 to 6.33)	2 (659)	3.34 (2.53 to 4.40)	5.67 (3.56 to 8.97)	As in the main analysis			2 (659)	3.11 (2.37 to 4.09)	4.68 (3.14 to 7.03)
Certolizumab pegol	Main	1 (178)	1.80 (1.24 to 2.39)	2.61 (1.37 to 5.01)	1 (178)	2.53 (1.47 to 3.98)	3.38 (1.59 to 7.15)	–	–	–	1 (178)	3.60 (2.02 to 5.74)	5.97 (2.39 to 15.03)
	Sensitivity	1 (178)	1.80 (1.24 to 2.39)	2.61 (1.37 to 5.01)	1 (178)	2.49 (1.46 to 3.87)	3.38 (1.59 to 7.15)	–	–	–	1 (178)	3.53 (2.00 to 5.58)	5.97 (2.39 to 15.03)
Etanercept	Main	5 (839)	2.23 (1.93 to 2.55)	4.23 (3.05 to 5.88)	3 (478)	2.75 (1.88 to 3.88)	3.86 (2.21 to 6.72)	2 (359)	3.43 (2.40 to 4.90)	5.04 (2.98 to 8.51)	3 (478)	3.17 (2.20 to 4.49)	4.74 (2.71 to 8.28)
	Sensitivity	3 (715)	2.17 (1.84 to 2.53)	3.98 (2.78 to 5.73)	2 (436)	2.65 (1.80 to 3.72)	3.72 (2.11 to 6.53)	As in the main analysis			2 (436)	3.03 (2.08 to 4.31)	4.50 (2.52 to 8.01)
Golimumab	Main	2 (429)	2.14 (1.75 to 2.53)	3.82 (2.47 to 5.86)	2 (429)	3.11 (2.24 to 4.26)	4.77 (2.85 to 7.98)	–	–	–	2 (429)	3.57 (2.51 to 5.00)	5.85 (3.31 to 10.28)
	Sensitivity	2 (429)	2.13 (1.74 to 2.53)	3.82 (2.47 to 5.86)	2 (429)	3.05 (2.21 to 4.13)	4.77 (2.85 to 7.98)	–	–	–	2 (429)	3.50 (2.48 to 4.88)	5.85 (3.31 to 10.28)
Infliximab	Main	2 (111)	2.45 (1.73 to 3.06)	5.54 (2.41 to 12.71)	–	–	–	1 (69)	5.59 (2.44 to 9.81)	14.71 (3.07 to 72.69)	1 (69)	4.86 (2.41 to 7.82)	12.07 (3.09 to 46.37)
	Sensitivity	2 (111)	2.44 (1.72 to 3.06)	5.54 (2.41 to 12.71)	–	–	–	As in the main analysis			1 (69)	4.72 (2.38 to 7.54)	12.07 (3.09 to 46.37)
Anti-TNFs as a class	Main	13 (2298)	2.21 (2.01 to 2.43)	4.12 (3.40 to 4.99)	8 (1744)	3.06 (2.52 to 3.76)	4.61 (3.51 to 6.05)	4 (510)	3.51 (2.55 to 4.86)	5.23 (3.31 to 8.27)	9 (1813)	3.37 (2.75 to 4.16)	5.22 (4.00 to 6.79)
	Sensitivity	11 (2174)	2.18 (1.97 to 2.42)	4.04 (3.32 to 4.92)	7 (1702)	2.99 (2.47 to 3.66)	4.57 (3.48 to 6.02)	As in the main analysis			8 (1771)	3.29 (2.68 to 4.07)	5.16 (3.94 to 6.72)

greater than the corresponding ASAS 20 estimates. For ASAS 50 there were two trials of etanercept (totalling 359 participants) and small single trials in adalimumab ($n = 82$) and infliximab ($n = 69$). A wider range of relative risks and CrIs resulted, ranging from 2.75 (adalimumab) to 5.59 (infliximab), which may be a consequence of the smaller numbers of patients studied. Only two trials, both of etanercept ($n = 359$), reported actual numbers of ASAS 70 responders. The pooling of these data showed that patients taking etanercept were more than three times more likely to be ASAS 70 responders than patients taking placebo (relative risk 3.59, 95% CrI 2.18 to 5.87).

For the nr-AxSpA population, each of the relative risks for certolizumab pegol and etanercept were based on single, quite large trials; the estimate for adalimumab was based on a similar number of patients (to etanercept and certolizumab) across two trials, whereas infliximab was represented by a single small trial ($n = 40$). ASAS 20 results were similar across treatments but for ASAS 40 heterogeneity of effect appeared evident; the smallest estimate was for etanercept and the largest estimate was seen in the small infliximab trial (*Table 8*). However, this infliximab trial was the only nr-AxSpA trial judged to be at high risk of bias. Only one trial (ABILITY-1[58]) reported ASAS 50 or ASAS 70 results. For ASAS 50 the relative risk was 4.23 (95% CrI 1.84 to 9.72; OR 5.96, 95% CrI 2.40 to 14.80). For ASAS 70 the relative risk was 4.58 (95% CrI 1.37 to 15.40; OR 5.42, 95% CrI 1.54 to 19.11).

Bath Ankylosing Spondylitis Disease Activity Index 50

For the AS population BASDAI 50 data were available for all five anti-TNFs; the number of participants studied varied widely, ranging from 69 patients in one infliximab trial to 659 patients in two adalimumab trials. Although a consistent beneficial effect was evident across treatments, some heterogeneity of effect could be seen with the relative risks ranging from 3.16 (adalimumab) to 4.86 (infliximab).

For the nr-AxSpA population the relative risks were lower than for the AS population being 2.52 (95% CrI 1.65 to 3.83, two trials) for adalimumab, 2.80 (95% CrI 1.71 to 4.47, one trial) for certolizumab and 1.92 (95% CrI 1.27 to 2.82, one trial) for etanercept (see *Table 8*).

Results of the AS sensitivity analyses were very similar to those of the main analyses (see *Table 7*).

Continuous outcomes at between 10 and 16 weeks

The results of the analyses of the continuous efficacy outcomes for patients with AS are presented in *Table 9*.

For the AS population, when compared with placebo, adalimumab ($n = 705$), certolizumab pegol ($n = 178$), etanercept ($n = 483$) and infliximab ($n = 132$) produced statistically significant reductions in disease activity, when assessed using BASDAI. The magnitude of the reductions in change from baseline BASDAI score ranged from 1.46 units (certolizumab pegol) to 2.28 units (infliximab). None of the three golimumab trials reported BASDAI as a continuous outcome. The number of data available for BASFI in patients with AS ranged from 132 patients in three infliximab trials, to 523 patients in five etanercept trials. When compared with placebo, all five anti-TNFs produced statistically significant improvements in function. The magnitude of the reductions in change from baseline BASFI score ranged from 1.1 units (certolizumab pegol) to 2.16 units (infliximab). When compared with placebo, statistically significant improvements in BASMI scores were found for AS patients taking adalimumab (mean difference in change from baseline –0.37 units, 95% CrI –0.50 to –0.23 units) and etanercept (mean difference in change from baseline –0.37 units, 95% CrI –0.65 to –0.09 units) but not for certolizumab pegol (mean difference in change from baseline –0.26 units, 95% CrI –0.55 to 0.03 units) and golimumab (mean difference in change from baseline –0.11 units, 95% CrI –0.26 to 0.04 units). Results for SF-36 MCS, SF-36 PCS and ethesitis [Maastricht Ankylosing Spondylitis Enthesitis Score (MASES)] are presented in *Table 9*.

TABLE 8 Results vs. placebo for nr-AxSpA population: response outcomes at 10–16 weeks

Intervention	ASAS 20			ASAS 40			BASDAI 50		
	Number of trials (number of patients)	Relative risk (95% CrI)	OR (95% CrI)	Number of trials (number of patients)	Relative risk (95% CrI)	OR (95% CrI)	Number of trials (number of patients)	Relative risk (95% CrI)	OR (95% CrI)
Adalimumab	2 (188)	1.92 (1.47 to 2.56)	3.71 (2.02 to 6.75)	2 (188)	3.14 (1.99 to 4.68)	5.04 (2.44 to 10.32)	2 (188)	2.52 (1.65 to 3.83)	3.97 (1.97 to 7.86)
Certolizumab pegol	1 (147)	1.59 (1.10 to 2.21)	2.32 (1.15 to 4.67)	1 (147)	3.04 (1.74 to 4.81)	4.75 (2.01 to 11.17)	1 (147)	2.80 (1.71 to 4.47)	4.92 (2.09 to 11.58)
Etanercept	1 (215)	1.46 (1.08 to 1.94)	1.94 (1.13 to 3.37)	1 (215)	2.07 (1.26 to 3.20)	2.55 (1.32 to 4.92)	1 (215)	1.92 (1.27 to 2.82)	2.45 (1.37 to 4.43)
Infliximab	–	–	–	1 (40)	3.63 (1.41 to 6.44)	6.85 (1.52 to 31.03)	–	–	–
Anti-TNFs as a class	4 (550)	1.65 (1.37 to 2.04)	2.52 (1.78 to 3.59)	5 (590)	2.74 (2.08 to 3.62)	3.92 (2.61 to 5.91)	4 (550)	2.31 (1.76 to 3.10)	3.33 (2.24 to 4.96)

TABLE 9 Results vs. placebo for AS population: continuous outcomes at 10–16 weeks

		BASDAI score		BASFI score	
Intervention	Type of analysis	Number of trials (number of patients)	Mean difference in change from baseline (95% CrI)	Number of trials (number of patients)	Mean difference in change from baseline (95% CrI)
Adalimumab	Main	3 (705)	−1.55 (−1.88 to −1.22)	2 (390)	−1.25 (−1.63 to −0.87)
	Sensitivity	2 (659)	−1.55 (−1.89 to −1.21)	1 (344)	−1.28 (−1.68 to −0.88)
Certolizumab pegol	Main	1 (178)	−1.46 (−2.17 to −0.74)	1 (178)	−1.10 (−1.83 to −0.37)
	Sensitivity	Same as the main analysis		Same as the main analysis	
Etanercept	Main	4 (483)	−1.75 (−2.14 to −1.37)	5 (523)	−1.43 (−1.82 to −1.04)
	Sensitivity	2 (359)	−1.72 (−2.16 to −1.29)	2 (359)	−1.29 (−1.76 to −0.84)
Golimumab	Main	–	–	2 (429)	−1.45 (−1.84 to −1.05)
	Sensitivity	–	–	Same as the main analysis	
Infliximab	Main	3 (132)	−2.28 (−3.18 to −1.38)	3 (132)	−2.16 (−3.18 to −1.12)
	Sensitivity	2 (111)	−2.18 (−3.14 to −1.21)	2 (111)	−1.94 (−3.07 to −0.80)
Anti-TNFs as a class	Main	11 (1498)	−1.66 (−1.88 to −1.43)	13 (1652)	−1.38 (−1.59 to −1.18)
	Sensitivity	7 (1305)	−1.63 (−1.88 to −1.39)	8 (1419)	−1.34 (−1.57 to −1.12)

MASES, Maastricht Ankylosing Spondylitis Enthesitis Score.

BASMI score		SF-36 PCS score		SF-36 MCS score		MASES	
Number of trials (number of patients)	Mean difference in change from baseline (95% CrI)	Number of trials (number of patients)	Mean difference in change from baseline (95% CrI)	Number of trials (number of patients)	Mean difference in change from baseline (95% CrI)	Number of trials (number of patients)	Mean difference in change from baseline (95% CrI)
2 (659)	–0.37 (–0.50 to –0.23)	2 (659)	3.53 (2.37 to 4.68)	2 (659)	1.41 (–0.19 to 3.02)	2 (659)	–0.50 (–0.89 to –0.11)
Same as the main analysis		Same as the main analysis		Same as the main analysis		Same as the main analysis	
1 (178)	–0.26 (–0.55 to 0.03)	1 (178)	5.64 (3.64 to 7.66)	1 (178)	1.25 (–2.08 to 4.61)	–	–
Same as the main analysis		Same as the main analysis		Same as the main analysis		–	–
1 (82)	–0.37 (–0.65 to –0.09)	–		–	–	–	–
Same as the main analysis		–		–	–	–	–
2 (429)	–0.11 (–0.26 to 0.04)	2 (429)	5.06 (3.71 to 6.40)	2 (429)	2.75 (1.08 to 4.40)	1 (216)	–0.70 (–1.53 to 0.11)
Same as the main analysis		Same as the main analysis		Same as the main analysis		Same as the main analysis	
–	–	–	–	–	–	–	–
–	–	–	–	–	–	–	–
6 (1348)	–0.27 (–0.36 to –0.18)	5 (1266)	4.40 (3.60 to 5.21)	5 (1266)	1.93 (0.12 to 3.72)	3 (875)	–0.54 (–0.89 to –0.19)
Same as the main analysis		Same as the main analysis		Same as the main analysis		Same as the main analysis	

For the nr-AxSpA population, a heterogeneity of effect on BASDAI and BASFI appears evident from the relative risks of the individual anti-TNFs. The smallest estimates were for etanercept and the largest estimates were seen in the small infliximab trial, although this trial was the only nr-AxSpA trial judged to be at high risk of bias (*Table 10*).

Results of the AS sensitivity analyses were very similar to those of the main analyses (see *Table 9*).

When the mean baseline BASDAI and BASFI are presented by treatment response at week 12 (or 14 for golimumab) for three of the five anti-TNFs (see *Appendix 6*), it can be seen that in patients with AS and patients with nr-AxSpA, on average baseline BASDAI does not differ greatly between responders and non-responders to either placebo or active anti-TNF therapy. In patients with AS or nr-AxSpA from the trials of adalimumab [ATLAS (Adalimumab Trial Evaluating Long-term Efficacy and Safety for Ankylosing Spondylitis)[61] and ABILITY-1[58]] and golimumab (GO-RAISE[90]) on average baseline BASFI was higher in non-responders compared with responders. However, this was not seen in the etanercept trials.

Individual anti-tumour necrosis factors compared with each other

For efficacy outcomes, all of the comparisons that could be made between different anti-TNFs at 10–16 weeks resulted in no statistically significant differences between treatments. For the full results see *Appendix 7*.

One small trial, which could not be included in the meta-analysis (see *Study characteristics*), compared infliximab with etanercept in a 2-year unblinded randomised study of 50 AS patients.[88] At 12 weeks there were statistically significant differences between groups in terms of BASDAI score (3.5 vs. 5.6; $p < 0.005$) and BASFI score (3.5 vs. 5; $p < 0.005$), favouring treatment with infliximab. By week 48, the BASDAI and BASFI scores were almost identical across the treatment groups (data were only presented graphically). In addition, at 12 weeks 19 of 25 infliximab patients were ASAS 20 responders compared with 15 of 25 etanercept patients (not a statistically significant difference). This study concluded that infliximab produces a more rapid clinical improvement, but, at the end of the study, treatment with both etanercept and infliximab was effective and safe. The results of this trial may explain why at 10–16 weeks the meta-analysis results for infliximab were a little better than those of the other anti-TNFs.

Another trial which could not be included in the meta-analysis compared infliximab with an infliximab biosimilar called CT-P13 in 250 AS patients.[110] The ASAS 40 response rates at week 14 were 42% for CT-P13 and 46% for infliximab [OR 0.85; 95% confidence interval (CI) 0.51 to 1.42] and at week 30 they were 52% for CT-P13 and 47% for infliximab (OR 1.19, 95% CI 0.70 to 2.00). At week 14 BASDAI median change from baseline scores were identical (–2.7) and at week 30 they differed slightly (–3.1 CT-P13 vs. –2.5 infliximab). For BASFI the median change from baseline scores were –2.2 CT-P13 versus –2.4 infliximab at week 14 and –2.6 CT-P13 versus –2.2 infliximab at week 30. The study concluded that CT-P13 had a comparable efficacy and safety profile with that of infliximab.

Anti-tumour necrosis factors as a class compared with placebo

Within this section the class effect, calculated as a common effect across all the TNF-inhibitors under consideration, assumes a single treatment effect for all the TNF-inhibitors. It is calculated as the pooled treatment effect using a fixed effect model. The common class-effect model may possibly underestimate the uncertainty around the treatment effect estimate. As explained in *Chapter 5*, if the differences between treatments is a result of systematic differences in study design between treatments then an exchangeable class-effect model may be appropriate. However, if in fact there is a true difference between treatments, such as between infliximab and the other TNF-inhibitors, then an exchangeable class-effect model may overestimate the uncertainty around the mean class-effect estimates. As the common class-effect model had a lower DIC than the exchangeable class-effect model, this is the model evaluated in this chapter. The economic model in *Chapter 6* explores the assumption that treatment effect differences are because of differences in study design between treatments.

TABLE 10 Results vs. placebo for nr-AxSpA population: continuous outcomes at between 10 and 16 weeks

Intervention	BASDAI score		BASFI score		BASMI score		SF-36 PCS score		SF-36 MCS score	
	Number of trials (number of patients)	Mean difference in change from baseline (95% CrI)	Number of trials (number of patients)	Mean difference in change from baseline (95% CrI)	Number of trials (number of patients)	Mean difference in change from baseline (95% CrI)	Number of trials (number of patients)	Mean difference in change from baseline (95% CrI)	Number of trials (number of patients)	Mean difference in change from baseline (95% CrI)
Adalimumab	2 (188)	–1.23 (–1.83 to –0.62)	2 (188)	–0.90 (–1.44 to –0.36)	2 (188)	–0.02 (–0.24 to 0.20)	2 (188)	4.98 (2.74 to 7.20)	2 (188)	1.13 (–1.86 to 4.13)
Certolizumab pegol	1 (147)	–1.85 (–2.83 to –0.88)	1 (147)	–1.90 (–2.87 to –0.94)	1 (147)	–0.55 (–0.89 to –0.20)	1 (147)	6.99 (4.23 to 9.76)	1 (147)	4.01 (0.44 to 7.53)
Etanercept	1 (215)	–0.70 (–1.54 to 0.12)	1 (215)	–0.60 (–1.16 to –0.06)	–	–	–	–	–	–
Infliximab	1 (40)	–2.67 (–4.21 to –1.13)	1 (40)	–2.24 (–3.67 to –0.80)	1 (40)	0.00 (–0.44 to 0.44)	1 (40)	2.10 (–0.21 to 4.37)	–	–
Anti-TNFs as a class	5 (590)	–1.32 (–1.74 to –0.90)	5 (590)	–0.99 (–1.34 to –0.64)	4 (375)	–0.15 (–0.32 to 0.02)	4 (375)	4.41 (3.04 to 5.81)	3 (335)	2.33 (0.07 to 4.62)

Binary responder outcomes at between 10 and 16 weeks

Assessment in Ankylosing Spondylitis improvement criteria – ASAS 20, ASAS 40, ASAS 50 and ASAS 70 When compared with placebo, anti-TNFs as a common class were more than twice as likely to result in patients with AS achieving an ASAS 20 response (relative risk 2.21, 95% CrI 2.01 to 2.43; 13 trials, see *Table 7*). Anti-TNFs were also around three times as likely to result in patients achieving an ASAS 40 response (relative risk 3.06, 95% CrI 2.52 to 3.76; eight trials) and three and a half times as likely to result in patients achieving an ASAS 50 response (relative risk 3.51, 95% CrI 2.55 to 4.86; four trials). Only two trials, both of etanercept, reported data suitable for the ASAS 70 analysis; the results are presented in *Individual anti-TNFs compared with placebo*. There was little evidence of heterogeneity for ASAS 20 ($I^2 = 16\%$) and ASAS 40 ($I^2 = 27\%$) but heterogeneity was evident for ASAS 50 ($I^2 = 52\%$). For ASAS 50, three of the four trials were small (i.e. fewer than 100 participants), which may partly explain the heterogeneity estimate.

For the nr-AxSpA population anti-TNFs as a common class were statistically significantly more effective than placebo, although the relative risks being lower than for the AS population. For ASAS 20 the relative risk was 1.65 (95% CrI 1.37 to 2.04; four trials) and for ASAS 40 the relative risk was 2.74 (95% CrI 2.08 to 3.62; five trials). Only one trial presented ASAS 50 and ASAS 70 results (see *Clinical effectiveness results: efficacy results from randomised controlled trials*). A heterogeneity estimate could be calculated for only ASAS 40 ($I^2 = 49\%$).

BASDAI 50 Anti-TNFs as a common class resulted in patients with AS being more than three times more likely to achieve a BASDAI 50 response than patients taking placebo (relative risk 3.37, 95% CrI 2.75 to 4.16; nine trials). There was little evidence of heterogeneity ($I^2 = 21\%$).

For the nr-AxSpA population, anti-TNFs as a common class were also statistically significantly more effective than placebo in terms of achieving a BASDAI 50, although the relative risk was lower than for the AS population (relative risk 2.31, 95% CrI 1.76 to 3.10; four trials). Results of the AS sensitivity analyses were very similar to the main analyses (see *Table 7*).

Binary responder outcomes at between 24 and 30 weeks

Four AS trials reported outcomes at between 24 and 30 weeks (see *Table 2*). Anti-TNFs as a common class were statistically significantly more effective than placebo at 24–30 weeks; for ASAS 20 the relative risk was 1.69 (95% CrI 1.30 to 2.14; four trials). No studies reported BASDAI 50 or ASAS 70 results, and only single studies reported on ASAS 40 (relative risk 4.01, 95% CrI 2.13 to 7.55)[102] and ASAS 50 (relative risk 4.17, 95% CrI 2.45 to 7.12).[72]

Continuous outcomes at between 10 and 16 weeks

When considered together as a group compared with placebo (see *Table 9*), treatment with an anti-TNF in patients with AS produced statistically significant improvements (calculated using mean difference in change from baseline) in the following areas: disease activity (BASDAI mean difference –1.66 units, 95% CrI –1.88 to –1.43 units; 11 trials); function (BASFI mean difference –1.38 units, 95% CrI –1.59 to –1.18 units; 13 trials); spinal mobility (BASMI mean difference –0.27 units, 95% CrI –0.36 to –0.18 units); physical health (SF-36 PCS mean difference 4.40 units, 95% CrI 3.60 to 5.21 units; five trials); mental health (SF-36 MCS mean difference 1.96 units, 95% CrI 0.87 to 3.05 units; five trials); and enthesitis (MASES mean difference –0.54 units, 95% CrI –0.89 to –0.19 units; three trials). There was little evidence of heterogeneity for BASDAI ($I^2 = 21\%$) and BASFI ($I^2 = 10\%$), but evidence of substantial heterogeneity for BASMI ($I^2 = 77\%$), SF-36 PCS ($I^2 = 76\%$), SF-36 MCS ($I^2 = 47\%$) and MASES ($I^2 = 91\%$).

In the nr-AxSpA population the mean differences achieved with anti-TNFs (see *Table 10*) were also statistically significant, although slightly lower than for the AS population. For BASDAI the mean difference was –1.32 units (95% CrI –1.74 to –0.90; $I^2 = 69\%$) and for BASFI the mean difference was –0.99 units (95% CrI –1.34 to –0.64 units; $I^2 = 83\%$) but there was evidence of substantial heterogeneity. The results for SF-36 MCS and SF-36 PCS were similar to those for AS (see *Table 10*).

Results of the AS sensitivity analyses were very similar to the main analyses (see *Table 9*). As the results of the independent treatment effects showed a trend that infliximab had a greater, although not statistically significant, effect on the change in BASDAI and BASFI from baseline, an additional sensitivity analysis was conducted for which infliximab was assumed to be different from the rest of the anti-TNFs. The results are presented in *Table 11*. The low weight of evidence available for infliximab ensures that the class effect for the other anti-TNFs does not change greatly. Although it is possible that infliximab has a greater effect than the other anti-TNFs at least at 12 weeks, there is no strong evidence from these analyses to suggest that it does.

Continuous outcomes at between 24 and 30 weeks

Four AS trials reported outcomes at between 24 and 30 weeks (see *Table 2*). The mean differences in change from baseline were −1.98 units (95% CrI −2.27 to −1.68 units, four trials) for BASDAI, −0.87 units (95% CrI −1.11 to −0.62 units; three trials)[72,100,102] for BASFI, and −1.00 unit (95% CrI −1.19 to −0.81 units; two studies)[97,102] for BASMI. One study reported SF-36 outcomes, with differences of 9.40 units (95% CrI 7.88 to 10.92 units) for SF-36 PCS and 0.70 units (95% CrI −1.36 to 2.76 units) for SF-36 MCS.[102]

Outcomes not included in the meta-analyses

Very few data were available on peripheral symptoms (other than enthesitis, see the MASES results in *Table 9*) or symptoms of extra-articular manifestations. One trial reported five cases of inflammatory bowel disease flare up to the 24-week time point: three occurred in patients on etanercept and two in patients on placebo.[72] Another study reported that there were no cases of inflammatory bowel disease at 12 weeks.[86] Incidence of uveitis was also reported in one trial; up to the 24-week time point there were three cases in the etanercept arm and eight cases in the placebo arm.[72]

One trial (ABILITY-1[58]) reported statistically significantly improved quality of life, using EQ-5D index scores, in patients taking adalimumab [change from baseline 0.15 units (SD 0.30 units)] when compared with those taking placebo [change from baseline 0.06 units (SD 0.28 units)]. A study of adalimumab reported no statistically significant difference in EQ-5D between groups at 12 weeks (0.78 units for adalimumab vs. 0.72 units for placebo; $p = 0.32$).[51]

For Ankylosing Spondylitis Quality of Life (ASQoL), a quality of life instrument specific to AS, ATLAS[61] was the only trial which reported results together with SDs or standard errors (SEs); significant improvements were found favouring treatment with adalimumab at week 12 [mean change from baseline −3.2 units (SD 0.3 units) for adalimumab vs. −1 unit (SD 0.4 units) for placebo].[62] Similar statistically significant results were reported in an etanercept trial at 12 weeks (mean change from baseline −3.3 units for etanercept vs. −0.1 units for placebo; $p = 0.02$)[71] and in an infliximab trial at 16 weeks (mean change from baseline −6.2 units for infliximab vs. −1 unit for placebo; $p = 0.007$).[50] Another small study of infliximab did not find a significant difference between groups at 30 weeks ($p = 0.14$).[100]

TABLE 11 The difference in change from baseline for BASDAI and BASFI scores assuming all TNFs have the same effect and assuming infliximab may be different

	BASDAI score		BASFI score	
Category of intervention	Mean	95% CrI	Mean	95% CrI
All TNFs	−1.66	−1.88 to −1.43	−1.38	−1.59 to −1.18
TNFs other than infliximab	−1.62	−1.85 to −1.38	−1.35	−1.56 to −1.14
Infliximab	−2.28	−3.18 to −1.38	−2.15	−3.18 to −1.11

'Placebo' response in ankylosing spondylitis and non-radiographic axial spondyloarthritis

To inform insight into the extent of any 'placebo' effects (outlined in *Chapter 1, Description of health problem*), *Table 12* compares the placebo response rates in trials which reported ASAS 20 results and at least one of ASAS 40 or BASDAI 50 results. These data highlight the relatively high rates of ASAS 20 response (median 31%, range 21–40%) when compared with ASAS 40 response (median 15%, range 10–23%) and BASDAI 50 response (median 16%, range 5–24%).

However, the extent of the 'placebo' response on the ASAS 20 results might result in an underestimation of anti-TNF efficacy, notably when ASAS 20 is the only ASAS improvement outcome reported in a trial. An increase in the likelihood of being a responder (i.e. the relative risks when compared with placebo) when moving up the ASAS thresholds seems apparent from the results in *Clinical effectiveness results: efficacy results from randomised controlled trials*. This might be explained by considering the subset of patients who achieve an ASAS 20 response largely because of regression to the mean (i.e. because of natural variation in repeated data measurements, such as patients transitioning from flare at randomisation to no flare at 12 weeks). For those patients who experience regression to the mean after taking an anti-TNF, the true benefit of treatment may be hidden in the ASAS 20 outcome for some patients, and the proportion of ASAS 20 responders might therefore differ only moderately between the anti-TNF and placebo groups. As the bar for response is raised, from ASAS 20 through to ASAS 70, this difference in the proportion of responders between active treatment and placebo groups is likely to increase as an effect because regression to the mean becomes less probable. The diluting effect of a placebo response on the relative risks therefore diminishes as the ASAS thresholds increase (and more informative estimates of treatment benefit can be seen). Regardless of the reason, these results highlight the limited applicability of ASAS 20 as a clinically informative outcome measure. ASAS 20 was nevertheless the most commonly reported responder outcome across the trials.

Summary of the randomised controlled trial clinical efficacy results

For both the AS and nr-AxSpA populations the results of the meta-analyses demonstrated that anti-TNFs produce statistically significant and clinically relevant benefits to patients in terms of improving function and reducing disease activity. The common class-effect model used may have underestimated the uncertainty in the effect estimates. Although there is a possibility that infliximab is more effective than other TNF inhibitors, at least at 12 weeks, there is no strong evidence to support this. For the disease activity, function and responder outcomes, the class-efficacy estimates were consistently slightly smaller for nr-AxSpA than for AS, most noticeably for BASFI and BASDAI 50.

The included RCTs were generally subject to low risks of bias and no important variation in baseline characteristics was evident, with the exception of CRP levels: in the nr-AxSpA trial populations CRP levels were much lower than in the AS populations. Although heterogeneity of CRP levels was evident across both the AS trials and the nr-AxSpA trials, in almost all the AS trials the CRP levels were higher than the 14 mg/l threshold identified as being a key predictor of treatment response (in AS, higher CRP levels are associated with an increased likelihood of BASDAI 50 response).[112] In the nr-AxSpA trials only the RAPID-axSpA[64] population came close to this cut-off point. These lower CRP levels may therefore have had an impact on the efficacy estimates for the nr-AxSpA population.

Statistical heterogeneity was more apparent in the nr-AxSpA analyses than in the AS analyses. This may be a result of both clinical heterogeneity in the nr-AxSpA trials (such as variation in CRP thresholds, or the proportion of MRI positive patients) and the fact that fewer studies were available for analysis. In the light of the statistical heterogeneity across the nr-AxSpA trials, both the reliability of the nr-AxSpA-pooled estimates and their true relevance to patients seen in clinical practice are questionable.

The clinical relevance of the efficacy of anti-TNFs can be evaluated in part by considering the literature on minimum clinically important differences (MCIDs) or minimum clinically important improvements. In a study of 125 AS patients, Pavy *et al.*[113] reported a MCID of 1 unit (or a 20% relative change) for BASDAI and

TABLE 12 Comparison of placebo response rates in trials reporting ASAS 20 results together with ASAS 40 or BASDAI 50 results

Population and study	Placebo compared with	Time point (weeks)	Number of patients on placebo	Number of responders			% of responders			Difference in response (%)		
				ASAS 20	ASAS 40	BASDAI 50	ASAS 20	ASAS 40	BASDAI 50	ASAS 20 vs. ASAS 40	ASAS 20 vs. BASDAI 50	ASAS 40 vs. BASDAI 50
Nr-axSpA[51]	Adalimumab	12	24	6	3	5	25	13	21	13	4	–8
AS[55]	Adalimumab	12	115	35	11	19	30	10	17	21	14	–7
Nr-axSpA[58]	Adalimumab	12	73	23	10	10	32	14	14	18	18	0
AS[61]	Adalimumab	12	107	22	14	17	21	13	16	7	5	–3
AS[64]	Certolizumab	12	57	21	11	6	37	19	11	18	26	9
Nr-axSpA[64]	Certolizumab	12	50	20	8	8	40	16	16	24	24	0
AS[74]	Etanercept	12	43	14	10	10	33	23	23	9	9	0
Nr-axSpA[76]	Etanercept	12	109	39	17	26	36	16	24	20	12	–8
AS[85]	Etanercept	12	51	19	11	10	37	22	20	16	18	2
AS[90]	Golimumab	14	78	17	12	12	22	15	15	6	6	0
AS[95]	Golimumab	14	105	26	10	5	25	10	5	15	20	5
AS[98]	Infliximab	12	35	10	–	3	29	–	9	–	20	–

Any minor discrepancies in the difference in response columns are because of rounding.

0.7 units (17.5% relative change) for BASFI. All the effect estimates from this review for both BASDAI and BASFI were considerably higher than these MCIDs. The small effect on spinal mobility (a group effect reduction of around 0.3 BASMI units) appears unlikely to be clinically important.

Summary of some key issues arising from the Food and Drug Administration assessments of the ABILITY-1[58] and RAPID-axSpA[64] trials

The FDA Arthritis Advisory Committee met in July 2013 to discuss licence applications for adalimumab for patients with active nr-AxSpA (with objective signs of inflammation) and certolizumab pegol for patients with active axSpA, including patients with AS.[114] An important issue which arose in both trials was the differences in diagnoses arising from radiograph images evaluated centrally compared with images being evaluated locally. The implications for efficacy were explored via further analyses.

RAPID-axSpA[64] trial (certolizumab pegol)

This trial aimed to recruit both AS and nr-AxSpA patients.[64] The nr-AxSpA patients had to have a positive MRI or an elevated CRP level; the definition used for CRP level elevation was 7.9 mg/l.

Comparison of ankylosing spondylitis and non-radiographic axial spondyloarthritis population characteristics In AS males predominated (72%), whereas in nr-AxSpA the male-to-female ratio was roughly equal. The AS population had a mean age of 41.5 years, which was around four years older than the nr-AxSpA population. Baseline BASFI, BASMI and CRP levels suggested more functional and mobility impairment and more inflammation in the AS group when compared with the nr-AxSpA group. However, baseline back pain severity and BASDAI scores were similar between the AS and nr-AxSpA subgroups (*Table 13*).

Methods used to evaluate radiograph images In the trial, many patients had their disease reclassified when radiographs were evaluated centrally, rather than being evaluated locally. Two readers were involved in the central evaluation of the radiographs, they were blinded to both the assigned subgroup and the treatment group; a third reader was used in cases of disagreement. Twenty-one per cent of locally classified AS patients were reclassified as nr-AxSpA by central readers and 51% of locally classified nr-AxSpA patients were reclassified as AS by the central readers. Based on the central assessments 184 patients had AS and 98 patients had nr-AxSpA. Central reads could not be made for 43 patients as radiographs were not available (37 AS patients and six nr-AxSpA patients).

ABILITY-1[58] trial (adalimumab)

This trial intended to recruit only nr-AxSpA patients, although this included patients ($n = 43$) who had nr-AxSpA but neither a positive MRI nor an elevated CRP level.[58] The population with these 43 patients excluded is referred to as the 'adalimumab target population' (ATP). As in the RAPID-axSpA[64] trial, central rereading of radiographs was performed (in addition to local evaluation), although this was only done for per-protocol patients who also reached week 104 [$n = 102$ (out of 185) patients]. Thirty-eight of the 102 patients were identified as having AS rather than nr-AxSpA. The FDA statistician analysed the results in these 38 patients and compared them to those for patients with centrally confirmed nr-AxSpA. The FDA document reported results for the subpopulations based on local or central diagnosis, including ATP analyses.

Comparison of ankylosing spondylitis and non-radiographic axial spondyloarthritis results and impact of reclassification in the trials

For certolizumab pegol the FDA statistical review stated that 'efficacy findings were consistent in both AS and nr-AxSpA subpopulations regardless of the discrepancy in pelvic X-ray readings at local or central lab for modified New York criteria'[115] (*Table 14*).

For ABILITY-1[58] a notably higher proportion of patients in the AS subgroup responded to adalimumab (ASAS 40) than placebo compared with patients with confirmed nr-AxSpA. This suggests that the treatment benefit in the whole trial population may be driven by benefit in AS patients rather than in nr-AxSpA patients, skewing the results for the ATP (see *Table 14*). It should be noted, however, that this may be an atypical AS population; the trial had intended to recruit only nr-AxSpA patients.

TABLE 13 Baseline characteristics of trials analysed by the FDA

Trial and population	Characteristic									
	Age (years), mean	% male	Duration (years) of symptoms	Weight (kg), mean	% HLA-B27 positive	% on NSAIDs	CRP level	% MRI positive	BASDAI score, mean	BASFI score, mean
ABILITY-1,[58] nr-AxSpA ($n = 142$)	38	46	Median 8, mean 11	80	80	81	Median ≈4, mean 9	51	6	4.7
RAPID-axSpA,[64] nr-AxSpA ($n = 147$)	37	48	Median 5.5, mean 8.6	82	75	84	Median 11.9, mean 16	54	6.5	4.9
RAPID-axSpA,[64] AS ($n = 178$)	42	73	Median 9.1, mean 11.9	82	82	91	Median 14.3, mean 21.3	N/A	6.4	5.7

N/A, not applicable.

TABLE 14 Food and Drug Administration analyses: percentage differences from placebo, by method of diagnosis

Outcomes at week 12	ABILITY-1[58] ATP population		RAPID-axSpA[64]							
	Local laboratory nr-AxSpA[a]	Central laboratory nr-AxSpA[b]	Local laboratory nr-AxSpA[c]		Central laboratory nr-AxSpA[d]		Local laboratory AS[e]		Central laboratory AS[f]	
	Adalimumab, 40 mg, % (95% CI)	Adalimumab, 40 mg, % (95% CI)	Certolizumab, 200 mg, % (95% CI)	Certolizumab, 400 mg, % (95% CI)	Certolizumab, 200 mg, % (95% CI)	Certolizumab, 400 mg, % (95% CI)	Certolizumab, 200 mg, % (95% CI)	Certolizumab, 400 mg, % (95% CI)	Certolizumab, 200 mg, % (95% CI)	Certolizumab, 400 mg, % (95% CI)
ASAS 20	28 (12 to 44)	15 (–14 to 44)	19 (1 to 38)	23 (4 to 42)	23 (2 to 44)	23 (1 to 44)	20 (3 to 37)	27 (10 to 45)	17 (1 to 33)	23 (7 to 39)
ASAS 40	27 (13 to 41)	11 (–16 to 38)	32 (14 to 49)	31 (14 to 48)	18 (0 to 36)	27 (8 to 47)	21 (5 to 36)	31 (14 to 47)	28 (13 to 43)	33 (17 to 48)
BASDAI 50	25 (11 to 39)	19 (–8 to 46)	–	–	–	–	–	–	–	–

a Adalimumab $n = 69$, placebo $n = 73$.
b Adalimumab $n = 25$, placebo $n = 20$.
c Certolizumab 200 mg, $n = 46$; certolizumab 400 mg, $n = 51$; placebo, $n = 50$.
d Certolizumab 200 mg, $n = 39$; certolizumab 400 mg, $n = 35$; placebo, $n = 39$.
e Certolizumab 200 mg, $n = 65$; certolizumab 400 mg, $n = 56$; placebo, $n = 57$.
f Certolizumab 200 mg, $n = 74$; certolizumab 400 mg, $n = 71$; placebo, $n = 67$.

Owing to the fact that only a select group of patients could be subject to central confirmation of their nr-AxSpA status, the FDA statistician explored assumptions around the proportion of true nr-AxSpA patients in the whole trial population. Given that the treatment difference in the non-centrally read patients was 23%:

- Assuming that all non-centrally read patients were true negatives and therefore including them in the analysis with the centrally read negatives, the treatment difference for the centrally read and non-centrally read negatives was 15%.
- Assuming that a fraction (i.e. 63%) of non-centrally read patients were true negatives and including only this fraction of non-centrally read patients with the centrally read negatives, the treatment difference was 14%.

The FDA document stated that:

> *Because there was a differential treatment effect between the centrally-read positive and centrally-read negative, it is safe to assume that the difference of 23% is an overestimate of the treatment effect because this includes both positive and negative x-ray groups. If there is a fraction of patients who are negative in the non-centrally-read group, treatment difference among this negative group would be smaller. Therefore, the treatment difference for negative x-rays (i.e. centrally-read and non-centrally-read) should be at most 15%. Based on the data provided, the estimate of the treatment effect in ASAS40 response for nr-AxSpA should be no bigger than 15%*
>
> *FDA Briefing Package*[116]

Overall, the results suggest reduced efficacy of anti-TNFs in the centrally diagnosed nr-AxSpA population compared with the locally diagnosed population. Nevertheless, there was noticeable variation across the two trials. In RAPID-axSpA[64] (certolizumab) the difference between the central and local populations appears small (and is not evident for 400-mg vs. placebo results). Conversely, in ABILITY-1[58] (adalimumab) the locally diagnosed population had notably more responders than the centrally diagnosed population, although the treatment group sample sizes were small.

Long-term efficacy results from open-label extensions of randomised controlled trials

Of the 24 included RCTs, 17 reported data from an open-label extension phase. Results for all studies are presented in *Appendix 8*. Considerable effort has been put into patient follow-up in anti-TNF trials with the result that data up to 5 years are available (there are data up to 8 years for infliximab but these included an involuntary treatment break which is not discussed further). The longest follow-up durations in patients with AS by anti-TNF are adalimumab 260 weeks, etanercept 264 weeks, infliximab 156 weeks, golimumab 268 weeks and certolizumab pegol 96 weeks. However, the data were reported across numerous publications and in various formats. Results were reported as observed, as completer analyses, using imputation (and rarely LOCF) for non-responders and LOCF for missing continuous data, but these related to differing populations (at varying time points): all patients randomised, all patients who took active drug at any point in the study or all patients who took active drug just during the open-label phase. The follow-up protocols were not clearly reported, with stopping rules unclear, but it appears that not all non-responders discontinued therapy. Therefore, the results may not reflect clinical practice should response be required for treatment continuation.

Table 15 presents the results based on non-responder imputation (NRI) analyses for the main studies when these results could be extracted. For AS the results show that across all the anti-TNFs after approximately 2 years of treatment, around half of patients are still achieving a good level of response to therapy. The results for golimumab look particularly strong with around 60% of all randomised patients achieving ASAS 40 and BASDAI 50 after 5 years. However, this is probably not reflective of clinical practice, as many of the normal weight patients took the 100-mg dose of golimumab rather than the 50-mg dose: the licence permits the use of 100-mg dose only in patients with a body weight of more than 100 kg who do not achieve an adequate clinical response after three or four doses. The equivalent results for adalimumab and etanercept are approximately 30% and 50%, although it is unknown if the difference may be because of variations in follow-up protocols rather than true treatment difference.

TABLE 15 Treatment effect over time (AS only) (results calculated using non-responder imputation)

Outcome	Trial	52 weeks, *n*/*N* (%)	104 weeks, *n*/*N* (%)	156 weeks, *n*/*N* (%)	5 years (approximately 264 weeks), *n*/*N* (%)
Adalimumab					
ASAS 20	ATLAS[61]	193/311 (62)[a]	135/311 (43)[a]	–	111/311 (36)[a]
ASAS 40	ATLAS[61]	138/311 (44)[a]	109/311 (35)[a]	–	88/311 (28)[a]
BASDAI 50	ATLAS[61]	167/311 (54)[a]	122/311 (39)[a]	–	96/311 (31)[a]
Certolizumab					
ASAS 20	RAPID-axSpA[64] (AS)	(48 weeks) 89/121 (74)[b]	(96 weeks) 78/121 (64)[b]	–	–
ASAS 40	RAPID-axSpA[64] (AS)	(48 weeks) 70/121 (58)[b]	(96 weeks) 61/121 (50)[b]	–	–
BASDAI 50	–	–	–	–	–
Etanercept					
ASAS 20	Calin 2004[83]	–	(108 weeks) 52/81 (64)[c]	–	–
ASAS 40	Calin 2004[83]	–	(108 weeks) 44/81 (54)[c]	–	40/81 (49)[c]
BASDAI 50	Calin 2004[83]	–	(108 weeks) 42/81 (52)[c]	–	39/81 (48)[c]
Golimumab					
ASAS 20	GO-RAISE[90]	–	235/356 (66)[b]	(160 weeks) 246/356 (69)[b]	235/356 (66)[b]
ASAS 40	GO-RAISE[90]	–	203/356 (57)[b]	(160 weeks) 208/356 (58)[b]	203/356 (57)[b]
BASDAI 50	GO-RAISE[90]	–	199/356 (58)[b]	–	199/356 (58)[b]
Infliximab					
ASAS 20	PLANETAS 2013[110]	(78 weeks) 125/174 (72)[c,d]	(102 weeks) 127/174 (73)[c,d]	–	–
ASAS 40	–	(78 weeks) 93/174 (53)[c,d]	(102 weeks) 101/174 (58)[c,d]	–	–
ASAS 40	ASSERT 2005[102]	(102 weeks)	33/78 (42)[b,e]	–	–
BASDAI 50	Braun 2002[98]	(54 weeks) 33/69 (48)[b]	(102 weeks) 30/69 (43)[b]	–	–

ASSERT, Ankylosing Spondylitis Study for the Evaluation of Recombinant Infliximab Therapy; NRI, non-responder imputation.
a NRI imputed result calculated using number of patients who had received at least one dose of active as denominator.
b NRI imputed result calculated using number of patients randomised as denominator.
c NRI imputed result calculated using number of patients who had received active during open-label phase as denominator.
d CT-P13 and infliximab combined.
e Only the subset of patients who took the 5-mg dose of infliximab (remaining patients took 5 or 7.5 mg).

The long-term follow-up for nr-AxSpA patients (*Table 16*) shows continued high proportions of responders. At 1 year around half of patients are achieving an ASAS 40 or BASDAI 50 level response and with certolizumab this is maintained at 2 years and with adalimumab at 3 years.

When the long-term data are presented as observed or completer analyses, the long-term results are similarly good; withdrawal rates are not high and a high proportion of those who remain on treatment continue to achieve a good response, see the example data available from one trial of adalimumab and one of certolizumab pegol (*Table 17*).

At long-term follow-up mean final values or mean change from baseline for BASDAI, BASFI and BASMI, when reported, were generally maintained at clinically meaningful levels.

For adalimumab, data from the large ATLAS trial[61] showed that mean changes from baseline at 1, 2 and 3 years remain stable and clinically meaningful at around –3.7 units for BASDAI and at around –2.9 units for BASFI. Similarly, the mean final value for BASMI remains at a level indicative of clinically significant

TABLE 16 Treatment effect over time (nr-AxSpA only)

Outcome	Trial	52 weeks, *n/N* (%)	104 weeks, *n/N* (%)	156 weeks, *n/N* (%)	5 years (approximately 264 weeks), *n/N* (%)
Adalimumab					
ASAS 20	ABILITY-1[58]	–	–	83/142 (58)[a]	–
ASAS 40	Haibel 2008[52]	23/46 (50)[b]	–	–	–
ASAS 40	ABILITY-1[58]	(68 weeks) 77/142 (54)[a]	–	67/142 (47)[a]	–
BASDAI 50	Haibel 2008[52]	24/46 (52)	–	–	–
BASDAI 50	ABILITY-1[58]	(68 weeks) 74/142 (52)[a]	–	70/142 (49)[a]	–
Certolizumab					
ASAS 20	RAPID-axSpA[64] (AS)	(48 weeks) 68/97 (70)[b]	(96 weeks) 59/97 (61)[b]	–	–
ASAS 40	RAPID-axSpA[64] (AS)	(48 weeks) 56/97 (58)[b]	(96 weeks) 49/97 (51)[b]	–	–
BASDAI 50	–	–	–	–	–
Etanercept					
ASAS 20	Dougados 2014[76]	(48 weeks) (CiC information has been removed)	–	–	–
ASAS 40	–	(48 weeks) 108/205 (53)[b]	–	–	–
BASDAI 50	–	(48 weeks) (CiC information has been removed)	–	–	–

CiC, commercial-in-confidence.
a NRI imputed result calculated using number of patients at week 12 as denominator.
b NRI imputed result calculated using number of patients randomised as denominator.
Results calculated using NRI.

TABLE 17 Observed or completer analysis results

Trial, anti-TNF (population)	Time point	Type of analysis	ASAS 20, *n*/*N* (%)	ASAS 40, *n*/*N* (%)	BASDAI 50, *n*/*N* (%)
ATLAS 2006,[61] adalimumab (AS)	52 weeks	Observed	193/276 (70)	138/276 (50)	167/276 (61)
	104 weeks	Observed	135/173 (78)	109/173 (63)	122/173 (71)
	5 years	Completer	111/125 (89)	88/125 (70)	96/124 (77)
RAPID-axSpA,[64] certolizumab pegol; all (AS)	96 weeks	Observed	78/93 (84)	61/93 (66)	–
RAPID-axSpA,[64] certolizumab pegol; all (nr-AxSpA)	96 weeks	Observed	59/74 (80)	49/74 (66)	–

treatment benefit (3.1 to 3.7 units). At 5 years the mean final values are BASDAI 1.8 units, BASFI 2.1 units, and BASMI 3.7 units. Clearly these results relate only to those patients who have remained on adalimumab in the long-term (40% of those who started adalimumab). They do, however, demonstrate continued benefit in a significant proportion of patients.

For certolizumab, results for these outcomes are available up to 96 weeks. At this time point the mean BASDAI and BASFI are indicative of clinically significant treatment benefit (both around 3 units).

The long-term data from Calin *et al.*[83] for etanercept, with 81 patients at 2 years and 59 (73%) remaining at 5 years also report mean BASDAI and BASFI scores of around 3.

From GO-RAISE[90] at 2 years for those who took golimumab throughout the trial and follow-up ($n = 138$), median BASDAI score was around 3 and median BASFI score was around 2. These values are from a LOCF analysis of all patients randomised to golimumab 50 mg.

For infliximab, the Braun *et al.*[98] and follow-up studies found, from 1 to 3 years, a stable mean BASDAI score of around 2.6, a stable mean BASFI score of around 3 and a stable mean BASMI score of around 2.7.

Overall, the reported data (although not particularly robust) do indicate that significant proportions of patients continue to derive real benefit from continued use of anti-TNFs. There is nothing to indicate any difference between them.

Almost no data were available regarding radiographic progression of bony disease in patients with AS. Furthermore, it should be noted that radiographic changes and progression of these take many years to appear and radiography is an insensitive tool by which to evaluate the progression of AS. Therefore, evidence, particularly that from relatively short-term studies, has to be interpreted with caution. The limited evidence includes mSASSS change from baseline, reported for golimumab from the GO-RAISE[90] study at 4 years (208 weeks): 1.3 units (SD 4.1 units) based on the 111 of 138 patients randomised to 50 mg. As results from untreated cohorts suggest a progression rate of 2 units/2 years, a rate of 1.3 units (or even 2 units) over 4 years seems beneficial. For further discussion of this issue see *Effect of anti-tumour necrosis factors on radiographic progression*. MASES was reported only for adalimumab from ATLAS;[61] in patients remaining on therapy at 2 years the mean change from baseline was 2.2 units ($n = 217$).

For nr-AxSpA patients long-term data for the continuous outcomes was limited to 1 year's follow-up. For adalimumab, data were available from only one small study (Haibel 2008,[52] $n = 46$): BASDAI change from baseline 2.8 units (95% CI 2.1 to 3.6 units); BASFI change from baseline 2 units (95% CI 1.4 to 2.6 units); BASMI change from baseline –0.4 units (95% CI –0.7 to –0.04 units); and MASES change from baseline of 0.9 units (95% CI –0.02 to 1.9 units). In addition, of 26 patients with magnetic resonance images at baseline and 52 weeks' follow-up, none showed a change in sclerosis or in erosions. For etanercept, data were available on 205 patients randomised to etanercept or placebo and then on long-term etanercept (Dougados 2014[76]): [commercial-in-confidence (CiC) information has been removed]. For certolizumab, LOCF analysis at 96 weeks ($n = 97$) gave a BASDAI final value score of 3.0, and a BASFI score of 2.4. Overall, the 1-year results in nr-AxSpA patients are similar to each other and also reflect those seen in AS patients. Again, the short-term nature of this follow-up relative to the 8–10 years over which radiographic changes develop must be borne in mind.

Findings from anti-tumour necrosis factor patient registry studies

Effect of anti-tumour necrosis factors on radiographic progression

A total of seven studies were identified that provided some comparative results on the effect of anti-TNFs on radiographic progression (*Table 18*).

TABLE 18 Effect of anti-TNFs on radiographic progression

Study	Methods	Results
van der Heijde *et al.* 2009[117]	Study used 2-year data from active treatment arms of two adalimumab trials (total $n = 397$) and compared them with OASIS cohort[118] (186 with radiographs at 2 years). Note: primary analysis set = 307 adalimumab (minimum of 1.5 years exposure to drug) and 169 anti-TNF naive (OASIS)	There were significant differences between adalimumab and OASIS[118] patients at baseline for BASDAI, BASFI and other measures. Increase in mSASSS was very similar in the two groups: adalimumab mean 0.8 (SD 2.6) and OASIS mean 0.9 (SD 3.3). When only patients who would have qualified for the adalimumab trials were included in the OASIS cohort ($n = 77$) the results were not changed. Note: in the light of these van der Heijde results, it would have been good to test effect of baseline BASDAI (mean 6.2 in adalimumab cohort and 3.4 in OASIS), as without treatment progression in adalimumab cohort would have been expected to be higher than in the OASIS one, so there might have been some effect of adalimumab
van der Heijde, *et al.* 2008[103]	Study compared 2-year data from infliximab trial (ASSERT[102]) ($n = 201$) with that from OASIS[118] ($n = 192$). OASIS patients not treated with any anti-TNF	There were significant differences between infliximab and OASIS[118] patients at baseline for BASDAI, BASFI and other measures (higher disease activity and worse function in trial patients). Mean increase in mSASSS was very similar in the two groups: infliximab 0.9 (SD 2.6) and OASIS 1.0 (SD 3.2). When only patients who would have qualified for the infliximab trials were included in the OASIS cohort ($n = 70$), the results changed very little [mean mSASSS increase 1.2 (SD 3.9)]
van der Heijde, *et al.* 2008[119]	Study compared 2-year data from etanercept trial (Davis *et al.*[72]) ($n = 257$) with that from OASIS[118] ($n = 175$). OASIS patients not treated with any anti-TNF	There were significant differences between infliximab and OASIS[118] patients at baseline for BASDAI, BASFI and other measures (higher disease activity and worse function in trial patients). Mean increase in mSASSS was very similar in the two groups: etanercept 0.91 (SD 2.5) and OASIS 0.95 (SD 3.2). When only patients who would have qualified for the etanercept trials were included in the OASIS cohort ($n = 76$), the results changed very little [mean mSASSS increase 1.3 (SD 3.6)]
Braun *et al.* 2014[120]	Long-term data on golimumab (2- and 4-year radiographic data) ($n = 233$). No comparison with OASIS[118] made	Mean increase in mSASSS to 2 years was 0.9 (SD 2.7) (50 mg) and 0.9 (SD 3.9) (100 mg). Mean increase in mSASSS to 4 years was 1.3 (SD 4.1) (50 mg) and 2.0 (SD 5.6) (100 mg). Note: 2-year results are very similar to those with other anti-TNFs and OASIS,[118] that is there is no benefit of golimumab evident

TABLE 18 Effect of anti-TNFs on radiographic progression (*continued*)

Study	Methods	Results
Haroon *et al.* 2013[121]	Cohort study ($n = 334$) with at least two spinal radiographs at 2-year intervals (patients with total spinal fusion at baseline excluded). Logistic regression analysis tested for baseline mSASSS, ESR, BASDAI, smoking, male vs. female, age at onset, disease duration, HLA-B27, anti-TNF use and NSAID index. Further analysis tested factors that could influence exposure to anti-TNFs using propensity matching	In total, 201 out of 334 patients had received anti-TNFs for a mean of 2.5 years (SD 2.6 years). No radiographic abnormality of the spine was seen at baseline in 144 patients (43%) and 102 patients (30.5%) showed no progression (> 1 mSASSS unit/year). Mulitvariate regression found baseline mSASSS (OR 1.06, 95% CI 1.04 to 1.08), ESR and smoking significantly increased and anti-TNF use significantly increased odds of radiographic progression (OR 0.47, 95% CI 0.24 to 0.94). Further analysis using the 142 patients who could be included post propensity matching confirmed these findings except for ESR: baseline mSASSS (OR 1.05, 95% CI 1.02 to 1.08) and anti-TNF (OR 0.30, 95% CI 0.11 to 0.78). Note: the association with anti-TNF use is explained by the more severe patients with radiographic changes at baseline being treated with anti-TNFs
Barialiakos *et al.* 2014[122]	Comparison of long-term (8 years) treatment with infliximab with historical cohort (infliximab $n = 22$ and Herne cohort $n = 34$)	Progression as assessed by mSASSS increased equally in infliximab treated patients and in the Herne cohort from baseline to 2, 4 and 6 years but while progression increased only slightly in the infliximab group between 6 and 8 years it increased greatly in the Herne cohort so that at 8 years there was a difference in infliximab's favour of 4.5 mSASSS ($p = 0.047$). Result was adjusted for baseline mSASSS. Other factors (age, symptom duration, BASDAI, BASFI) not significant confounders
Barialiakos 2007[123]	4-year radiographic progression in AS patients treated with infliximab ($n = 33$). Crude comparison made with OASIS cohort[118] results at 4 years	At baseline, mean mSASSS was 11.6 (15.3 SD), mean BASDAI was 6.6 (1.4 SD) and mean BASFI was 3.5 (1.9 SD). Progression assessed by mSASSS. Mean change over 4 years was 1.6 (SD 2.6) mSASSS units. Published results for OASIS are 4.4 units in 4 years

ASSERT, Ankylosing Spondylitis Study for the Evaluation of Recombinant Infliximab Therapy; OASIS, Outcomes in Ankylosing Spondylitis International Study.

Four studies reported on disease progression over 2 years of follow-up in terms of mSASSS in patients taking adalimumab,[117] infliximab,[103] etanercept[119] and golimumab.[120] All four open-label, uncontrolled follow-up studies found that mSASSS increased by a mean of around 0.9 units over 2 years. Three of these studies compared their rates with those from the Outcomes in Ankylosing Spondylitis International Study (OASIS) cohort[118] (of patients not taking an anti-TNF) and found no difference (mean rate over 2 years for OASIS was 0.9 units, *Table 19*). As stated in the previous section, radiographic changes and progression of these take many years to appear and, therefore, the evidence from these relatively short-term studies should be interpreted with caution.

TABLE 19 Summary of long-term results for mSASSS change

	Increase in mSASSS over 2 years, patients on an anti-TNF		Increase in mSASSS over 2 years, patients from OASIS cohort[118]	
Trial, anti-TNF	*n*	Mean (SD)	*n*	Mean (SD)
van der Heijde 2009,[117] adalimumab	397	0.8 (SD 2.6)	186	0.9 (SD 3.3)
van der Heijde 2008, etanercept[119]	257	0.91 (SD 2.45)	175	0.95 (SD 3.2)
Infliximab[103]	201	0.9 (SD 2.6)	192	1.0 (SD 3.2)
Golimumab[120]	111	50 mg, 0.9 (SD 2.7)	–	–
	122	100 mg, 0.9 (SD 3.9)	–	–

Comparison of the rates calculated from the OASIS cohort[118] in these studies with those from the studies by Ramiro[124,125] highlight a discrepancy; the latter reported rates of 2 mSASSS units every 2 years, rather than the 0.9 units/2 years used here to compare with individual anti-TNFs.

Two very small studies of infliximab reported some inhibiting effect on radiographic progression.[122,123] The first[123] compared findings in 22 infliximab patients with 34 from the HERNE cohort, over 2, 4, 6 and 8 years. Progression as assessed by mSASSS increased equally in infliximab treated patients and in the untreated HERNE cohort from baseline to 2, 4 and 6 years but then while progression increased only slightly in the infliximab group between 6 and 8 years it increased greatly in the HERNE cohort so that at 8 years there was a difference in infliximab's favour of 4.5 mSASSS units. The result was adjusted for baseline mSASSS (other factors, age, symptom duration, BASDAI, BASFI, etc., were not statistically significant confounders). The other study of 33 patients[123] found the mean progression over 4 years was 1.6 mSASSS units (SD 2.6 units), lower than the 4.4 units seen in the untreated OASIS cohort[118] at 4 years.

Another study examined a cohort of 334 patients with at least two spinal radiographs at 2-year intervals (patients with total spinal fusion at baseline were excluded).[121] In this study 201 out of 334 patients had received anti-TNFs for a mean of 2.5 years (SD 2.6 years) and no radiographic abnormality of the spine was seen at baseline in 144 patients (43%). At follow-up 102 patients (30.5%) showed no progression ($\geq$ 1 mSASSS unit/year). Mulitvariate regression found baseline mSASSS (OR 1.06, 95% CI 1.04 to 1.08), ESR and smoking significantly increased the odds of radiographic progression, but anti-TNF use was significantly associated with a > 50% reduction in the (adjusted) odds of progression (0.47, 95% CI 0.24 to 0.94). Further analysis that tested factors that could influence exposure to anti-TNFs using propensity matching confirmed the association with mSASSS and found a stronger association with anti-TNF use (OR 0.30, 95% CI 0.11 to 0.78).

In conclusion, there is evidence of disease progression over time, although the disease course is highly variable. Best estimates of yearly disease progression rates without anti-TNF therapy are around 1.0 mSASSS units and 0.035–0.07 BASFI units. Whether or not there is any impact of anti-TNF treatment is unclear; a beneficial effect can neither be assumed, and nor, given the short-term nature of the follow-up and the insensitivity of radiography as a tool for the evaluation of disease progression in AS, can one be discounted.

Drug survival and anti-tumour necrosis factor switching

The EndNote Library generated by the searches for RCTs of all the anti-TNFs were separately screened to identify patient registry studies of any or all of the anti-TNFs. This was possible because the search strategy for RCTs was very sensitive and will have identified any clinical study including any of the named anti-TNFs.

A total of 25 potentially relevant studies were screened fully and 12 publications that reported some data on drug survival or the efficacy of anti-TNFs after switching were identified (see *Table 20* for summary details of each). Across the 12 studies, the sources of data were either retrospective cohort studies or prospective registers (although analysis plans may have been retrospective), from a range of regions: USA (two studies), Canada (one study) and Europe (nine studies). No data from a UK-based cohort were available. Most of the cohorts and registries included experience with the three oldest anti-TNFs: infliximab, etanercept and adalimumab. One study (of the RHAPSODY cohort) included results from 326 patients treated with adalimumab as second anti-TNF after infliximab or etanercept. Small numbers of patients provided data on golimumab (three studies) and even smaller numbers on certolizumab (two studies). The population in 10 of the 12 studies was AS, although the diagnostic criteria used to specify AS were rarely given. One study provided results specifically for nr-AxSpA and one study provided results for axial SpA (nr-AxSpA or AS).

TABLE 20 Registry studies reporting data on drug survival and anti-TNF switching

Citation	Study/registry and method	*n* (duration, where stated)	Population	Anti-TNFs included	Drug survival	Efficacy on switching
Bonafede 2012[126]	Market Scan (administrative claims data) 2005–9, USA; retrospective	308 (360 days)	AS	Etanercept, adalimumab and infliximab	*n* (%) who stopped treatment and did not switch/who switched: • etanercept: (*n* = 149) 42 (28%)/12 (8%) • adalimumab: (*n* = 103) 36 (35%)/11 (11%) • infliximab: (*n* = 46) 14 (30%)/6 (13%)	NR
Choquette 2013[127] (abstract only)	Rhumadata register, Canada; unknown	119 (5 years)	AS, previous NSAIDs and BASDAI score of ≥ 4	Etanercept, adalimumab and infliximab	*n* who remained on same anti-TNF was 80% at 1 year; 70% at 2 years; and 55% at 5 years (no difference between anti-TNFs)	NR
Gulfe 2014[128]	SSATG registry, Sweden; prospective	112 (2 years)	Nr-axSpA not AS, demographic summary available	Etanercept, adalimumab, infliximab, golimumab and certolizumab	Kaplan–Meier estimates drug survival was 76% at 1 year and 65% at 2 years	NR
Nell-Duxneuner 2012[129]	Drug reimbursement data, Austria; retrospective	694 (2 years)	AS	Etanercept, adalimumab and infliximab	Starting in 2007 drug survival was: • etanercept: 0.83 (1 year) and 0.58 (2 years) • adalimumab: 0.70 (1 year) and 0.55 (2 years) • infliximab: 0.71 (1 year) and 0.54 (2 years)	NR

continued

TABLE 20 Registry studies reporting data on drug survival and anti-TNF switching (*continued*)

Citation	Study/registry and method	*n* (duration, where stated)	Population	Anti-TNFs included	Drug survival	Efficacy on switching
Yeaw 2014[130]	LifeLink Health Plan Claims database 2004–10, USA; retrospective	632	AS patients who had discontinued an anti-TNF	Etanercept, adalimumab and infliximab	% who restart within 360 days after stopping: • etanercept: 59% (*n* = 376) • adalimumab: 45% (*n* = 134) • infliximab: 39% (*n* = 122) % who switch to another anti-TNF or biologic: • etanercept: 17% (*n* = 376) • adalimumab: 13% (*n* = 134) • infliximab: 24% (*n* = 122) % who switch to non-biologic: • etanercept: 5% (*n* = 376) • adalimumab: 8% (*n* = 134) • infliximab: 6% (*n* = 122) % who switch to no new treatment: • etanercept: 18% (*n* = 376) • adalimumab: 34% (*n* = 134) • infliximab: 30% (*n* = 122)	NR

Citation	Study/registry and method	*n* (duration, where stated)	Population	Anti-TNFs included	Drug survival	Efficacy on switching
Scire *et al.* 2013[131]	MonitorNet database (Italian Society of Rhuematology) to 2012, Italy; multiple imputation used for missing data	498	AS	Etanercept, adalimumab and infliximab	Unadjusted Kaplan–Meier estimates of drug survival at: • 1 year: 0.87 (95% CI 0.83 to 0.89) • 2 years: 0.72 (95% CI 0.67 to 0.77) • 3 years: 0.69 (95% CI 0.63 to 0.74) Adjusted HR discontinuation rate (median follow-up 17 months) 0.59 (95% CI 0.46 to 0.75) (adjusted for age, sex, number of comorbidities, disease duration, number of previous DMARDs, concurrent DMARDS, baseline BASDAI score and BASFI score)	NR
Zufferey 2014[132]	Single centre in Switzerland (Centre Hospitalier Universitaire Vaudois) 2011–12; retrospective	112, of whom 77 were AS (follow-up at 12 and 24 months)	SpA (AxSpA and AS)	Etanercept, adalimumab, infliximab and golimumab	Median drug survival across all anti-TNFs 12 months (IQR 7–19 months) for AxSpA and 8 months (IQR 6–13 months) for AS Drug survival for AS: • 1 year 49% • 2 years 36% No difference between anti-TNFs	NR
Pavelka 2009[133]	ATTRA national registry, Czech Republic; prospective	310 (1 year)	AS (note mean BASDAI score 6.4 at baseline)	Etanercept, adalimumab and infliximab	Drug survival at 1 year was 84%; at 2 years was 76%; and at 3 years was 72%	NR

continued

TABLE 20 Registry studies reporting data on drug survival and anti-TNF switching (*continued*)

Citation	Study/registry and method	*n* (duration, where stated)	Population	Anti-TNFs included	Drug survival	Efficacy on switching
Lie 2011[134]	NOR-DMARD register 2000–9, Norway; prospective	514 (2 years)	AS	Etanercept, adalimumab and infliximab	In total 77 patients switched from first anti-TNF; 437 did not. In the 77 switchers, median drug survival on first anti-TNF was 266 days on the first anti-TNF (range 1–1392) and the second anti-TNF was a median of 77 days (range 0–1608 after the first was stopped). Finding may just be a consequence of the stopping rules in Denmark (patients given around 6 months to achieve a response) % on treatment after 1 and 2 years: First anti-TNF: 76% and 65% Second anti-TNF: 67% and 60%	Non-switchers: response to first anti-TNF at 3 months ($n = 362$): • BASDAI 50, *n/N*: 105/362 • ASAS 20, *n/N*: 106/202 • ASAS 40, *n/N*: 76/202 • Median (IQR) BASFI score: 2.3 (0.7–4.0) Median (IQR) BASDAI score: 2.6 (1.3–4.4) Switchers: response to first anti-TNF at 3 months: • BASDAI 50, *n/N*: 6/63 • ASAS 20, *n/N*: 11/23 • ASAS 40, *n/N*: 7/23 Median (IQR) BASFI, score: 4.7 (1.5–6.0) ($n = 63$) Median (IQR) BASDAI, score: 4.8 (3.3–7.01) ($n = 63$)

Citation	Study/registry and method	*n* (duration, where stated)	Population	Anti-TNFs included	Drug survival	Efficacy on switching
Glintborg 2010[112]	DANIBO registry, Denmark; prospective	842 (8 years)	AS	Etanercept, adalimumab and infliximab	Median drug survival was 4.3 years (unadjusted 1- and 2-year retention rates 74% and 63%) which was similar across three anti-TNFs; only male sex, low baseline VAS fatigue and high CRP level (> 14 mg/l) were associated with better drug survival	Response to second anti-TNF at 3 months: • BASDAI 50, *n/N*: 13/62 • ASAS 20, *n/N*: 18/45 • ASAS 40, *n/N*: 14/45 Median (IQR) BASFI score: 3.3 (1.6–5.7) (*n* = 62) Median (IQR) BASDAI score: 4.1 (1.9–6.1) (*n* = 62) Data also available by reason for withdrawal NR

continued

TABLE 20 Registry studies reporting data on drug survival and anti-TNF switching (*continued*)

Citation	Study/registry and method	*n* (duration, where stated)	Population	Anti-TNFs included	Drug survival	Efficacy on switching
Glintborg 2012[135]	DANBIO registry, Denmark; prospective	1436 (432 switchers; median 2.4 years)	AS (switchers only; had received at least two anti-TNFs during follow-up)	Etanercept, adalimumab, infliximab and golimumab (certolizumab and other biologics had less than 1% between them and only to first treatment course)	Median (95% CI) years of drug survival (*n*) (% on treatment after 2 years) for sequential anti-TNFs: • First anti-TNF: 3.1 (2.6 to 3.7), *n* = 1436 (58%) • Second anti-TNF: 1.6 (1.0 to 2.2), *n* = 432 (47%) • Third anti-TNF: 1.8 (0.9 to 2.7), *n* = 137 (49%)	Median (IQR) BASDAI at 3 months for sequential anti-TNFs: • First anti-TNF (*n* = 1436): 2.8 (1.1–4.8) • Second anti-TNF (*n* = 432): 3.6 (1.9–6.4) • Third anti-TNF (*n* = 137): 5.1 (3.6–6.7) Median (IQR) BASFI at 3 months for sequential anti-TNFs: • First anti-TNF (*n* = 1436): 2.8 (1.1–4.8) • Second anti-TNF (*n* = 432): 3.6 (1.7–6.0) • Third anti-TNF (*n* = 137): 5.1 (3.0–7.3) % BASDAI 50/20 mm responders at 6 months (NR at 3 months): • First anti-TNF: 54% • Second anti-TNF: 37% • Third anti-TNF: 30%
Rudwaleit 2009[136]	RHAPSODY, European cohort; prospective uncontrolled cohort of patients treated with adalimumab	1250 (12-week response data only)	AS	Adalimumab	NR	12-week response rates: anti-TNF naive (*n* = 924); BASDAI 50–63%; ASAS 40–59%; anti-TNF exposed (etanercept and/or infliximab, *n* = 326); BASDAI 50–41%; and ASAS 40–38% Logistic regression with backward elimination found that younger age, higher CRP level, HLA-B27 positive and anti-TNF naivety all predictive of better response (table 1[134])

HR, hazard ratio; IQR, interquartile range; NR, not reported; RHAPSODY, Review of safety and effectiveness witH Adalimumab in Patients with active ankylosing SpOnDYlitis; SSATG, Southern Sweden Anti-rheumatic Therapy Group.

Drug survival on first anti-TNF for all anti-TNFs was around 70–80% at 1 year, around 65–75% at 2 years, around 70% at 3 years and 55% at 5 years. Little difference between the three older anti-TNFs was identified, although one analysis using Cox proportional hazard estimates found statistically lower rates of discontinuation with etanercept and adalimumab compared with infliximab.[131]

The median drug survival in AS patients across all anti-TNFs reported varied (*Table 21*). Based on the largest registry (DANBIO)[135] the median drug survival for a first anti-TNF was 3.1 years (95% CI 2.6 to 3.7 years) ($n = 1436$), with 58% of patients remaining on treatment at 2 years. Median drug survival for a second anti-TNF was 1.6 years (95% CI 1.0 to 2.2 years) ($n = 432$), with 47% of patients remaining on treatment at 2 years, and for a third, 1.8 years (95% CI 0.9 to 2.7 years) ($n = 137$) (49% on treatment at 2 years).

The efficacy of second or third anti-TNFs after switching in AS patients was reported in only a small number of studies. One analysis based on the NOR-DMARD registry[134] showed how the response rate and BASDAI and BASFI achieved at 3 months in patients who remain on their first therapy is (not surprisingly) better than in patients who switch. Median BASDAI and BASFI achieved with a second anti-TNF were not as low (not as good) as was achieved with a first anti-TNF in non-switchers. An analysis of the DANBIO registry indicated that response (BASDAI 50) at 6 months reduced with subsequent anti-TNFs, as did the median improvement in BASDAI and BASFI achieved (*Table 22*).[135] These results are supported by the RHAPSODY study that found higher response rates with adalimumab in anti-TNF naive patients (BASDAI 50–63%; ASAS 40–59%)($n = 924$) than in anti-TNF exposed (BASDAI 50–41%; ASAS 40–38%) ($n = 326$).[136]

The registries and cohort studies provided no data on the efficacy of anti-TNFs as second or third, after switching in nr-AxSpA patients.

TABLE 21 Drug survival results from analysis of DANBIO registry[135]

	Drug survival for sequential anti-TNFs	
Anti-TNF	Median (95% CI)	% on treatment after 2 years
First ($n = 1436$)	3.1 (2.6 to 3.7)	58
Second ($n = 432$)	1.6 (1.0 to 2.2)	47
Third ($n = 137$)	1.8 (0.9 to 2.7)	49

TABLE 22 Efficacy results from analysis of DANBIO registry[135]

Anti-TNF	% BASDAI score 50/20 mm responders at 6 months (at 3 NR)	BASDAI score at 0 months for sequential anti-TNFs, median (IQR)	BASDAI score at 3 months for sequential anti-TNFs, median (IQR)	BASFI score at 0 months for sequential anti-TNFs, median (IQR)	BASFI score at 3 months for sequential anti-TNFs, median (IQR)
First ($n = 1436$)	54	5.9 (4.5–7.1)	2.8 (1.1–4.8)	5.0 (3.4–6.7)	2.8 (1.1–4.8)
Second ($n = 432$)	37	5.6 (3.8–7.3)	3.6 (1.9–6.4)	5.2 (3.5–7.0)	3.6 (1.7–6.0)
Third ($n = 137$)	30	6.4 (4.8–7.9)	5.1 (3.6–6.7)	6.4 (4.2–7.9)	5.1 (3.0–7.3)

IQR, interquartile range.

In summary, sequential treatment with anti-TNFs can be worthwhile in patients with AS but the response rates and benefits are reduced with second and third anti-TNFs, with the proportion of BASDAI 50 responders falling approximately 10% with each subsequent anti-TNF and the median BASDAI and BASFIs achieved increasing (worsening). The lower efficacy of a second anti-TNF relative to a first is reflected in lower median drug survival and proportion of patients remaining on therapy at 2 years. Interestingly, despite a further reduction in response and efficacy with a third anti-TNF, drug survival does not fall, suggesting that at this stage in their treatment history patients may continue with a less than optimally effective anti-TNF given any better alternative.

Clinical effectiveness results: adverse events

Randomised trials

We focused on the following outcomes, known to have possible associations with anti-TNF treatment: serious infections, tuberculosis (including tuberculosis reactivation), injection/infusion site reactions, congestive heart failure, cancer, non-melanoma skin cancer, SAEs and withdrawals due to SAEs. For the randomised phases of the trials included in the review, the reporting of AE data was generally limited. For three of the 24 trials no information on AEs was available.[55,56,74] Several trials provided AE data only at time points after which placebo patients may have switched to receive an anti-TNF (so true placebo comparisons were not available).

Analysable data on injection/infusion site reactions were available for 10 trials, although these studies were only of etanercept or infliximab. The data for certolizumab, golimumab and adalimumab trials either were not reported or were only provided at time points after which placebo patients could 'escape' to receive an anti-TNF; these data would not allow for an accurate comparison with placebo. Results for injection/infusion site reactions analyses from this review for etanercept and infliximab showed a statistically significant increase in reactions associated with etanercept (relative risk 2.69, 95% CrI 1.82 to 3.89) compared with placebo but no significant difference between infliximab and placebo. Compared with each other, the risk of an injection/infusion site reaction was statistically significantly higher with etanercept than with infliximab (relative risk 2.27, 95% CrI 1.01 to 5.37). Incidence of serious infections was reported in only eight trials, although such events were rare (nine cases in total). Of the eight trials which reported incidence of tuberculosis, only four cases were identified; three cases were reported in the longest study, the 54-week trial which compared infliximab with an infliximab biosimilar (CT-P13).[110] Four trials reported on congestive heart failure (no cases reported), six trials reported on cancer (one case) and three trials reported on non-melanoma skin cancer (two cases, one in each group of the ABILITY-1[58] trial). In most trials few SAEs were reported; group rates ranged from 0% to around 9%. Similarly, most trials had few withdrawals because of AEs; rates ranged from 0% to around 12%. Full results are reported in *Appendix 9*.

Large systematic reviews

Overall, the number and size of trials, and the short duration of their placebo-controlled phases, were too limited to provide enough data for meaningful analyses of AE. This common problem, of having too few data to evaluate AEs, underpinned the rationale for a Cochrane review (and network meta-analysis) of AEs of nine biologics in adults with any disease, except HIV/AIDS.[137] In order to provide a better understanding of toxicity, data were pooled across diseases by assuming a similar rate of AEs (across diseases). For the present assessment, estimates of AE rates have therefore been derived from the Cochrane review, which included 160 RCTs ($n = 48,676$) and 46 open-label extension studies ($n = 11,954$). The median durations were 6 months for RCTs and 13 months for open-label extension studies. The biologics included were abatacept, adalimumab, anakinra, certolizumab pegol, etanercept, golimumab, infliximab, rituzimab and tocilizumab. The anti-TNFs included in the present assessment were studied in 115 (72%) of the RCTs and 40 (87%) of the open-label studies included in the Cochrane review. Most studies assessed etanercept or infliximab in cancer or rheumatoid arthritis patients; 10 RCTs were of AS (fewer than in this assessment, as in the Cochrane review databases were searched up until January 2010). The biologics were evaluated both as a group and as individual interventions. The results from the RCTs, what the review classified as

'major' outcomes, are in *Tables 23* and *24*. Biologics as a group were associated with statistically significantly higher rates of total AEs, withdrawals because of AEs, serious infections and tuberculosis reactivation compared with control treatments. When the individual anti-TNFs were analysed separately, compared with control treatments only infliximab and certolizumab were statistically significantly associated with AEs: infliximab with higher rates of total AEs [number needed to harm (NNH) 13, 95% CrI 8 to 505] and withdrawals because of AEs (NNH 10, 95% CrI 5 to 30), and certolizumab pegol with higher rates of serious infections (NNH 12, 95% CrI 4 to 79) and SAEs (NNH 18, 95% CrI 9 to 162) (see *Table 24*).

For total AEs, the Cochrane review team judged the strength of evidence to be high; for SAEs, withdrawals because of AEs and serious infections, the strength of evidence was judged to be moderate; and for tuberculosis reactivation, lymphoma and congestive heart failure, the strength of evidence was judged to be low. For tuberculosis reactivation, lymphoma and congestive heart failure, the network meta-analysis statistical models did not converge (because of low numbers of events) therefore estimates for individual anti-TNFs were not available. Outcomes which were classed in the review as 'minor' were not analysed by the review authors because of the low numbers of events and the complexity of the analyses for the major outcomes. The minor outcomes included cardiac AEs, infusion and injection site reactions, allergic reactions, neurological outcomes, deaths, all cancers, serious lung infections or pneumonia, fungal infections and opportunistic infections. For the purposes of the present assessment, further large studies on cancer risk were therefore sought. An individual patient data meta-analysis of 22,904 adults (from 74 RCTs) which assessed the cancer risk of taking adalimumab, etanercept or infliximab in the short term (median duration < 6 months) was identified.[138] Although funded by manufacturers, this study was requested by the European Medicines Agency and was planned and conducted by independent researchers working with an independent academic steering committee. For all three anti-TNFs as a group, there was no increase in risk of cancers excluding non-melanoma skin cancer (relative risk 0.99, 95% CI 0.61 to 1.68) but there was a doubling in the risk of non-melanoma skin cancer associated with taking an anti-TNF (relative risk 2.02, 95% CI 1.11 to 3.95). Evaluation of drug-specific effects was hampered by statistical precision and by differences in baseline cancer risk and reporting detail across trials.[138]

Another review of AE effects of etanercept, adalimumab and infliximab was based on systematic searches for systematic reviews of the safety of biologic agents.[139] Six reviews that were sufficiently rigorous to meet the Database of Abstracts of Reviews of Effects inclusion criteria were included in the overview. This review also included large RCTs and non-randomised studies (≥ 500 patients), and was focused on serious potential AEs, such as serious infections, reactivation of latent tuberculosis and cancer.[139] *Table 25*, which summarises the rates of SAEs among the included non-randomised studies and large RCTs, indicates that

TABLE 23 Cochrane summary of findings table for biologics as a class (reproduced with permission from Singh *et al.*[137])

AE	Risk with comparator, per 1000 patients unless otherwise stated	Risk with intervention, per 1000 patients, unless otherwise stated (95% CrI)	OR (95% CrI)	Number of participants (studies)
SAEs	118	127 (115 to 142)	1.09 (0.97 to 1.24)	21,152 (76)
Total AEs	724	770 (741 to 797)	1.28 (1.09 to 1.50)	14,959 (48)
Withdrawal due to AEs	98	137 (115 to 168)	1.47 (1.20 to 1.86)	22,636 (83)
Serious infections	26	35 (27 to 46)	1.37 (1.04 to 1.82)	21,853 (70)
Tuberculosis reactivation	4 per 10,000	20 per 10,000	4.68 (1.18 to 18.6)	30,671 (71)
Lymphoma	9 per 10,000	1	0.53 (0.17 to 1.66)	21,260 (52)
Congestive heart failure	8	6 (1 to 21)	0.69 (0.18 to 2.69)	8847 (24)

TABLE 24 Cochrane summary of findings table for individual anti-TNFs (adapted from Singh *et al.*[137])

Anti-TNF	Risk with comparator, per 1000 patients unless otherwise stated	Risk with intervention, per 1000 patients, unless otherwise stated (95% CrI)	OR (95% CrI)	Number of participants (studies)
SAEs				
Adalimumab	118	114 (90 to 145)	0.96 (0.74 to 1.27)	4662 (15)
Certolizumab	118	174 (124 to 237)	1.57 (1.06 to 2.32)	2421 (6)
Etanercept	118	142 (111 to 184)	1.24 (0.93 to 1.69)	3931 (21)
Golimumab	118	123 (82 to 184)	1.05 (0.67 to 1.69)	1564 (8)
Infliximab	118	133 (102 to 174)	1.15 (0.85 to 1.57)	3403 (14)
Total AEs				
Adalimumab	724	730 (637 to 802)	1.03 (0.67 to 1.54)	3266 (10)
Certolizumab	724	754 (651 to 837)	1.17 (0.71 to 1.95)	1829 (5)
Etanercept	724	784 (677 to 866)	1.38 (0.80 to 2.46)	1600 (7)
Golimumab	724	765 (672 to 839)	1.24 (0.78 to 1.98)	1187 (6)
Infliximab	724	803 (726 to 860)	1.55 (1.01 to 2.35)	2330 (9)
Withdrawal due to AEs				
Adalimumab	98	128 (81 to 194)	1.35 (0.82 to 2.22)	5268 (18)
Certolizumab	98	125 (70 to 226)	1.32 (0.69 to 2.69)	2421 (6)
Etanercept	98	124 (82 to 191)	1.30 (0.82 to 2.17)	5189 (25)
Golimumab	98	127 (64 to 241)	1.34 (0.63 to 2.92)	1549 (7)
Infliximab	98	203 (132 to 310)	2.34 (1.40 to 4.14)	2973 (15)
Serious infections				
Adalimumab	26	32 (17 to 60)	1.23 (0.65 to 2.40)	4847 (15)
Certolizumab	26	113 (39 to 330)	4.75 (1.52 to 18.45)	1683 (4)
Etanercept	26	33 (19 to 61)	1.29 (0.72 to 2.45)	4630 (19)
Golimumab	26	29 (12 to 65)	1.11 (0.45 to 2.59)	1334 (6)
Infliximab	26	36 (20 to 65)	1.41 (0.75 to 2.62)	2652 (13)
Tuberculosis reactivation				
All nine biologics	4 per 10,000	20 per 10,000	4.68 (1.18 to 18.60)	30,671 (71)
Lymphoma				
All nine biologics	9 per 10,000	1	0.53 (0.17 to 1.66)	21,260 (52)
Congestive heart failure				
All nine biologics	8	6 (1 to 21)	0.69 (0.18 to 2.69)	8847 (24)

TABLE 25 Prevalence ranges of SAEs from non-randomised studies and RCTs (reproduced from Rodgers *et al.*[139])

Drug	Serious infections (%)	Cancer (%)	Tuberculosis (%)	Mortality (%)	Withdrawals due to AE (%)
Etanercept	0.6–13.2	1–5.7	0–1.4	0–3.1	0–13.6
Infliximab	0.8–13.8	0.16–5.1	0.06–4.6	0.06–2.0	6.4–12.8
Adalimumab	0.4–5.1	0.1–1.1	0–0.4	0.5–0.9	5.8–10.7

the rates of SAEs cover a broadly similar range across the three different biologic agents. However, all estimates were derived from a highly heterogeneous group of studies in terms of participants (e.g. inflammatory condition or disease severity), study design (e.g. length of follow-up) and treatment regimens (e.g. dose and frequency). Consequently, reliable estimates of the relative rate of SAEs for each drug could not be made.

Withdrawal rates due to AEs were typically < 10% for all drugs, with the highest reported single estimate being 13.6% for one etanercept study. This suggested that the majority of patients can tolerate biologic treatment in the medium term, although again the estimates were derived from a highly heterogeneous group of studies; therefore, the possibility of poorer tolerability in specific patient groups was not ruled out.

Open-label extensions of randomised trials

Of the longer-term follow-up studies included in our present review we evaluated those reporting AEs after 6 months (as the Cochrane review covered events occurring up to 6 months); 13 trial cohorts had studies which reported data after 6 months. Both the type of AEs assessed, and the periods over which they were assessed, varied across studies. *Table 26* compares results for studies with at least around 2 years of follow-up. The ATLAS[61] and GO-RAISE[90] trials both had extension study publications at the 2-year and 5-year time points.[140–143] Both cohorts were analysed using mITT data, in which patients had to have received at least one dose of treatment. This amounted to 99% of the randomised patients in both studies (311 out of 315 in ATLAS[61] and 353 out of 356 in GO-RAISE[90]). Davis[72,73] reported results for the 257 patients who enrolled in a 168-week open-label study following week 24 of the randomised phase; 277 patients had taken part in the earlier randomised study. All 257 patients in the open-label study had received at least one dose of etanercept.[144,145] The Calin trial[83–85] randomised 84 patients, with 81 patients enrolling in the open-label extension study. Results were presented separately for the 12-week to 2-year time points and the 2- to 5-year time points.[146,147] RAPID-axSpA[64] data at 96 weeks were reported in the manufacturer's submission. These data related to the mITT population: 315 (97%) of the 325 originally randomised patients.

The 2-year study of the ASSERT (Ankylosing Spondylitis Study for the Evaluation of Recombinant Infliximab Therapy)[102] (infliximab) cohort allowed dose escalation whereby, from week 36, patients with BASDAI scores of ≥ 3 could increase their dose to 7.5 mg/kg, which is a currently unlicensed dose. Results for the 5 mg/kg group (74 patients) between weeks 24 and 102 have therefore been presented in *Table 26*. The Braun cohort[148] was followed up for 8 years, but it was a small study which reported only SAEs and withdrawals due to SAEs.

Table 26 illustrates that rates of SAEs, cancer and serious infections were similar across all four anti-TNFs when using incidence per 100 patient-years as estimates. At 5 years, SAEs appeared more prevalent with adalimumab (45%) than golimumab (20%), although it is possible this difference is because of the way the data were reported; it was unclear whether the ATLAS[61] data related to the total number of SAEs or to the number of patients experiencing a SAE. At 2 years, the incidence of injection site reactions was higher in patients taking etanercept than in patients taking adalimumab, golimumab or certolizumab pegol. Withdrawal rates due to AEs were broadly similar across treatments. The reporting of tuberculosis and congestive heart failure was limited.

TABLE 26 Studies with adverse event data at around 2 years (or later)

	Adalimumab		Golimumab		Etanercept		Certolizumab		Infliximab
	ATLAS[61]		GO-RAISE[90]		Davis[72,73]	Calin[83–85]	RAPID-axSpA65		[a]ASSERT[102]
Event outcome	2 years ($n = 311$)	5 years ($n = 311$)	2 years ($n = 353$)	5 years ($n = 353$)	24–192 weeks ($n = 257$)[b]	12–108 weeks ($n = 81$)[b]	2–5 years ($n = 59$)	96 weeks ($n = 315$)	24–102 weeks ($n = 74$)
SAEs	48 events (15%) 10.5/100 PY	140 events (45%) 11.7/100 PY	40 events (11%)	72 events (20%)	33 events (13%) 8/100 PY	19 events (23%)	21/100 PY	AiC information has been removed	15 events (20%)
Withdrawals because of AEs	24 (8%) events 4.5/100 PY	–	19 events (5%)	32 events (9%) 2.13/100 PY	14 events (5%)	15 events (19%)	7 events (12%)	AiC information has been removed	–
Serious infections	6 (2%) events 1.1/100 PY	17 events (5%) 1.4/100 PY	11 events (3%)	21 events (6%) 2.1/100 PY	6 events (2%) 2/100 PY	5 events (6%)	3 events (5%) 3/100 PY	AiC information has been removed	3 events (4%)
Cancer	4 events (1%) 0.7/100 PY	3 events (1%) 0.2/100 PY	2 events (0.6%)	3 events (0.8%) 0.21/100 PY	–	4 events (5%)	3 events (5%)	–	1 events (1%)
NMSC	0.4/100 PY	–	–	–	–	–	–	–	–
Congestive heart failure	0 events	2 events (0.6%) 0.2/100 PY	–	–	–	–	–	–	–
Injection site reactions	42 events (14%) 17.6/100 PY	–	38 events (11%)	43 events (12%)	57 events (22%)	30 events (37%)	7 events (12%)	AiC information has been removed	9 events (12%)
Tuberculosis	0 events	0 events	–	–	–	0 events	0 events	AiC information has been removed	–

AiC, academic-in-confidence; ASSERT, Ankylosing Spondylitis Study for the Evaluation of Recombinant Infliximab Therapy; NMSC, non-melanoma skin cancer; PY, person-year.
a 5 mg/kg group.
b Weeks from randomisation.

Summary of adverse event data

Data from large systematic reviews, which included patients with a wide range of diseases, suggest that, in the short term, anti-TNFs as a group are associated with significantly higher rates of serious infections, tuberculosis reactivation, non-melanoma skin cancer, total AEs and withdrawals because of AEs than control treatments. Specifically, infliximab is associated with significantly higher rates of total AEs and withdrawals because of AEs, and certolizumab pegol is associated with significantly higher rates of serious infections and SAEs. Analyses from the present review showed etanercept to be statistically significantly more likely to result in an injection/infusion site reaction compared with infliximab, although analysable data on such reactions were not reported for the three other anti-TNFs. Evaluations of longer-term data are more scarce, although suggest similar safety profiles across anti-TNFs. Data from the open-label studies included in this review also do not suggest that there are important differences between treatments, other than a higher incidence of injection site reactions following treatment with etanercept. These open-label data are, however, limited by the small sample sizes and non-randomised study designs.

Review of natural history of ankylosing spondylitis and non-radiographic axial spondyloarthritis

In order to get some understanding of what happens to patients who, although eligible for anti-TNF therapy for their AS or nr-AxSpA, do not receive it, we conducted a rapid review of relevant literature. This was not a systematic review but one that started with the library of papers found by the main searches for RCTs of the anti-TNFs and then followed relevant citations to papers on AS and axSpA in patients not receiving an anti-TNF. Potentially relevant papers were those that reported on the pattern of disease, AS or nr-AxSpA or axSpA, without treatment with anti-TNFs over time. This process identified a number of relevant registries: OASIS,[118] Scotland and Ireland Registry for Ankylosing Spondylitis (SIRAS), Devenir des Spondylarthropathies Indifferenciées Récentes, Esperanza, Spanish Registry of spondyloarthritis, German Spondyloarthritis Inception Cohort (GESPIC) and St Mary RheumaToid Arthritis (SMART). Additional searches of MEDLINE were conducted using these specific registry names. All relevant studies identified through this process are presented in *Table 27*.

The studies collectively explore the associations between the various components of axSpA: disease activity, structural damage and spinal mobility. The exploration of the ASSERT trial baseline data[24,124] reveals that HRQoL as determined by SF-36 physical and mental components, is determined by BASFI and BASDAI; BASFI is determined by BASDAI, mSASSS and BASMI (spinal mobility); and BASMI is independently determined both by irreversible spinal damage (mSASSS) in late disease and reversible spinal damage (MRI) in early disease.

The studies identified that from a clinical practice and patients' point of view disease progression in terms of BASFI, a measure of the patient's functional ability, is very important. A number of studies on the disease progression of AS have been based on the European OASIS cohort[118] (a consecutive cohort, started in 1996, although there were no further specific eligibility criteria); the total cohort numbers 217 patients. One of these, a study by Landewe *et al.*,[10] demonstrated that physical function impairment (BASFI) is independently affected by both disease activity (BASDAI) and bony progression, usually assessed using mSASSS despite this being a measure of bony growth in the spine only (and not in the sacroiliac joints). Other studies by Ramiro[124,125] have demonstrated that radiographic progression, increases on average by around 2 mSASSS units every 2 years.[124,125] However, this progression is highly variable; the average patient with inactive disease [Ankylosing Spondylitis Disease Activity Score (ASDAS) 0] would progress by 5 mSASSS units over 12 years compared with a patient with 'very active disease' (ASDAS 4) who would have 19 units of progression.[124] In addition, of 68 patients who were followed for 12 years, 18% had no progression on mSASSS.[125] The variability is also demonstrated by the results based on a different cohort: a single German clinic ($n = 146$).[12] Baseline characteristics were similar to those in the OASIS cohort[118] (see *Table 27*). Mean follow-up was 3.8 years (SD 1.7 years) and mean mSASSS change was 1.3 units/year (SD 2.5 units/year) with a range of 0–22.8 mSASSS units. Thirty-four (23%) patients showed no progression.

TABLE 27 Natural history of axSpA: relevant outcomes and impact of anti-TNFs

Study	Description	Population characteristics	Summary of findings
Landewe 2009[10]	Examined the relationship between disease activity, radiographic damage and physical function in AS. Based on (European) OASIS cohort[118] baseline and 2-year data. $n = 217$ consecutive (from 1996) patients with AS (no specific criteria). BASDAI score mean 3.4 (SD 2.1), with 38% ≥ 4	BASFI score mean 3.4 (SD 2.6), 41% ≥ 4. mSASSS median 5, 69% > 0. Note: does mSASSS < 0 mean nr-AxSpA? None of the patients in the cohort had used anti-TNFs. Subgroup ($n = 188$) baseline BASDAI score of ≤ 6	Univariate correlation between baseline mSASSS and BASFI score = 0.45 (Spearman's rank correlation coefficient), but this was modified by baseline BASDAI: • BASDAI score of 0–2 ($n = 68$) = 0.68 • BASDAI score of > 2–4 ($n = 60$) = 0.58 • BASDAI score of > 4–6 ($n = 60$) = 0.43 • BASDAI score of > 6–8 ($n = 22$) = 0.40 • BASDAI score of > 8–10 ($n = 7$) = –0.20 Suggests a ceiling effect of BASFI. Owing to the high level of correlation between BASDAI and BASFI, a correlation between mSASSS and BASFI cannot be demonstrated at the highest level of BASDAI. There was a multivariate relationship between BASDAI and mSASSS with BASFI using baseline and 2-year data (but not longitudinal?) ($n = 188$, baseline BASDAI score of ≤ 6 only). Regression coefficients found that both BASDAI and mSASSS are statistically significant ($p < 0.001$) explanatory variables for BASFI (0.73 and 0.057 units respectively)
Ramiro 2014[124]	Analysed long-term relationship between disease activity (ASDAS, BASDAI) and radiographic damage (mSASSS) in AS. Used OASIS cohort[118] over 12 years	Subgroup used patients ($n = 184$) who had at least two sets of radiographs. Baseline characteristics of this subgroup: BASDAI score mean 3.4 (SD 2.0); mSASSS mean 10.8 (SD 15.2), 81% > 0. None of the patients had used anti-TNFs	On average patients had a progression of 1.9 mSASSS units/2 years. This varied with baseline ASDAS: • ASDAS < 1.3 progress = 0.7 mSASSS units/2 years • ASDAS > 3.5 progress = 3.1 mSASSS units/2 years The relationship with BASDAI was similar: • Baseline BASDAI score of < 4, 1.5 mSASSS units/2 years • BASDAI score of ≥ 4, 2.7 mSASSS units/2 years • BASDAI score of > 6 units, 2.0 mSASSS units/2 years The analysis found that the average patient with inactive disease (ASDAS score 1.0) would progress by 5 mSASSS units over 12 years compared with a patient with 'very active disease' (ASDAS score of 4) would have 19 units of progression
Ramiro 2013[125]	Earlier analysis of OASIS cohort[118] 12-year data to describe the evolution of radiographic abnormalities in AS patients	Subgroup used ($n = 186$) who had at least two sets of radiographs. Baseline characteristics of this subgroup: BASDAI score mean 3.4 (SD 2.0); mSASSS mean 11.6 (SD 16.2). None of the patients had used anti-TNFs	Long-term radiographic progression in AS highly variable at the patient level, but is more severe in men who are HLA-B27 positive. Over whole follow-up, 24% of patients (and 18% of the 68 patients who were followed for 12 years) had no progression on mSASSS. Duration of disease is not relevant. At the group level, progress is linear at 2 mSASSS units/2 years

TABLE 27 Natural history of axSpA: relevant outcomes and impact of anti-TNFs (*continued*)

Study	Description	Population characteristics	Summary of findings
Baraliakos 2009[12]	Natural course of radiographic progression in AS. Retrospective cohort, single clinic (Herne, Germany), 1993–2005. Mean follow-up 3.8 years (SD 1.7 years)	$n = 146$ anti-TNF naive patients. Baseline: • mSASSS mean 20.5 (SD 14.4) • BASDAI score mean 4.4 (SD 1.9) (range 0.5–7.3) • BASFI score mean 3.8 (SD 2.6) (range 1.0–8.4)	Mean mSASSS change was 1.3 units/year (SD 2.5 units/year). Note: range was mSASSS 0–22.8. Thirty-four (23%) patients showed no progression
Dean 2014,[149] poster at BSR meeting	SIRAS cohort. Study of BASDAI over time	BASDAI score at diagnosis data available for only 240 patients (out of the 1210 patient cohort). Baseline BASDAI score (at diagnosis) of 4.9 (SD 2.3). High disease activity group BASDAI score of 6.3 (SD 1.4) and low disease activity group BASDAI score of 2.5 (SD 1.3)	Baseline BASDAI remained fairly stable over time: across the whole cohort and in the high and low disease activity groups. The subgroup treated with anti-TNFs had higher mean BASDAI score [5.7 (SD 2.0)] than non-biologic patients [4.2 (SD 2.5)] and this remained so until around a year after treatment with anti-TNFs began, when mean BASDAI fell to the level of the non-biologic patients
Healey 2013[14]	Cohort study, single centre, England. Followed patients over 10 years [$n = 69$ who provided assessments at baseline (1998) and at 10 years (2008)]. Assessments using RLDQ, BASDAI, ASQoL and EQ-5D (and others)	At study entry patients were 84% male, mean age 49 years, disease duration 15.5 years, symptom duration 21.4 years. 1.5% on an anti-TNF at 10 years	Only RLDQ changed significantly over time for assessment 1 (1998) and 2 (2008): • RLDQ: mean 10.4 (SD 8.3); mean 13.6 (SD 10.9); $p = 0.002$ • BASDAI: mean 4.1 (SD 2.5); mean 4.4 (SD 2.7); $p = 0.36$ • ASQoL: mean 6.4 (SD 6.3); mean 7.5 (SD 6.4); $p = 0.15$ • EQ-5D: mean 0.64 (SD 0.28); mean 0.61 (SD 0.30); $p = 0.45$ However, as RLDQ (range 0–48) is a measure of function (comparable with BASFI) it does indicate progression with time even in these AS patients whose disease at study entry was already well established
Stone 2007[150]	Analysis of longitudinal data from SMART (Bath, UK) data set ($n = 224$). Regression analysis of BASDAI score on symptom duration and BASFI score adjusted for BASDAI score > 4 at baseline. Duration of follow-up was unclear	Overall, 68% had a baseline BASDAI score of ≥ 4. Mean symptom duration was 28.8 years	Only 20% experienced a significant change in BASDAI score over time (13% a decrease; 7% an increase). BASFI score increases over time by 0.035 units/symptom-year. In patients with baseline BASDAI score of ≥ 4, those who would be treated with anti-TNFs, the increase over time is 0.039 units/symptom-year
Machado 2010[151]	Baseline data from ASSERT.[102] Analysis of relation between mSASSS and MRI inflammation and BASMI	$n = 214$ AS patients (mNY criteria). Baseline median (IQR): • BASMI score of 4.6 (3.6–5.8) • BASDAI score of 6.5 (5.3–7.0) • CRP level 1.5 mg/dl (0.7–2.9 mg/dl) • mSASSS 13.8 (4.5–29.1)	Concluded that spinal mobility (BASMI) independently determined both by irreversible (mSASSS) and reversible spinal damage (MRI), the former in late disease and the latter in early disease

continued

TABLE 27 Natural history of axSpA: relevant outcomes and impact of anti-TNFs (*continued*)

Study	Description	Population characteristics	Summary of findings
Machado 2011[24]	Baseline data from ASSERT.[102] Analysis of relation between SF-36 and BASFI and BASDAI, ASDAS, CRP level, mSASSS, MRI inflammation and BASMI	$n = 214$ AS patients (mNY criteria)	Regression coefficients for associations reported in the publication. Briefly, SF-36 is determined by BASFI and BASDAI; and BASFI is determined by BASDAI, mSASSS and BASMI
Kobelt 2004[152]	A modelling study of infliximab but refers to large UK observational data set and generates an estimate for BASFI over time. Survey in 2002 ($n = 1413$). Value generated from patients who were captured in two surveys at two time points, 1992/1994 and November 2002, approximately 8 years apart ($n = 1100$). Data from a cohort of 493 patients who had been followed up for more than 3 years were used as a check for the result based on the survey	–	From the whole survey ($n = 1413$) mean BASDAI score 4.2 (SD 2.3) and mean BASFI score 4.4 (SD 2.8). The population was broader than that eligible for anti-TNFs, with 47% having a BASDAI score of < 4. It appears (but is unclear) that this is the BASDAI at the later time (2002) point not the earlier (1992/4). Estimate of annual BASFI progression was 0.07 points. Note: progression was faster (0.1 points) in patients with BASFI score of < 4 at baseline, but was stable (0?) in patients with BASFI score above 7. (Ceiling effect of BASFI?) When only patients with BASDAI score of ≥ 4 included BASFI progression was estimated as 0.054. Data from the cohort study generated similar findings; however, the number was not actually reported for whole survey. BASFI progression was 0.059 for patients with a BASDAI score of ≥ 4
Nr-axSpA			
Kiltz 2012[153]	Comparison of characteristics of patients with AS and nr-AxSpA. Cohort of 100 patients seen in 2010 in Herne clinic, Germany. Analysis tested if the proportion of patients reaching pre-specified cut-off criteria (markers of disease severity) differed between AS and nr-AxSpA	Consecutive, diagnosed with axSpA. None of the patients had used anti-TNFs. $n = 100$ AxSpA: $n = 44$ nr-AxSpA and $n = 56$ AS • Median BASDAI score, 4.3 (AS) and 3.6 (nr-AxSpA); $p = 0.2$ • Median BASFI score, 2.9 (AS) and 1.5 (nr-AxSpA); $p = 0.05$ • Median CRP level, 8.0 mg/l (AS) and 3.8 mg/l (nr-AxSpA); $p < 0.001$ • Median mSASSS 3.0 (AS) and 1.1 (nr-AxSpA); $p < 0.007$	Differences were statistically significant for ASDAS, CRP level, mSASSS and number of inflamed lesions. Proportion of males also significantly different. Results: • % male (p-value), 31.8% (nr-AxSpA) and 76.8% (AS); $p < 0.001$ • BASDAI score of ≥ 4: 43% (nr-AxSpA) and 53.5% (AS); $p = 0.1$ • BASFI score of ≥ 3: 34.1% (nr-AxSpA) and 46.4% (AS); $p = 0.08$ • ASDAS > 2: 54.5% (nr-AxSpA) and 78.6% (AS); $p = 0.01$ • CRP level > 5 mg/l: 29.5% (nr-AxSpA) and 69.1% (AS); $p < 0.001$ • mSASSS ≥ 3: 27.3%(nr-AxSpA) and 51.9% (AS); $p = 0.01$ • Number of inflamed lesions per patient ≥ 3: 9.1% (nr-AxSpA) and 46.4% (AS); $p = 0.01$

TABLE 4 Risk of bias assessment results

Trial	Bias domain							
	1. Sequence generation	2. Allocation concealment	3. Important baseline imbalance	Selection bias based on 1, 2, and 3	4. Blinding of participants and personnel	5. Blinding of outcome assessment	6. Incomplete outcome data	7. Selective reporting
	Risk of bias judgement							
Adalimumab vs. placebo								
Haibel 2008[52]	Unclear	Unclear	Unclear	Unclear	Low	Low	Low	Low
Hu 2012[55]	Unclear	Unclear	Low	Low	Unclear	Low	Unclear	Unclear
Huang 2014[56]	Low	Low	Low	Low	Low	Low	Low	Low
Lambert 2007[57]	Unclear	Unclear	Low	Low	Unclear	Low	Low	Low
ABILITY-1 2013[58]	Low	Low	Low	Low	Low	Low	Low	Low
ATLAS 2006[61]	Unclear	Unclear	Low	Low	Low	Low	Low	Low
Certolizumab pegol vs. placebo								
RAPID-axSpA 2014[64]	Low	Low	Low	Low	Unclear	Low	Low	Low
Etanercept vs. placebo								
Barkham 2010[71]	Unclear	Unclear	Unclear	Unclear	Low	Low	Unclear	Low
Davis 2003[72]	Unclear	Unclear	Low	Low	Low	Low	Low	Low
Dougados 2011[74]	Unclear	Unclear	Low	Low	Low	Low	Low	Low
Dougados 2014[76]	Low	Low	Low	Low	Low	Low	Low	Low
Gorman 2002[79]	Low	Low	High[a]	High[a]	Low	Low	Low	Low
Calin 2004[83]	Unclear	Unclear	High	High	Low	Low	Low	Low
van der Heijde 2006[86]	Unclear	Unclear	Unclear	Unclear	Low	Low	Low	Low

continued

TABLE 4 Risk of bias assessment results (*continued*)

	Bias domain							
	1. Sequence generation	2. Allocation concealment	3. Important baseline imbalance	Selection bias based on 1, 2, and 3	4. Blinding of participants and personnel	5. Blinding of outcome assessment	6. Incomplete outcome data	7. Selective reporting
Trial	Risk of bias judgement							
Etanercept vs. infliximab								
Giardina 2010[88]	High	High	Low	Low	High	High	Low	Low
Golimumab vs. placebo								
GO-RAISE 2008[90]	Low	Low	Low	Low	Low	Low	Low	Low
Bao 2014[95]	Unclear	Unclear	Unclear	Unclear	Low	Low	Unclear	Low
Tam 2014[97]	Low	Unclear	Unclear	Unclear	Low	Low	Low	Low
Infliximab vs. placebo								
Barkham 2009[50]	Unclear	Unclear	High	High	Low	Low	Low	Low
Braun 2002[98]	Low	Low	Low	Low	Low	Low	Low	Low
Marzo-Ortega 2005[100]	Low	Low	Unclear	Low	Low	Low	Unclear	Low
Van den Bosch 2002[101]	Unclear	Unclear	High	High	Low	Low	Low	Low
ASSERT[102]	Unclear	Unclear	Low	Low	Low	Low	Low	Low
Infliximab vs. biosimilar infliximab (InflectraCT-P13)								
PLANETAS 2013[110]	Low	Low	Unclear	Low	Low	Low	Low	Low

ASSERT, Ankylosing Spondylitis Study for the Evaluation of Recombinant Infliximab Therapy; ATLAS, Adalimumab Trial Evaluating Long-term Efficacy and Safety for Ankylosing Spondylitis.
a Judged to be likely to be because of chance.

The best model

Models were run when it was assumed that:

- there were different independent treatment effects
- there was just one treatment class effect.

In addition, fixed-effect and random-effects models were run when there were sufficient data. These models relate to models A1, A2, A3 and A4 in *Chapter 5*. For the non-radiographic population, there were very few studies and therefore only fixed-effect analyses were conducted.

The DIC and I^2 results for each outcome (other than injection site reactions) are shown for the AS population in *Table 5* and for the nr-AxSpA population in *Table 6*. The lower the DIC for a given outcome, the better the model fit. I^2 varies between 0% and 100%, with 0% representing no heterogeneity in the results and 100% indicating that all of the variation in the results can be explained by heterogeneity. The greater the value of I^2, the more likely it is that a random-effects model would be a better fit. But this is not always the case, as if there are few studies then there will be significant uncertainty around the between-study variance and therefore the I^2 also. Random-effect model results and I^2 results are not presented for some outcomes because of sensitivity to prior distributions in the model.

Overall, assuming a class effect for the treatments produced a better-fitting model than assuming independent treatment effects. In addition, a fixed-effect analysis was more often than not appropriate. The mean and median effects of the two analyses were also similar. Hence the fixed-effect results are reported in this chapter; these represent a common class effect.

For AS, the common class-effect model was found to be a much better fit than the independent treatment effect model. As described in *Chapter 5* the exchangeable class-effect model, not explored here, also fitted the data well, although not so well as the common class-effect model. It should be noted here that the common class-effect model may possibly underestimate the uncertainty around the treatment effect estimate. As explained in *Chapter 5*, if the differences between treatments are a result of systematic

TABLE 5 The AS population model DIC statistics

	Independent effects		Class effect		
Outcome	Fixed effect	Random effects	Fixed effect	Random effects	I^2 (%)
Related model in *Chapter 5*	A1	A2	A3	A4	–
BASDAI 50	16.82	–	10.86	12.71	21
BASDAI	16.76	18.22	13.53	15.12	21
BASFI	18.96	20.87	14.79	16.80	10
ASAS 20	10.68	17.05	9.98	8.73	16
ASAS 40	10.36	14.07	8.50	10.29	27
ASAS 50	8.38	–	6.68	8.11	52
ASAS 70	2.92	–	–	–	–
BASMI	–0.87	–	0.12	–3.01	77
SF-36 PCS	19.64	–	20.20	17.71	76
MASES	5.99	–	4.17	–	–
SF-36 MCS	19.20	–	16.67	18.26	47

MASES, Maastricht Ankylosing Spondylitis Enthesitis Score.

TABLE 6 The nr-AxSpA population model DIC statistics

	Independent effects		Class effect		
Outcome	Fixed effect	Random effects	Fixed effect	Random effects	I^2 (%)
Related model in *Chapter 5*	A1	A2	A3	A4	–
BASDAI 50	6.74	–	4.85	–	–
BASDAI	10.80	–	11.07	11.51	69
BASFI	11.45	–	13.74	10.70	83
ASAS 20	6.72	–	5.23	–	–
ASAS 40	11.17	–	7.96	9.30	49
ASAS 50	–	–	–	–	–
ASAS 70	–	–	–	–	–
BASMI	1.80	–	4.74	2.42	89
SF-36 PCS	16.67	–	20.18	–	–
MASES	–	–	–	–	–
SF-36 MCS	14.61	–	14.08	–	–

MASES, Maastricht Ankylosing Spondylitis Enthesitis Score.

differences in study design between treatments, then an exchangeable class-effect model may be appropriate. However, if in fact there is a true difference between treatments, such as between infliximab and the other TNF-inhibitors, then an exchangeable class-effect model may overestimate the uncertainty around the effect estimates. As the common class-effect model had a lower DIC than the exchangeable class-effect model, this is the model evaluated in this chapter. The economic model explores the assumption that treatment effect differences are in fact because of systematic differences in study design between treatments.

As there was very little difference between the results in which change from baseline was imputed assuming a within-study correlation of 0.3 or 0.7, only the results assuming a within-study correlation of 0.3 are reported here. A comparison of the results assuming different within-study correlations is presented in *Appendix 5*.

Individual anti-tumour necrosis factors compared with placebo

Binary responder outcomes at between 10 and 16 weeks

The results of the analyses of the responder outcomes between 10 and 16 weeks for patients with AS are presented in *Table 7*.

Assessment in Ankylosing Spondylitis improvement criteria: Assessment in Ankylosing Spondylitis 20, Assessment in Ankylosing Spondylitis 40, Assessment in Ankylosing Spondylitis 50 and Assessment in Ankylosing Spondylitis 70

For the AS population ASAS 20 data were available for all five anti-TNFs, although the number of participants studied varied considerably, ranging from 839 patients in five etanercept trials to 111 patients in two infliximab trials. A consistent effect was evident across the treatments with the pooled relative risks ranging from 1.80 (certolizumab pegol) to 2.45 (infliximab). ASAS 40 data were available for four anti-TNFs (no data were available for infliximab); the number of data available ranged from 178 patients in one certolizumab trial to 659 patients in two adalimumab trials. Again, a consistent effect was found, with relative risks ranging from 2.53 (certolizumab pegol) to 3.42 (adalimumab); all the relative risks were

TABLE 7 Results versus placebo for AS population: response outcomes at between 10 and 16 weeks

Intervention	Type of analysis	ASAS 20			ASAS 40			ASAS 50			BASDAI 50		
		Number of trials (number of patients)	Relative risk (95% CrI)	OR (95% CrI)	Number of trials (number of patients)	Relative risk (95% CrI)	OR (95% CrI)	Number of trials (number of patients)	Relative risk (95% CrI)	OR (95% CrI)	Number of trials (number of patients)	Relative risk (95% CrI)	OR (95% CrI)
Adalimumab	Main	3 (741)	2.28 (1.98 to 2.62)	4.52 (3.23 to 6.33)	2 (659)	3.42 (2.57 to 4.55)	5.67 (3.56 to 8.97)	1 (82)	2.75 (1.11 to 5.45)	3.58 (1.12 to 11.17)	2 (659)	3.16 (2.40 to 4.16)	4.68 (3.14 to 7.03)
	Sensitivity	3 (741)	2.27 (1.97 to 2.62)	4.52 (3.23 to 6.33)	2 (659)	3.34 (2.53 to 4.40)	5.67 (3.56 to 8.97)	As in the main analysis			2 (659)	3.11 (2.37 to 4.09)	4.68 (3.14 to 7.03)
Certolizumab pegol	Main	1 (178)	1.80 (1.24 to 2.39)	2.61 (1.37 to 5.01)	1 (178)	2.53 (1.47 to 3.98)	3.38 (1.59 to 7.15)	–	–	–	1 (178)	3.60 (2.02 to 5.74)	5.97 (2.39 to 15.03)
	Sensitivity	1 (178)	1.80 (1.24 to 2.39)	2.61 (1.37 to 5.01)	1 (178)	2.49 (1.46 to 3.87)	3.38 (1.59 to 7.15)	–	–	–	1 (178)	3.53 (2.00 to 5.58)	5.97 (2.39 to 15.03)
Etanercept	Main	5 (839)	2.23 (1.93 to 2.55)	4.23 (3.05 to 5.88)	3 (478)	2.75 (1.88 to 3.88)	3.86 (2.21 to 6.72)	2 (359)	3.43 (2.40 to 4.90)	5.04 (2.98 to 8.51)	3 (478)	3.17 (2.20 to 4.49)	4.74 (2.71 to 8.28)
	Sensitivity	3 (715)	2.17 (1.84 to 2.53)	3.98 (2.78 to 5.73)	2 (436)	2.65 (1.80 to 3.72)	3.72 (2.11 to 6.53)	As in the main analysis			2 (436)	3.03 (2.08 to 4.31)	4.50 (2.52 to 8.01)
Golimumab	Main	2 (429)	2.14 (1.75 to 2.53)	3.82 (2.47 to 5.86)	2 (429)	3.11 (2.24 to 4.26)	4.77 (2.85 to 7.98)	–	–	–	2 (429)	3.57 (2.51 to 5.00)	5.85 (3.31 to 10.28)
	Sensitivity	2 (429)	2.13 (1.74 to 2.53)	3.82 (2.47 to 5.86)	2 (429)	3.05 (2.21 to 4.13)	4.77 (2.85 to 7.98)	–	–		2 (429)	3.50 (2.48 to 4.88)	5.85 (3.31 to 10.28)
Infliximab	Main	2 (111)	2.45 (1.73 to 3.06)	5.54 (2.41 to 12.71)	–	–	–	1 (69)	5.59 (2.44 to 9.81)	14.71 (3.07 to 72.69)	1 (69)	4.86 (2.41 to 7.82)	12.07 (3.09 to 46.37)
	Sensitivity	2 (111)	2.44 (1.72 to 3.06)	5.54 (2.41 to 12.71)	–	–	–	As in the main analysis			1 (69)	4.72 (2.38 to 7.54)	12.07 (3.09 to 46.37)
Anti-TNFs as a class	Main	13 (2298)	2.21 (2.01 to 2.43)	4.12 (3.40 to 4.99)	8 (1744)	3.06 (2.52 to 3.76)	4.61 (3.51 to 6.05)	4 (510)	3.51 (2.55 to 4.86)	5.23 (3.31 to 8.27)	9 (1813)	3.37 (2.75 to 4.16)	5.22 (4.00 to 6.79)
	Sensitivity	11 (2174)	2.18 (1.97 to 2.42)	4.04 (3.32 to 4.92)	7 (1702)	2.99 (2.47 to 3.66)	4.57 (3.48 to 6.02)	As in the main analysis			8 (1771)	3.29 (2.68 to 4.07)	5.16 (3.94 to 6.72)

greater than the corresponding ASAS 20 estimates. For ASAS 50 there were two trials of etanercept (totalling 359 participants) and small single trials in adalimumab ($n = 82$) and infliximab ($n = 69$). A wider range of relative risks and CrIs resulted, ranging from 2.75 (adalimumab) to 5.59 (infliximab), which may be a consequence of the smaller numbers of patients studied. Only two trials, both of etanercept ($n = 359$), reported actual numbers of ASAS 70 responders. The pooling of these data showed that patients taking etanercept were more than three times more likely to be ASAS 70 responders than patients taking placebo (relative risk 3.59, 95% CrI 2.18 to 5.87).

For the nr-AxSpA population, each of the relative risks for certolizumab pegol and etanercept were based on single, quite large trials; the estimate for adalimumab was based on a similar number of patients (to etanercept and certolizumab) across two trials, whereas infliximab was represented by a single small trial ($n = 40$). ASAS 20 results were similar across treatments but for ASAS 40 heterogeneity of effect appeared evident; the smallest estimate was for etanercept and the largest estimate was seen in the small infliximab trial (*Table 8*). However, this infliximab trial was the only nr-AxSpA trial judged to be at high risk of bias. Only one trial (ABILITY-1[58]) reported ASAS 50 or ASAS 70 results. For ASAS 50 the relative risk was 4.23 (95% CrI 1.84 to 9.72; OR 5.96, 95% CrI 2.40 to 14.80). For ASAS 70 the relative risk was 4.58 (95% CrI 1.37 to 15.40; OR 5.42, 95% CrI 1.54 to 19.11).

Bath Ankylosing Spondylitis Disease Activity Index 50

For the AS population BASDAI 50 data were available for all five anti-TNFs; the number of participants studied varied widely, ranging from 69 patients in one infliximab trial to 659 patients in two adalimumab trials. Although a consistent beneficial effect was evident across treatments, some heterogeneity of effect could be seen with the relative risks ranging from 3.16 (adalimumab) to 4.86 (infliximab).

For the nr-AxSpA population the relative risks were lower than for the AS population being 2.52 (95% CrI 1.65 to 3.83, two trials) for adalimumab, 2.80 (95% CrI 1.71 to 4.47, one trial) for certolizumab and 1.92 (95% CrI 1.27 to 2.82, one trial) for etanercept (see *Table 8*).

Results of the AS sensitivity analyses were very similar to those of the main analyses (see *Table 7*).

Continuous outcomes at between 10 and 16 weeks

The results of the analyses of the continuous efficacy outcomes for patients with AS are presented in *Table 9*.

For the AS population, when compared with placebo, adalimumab ($n = 705$), certolizumab pegol ($n = 178$), etanercept ($n = 483$) and infliximab ($n = 132$) produced statistically significant reductions in disease activity, when assessed using BASDAI. The magnitude of the reductions in change from baseline BASDAI score ranged from 1.46 units (certolizumab pegol) to 2.28 units (infliximab). None of the three golimumab trials reported BASDAI as a continuous outcome. The number of data available for BASFI in patients with AS ranged from 132 patients in three infliximab trials, to 523 patients in five etanercept trials. When compared with placebo, all five anti-TNFs produced statistically significant improvements in function. The magnitude of the reductions in change from baseline BASFI score ranged from 1.1 units (certolizumab pegol) to 2.16 units (infliximab). When compared with placebo, statistically significant improvements in BASMI scores were found for AS patients taking adalimumab (mean difference in change from baseline –0.37 units, 95% CrI –0.50 to –0.23 units) and etanercept (mean difference in change from baseline –0.37 units, 95% CrI –0.65 to –0.09 units) but not for certolizumab pegol (mean difference in change from baseline –0.26 units, 95% CrI –0.55 to 0.03 units) and golimumab (mean difference in change from baseline –0.11 units, 95% CrI –0.26 to 0.04 units). Results for SF-36 MCS, SF-36 PCS and ethesitis [Maastricht Ankylosing Spondylitis Enthesitis Score (MASES)] are presented in *Table 9*.

TABLE 8 Results vs. placebo for nr-AxSpA population: response outcomes at 10–16 weeks

Intervention	ASAS 20			ASAS 40			BASDAI 50		
	Number of trials (number of patients)	Relative risk (95% CrI)	OR (95% CrI)	Number of trials (number of patients)	Relative risk (95% CrI)	OR (95% CrI)	Number of trials (number of patients)	Relative risk (95% CrI)	OR (95% CrI)
Adalimumab	2 (188)	1.92 (1.47 to 2.56)	3.71 (2.02 to 6.75)	2 (188)	3.14 (1.99 to 4.68)	5.04 (2.44 to 10.32)	2 (188)	2.52 (1.65 to 3.83)	3.97 (1.97 to 7.86)
Certolizumab pegol	1 (147)	1.59 (1.10 to 2.21)	2.32 (1.15 to 4.67)	1 (147)	3.04 (1.74 to 4.81)	4.75 (2.01 to 11.17)	1 (147)	2.80 (1.71 to 4.47)	4.92 (2.09 to 11.58)
Etanercept	1 (215)	1.46 (1.08 to 1.94)	1.94 (1.13 to 3.37)	1 (215)	2.07 (1.26 to 3.20)	2.55 (1.32 to 4.92)	1 (215)	1.92 (1.27 to 2.82)	2.45 (1.37 to 4.43)
Infliximab	–	–	–	1 (40)	3.63 (1.41 to 6.44)	6.85 (1.52 to 31.03)	–	–	–
Anti-TNFs as a class	4 (550)	1.65 (1.37 to 2.04)	2.52 (1.78 to 3.59)	5 (590)	2.74 (2.08 to 3.62)	3.92 (2.61 to 5.91)	4 (550)	2.31 (1.76 to 3.10)	3.33 (2.24 to 4.96)

TABLE 9 Results vs. placebo for AS population: continuous outcomes at 10–16 weeks

		BASDAI score		BASFI score	
Intervention	Type of analysis	Number of trials (number of patients)	Mean difference in change from baseline (95% CrI)	Number of trials (number of patients)	Mean difference in change from baseline (95% CrI)
Adalimumab	Main	3 (705)	−1.55 (−1.88 to −1.22)	2 (390)	−1.25 (−1.63 to −0.87)
	Sensitivity	2 (659)	−1.55 (−1.89 to −1.21)	1 (344)	−1.28 (−1.68 to −0.88)
Certolizumab pegol	Main	1 (178)	−1.46 (−2.17 to −0.74)	1 (178)	−1.10 (−1.83 to −0.37)
	Sensitivity	Same as the main analysis		Same as the main analysis	
Etanercept	Main	4 (483)	−1.75 (−2.14 to −1.37)	5 (523)	−1.43 (−1.82 to −1.04)
	Sensitivity	2 (359)	−1.72 (−2.16 to −1.29)	2 (359)	−1.29 (−1.76 to −0.84)
Golimumab	Main	–	–	2 (429)	−1.45 (−1.84 to −1.05)
	Sensitivity	–	–	Same as the main analysis	
Infliximab	Main	3 (132)	−2.28 (−3.18 to −1.38)	3 (132)	−2.16 (−3.18 to −1.12)
	Sensitivity	2 (111)	−2.18 (−3.14 to −1.21)	2 (111)	−1.94 (−3.07 to −0.80)
Anti-TNFs as a class	Main	11 (1498)	−1.66 (−1.88 to −1.43)	13 (1652)	−1.38 (−1.59 to −1.18)
	Sensitivity	7 (1305)	−1.63 (−1.88 to −1.39)	8 (1419)	−1.34 (−1.57 to −1.12)

MASES, Maastricht Ankylosing Spondylitis Enthesitis Score.

BASMI score		SF-36 PCS score		SF-36 MCS score		MASES	
Number of trials (number of patients)	Mean difference in change from baseline (95% CrI)	Number of trials (number of patients)	Mean difference in change from baseline (95% CrI)	Number of trials (number of patients)	Mean difference in change from baseline (95% CrI)	Number of trials (number of patients)	Mean difference in change from baseline (95% CrI)
2 (659)	–0.37 (–0.50 to –0.23)	2 (659)	3.53 (2.37 to 4.68)	2 (659)	1.41 (–0.19 to 3.02)	2 (659)	–0.50 (–0.89 to –0.11)
Same as the main analysis		Same as the main analysis		Same as the main analysis		Same as the main analysis	
1 (178)	–0.26 (–0.55 to 0.03)	1 (178)	5.64 (3.64 to 7.66)	1 (178)	1.25 (–2.08 to 4.61)	–	–
Same as the main analysis		Same as the main analysis		Same as the main analysis		–	–
1 (82)	–0.37 (–0.65 to –0.09)	–		–	–	–	–
Same as the main analysis		–		–	–	–	–
2 (429)	–0.11 (–0.26 to 0.04)	2 (429)	5.06 (3.71 to 6.40)	2 (429)	2.75 (1.08 to 4.40)	1 (216)	–0.70 (–1.53 to 0.11)
Same as the main analysis		Same as the main analysis		Same as the main analysis		Same as the main analysis	
–	–	–	–	–	–	–	–
–	–	–	–	–	–	–	–
6 (1348)	–0.27 (–0.36 to –0.18)	5 (1266)	4.40 (3.60 to 5.21)	5 (1266)	1.93 (0.12 to 3.72)	3 (875)	–0.54 (–0.89 to –0.19)
Same as the main analysis		Same as the main analysis		Same as the main analysis		Same as the main analysis	

For the nr-AxSpA population, a heterogeneity of effect on BASDAI and BASFI appears evident from the relative risks of the individual anti-TNFs. The smallest estimates were for etanercept and the largest estimates were seen in the small infliximab trial, although this trial was the only nr-AxSpA trial judged to be at high risk of bias (*Table 10*).

Results of the AS sensitivity analyses were very similar to those of the main analyses (see *Table 9*).

When the mean baseline BASDAI and BASFI are presented by treatment response at week 12 (or 14 for golimumab) for three of the five anti-TNFs (see *Appendix 6*), it can be seen that in patients with AS and patients with nr-AxSpA, on average baseline BASDAI does not differ greatly between responders and non-responders to either placebo or active anti-TNF therapy. In patients with AS or nr-AxSpA from the trials of adalimumab [ATLAS (Adalimumab Trial Evaluating Long-term Efficacy and Safety for Ankylosing Spondylitis)[61] and ABILITY-1[58]] and golimumab (GO-RAISE[90]) on average baseline BASFI was higher in non-responders compared with responders. However, this was not seen in the etanercept trials.

Individual anti-tumour necrosis factors compared with each other

For efficacy outcomes, all of the comparisons that could be made between different anti-TNFs at 10–16 weeks resulted in no statistically significant differences between treatments. For the full results see *Appendix 7*.

One small trial, which could not be included in the meta-analysis (see *Study characteristics*), compared infliximab with etanercept in a 2-year unblinded randomised study of 50 AS patients.[88] At 12 weeks there were statistically significant differences between groups in terms of BASDAI score (3.5 vs. 5.6; $p < 0.005$) and BASFI score (3.5 vs. 5; $p < 0.005$), favouring treatment with infliximab. By week 48, the BASDAI and BASFI scores were almost identical across the treatment groups (data were only presented graphically). In addition, at 12 weeks 19 of 25 infliximab patients were ASAS 20 responders compared with 15 of 25 etanercept patients (not a statistically significant difference). This study concluded that infliximab produces a more rapid clinical improvement, but, at the end of the study, treatment with both etanercept and infliximab was effective and safe. The results of this trial may explain why at 10–16 weeks the meta-analysis results for infliximab were a little better than those of the other anti-TNFs.

Another trial which could not be included in the meta-analysis compared infliximab with an infliximab biosimilar called CT-P13 in 250 AS patients.[110] The ASAS 40 response rates at week 14 were 42% for CT-P13 and 46% for infliximab [OR 0.85; 95% confidence interval (CI) 0.51 to 1.42] and at week 30 they were 52% for CT-P13 and 47% for infliximab (OR 1.19, 95% CI 0.70 to 2.00). At week 14 BASDAI median change from baseline scores were identical (–2.7) and at week 30 they differed slightly (–3.1 CT-P13 vs. –2.5 infliximab). For BASFI the median change from baseline scores were –2.2 CT-P13 versus –2.4 infliximab at week 14 and –2.6 CT-P13 versus –2.2 infliximab at week 30. The study concluded that CT-P13 had a comparable efficacy and safety profile with that of infliximab.

Anti-tumour necrosis factors as a class compared with placebo

Within this section the class effect, calculated as a common effect across all the TNF-inhibitors under consideration, assumes a single treatment effect for all the TNF-inhibitors. It is calculated as the pooled treatment effect using a fixed effect model. The common class-effect model may possibly underestimate the uncertainty around the treatment effect estimate. As explained in *Chapter 5*, if the differences between treatments is a result of systematic differences in study design between treatments then an exchangeable class-effect model may be appropriate. However, if in fact there is a true difference between treatments, such as between infliximab and the other TNF-inhibitors, then an exchangeable class-effect model may overestimate the uncertainty around the mean class-effect estimates. As the common class-effect model had a lower DIC than the exchangeable class-effect model, this is the model evaluated in this chapter. The economic model in *Chapter 6* explores the assumption that treatment effect differences are because of differences in study design between treatments.

TABLE 10 Results vs. placebo for nr-AxSpA population: continuous outcomes at between 10 and 16 weeks

Intervention	BASDAI score		BASFI score		BASMI score		SF-36 PCS score		SF-36 MCS score	
	Number of trials (number of patients)	Mean difference in change from baseline (95% CrI)	Number of trials (number of patients)	Mean difference in change from baseline (95% CrI)	Number of trials (number of patients)	Mean difference in change from baseline (95% CrI)	Number of trials (number of patients)	Mean difference in change from baseline (95% CrI)	Number of trials (number of patients)	Mean difference in change from baseline (95% CrI)
Adalimumab	2 (188)	–1.23 (–1.83 to –0.62)	2 (188)	–0.90 (–1.44 to –0.36)	2 (188)	–0.02 (–0.24 to 0.20)	2 (188)	4.98 (2.74 to 7.20)	2 (188)	1.13 (–1.86 to 4.13)
Certolizumab pegol	1 (147)	–1.85 (–2.83 to –0.88)	1 (147)	–1.90 (–2.87 to –0.94)	1 (147)	–0.55 (–0.89 to –0.20)	1 (147)	6.99 (4.23 to 9.76)	1 (147)	4.01 (0.44 to 7.53)
Etanercept	1 (215)	–0.70 (–1.54 to 0.12)	1 (215)	–0.60 (–1.16 to –0.06)	–	–	–	–	–	–
Infliximab	1 (40)	–2.67 (–4.21 to –1.13)	1 (40)	–2.24 (–3.67 to –0.80)	1 (40)	0.00 (–0.44 to 0.44)	1 (40)	2.10 (–0.21 to 4.37)	–	–
Anti-TNFs as a class	5 (590)	–1.32 (–1.74 to –0.90)	5 (590)	–0.99 (–1.34 to –0.64)	4 (375)	–0.15 (–0.32 to 0.02)	4 (375)	4.41 (3.04 to 5.81)	3 (335)	2.33 (0.07 to 4.62)

Binary responder outcomes at between 10 and 16 weeks

Assessment in Ankylosing Spondylitis improvement criteria – ASAS 20, ASAS 40, ASAS 50 and ASAS 70 When compared with placebo, anti-TNFs as a common class were more than twice as likely to result in patients with AS achieving an ASAS 20 response (relative risk 2.21, 95% CrI 2.01 to 2.43; 13 trials, see *Table 7*). Anti-TNFs were also around three times as likely to result in patients achieving an ASAS 40 response (relative risk 3.06, 95% CrI 2.52 to 3.76; eight trials) and three and a half times as likely to result in patients achieving an ASAS 50 response (relative risk 3.51, 95% CrI 2.55 to 4.86; four trials). Only two trials, both of etanercept, reported data suitable for the ASAS 70 analysis; the results are presented in *Individual anti-TNFs compared with placebo*. There was little evidence of heterogeneity for ASAS 20 ($I^2 = 16\%$) and ASAS 40 ($I^2 = 27\%$) but heterogeneity was evident for ASAS 50 ($I^2 = 52\%$). For ASAS 50, three of the four trials were small (i.e. fewer than 100 participants), which may partly explain the heterogeneity estimate.

For the nr-AxSpA population anti-TNFs as a common class were statistically significantly more effective than placebo, although the relative risks being lower than for the AS population. For ASAS 20 the relative risk was 1.65 (95% CrI 1.37 to 2.04; four trials) and for ASAS 40 the relative risk was 2.74 (95% CrI 2.08 to 3.62; five trials). Only one trial presented ASAS 50 and ASAS 70 results (see *Clinical effectiveness results: efficacy results from randomised controlled trials*). A heterogeneity estimate could be calculated for only ASAS 40 ($I^2 = 49\%$).

BASDAI 50 Anti-TNFs as a common class resulted in patients with AS being more than three times more likely to achieve a BASDAI 50 response than patients taking placebo (relative risk 3.37, 95% CrI 2.75 to 4.16; nine trials). There was little evidence of heterogeneity ($I^2 = 21\%$).

For the nr-AxSpA population, anti-TNFs as a common class were also statistically significantly more effective than placebo in terms of achieving a BASDAI 50, although the relative risk was lower than for the AS population (relative risk 2.31, 95% CrI 1.76 to 3.10; four trials). Results of the AS sensitivity analyses were very similar to the main analyses (see *Table 7*).

Binary responder outcomes at between 24 and 30 weeks

Four AS trials reported outcomes at between 24 and 30 weeks (see *Table 2*). Anti-TNFs as a common class were statistically significantly more effective than placebo at 24–30 weeks; for ASAS 20 the relative risk was 1.69 (95% CrI 1.30 to 2.14; four trials). No studies reported BASDAI 50 or ASAS 70 results, and only single studies reported on ASAS 40 (relative risk 4.01, 95% CrI 2.13 to 7.55)[102] and ASAS 50 (relative risk 4.17, 95% CrI 2.45 to 7.12).[72]

Continuous outcomes at between 10 and 16 weeks

When considered together as a group compared with placebo (see *Table 9*), treatment with an anti-TNF in patients with AS produced statistically significant improvements (calculated using mean difference in change from baseline) in the following areas: disease activity (BASDAI mean difference –1.66 units, 95% CrI –1.88 to –1.43 units; 11 trials); function (BASFI mean difference –1.38 units, 95% CrI –1.59 to –1.18 units; 13 trials); spinal mobility (BASMI mean difference –0.27 units, 95% CrI –0.36 to –0.18 units); physical health (SF-36 PCS mean difference 4.40 units, 95% CrI 3.60 to 5.21 units; five trials); mental health (SF-36 MCS mean difference 1.96 units, 95% CrI 0.87 to 3.05 units; five trials); and enthesitis (MASES mean difference –0.54 units, 95% CrI –0.89 to –0.19 units; three trials). There was little evidence of heterogeneity for BASDAI ($I^2 = 21\%$) and BASFI ($I^2 = 10\%$), but evidence of substantial heterogeneity for BASMI ($I^2 = 77\%$), SF-36 PCS ($I^2 = 76\%$), SF-36 MCS ($I^2 = 47\%$) and MASES ($I^2 = 91\%$).

In the nr-AxSpA population the mean differences achieved with anti-TNFs (see *Table 10*) were also statistically significant, although slightly lower than for the AS population. For BASDAI the mean difference was –1.32 units (95% CrI –1.74 to –0.90; $I^2 = 69\%$) and for BASFI the mean difference was –0.99 units (95% CrI –1.34 to –0.64 units; $I^2 = 83\%$) but there was evidence of substantial heterogeneity. The results for SF-36 MCS and SF-36 PCS were similar to those for AS (see *Table 10*).

Results of the AS sensitivity analyses were very similar to the main analyses (see *Table 9*). As the results of the independent treatment effects showed a trend that infliximab had a greater, although not statistically significant, effect on the change in BASDAI and BASFI from baseline, an additional sensitivity analysis was conducted for which infliximab was assumed to be different from the rest of the anti-TNFs. The results are presented in *Table 11*. The low weight of evidence available for infliximab ensures that the class effect for the other anti-TNFs does not change greatly. Although it is possible that infliximab has a greater effect than the other anti-TNFs at least at 12 weeks, there is no strong evidence from these analyses to suggest that it does.

Continuous outcomes at between 24 and 30 weeks

Four AS trials reported outcomes at between 24 and 30 weeks (see *Table 2*). The mean differences in change from baseline were −1.98 units (95% CrI −2.27 to −1.68 units, four trials) for BASDAI, −0.87 units (95% CrI −1.11 to −0.62 units; three trials)[72,100,102] for BASFI, and −1.00 unit (95% CrI −1.19 to −0.81 units; two studies)[97,102] for BASMI. One study reported SF-36 outcomes, with differences of 9.40 units (95% CrI 7.88 to 10.92 units) for SF-36 PCS and 0.70 units (95% CrI −1.36 to 2.76 units) for SF-36 MCS.[102]

Outcomes not included in the meta-analyses

Very few data were available on peripheral symptoms (other than enthesitis, see the MASES results in *Table 9*) or symptoms of extra-articular manifestations. One trial reported five cases of inflammatory bowel disease flare up to the 24-week time point: three occurred in patients on etanercept and two in patients on placebo.[72] Another study reported that there were no cases of inflammatory bowel disease at 12 weeks.[86] Incidence of uveitis was also reported in one trial; up to the 24-week time point there were three cases in the etanercept arm and eight cases in the placebo arm.[72]

One trial (ABILITY-1[58]) reported statistically significantly improved quality of life, using EQ-5D index scores, in patients taking adalimumab [change from baseline 0.15 units (SD 0.30 units)] when compared with those taking placebo [change from baseline 0.06 units (SD 0.28 units)]. A study of adalimumab reported no statistically significant difference in EQ-5D between groups at 12 weeks (0.78 units for adalimumab vs. 0.72 units for placebo; $p = 0.32$).[51]

For Ankylosing Spondylitis Quality of Life (ASQoL), a quality of life instrument specific to AS, ATLAS[61] was the only trial which reported results together with SDs or standard errors (SEs); significant improvements were found favouring treatment with adalimumab at week 12 [mean change from baseline −3.2 units (SD 0.3 units) for adalimumab vs. −1 unit (SD 0.4 units) for placebo].[62] Similar statistically significant results were reported in an etanercept trial at 12 weeks (mean change from baseline −3.3 units for etanercept vs. −0.1 units for placebo; $p = 0.02$)[71] and in an infliximab trial at 16 weeks (mean change from baseline −6.2 units for infliximab vs. −1 unit for placebo; $p = 0.007$).[50] Another small study of infliximab did not find a significant difference between groups at 30 weeks ($p = 0.14$).[100]

TABLE 11 The difference in change from baseline for BASDAI and BASFI scores assuming all TNFs have the same effect and assuming infliximab may be different

	BASDAI score		BASFI score	
Category of intervention	Mean	95% CrI	Mean	95% CrI
All TNFs	−1.66	−1.88 to −1.43	−1.38	−1.59 to −1.18
TNFs other than infliximab	−1.62	−1.85 to −1.38	−1.35	−1.56 to −1.14
Infliximab	−2.28	−3.18 to −1.38	−2.15	−3.18 to −1.11

'Placebo' response in ankylosing spondylitis and non-radiographic axial spondyloarthritis

To inform insight into the extent of any 'placebo' effects (outlined in *Chapter 1, Description of health problem*), *Table 12* compares the placebo response rates in trials which reported ASAS 20 results and at least one of ASAS 40 or BASDAI 50 results. These data highlight the relatively high rates of ASAS 20 response (median 31%, range 21–40%) when compared with ASAS 40 response (median 15%, range 10–23%) and BASDAI 50 response (median 16%, range 5–24%).

However, the extent of the 'placebo' response on the ASAS 20 results might result in an underestimation of anti-TNF efficacy, notably when ASAS 20 is the only ASAS improvement outcome reported in a trial. An increase in the likelihood of being a responder (i.e. the relative risks when compared with placebo) when moving up the ASAS thresholds seems apparent from the results in *Clinical effectiveness results: efficacy results from randomised controlled trials*. This might be explained by considering the subset of patients who achieve an ASAS 20 response largely because of regression to the mean (i.e. because of natural variation in repeated data measurements, such as patients transitioning from flare at randomisation to no flare at 12 weeks). For those patients who experience regression to the mean after taking an anti-TNF, the true benefit of treatment may be hidden in the ASAS 20 outcome for some patients, and the proportion of ASAS 20 responders might therefore differ only moderately between the anti-TNF and placebo groups. As the bar for response is raised, from ASAS 20 through to ASAS 70, this difference in the proportion of responders between active treatment and placebo groups is likely to increase as an effect because regression to the mean becomes less probable. The diluting effect of a placebo response on the relative risks therefore diminishes as the ASAS thresholds increase (and more informative estimates of treatment benefit can be seen). Regardless of the reason, these results highlight the limited applicability of ASAS 20 as a clinically informative outcome measure. ASAS 20 was nevertheless the most commonly reported responder outcome across the trials.

Summary of the randomised controlled trial clinical efficacy results

For both the AS and nr-AxSpA populations the results of the meta-analyses demonstrated that anti-TNFs produce statistically significant and clinically relevant benefits to patients in terms of improving function and reducing disease activity. The common class-effect model used may have underestimated the uncertainty in the effect estimates. Although there is a possibility that infliximab is more effective than other TNF inhibitors, at least at 12 weeks, there is no strong evidence to support this. For the disease activity, function and responder outcomes, the class-efficacy estimates were consistently slightly smaller for nr-AxSpA than for AS, most noticeably for BASFI and BASDAI 50.

The included RCTs were generally subject to low risks of bias and no important variation in baseline characteristics was evident, with the exception of CRP levels: in the nr-AxSpA trial populations CRP levels were much lower than in the AS populations. Although heterogeneity of CRP levels was evident across both the AS trials and the nr-AxSpA trials, in almost all the AS trials the CRP levels were higher than the 14 mg/l threshold identified as being a key predictor of treatment response (in AS, higher CRP levels are associated with an increased likelihood of BASDAI 50 response).[112] In the nr-AxSpA trials only the RAPID-axSpA[64] population came close to this cut-off point. These lower CRP levels may therefore have had an impact on the efficacy estimates for the nr-AxSpA population.

Statistical heterogeneity was more apparent in the nr-AxSpA analyses than in the AS analyses. This may be a result of both clinical heterogeneity in the nr-AxSpA trials (such as variation in CRP thresholds, or the proportion of MRI positive patients) and the fact that fewer studies were available for analysis. In the light of the statistical heterogeneity across the nr-AxSpA trials, both the reliability of the nr-AxSpA-pooled estimates and their true relevance to patients seen in clinical practice are questionable.

The clinical relevance of the efficacy of anti-TNFs can be evaluated in part by considering the literature on minimum clinically important differences (MCIDs) or minimum clinically important improvements. In a study of 125 AS patients, Pavy *et al.*[113] reported a MCID of 1 unit (or a 20% relative change) for BASDAI and

TABLE 12 Comparison of placebo response rates in trials reporting ASAS 20 results together with ASAS 40 or BASDAI 50 results

Population and study	Placebo compared with	Time point (weeks)	Number of patients on placebo	Number of responders			% of responders			Difference in response (%)		
				ASAS 20	ASAS 40	BASDAI 50	ASAS 20	ASAS 40	BASDAI 50	ASAS 20 vs. ASAS 40	ASAS 20 vs. BASDAI 50	ASAS 40 vs. BASDAI 50
Nr-axSpA[51]	Adalimumab	12	24	6	3	5	25	13	21	13	4	–8
AS[55]	Adalimumab	12	115	35	11	19	30	10	17	21	14	–7
Nr-axSpA[58]	Adalimumab	12	73	23	10	10	32	14	14	18	18	0
AS[61]	Adalimumab	12	107	22	14	17	21	13	16	7	5	–3
AS[64]	Certolizumab	12	57	21	11	6	37	19	11	18	26	9
Nr-axSpA[64]	Certolizumab	12	50	20	8	8	40	16	16	24	24	0
AS[74]	Etanercept	12	43	14	10	10	33	23	23	9	9	0
Nr-axSpA[76]	Etanercept	12	109	39	17	26	36	16	24	20	12	–8
AS[85]	Etanercept	12	51	19	11	10	37	22	20	16	18	2
AS[90]	Golimumab	14	78	17	12	12	22	15	15	6	6	0
AS[95]	Golimumab	14	105	26	10	5	25	10	5	15	20	5
AS[98]	Infliximab	12	35	10	–	3	29	–	9	–	20	–

Any minor discrepancies in the difference in response columns are because of rounding.

0.7 units (17.5% relative change) for BASFI. All the effect estimates from this review for both BASDAI and BASFI were considerably higher than these MCIDs. The small effect on spinal mobility (a group effect reduction of around 0.3 BASMI units) appears unlikely to be clinically important.

Summary of some key issues arising from the Food and Drug Administration assessments of the ABILITY-1[58] and RAPID-axSpA[64] trials

The FDA Arthritis Advisory Committee met in July 2013 to discuss licence applications for adalimumab for patients with active nr-AxSpA (with objective signs of inflammation) and certolizumab pegol for patients with active axSpA, including patients with AS.[114] An important issue which arose in both trials was the differences in diagnoses arising from radiograph images evaluated centrally compared with images being evaluated locally. The implications for efficacy were explored via further analyses.

RAPID-axSpA[64] trial (certolizumab pegol)

This trial aimed to recruit both AS and nr-AxSpA patients.[64] The nr-AxSpA patients had to have a positive MRI or an elevated CRP level; the definition used for CRP level elevation was 7.9 mg/l.

Comparison of ankylosing spondylitis and non-radiographic axial spondyloarthritis population characteristics In AS males predominated (72%), whereas in nr-AxSpA the male-to-female ratio was roughly equal. The AS population had a mean age of 41.5 years, which was around four years older than the nr-AxSpA population. Baseline BASFI, BASMI and CRP levels suggested more functional and mobility impairment and more inflammation in the AS group when compared with the nr-AxSpA group. However, baseline back pain severity and BASDAI scores were similar between the AS and nr-AxSpA subgroups (*Table 13*).

Methods used to evaluate radiograph images In the trial, many patients had their disease reclassified when radiographs were evaluated centrally, rather than being evaluated locally. Two readers were involved in the central evaluation of the radiographs, they were blinded to both the assigned subgroup and the treatment group; a third reader was used in cases of disagreement. Twenty-one per cent of locally classified AS patients were reclassified as nr-AxSpA by central readers and 51% of locally classified nr-AxSpA patients were reclassified as AS by the central readers. Based on the central assessments 184 patients had AS and 98 patients had nr-AxSpA. Central reads could not be made for 43 patients as radiographs were not available (37 AS patients and six nr-AxSpA patients).

ABILITY-1[58] trial (adalimumab)

This trial intended to recruit only nr-AxSpA patients, although this included patients ($n = 43$) who had nr-AxSpA but neither a positive MRI nor an elevated CRP level.[58] The population with these 43 patients excluded is referred to as the 'adalimumab target population' (ATP). As in the RAPID-axSpA[64] trial, central rereading of radiographs was performed (in addition to local evaluation), although this was only done for per-protocol patients who also reached week 104 [$n = 102$ (out of 185) patients]. Thirty-eight of the 102 patients were identified as having AS rather than nr-AxSpA. The FDA statistician analysed the results in these 38 patients and compared them to those for patients with centrally confirmed nr-AxSpA. The FDA document reported results for the subpopulations based on local or central diagnosis, including ATP analyses.

Comparison of ankylosing spondylitis and non-radiographic axial spondyloarthritis results and impact of reclassification in the trials

For certolizumab pegol the FDA statistical review stated that 'efficacy findings were consistent in both AS and nr-AxSpA subpopulations regardless of the discrepancy in pelvic X-ray readings at local or central lab for modified New York criteria'[115] (*Table 14*).

For ABILITY-1[58] a notably higher proportion of patients in the AS subgroup responded to adalimumab (ASAS 40) than placebo compared with patients with confirmed nr-AxSpA. This suggests that the treatment benefit in the whole trial population may be driven by benefit in AS patients rather than in nr-AxSpA patients, skewing the results for the ATP (see *Table 14*). It should be noted, however, that this may be an atypical AS population; the trial had intended to recruit only nr-AxSpA patients.

TABLE 13 Baseline characteristics of trials analysed by the FDA

Trial and population	Characteristic										
	Age (years), mean	% male	Duration (years) of symptoms	Weight (kg), mean	% HLA-B27 positive	% on NSAIDs	CRP level	% MRI positive	BASDAI score, mean	BASFI score, mean
ABILITY-1,[58] nr-AxSpA ($n = 142$)	38	46	Median 8, mean 11	80	80	81	Median ≈4, mean 9	51	6	4.7
RAPID-axSpA,[64] nr-AxSpA ($n = 147$)	37	48	Median 5.5, mean 8.6	82	75	84	Median 11.9, mean 16	54	6.5	4.9
RAPID-axSpA,[64] AS ($n = 178$)	42	73	Median 9.1, mean 11.9	82	82	91	Median 14.3, mean 21.3	N/A	6.4	5.7

N/A, not applicable.

TABLE 14 Food and Drug Administration analyses: percentage differences from placebo, by method of diagnosis

Outcomes at week 12	ABILITY-1[58] ATP population		RAPID-axSpA[64]							
	Local laboratory nr-AxSpA[a]	Central laboratory nr-AxSpA[b]	Local laboratory nr-AxSpA[c]		Central laboratory nr-AxSpA[d]		Local laboratory AS[e]		Central laboratory AS[f]	
	Adalimumab, 40 mg, % (95% CI)	Adalimumab, 40 mg, % (95% CI)	Certolizumab, 200 mg, % (95% CI)	Certolizumab, 400 mg, % (95% CI)	Certolizumab, 200 mg, % (95% CI)	Certolizumab, 400 mg, % (95% CI)	Certolizumab, 200 mg, % (95% CI)	Certolizumab, 400 mg, % (95% CI)	Certolizumab, 200 mg, % (95% CI)	Certolizumab, 400 mg, % (95% CI)
ASAS 20	28 (12 to 44)	15 (–14 to 44)	19 (1 to 38)	23 (4 to 42)	23 (2 to 44)	23 (1 to 44)	20 (3 to 37)	27 (10 to 45)	17 (1 to 33)	23 (7 to 39)
ASAS 40	27 (13 to 41)	11 (–16 to 38)	32 (14 to 49)	31 (14 to 48)	18 (0 to 36)	27 (8 to 47)	21 (5 to 36)	31 (14 to 47)	28 (13 to 43)	33 (17 to 48)
BASDAI 50	25 (11 to 39)	19 (–8 to 46)	–	–	–	–	–	–	–	–

a Adalimumab $n = 69$, placebo $n = 73$.
b Adalimumab $n = 25$, placebo $n = 20$.
c Certolizumab 200 mg, $n = 46$; certolizumab 400 mg, $n = 51$; placebo, $n = 50$.
d Certolizumab 200 mg, $n = 39$; certolizumab 400 mg, $n = 35$; placebo, $n = 39$.
e Certolizumab 200 mg, $n = 65$; certolizumab 400 mg, $n = 56$; placebo, $n = 57$.
f Certolizumab 200 mg, $n = 74$; certolizumab 400 mg, $n = 71$; placebo, $n = 67$.

Owing to the fact that only a select group of patients could be subject to central confirmation of their nr-AxSpA status, the FDA statistician explored assumptions around the proportion of true nr-AxSpA patients in the whole trial population. Given that the treatment difference in the non-centrally read patients was 23%:

- Assuming that all non-centrally read patients were true negatives and therefore including them in the analysis with the centrally read negatives, the treatment difference for the centrally read and non-centrally read negatives was 15%.
- Assuming that a fraction (i.e. 63%) of non-centrally read patients were true negatives and including only this fraction of non-centrally read patients with the centrally read negatives, the treatment difference was 14%.

The FDA document stated that:

> *Because there was a differential treatment effect between the centrally-read positive and centrally-read negative, it is safe to assume that the difference of 23% is an overestimate of the treatment effect because this includes both positive and negative x-ray groups. If there is a fraction of patients who are negative in the non-centrally-read group, treatment difference among this negative group would be smaller. Therefore, the treatment difference for negative x-rays (i.e. centrally-read and non-centrally-read) should be at most 15%. Based on the data provided, the estimate of the treatment effect in ASAS40 response for nr-AxSpA should be no bigger than 15%*
>
> *FDA Briefing Package*[116]

Overall, the results suggest reduced efficacy of anti-TNFs in the centrally diagnosed nr-AxSpA population compared with the locally diagnosed population. Nevertheless, there was noticeable variation across the two trials. In RAPID-axSpA[64] (certolizumab) the difference between the central and local populations appears small (and is not evident for 400-mg vs. placebo results). Conversely, in ABILITY-1[58] (adalimumab) the locally diagnosed population had notably more responders than the centrally diagnosed population, although the treatment group sample sizes were small.

Long-term efficacy results from open-label extensions of randomised controlled trials

Of the 24 included RCTs, 17 reported data from an open-label extension phase. Results for all studies are presented in *Appendix 8*. Considerable effort has been put into patient follow-up in anti-TNF trials with the result that data up to 5 years are available (there are data up to 8 years for infliximab but these included an involuntary treatment break which is not discussed further). The longest follow-up durations in patients with AS by anti-TNF are adalimumab 260 weeks, etanercept 264 weeks, infliximab 156 weeks, golimumab 268 weeks and certolizumab pegol 96 weeks. However, the data were reported across numerous publications and in various formats. Results were reported as observed, as completer analyses, using imputation (and rarely LOCF) for non-responders and LOCF for missing continuous data, but these related to differing populations (at varying time points): all patients randomised, all patients who took active drug at any point in the study or all patients who took active drug just during the open-label phase. The follow-up protocols were not clearly reported, with stopping rules unclear, but it appears that not all non-responders discontinued therapy. Therefore, the results may not reflect clinical practice should response be required for treatment continuation.

Table 15 presents the results based on non-responder imputation (NRI) analyses for the main studies when these results could be extracted. For AS the results show that across all the anti-TNFs after approximately 2 years of treatment, around half of patients are still achieving a good level of response to therapy. The results for golimumab look particularly strong with around 60% of all randomised patients achieving ASAS 40 and BASDAI 50 after 5 years. However, this is probably not reflective of clinical practice, as many of the normal weight patients took the 100-mg dose of golimumab rather than the 50-mg dose: the licence permits the use of 100-mg dose only in patients with a body weight of more than 100 kg who do not achieve an adequate clinical response after three or four doses. The equivalent results for adalimumab and etanercept are approximately 30% and 50%, although it is unknown if the difference may be because of variations in follow-up protocols rather than true treatment difference.

TABLE 15 Treatment effect over time (AS only) (results calculated using non-responder imputation)

Outcome	Trial	52 weeks, *n*/*N* (%)	104 weeks, *n*/*N* (%)	156 weeks, *n*/*N* (%)	5 years (approximately 264 weeks), *n*/*N* (%)
Adalimumab					
ASAS 20	ATLAS[61]	193/311 (62)[a]	135/311 (43)[a]	–	111/311 (36)[a]
ASAS 40	ATLAS[61]	138/311 (44)[a]	109/311 (35)[a]	–	88/311 (28)[a]
BASDAI 50	ATLAS[61]	167/311 (54)[a]	122/311 (39)[a]	–	96/311 (31)[a]
Certolizumab					
ASAS 20	RAPID-axSpA[64] (AS)	(48 weeks) 89/121 (74)[b]	(96 weeks) 78/121 (64)[b]	–	–
ASAS 40	RAPID-axSpA[64] (AS)	(48 weeks) 70/121 (58)[b]	(96 weeks) 61/121 (50)[b]	–	–
BASDAI 50	–	–	–	–	–
Etanercept					
ASAS 20	Calin 2004[83]	–	(108 weeks) 52/81 (64)[c]	–	–
ASAS 40	Calin 2004[83]	–	(108 weeks) 44/81 (54)[c]	–	40/81 (49)[c]
BASDAI 50	Calin 2004[83]	–	(108 weeks) 42/81 (52)[c]	–	39/81 (48)[c]
Golimumab					
ASAS 20	GO-RAISE[90]	–	235/356 (66)[b]	(160 weeks) 246/356 (69)[b]	235/356 (66)[b]
ASAS 40	GO-RAISE[90]	–	203/356 (57)[b]	(160 weeks) 208/356 (58)[b]	203/356 (57)[b]
BASDAI 50	GO-RAISE[90]	–	199/356 (58)[b]	–	199/356 (58)[b]
Infliximab					
ASAS 20	PLANETAS 2013[110]	(78 weeks) 125/174 (72)[c,d]	(102 weeks) 127/174 (73)[c,d]	–	–
ASAS 40	–	(78 weeks) 93/174 (53)[c,d]	(102 weeks) 101/174 (58)[c,d]	–	–
ASAS 40	ASSERT 2005[102]	(102 weeks)	33/78 (42)[b,e]	–	–
BASDAI 50	Braun 2002[98]	(54 weeks) 33/69 (48)[b]	(102 weeks) 30/69 (43)[b]	–	–

ASSERT, Ankylosing Spondylitis Study for the Evaluation of Recombinant Infliximab Therapy; NRI, non-responder imputation.
a NRI imputed result calculated using number of patients who had received at least one dose of active as denominator.
b NRI imputed result calculated using number of patients randomised as denominator.
c NRI imputed result calculated using number of patients who had received active during open-label phase as denominator.
d CT-P13 and infliximab combined.
e Only the subset of patients who took the 5-mg dose of infliximab (remaining patients took 5 or 7.5 mg).

The long-term follow-up for nr-AxSpA patients (*Table 16*) shows continued high proportions of responders. At 1 year around half of patients are achieving an ASAS 40 or BASDAI 50 level response and with certolizumab this is maintained at 2 years and with adalimumab at 3 years.

When the long-term data are presented as observed or completer analyses, the long-term results are similarly good; withdrawal rates are not high and a high proportion of those who remain on treatment continue to achieve a good response, see the example data available from one trial of adalimumab and one of certolizumab pegol (*Table 17*).

At long-term follow-up mean final values or mean change from baseline for BASDAI, BASFI and BASMI, when reported, were generally maintained at clinically meaningful levels.

For adalimumab, data from the large ATLAS trial[61] showed that mean changes from baseline at 1, 2 and 3 years remain stable and clinically meaningful at around –3.7 units for BASDAI and at around –2.9 units for BASFI. Similarly, the mean final value for BASMI remains at a level indicative of clinically significant

TABLE 16 Treatment effect over time (nr-AxSpA only)

Outcome	Trial	52 weeks, *n/N* (%)	104 weeks, *n/N* (%)	156 weeks, *n/N* (%)	5 years (approximately 264 weeks), *n/N* (%)
Adalimumab					
ASAS 20	ABILITY-1[58]	–	–	83/142 (58)[a]	–
ASAS 40	Haibel 2008[52]	23/46 (50)[b]	–	–	–
ASAS 40	ABILITY-1[58]	(68 weeks) 77/142 (54)[a]	–	67/142 (47)[a]	–
BASDAI 50	Haibel 2008[52]	24/46 (52)	–	–	–
BASDAI 50	ABILITY-1[58]	(68 weeks) 74/142 (52)[a]	–	70/142 (49)[a]	–
Certolizumab					
ASAS 20	RAPID-axSpA[64] (AS)	(48 weeks) 68/97 (70)[b]	(96 weeks) 59/97 (61)[b]	–	–
ASAS 40	RAPID-axSpA[64] (AS)	(48 weeks) 56/97 (58)[b]	(96 weeks) 49/97 (51)[b]	–	–
BASDAI 50	–	–	–	–	–
Etanercept					
ASAS 20	Dougados 2014[76]	(48 weeks) (CiC information has been removed)	–	–	–
ASAS 40	–	(48 weeks) 108/205 (53)[b]	–	–	–
BASDAI 50	–	(48 weeks) (CiC information has been removed)	–	–	–

CiC, commercial-in-confidence.
a NRI imputed result calculated using number of patients at week 12 as denominator.
b NRI imputed result calculated using number of patients randomised as denominator.
Results calculated using NRI.

TABLE 17 Observed or completer analysis results

Trial, anti-TNF (population)	Time point	Type of analysis	ASAS 20, *n*/*N* (%)	ASAS 40, *n*/*N* (%)	BASDAI 50, *n*/*N* (%)
ATLAS 2006,[61] adalimumab (AS)	52 weeks	Observed	193/276 (70)	138/276 (50)	167/276 (61)
	104 weeks	Observed	135/173 (78)	109/173 (63)	122/173 (71)
	5 years	Completer	111/125 (89)	88/125 (70)	96/124 (77)
RAPID-axSpA,[64] certolizumab pegol; all (AS)	96 weeks	Observed	78/93 (84)	61/93 (66)	–
RAPID-axSpA,[64] certolizumab pegol; all (nr-AxSpA)	96 weeks	Observed	59/74 (80)	49/74 (66)	–

treatment benefit (3.1 to 3.7 units). At 5 years the mean final values are BASDAI 1.8 units, BASFI 2.1 units, and BASMI 3.7 units. Clearly these results relate only to those patients who have remained on adalimumab in the long-term (40% of those who started adalimumab). They do, however, demonstrate continued benefit in a significant proportion of patients.

For certolizumab, results for these outcomes are available up to 96 weeks. At this time point the mean BASDAI and BASFI are indicative of clinically significant treatment benefit (both around 3 units).

The long-term data from Calin *et al.*[83] for etanercept, with 81 patients at 2 years and 59 (73%) remaining at 5 years also report mean BASDAI and BASFI scores of around 3.

From GO-RAISE[90] at 2 years for those who took golimumab throughout the trial and follow-up ($n = 138$), median BASDAI score was around 3 and median BASFI score was around 2. These values are from a LOCF analysis of all patients randomised to golimumab 50 mg.

For infliximab, the Braun *et al.*[98] and follow-up studies found, from 1 to 3 years, a stable mean BASDAI score of around 2.6, a stable mean BASFI score of around 3 and a stable mean BASMI score of around 2.7.

Overall, the reported data (although not particularly robust) do indicate that significant proportions of patients continue to derive real benefit from continued use of anti-TNFs. There is nothing to indicate any difference between them.

Almost no data were available regarding radiographic progression of bony disease in patients with AS. Furthermore, it should be noted that radiographic changes and progression of these take many years to appear and radiography is an insensitive tool by which to evaluate the progression of AS. Therefore, evidence, particularly that from relatively short-term studies, has to be interpreted with caution. The limited evidence includes mSASSS change from baseline, reported for golimumab from the GO-RAISE[90] study at 4 years (208 weeks): 1.3 units (SD 4.1 units) based on the 111 of 138 patients randomised to 50 mg. As results from untreated cohorts suggest a progression rate of 2 units/2 years, a rate of 1.3 units (or even 2 units) over 4 years seems beneficial. For further discussion of this issue see *Effect of anti-tumour necrosis factors on radiographic progression*. MASES was reported only for adalimumab from ATLAS;[61] in patients remaining on therapy at 2 years the mean change from baseline was 2.2 units ($n = 217$).

For nr-AxSpA patients long-term data for the continuous outcomes was limited to 1 year's follow-up. For adalimumab, data were available from only one small study (Haibel 2008,[52] $n = 46$): BASDAI change from baseline 2.8 units (95% CI 2.1 to 3.6 units); BASFI change from baseline 2 units (95% CI 1.4 to 2.6 units); BASMI change from baseline –0.4 units (95% CI –0.7 to –0.04 units); and MASES change from baseline of 0.9 units (95% CI –0.02 to 1.9 units). In addition, of 26 patients with magnetic resonance images at baseline and 52 weeks' follow-up, none showed a change in sclerosis or in erosions. For etanercept, data were available on 205 patients randomised to etanercept or placebo and then on long-term etanercept (Dougados 2014[76]): [commercial-in-confidence (CiC) information has been removed]. For certolizumab, LOCF analysis at 96 weeks ($n = 97$) gave a BASDAI final value score of 3.0, and a BASFI score of 2.4. Overall, the 1-year results in nr-AxSpA patients are similar to each other and also reflect those seen in AS patients. Again, the short-term nature of this follow-up relative to the 8–10 years over which radiographic changes develop must be borne in mind.

Findings from anti-tumour necrosis factor patient registry studies

Effect of anti-tumour necrosis factors on radiographic progression

A total of seven studies were identified that provided some comparative results on the effect of anti-TNFs on radiographic progression (*Table 18*).

TABLE 18 Effect of anti-TNFs on radiographic progression

Study	Methods	Results
van der Heijde *et al.* 2009[117]	Study used 2-year data from active treatment arms of two adalimumab trials (total $n = 397$) and compared them with OASIS cohort[118] (186 with radiographs at 2 years). Note: primary analysis set = 307 adalimumab (minimum of 1.5 years exposure to drug) and 169 anti-TNF naive (OASIS)	There were significant differences between adalimumab and OASIS[118] patients at baseline for BASDAI, BASFI and other measures. Increase in mSASSS was very similar in the two groups: adalimumab mean 0.8 (SD 2.6) and OASIS mean 0.9 (SD 3.3). When only patients who would have qualified for the adalimumab trials were included in the OASIS cohort ($n = 77$) the results were not changed. Note: in the light of these van der Heijde results, it would have been good to test effect of baseline BASDAI (mean 6.2 in adalimumab cohort and 3.4 in OASIS), as without treatment progression in adalimumab cohort would have been expected to be higher than in the OASIS one, so there might have been some effect of adalimumab
van der Heijde, *et al.* 2008[103]	Study compared 2-year data from infliximab trial (ASSERT[102]) ($n = 201$) with that from OASIS[118] ($n = 192$). OASIS patients not treated with any anti-TNF	There were significant differences between infliximab and OASIS[118] patients at baseline for BASDAI, BASFI and other measures (higher disease activity and worse function in trial patients). Mean increase in mSASSS was very similar in the two groups: infliximab 0.9 (SD 2.6) and OASIS 1.0 (SD 3.2). When only patients who would have qualified for the infliximab trials were included in the OASIS cohort ($n = 70$), the results changed very little [mean mSASSS increase 1.2 (SD 3.9)]
van der Heijde, *et al.* 2008[119]	Study compared 2-year data from etanercept trial (Davis *et al.*[72]) ($n = 257$) with that from OASIS[118] ($n = 175$). OASIS patients not treated with any anti-TNF	There were significant differences between infliximab and OASIS[118] patients at baseline for BASDAI, BASFI and other measures (higher disease activity and worse function in trial patients). Mean increase in mSASSS was very similar in the two groups: etanercept 0.91 (SD 2.5) and OASIS 0.95 (SD 3.2). When only patients who would have qualified for the etanercept trials were included in the OASIS cohort ($n = 76$), the results changed very little [mean mSASSS increase 1.3 (SD 3.6)]
Braun *et al.* 2014[120]	Long-term data on golimumab (2- and 4-year radiographic data) ($n = 233$). No comparison with OASIS[118] made	Mean increase in mSASSS to 2 years was 0.9 (SD 2.7) (50 mg) and 0.9 (SD 3.9) (100 mg). Mean increase in mSASSS to 4 years was 1.3 (SD 4.1) (50 mg) and 2.0 (SD 5.6) (100 mg). Note: 2-year results are very similar to those with other anti-TNFs and OASIS,[118] that is there is no benefit of golimumab evident

TABLE 18 Effect of anti-TNFs on radiographic progression (*continued*)

Study	Methods	Results
Haroon *et al.* 2013[121]	Cohort study ($n = 334$) with at least two spinal radiographs at 2-year intervals (patients with total spinal fusion at baseline excluded). Logistic regression analysis tested for baseline mSASSS, ESR, BASDAI, smoking, male vs. female, age at onset, disease duration, HLA-B27, anti-TNF use and NSAID index. Further analysis tested factors that could influence exposure to anti-TNFs using propensity matching	In total, 201 out of 334 patients had received anti-TNFs for a mean of 2.5 years (SD 2.6 years). No radiographic abnormality of the spine was seen at baseline in 144 patients (43%) and 102 patients (30.5%) showed no progression (> 1 mSASSS unit/year). Mulitvariate regression found baseline mSASSS (OR 1.06, 95% CI 1.04 to 1.08), ESR and smoking significantly increased and anti-TNF use significantly increased odds of radiographic progression (OR 0.47, 95% CI 0.24 to 0.94). Further analysis using the 142 patients who could be included post propensity matching confirmed these findings except for ESR: baseline mSASSS (OR 1.05, 95% CI 1.02 to 1.08) and anti-TNF (OR 0.30, 95% CI 0.11 to 0.78). Note: the association with anti-TNF use is explained by the more severe patients with radiographic changes at baseline being treated with anti-TNFs
Barialiakos *et al.* 2014[122]	Comparison of long-term (8 years) treatment with infliximab with historical cohort (infliximab $n = 22$ and Herne cohort $n = 34$)	Progression as assessed by mSASSS increased equally in infliximab treated patients and in the Herne cohort from baseline to 2, 4 and 6 years but while progression increased only slightly in the infliximab group between 6 and 8 years it increased greatly in the Herne cohort so that at 8 years there was a difference in infliximab's favour of 4.5 mSASSS ($p = 0.047$). Result was adjusted for baseline mSASSS. Other factors (age, symptom duration, BASDAI, BASFI) not significant confounders
Barialiakos 2007[123]	4-year radiographic progression in AS patients treated with infliximab ($n = 33$). Crude comparison made with OASIS cohort[118] results at 4 years	At baseline, mean mSASSS was 11.6 (15.3 SD), mean BASDAI was 6.6 (1.4 SD) and mean BASFI was 3.5 (1.9 SD). Progression assessed by mSASSS. Mean change over 4 years was 1.6 (SD 2.6) mSASSS units. Published results for OASIS are 4.4 units in 4 years

ASSERT, Ankylosing Spondylitis Study for the Evaluation of Recombinant Infliximab Therapy; OASIS, Outcomes in Ankylosing Spondylitis International Study.

Four studies reported on disease progression over 2 years of follow-up in terms of mSASSS in patients taking adalimumab,[117] infliximab,[103] etanercept[119] and golimumab.[120] All four open-label, uncontrolled follow-up studies found that mSASSS increased by a mean of around 0.9 units over 2 years. Three of these studies compared their rates with those from the Outcomes in Ankylosing Spondylitis International Study (OASIS) cohort[118] (of patients not taking an anti-TNF) and found no difference (mean rate over 2 years for OASIS was 0.9 units, *Table 19*). As stated in the previous section, radiographic changes and progression of these take many years to appear and, therefore, the evidence from these relatively short-term studies should be interpreted with caution.

TABLE 19 Summary of long-term results for mSASSS change

	Increase in mSASSS over 2 years, patients on an anti-TNF		Increase in mSASSS over 2 years, patients from OASIS cohort[118]	
Trial, anti-TNF	*n*	Mean (SD)	*n*	Mean (SD)
van der Heijde 2009,[117] adalimumab	397	0.8 (SD 2.6)	186	0.9 (SD 3.3)
van der Heijde 2008, etanercept[119]	257	0.91 (SD 2.45)	175	0.95 (SD 3.2)
Infliximab[103]	201	0.9 (SD 2.6)	192	1.0 (SD 3.2)
Golimumab[120]	111	50 mg, 0.9 (SD 2.7)	–	–
	122	100 mg, 0.9 (SD 3.9)	–	–

Comparison of the rates calculated from the OASIS cohort[118] in these studies with those from the studies by Ramiro[124,125] highlight a discrepancy; the latter reported rates of 2 mSASSS units every 2 years, rather than the 0.9 units/2 years used here to compare with individual anti-TNFs.

Two very small studies of infliximab reported some inhibiting effect on radiographic progression.[122,123] The first[123] compared findings in 22 infliximab patients with 34 from the HERNE cohort, over 2, 4, 6 and 8 years. Progression as assessed by mSASSS increased equally in infliximab treated patients and in the untreated HERNE cohort from baseline to 2, 4 and 6 years but then while progression increased only slightly in the infliximab group between 6 and 8 years it increased greatly in the HERNE cohort so that at 8 years there was a difference in infliximab's favour of 4.5 mSASSS units. The result was adjusted for baseline mSASSS (other factors, age, symptom duration, BASDAI, BASFI, etc., were not statistically significant confounders). The other study of 33 patients[123] found the mean progression over 4 years was 1.6 mSASSS units (SD 2.6 units), lower than the 4.4 units seen in the untreated OASIS cohort[118] at 4 years.

Another study examined a cohort of 334 patients with at least two spinal radiographs at 2-year intervals (patients with total spinal fusion at baseline were excluded).[121] In this study 201 out of 334 patients had received anti-TNFs for a mean of 2.5 years (SD 2.6 years) and no radiographic abnormality of the spine was seen at baseline in 144 patients (43%). At follow-up 102 patients (30.5%) showed no progression ($\geq$ 1 mSASSS unit/year). Mulitvariate regression found baseline mSASSS (OR 1.06, 95% CI 1.04 to 1.08), ESR and smoking significantly increased the odds of radiographic progression, but anti-TNF use was significantly associated with a > 50% reduction in the (adjusted) odds of progression (0.47, 95% CI 0.24 to 0.94). Further analysis that tested factors that could influence exposure to anti-TNFs using propensity matching confirmed the association with mSASSS and found a stronger association with anti-TNF use (OR 0.30, 95% CI 0.11 to 0.78).

In conclusion, there is evidence of disease progression over time, although the disease course is highly variable. Best estimates of yearly disease progression rates without anti-TNF therapy are around 1.0 mSASSS units and 0.035–0.07 BASFI units. Whether or not there is any impact of anti-TNF treatment is unclear; a beneficial effect can neither be assumed, and nor, given the short-term nature of the follow-up and the insensitivity of radiography as a tool for the evaluation of disease progression in AS, can one be discounted.

Drug survival and anti-tumour necrosis factor switching

The EndNote Library generated by the searches for RCTs of all the anti-TNFs were separately screened to identify patient registry studies of any or all of the anti-TNFs. This was possible because the search strategy for RCTs was very sensitive and will have identified any clinical study including any of the named anti-TNFs.

A total of 25 potentially relevant studies were screened fully and 12 publications that reported some data on drug survival or the efficacy of anti-TNFs after switching were identified (see *Table 20* for summary details of each). Across the 12 studies, the sources of data were either retrospective cohort studies or prospective registers (although analysis plans may have been retrospective), from a range of regions: USA (two studies), Canada (one study) and Europe (nine studies). No data from a UK-based cohort were available. Most of the cohorts and registries included experience with the three oldest anti-TNFs: infliximab, etanercept and adalimumab. One study (of the RHAPSODY cohort) included results from 326 patients treated with adalimumab as second anti-TNF after infliximab or etanercept. Small numbers of patients provided data on golimumab (three studies) and even smaller numbers on certolizumab (two studies). The population in 10 of the 12 studies was AS, although the diagnostic criteria used to specify AS were rarely given. One study provided results specifically for nr-AxSpA and one study provided results for axial SpA (nr-AxSpA or AS).

TABLE 20 Registry studies reporting data on drug survival and anti-TNF switching

Citation	Study/registry and method	*n* (duration, where stated)	Population	Anti-TNFs included	Drug survival	Efficacy on switching
Bonafede 2012[126]	Market Scan (administrative claims data) 2005–9, USA; retrospective	308 (360 days)	AS	Etanercept, adalimumab and infliximab	*n* (%) who stopped treatment and did not switch/who switched: • etanercept: (*n* = 149) 42 (28%)/12 (8%) • adalimumab: (*n* = 103) 36 (35%)/11 (11%) • infliximab: (*n* = 46) 14 (30%)/6 (13%)	NR
Choquette 2013[127] (abstract only)	Rhumadata register, Canada; unknown	119 (5 years)	AS, previous NSAIDs and BASDAI score of ≥ 4	Etanercept, adalimumab and infliximab	*n* who remained on same anti-TNF was 80% at 1 year; 70% at 2 years; and 55% at 5 years (no difference between anti-TNFs)	NR
Gulfe 2014[128]	SSATG registry, Sweden; prospective	112 (2 years)	Nr-axSpA not AS, demographic summary available	Etanercept, adalimumab, infliximab, golimumab and certolizumab	Kaplan–Meier estimates drug survival was 76% at 1 year and 65% at 2 years	NR
Nell-Duxneuner 2012[129]	Drug reimbursement data, Austria; retrospective	694 (2 years)	AS	Etanercept, adalimumab and infliximab	Starting in 2007 drug survival was: • etanercept: 0.83 (1 year) and 0.58 (2 years) • adalimumab: 0.70 (1 year) and 0.55 (2 years) • infliximab: 0.71 (1 year) and 0.54 (2 years)	NR

continued

TABLE 20 Registry studies reporting data on drug survival and anti-TNF switching (*continued*)

Citation	Study/registry and method	*n* (duration, where stated)	Population	Anti-TNFs included	Drug survival	Efficacy on switching
Yeaw 2014[130]	LifeLink Health Plan Claims database 2004–10, USA; retrospective	632	AS patients who had discontinued an anti-TNF	Etanercept, adalimumab and infliximab	% who restart within 360 days after stopping: • etanercept: 59% (*n* = 376) • adalimumab: 45% (*n* = 134) • infliximab: 39% (*n* = 122) % who switch to another anti-TNF or biologic: • etanercept: 17% (*n* = 376) • adalimumab: 13% (*n* = 134) • infliximab: 24% (*n* = 122) % who switch to non-biologic: • etanercept: 5% (*n* = 376) • adalimumab: 8% (*n* = 134) • infliximab: 6% (*n* = 122) % who switch to no new treatment: • etanercept: 18% (*n* = 376) • adalimumab: 34% (*n* = 134) • infliximab: 30% (*n* = 122)	NR

Citation	Study/registry and method	*n* (duration, where stated)	Population	Anti-TNFs included	Drug survival	Efficacy on switching
Scire *et al.* 2013[131]	MonitorNet database (Italian Society of Rhuematology) to 2012, Italy; multiple imputation used for missing data	498	AS	Etanercept, adalimumab and infliximab	Unadjusted Kaplan–Meier estimates of drug survival at: • 1 year: 0.87 (95% CI 0.83 to 0.89) • 2 years: 0.72 (95% CI 0.67 to 0.77) • 3 years: 0.69 (95% CI 0.63 to 0.74) Adjusted HR discontinuation rate (median follow-up 17 months) 0.59 (95% CI 0.46 to 0.75) (adjusted for age, sex, number of comorbidities, disease duration, number of previous DMARDs, concurrent DMARDS, baseline BASDAI score and BASFI score)	NR
Zufferey 2014[132]	Single centre in Switzerland (Centre Hospitalier Universitaire Vaudois) 2011–12; retrospective	112, of whom 77 were AS (follow-up at 12 and 24 months)	SpA (AxSpA and AS)	Etanercept, adalimumab, infliximab and golimumab	Median drug survival across all anti-TNFs 12 months (IQR 7–19 months) for AxSpA and 8 months (IQR 6–13 months) for AS Drug survival for AS: • 1 year 49% • 2 years 36% No difference between anti-TNFs	NR
Pavelka 2009[133]	ATTRA national registry, Czech Republic; prospective	310 (1 year)	AS (note mean BASDAI score 6.4 at baseline)	Etanercept, adalimumab and infliximab	Drug survival at 1 year was 84%; at 2 years was 76%; and at 3 years was 72%	NR

continued

TABLE 20 Registry studies reporting data on drug survival and anti-TNF switching (*continued*)

Citation	Study/registry and method	*n* (duration, where stated)	Population	Anti-TNFs included	Drug survival	Efficacy on switching
Lie 2011[134]	NOR-DMARD register 2000–9, Norway; prospective	514 (2 years)	AS	Etanercept, adalimumab and infliximab	In total 77 patients switched from first anti-TNF; 437 did not. In the 77 switchers, median drug survival on first anti-TNF was 266 days on the first anti-TNF (range 1–1392) and the second anti-TNF was a median of 77 days (range 0–1608 after the first was stopped). Finding may just be a consequence of the stopping rules in Denmark (patients given around 6 months to achieve a response) % on treatment after 1 and 2 years: First anti-TNF: 76% and 65% Second anti-TNF: 67% and 60%	Non-switchers: response to first anti-TNF at 3 months ($n = 362$): • BASDAI 50, *n/N*: 105/362 • ASAS 20, *n/N*: 106/202 • ASAS 40, *n/N*: 76/202 • Median (IQR) BASFI score: 2.3 (0.7–4.0) Median (IQR) BASDAI score: 2.6 (1.3–4.4) Switchers: response to first anti-TNF at 3 months: • BASDAI 50, *n/N*: 6/63 • ASAS 20, *n/N*: 11/23 • ASAS 40, *n/N*: 7/23 Median (IQR) BASFI, score: 4.7 (1.5–6.0) ($n = 63$) Median (IQR) BASDAI, score: 4.8 (3.3–7.01) ($n = 63$)

Citation	Study/registry and method	*n* (duration, where stated)	Population	Anti-TNFs included	Drug survival	Efficacy on switching
Glintborg 2010[112]	DANIBO registry, Denmark; prospective	842 (8 years)	AS	Etanercept, adalimumab and infliximab	Median drug survival was 4.3 years (unadjusted 1- and 2-year retention rates 74% and 63%) which was similar across three anti-TNFs; only male sex, low baseline VAS fatigue and high CRP level (> 14 mg/l) were associated with better drug survival	Response to second anti-TNF at 3 months: • BASDAI 50, *n*/*N*: 13/62 • ASAS 20, *n*/*N*: 18/45 • ASAS 40, *n*/*N*: 14/45 Median (IQR) BASFI score: 3.3 (1.6–5.7) (*n* = 62) Median (IQR) BASDAI score: 4.1 (1.9–6.1) (*n* = 62) Data also available by reason for withdrawal NR

continued

TABLE 20 Registry studies reporting data on drug survival and anti-TNF switching (*continued*)

Citation	Study/registry and method	*n* (duration, where stated)	Population	Anti-TNFs included	Drug survival	Efficacy on switching
Glintborg 2012[135]	DANBIO registry, Denmark; prospective	1436 (432 switchers; median 2.4 years)	AS (switchers only; had received at least two anti-TNFs during follow-up)	Etanercept, adalimumab, infliximab and golimumab (certolizumab and other biologics had less than 1% between them and only to first treatment course)	Median (95% CI) years of drug survival (*n*) (% on treatment after 2 years) for sequential anti-TNFs: • First anti-TNF: 3.1 (2.6 to 3.7), *n* = 1436 (58%) • Second anti-TNF: 1.6 (1.0 to 2.2), *n* = 432 (47%) • Third anti-TNF: 1.8 (0.9 to 2.7), *n* = 137 (49%)	Median (IQR) BASDAI at 3 months for sequential anti-TNFs: • First anti-TNF (*n* = 1436): 2.8 (1.1–4.8) • Second anti-TNF (*n* = 432): 3.6 (1.9–6.4) • Third anti-TNF (*n* = 137): 5.1 (3.6–6.7) Median (IQR) BASFI at 3 months for sequential anti-TNFs: • First anti-TNF (*n* = 1436): 2.8 (1.1–4.8) • Second anti-TNF (*n* = 432): 3.6 (1.7–6.0) • Third anti-TNF (*n* = 137): 5.1 (3.0–7.3) % BASDAI 50/20 mm responders at 6 months (NR at 3 months): • First anti-TNF: 54% • Second anti-TNF: 37% • Third anti-TNF: 30%
Rudwaleit 2009[136]	RHAPSODY, European cohort; prospective uncontrolled cohort of patients treated with adalimumab	1250 (12-week response data only)	AS	Adalimumab	NR	12-week response rates: anti-TNF naive (*n* = 924); BASDAI 50–63%; ASAS 40–59%; anti-TNF exposed (etanercept and/or infliximab, *n* = 326); BASDAI 50–41%; and ASAS 40–38% Logistic regression with backward elimination found that younger age, higher CRP level, HLA-B27 positive and anti-TNF naivety all predictive of better response (table 1[134])

HR, hazard ratio; IQR, interquartile range; NR, not reported; RHAPSODY, Review of safety and effectiveness witH Adalimumab in Patients with active ankylosing SpOnDYlitis; SSATG, Southern Sweden Anti-rheumatic Therapy Group.

Drug survival on first anti-TNF for all anti-TNFs was around 70–80% at 1 year, around 65–75% at 2 years, around 70% at 3 years and 55% at 5 years. Little difference between the three older anti-TNFs was identified, although one analysis using Cox proportional hazard estimates found statistically lower rates of discontinuation with etanercept and adalimumab compared with infliximab.[131]

The median drug survival in AS patients across all anti-TNFs reported varied (*Table 21*). Based on the largest registry (DANBIO)[135] the median drug survival for a first anti-TNF was 3.1 years (95% CI 2.6 to 3.7 years) ($n = 1436$), with 58% of patients remaining on treatment at 2 years. Median drug survival for a second anti-TNF was 1.6 years (95% CI 1.0 to 2.2 years) ($n = 432$), with 47% of patients remaining on treatment at 2 years, and for a third, 1.8 years (95% CI 0.9 to 2.7 years) ($n = 137$) (49% on treatment at 2 years).

The efficacy of second or third anti-TNFs after switching in AS patients was reported in only a small number of studies. One analysis based on the NOR-DMARD registry[134] showed how the response rate and BASDAI and BASFI achieved at 3 months in patients who remain on their first therapy is (not surprisingly) better than in patients who switch. Median BASDAI and BASFI achieved with a second anti-TNF were not as low (not as good) as was achieved with a first anti-TNF in non-switchers. An analysis of the DANBIO registry indicated that response (BASDAI 50) at 6 months reduced with subsequent anti-TNFs, as did the median improvement in BASDAI and BASFI achieved (*Table 22*).[135] These results are supported by the RHAPSODY study that found higher response rates with adalimumab in anti-TNF naive patients (BASDAI 50–63%; ASAS 40–59%)($n = 924$) than in anti-TNF exposed (BASDAI 50–41%; ASAS 40–38%) ($n = 326$).[136]

The registries and cohort studies provided no data on the efficacy of anti-TNFs as second or third, after switching in nr-AxSpA patients.

TABLE 21 Drug survival results from analysis of DANBIO registry[135]

	Drug survival for sequential anti-TNFs	
Anti-TNF	Median (95% CI)	% on treatment after 2 years
First ($n = 1436$)	3.1 (2.6 to 3.7)	58
Second ($n = 432$)	1.6 (1.0 to 2.2)	47
Third ($n = 137$)	1.8 (0.9 to 2.7)	49

TABLE 22 Efficacy results from analysis of DANBIO registry[135]

Anti-TNF	% BASDAI score 50/20 mm responders at 6 months (at 3 NR)	BASDAI score at 0 months for sequential anti-TNFs, median (IQR)	BASDAI score at 3 months for sequential anti-TNFs, median (IQR)	BASFI score at 0 months for sequential anti-TNFs, median (IQR)	BASFI score at 3 months for sequential anti-TNFs, median (IQR)
First ($n = 1436$)	54	5.9 (4.5–7.1)	2.8 (1.1–4.8)	5.0 (3.4–6.7)	2.8 (1.1–4.8)
Second ($n = 432$)	37	5.6 (3.8–7.3)	3.6 (1.9–6.4)	5.2 (3.5–7.0)	3.6 (1.7–6.0)
Third ($n = 137$)	30	6.4 (4.8–7.9)	5.1 (3.6–6.7)	6.4 (4.2–7.9)	5.1 (3.0–7.3)

IQR, interquartile range.

In summary, sequential treatment with anti-TNFs can be worthwhile in patients with AS but the response rates and benefits are reduced with second and third anti-TNFs, with the proportion of BASDAI 50 responders falling approximately 10% with each subsequent anti-TNF and the median BASDAI and BASFIs achieved increasing (worsening). The lower efficacy of a second anti-TNF relative to a first is reflected in lower median drug survival and proportion of patients remaining on therapy at 2 years. Interestingly, despite a further reduction in response and efficacy with a third anti-TNF, drug survival does not fall, suggesting that at this stage in their treatment history patients may continue with a less than optimally effective anti-TNF given any better alternative.

Clinical effectiveness results: adverse events

Randomised trials

We focused on the following outcomes, known to have possible associations with anti-TNF treatment: serious infections, tuberculosis (including tuberculosis reactivation), injection/infusion site reactions, congestive heart failure, cancer, non-melanoma skin cancer, SAEs and withdrawals due to SAEs. For the randomised phases of the trials included in the review, the reporting of AE data was generally limited. For three of the 24 trials no information on AEs was available.[55,56,74] Several trials provided AE data only at time points after which placebo patients may have switched to receive an anti-TNF (so true placebo comparisons were not available).

Analysable data on injection/infusion site reactions were available for 10 trials, although these studies were only of etanercept or infliximab. The data for certolizumab, golimumab and adalimumab trials either were not reported or were only provided at time points after which placebo patients could 'escape' to receive an anti-TNF; these data would not allow for an accurate comparison with placebo. Results for injection/infusion site reactions analyses from this review for etanercept and infliximab showed a statistically significant increase in reactions associated with etanercept (relative risk 2.69, 95% CrI 1.82 to 3.89) compared with placebo but no significant difference between infliximab and placebo. Compared with each other, the risk of an injection/infusion site reaction was statistically significantly higher with etanercept than with infliximab (relative risk 2.27, 95% CrI 1.01 to 5.37). Incidence of serious infections was reported in only eight trials, although such events were rare (nine cases in total). Of the eight trials which reported incidence of tuberculosis, only four cases were identified; three cases were reported in the longest study, the 54-week trial which compared infliximab with an infliximab biosimilar (CT-P13).[110] Four trials reported on congestive heart failure (no cases reported), six trials reported on cancer (one case) and three trials reported on non-melanoma skin cancer (two cases, one in each group of the ABILITY-1[58] trial). In most trials few SAEs were reported; group rates ranged from 0% to around 9%. Similarly, most trials had few withdrawals because of AEs; rates ranged from 0% to around 12%. Full results are reported in *Appendix 9*.

Large systematic reviews

Overall, the number and size of trials, and the short duration of their placebo-controlled phases, were too limited to provide enough data for meaningful analyses of AE. This common problem, of having too few data to evaluate AEs, underpinned the rationale for a Cochrane review (and network meta-analysis) of AEs of nine biologics in adults with any disease, except HIV/AIDS.[137] In order to provide a better understanding of toxicity, data were pooled across diseases by assuming a similar rate of AEs (across diseases). For the present assessment, estimates of AE rates have therefore been derived from the Cochrane review, which included 160 RCTs ($n = 48,676$) and 46 open-label extension studies ($n = 11,954$). The median durations were 6 months for RCTs and 13 months for open-label extension studies. The biologics included were abatacept, adalimumab, anakinra, certolizumab pegol, etanercept, golimumab, infliximab, rituzimab and tocilizumab. The anti-TNFs included in the present assessment were studied in 115 (72%) of the RCTs and 40 (87%) of the open-label studies included in the Cochrane review. Most studies assessed etanercept or infliximab in cancer or rheumatoid arthritis patients; 10 RCTs were of AS (fewer than in this assessment, as in the Cochrane review databases were searched up until January 2010). The biologics were evaluated both as a group and as individual interventions. The results from the RCTs, what the review classified as

'major' outcomes, are in *Tables 23* and *24*. Biologics as a group were associated with statistically significantly higher rates of total AEs, withdrawals because of AEs, serious infections and tuberculosis reactivation compared with control treatments. When the individual anti-TNFs were analysed separately, compared with control treatments only infliximab and certolizumab were statistically significantly associated with AEs: infliximab with higher rates of total AEs [number needed to harm (NNH) 13, 95% CrI 8 to 505] and withdrawals because of AEs (NNH 10, 95% CrI 5 to 30), and certolizumab pegol with higher rates of serious infections (NNH 12, 95% CrI 4 to 79) and SAEs (NNH 18, 95% CrI 9 to 162) (see *Table 24*).

For total AEs, the Cochrane review team judged the strength of evidence to be high; for SAEs, withdrawals because of AEs and serious infections, the strength of evidence was judged to be moderate; and for tuberculosis reactivation, lymphoma and congestive heart failure, the strength of evidence was judged to be low. For tuberculosis reactivation, lymphoma and congestive heart failure, the network meta-analysis statistical models did not converge (because of low numbers of events) therefore estimates for individual anti-TNFs were not available. Outcomes which were classed in the review as 'minor' were not analysed by the review authors because of the low numbers of events and the complexity of the analyses for the major outcomes. The minor outcomes included cardiac AEs, infusion and injection site reactions, allergic reactions, neurological outcomes, deaths, all cancers, serious lung infections or pneumonia, fungal infections and opportunistic infections. For the purposes of the present assessment, further large studies on cancer risk were therefore sought. An individual patient data meta-analysis of 22,904 adults (from 74 RCTs) which assessed the cancer risk of taking adalimumab, etanercept or infliximab in the short term (median duration < 6 months) was identified.[138] Although funded by manufacturers, this study was requested by the European Medicines Agency and was planned and conducted by independent researchers working with an independent academic steering committee. For all three anti-TNFs as a group, there was no increase in risk of cancers excluding non-melanoma skin cancer (relative risk 0.99, 95% CI 0.61 to 1.68) but there was a doubling in the risk of non-melanoma skin cancer associated with taking an anti-TNF (relative risk 2.02, 95% CI 1.11 to 3.95). Evaluation of drug-specific effects was hampered by statistical precision and by differences in baseline cancer risk and reporting detail across trials.[138]

Another review of AE effects of etanercept, adalimumab and infliximab was based on systematic searches for systematic reviews of the safety of biologic agents.[139] Six reviews that were sufficiently rigorous to meet the Database of Abstracts of Reviews of Effects inclusion criteria were included in the overview. This review also included large RCTs and non-randomised studies (≥ 500 patients), and was focused on serious potential AEs, such as serious infections, reactivation of latent tuberculosis and cancer.[139] *Table 25*, which summarises the rates of SAEs among the included non-randomised studies and large RCTs, indicates that

TABLE 23 Cochrane summary of findings table for biologics as a class (reproduced with permission from Singh *et al.*[137])

AE	Risk with comparator, per 1000 patients unless otherwise stated	Risk with intervention, per 1000 patients, unless otherwise stated (95% CrI)	OR (95% CrI)	Number of participants (studies)
SAEs	118	127 (115 to 142)	1.09 (0.97 to 1.24)	21,152 (76)
Total AEs	724	770 (741 to 797)	1.28 (1.09 to 1.50)	14,959 (48)
Withdrawal due to AEs	98	137 (115 to 168)	1.47 (1.20 to 1.86)	22,636 (83)
Serious infections	26	35 (27 to 46)	1.37 (1.04 to 1.82)	21,853 (70)
Tuberculosis reactivation	4 per 10,000	20 per 10,000	4.68 (1.18 to 18.6)	30,671 (71)
Lymphoma	9 per 10,000	1	0.53 (0.17 to 1.66)	21,260 (52)
Congestive heart failure	8	6 (1 to 21)	0.69 (0.18 to 2.69)	8847 (24)

TABLE 24 Cochrane summary of findings table for individual anti-TNFs (adapted from Singh *et al.*[137])

Anti-TNF	Risk with comparator, per 1000 patients unless otherwise stated	Risk with intervention, per 1000 patients, unless otherwise stated (95% CrI)	OR (95% CrI)	Number of participants (studies)
SAEs				
Adalimumab	118	114 (90 to 145)	0.96 (0.74 to 1.27)	4662 (15)
Certolizumab	118	174 (124 to 237)	1.57 (1.06 to 2.32)	2421 (6)
Etanercept	118	142 (111 to 184)	1.24 (0.93 to 1.69)	3931 (21)
Golimumab	118	123 (82 to 184)	1.05 (0.67 to 1.69)	1564 (8)
Infliximab	118	133 (102 to 174)	1.15 (0.85 to 1.57)	3403 (14)
Total AEs				
Adalimumab	724	730 (637 to 802)	1.03 (0.67 to 1.54)	3266 (10)
Certolizumab	724	754 (651 to 837)	1.17 (0.71 to 1.95)	1829 (5)
Etanercept	724	784 (677 to 866)	1.38 (0.80 to 2.46)	1600 (7)
Golimumab	724	765 (672 to 839)	1.24 (0.78 to 1.98)	1187 (6)
Infliximab	724	803 (726 to 860)	1.55 (1.01 to 2.35)	2330 (9)
Withdrawal due to AEs				
Adalimumab	98	128 (81 to 194)	1.35 (0.82 to 2.22)	5268 (18)
Certolizumab	98	125 (70 to 226)	1.32 (0.69 to 2.69)	2421 (6)
Etanercept	98	124 (82 to 191)	1.30 (0.82 to 2.17)	5189 (25)
Golimumab	98	127 (64 to 241)	1.34 (0.63 to 2.92)	1549 (7)
Infliximab	98	203 (132 to 310)	2.34 (1.40 to 4.14)	2973 (15)
Serious infections				
Adalimumab	26	32 (17 to 60)	1.23 (0.65 to 2.40)	4847 (15)
Certolizumab	26	113 (39 to 330)	4.75 (1.52 to 18.45)	1683 (4)
Etanercept	26	33 (19 to 61)	1.29 (0.72 to 2.45)	4630 (19)
Golimumab	26	29 (12 to 65)	1.11 (0.45 to 2.59)	1334 (6)
Infliximab	26	36 (20 to 65)	1.41 (0.75 to 2.62)	2652 (13)
Tuberculosis reactivation				
All nine biologics	4 per 10,000	20 per 10,000	4.68 (1.18 to 18.60)	30,671 (71)
Lymphoma				
All nine biologics	9 per 10,000	1	0.53 (0.17 to 1.66)	21,260 (52)
Congestive heart failure				
All nine biologics	8	6 (1 to 21)	0.69 (0.18 to 2.69)	8847 (24)

TABLE 25 Prevalence ranges of SAEs from non-randomised studies and RCTs (reproduced from Rodgers *et al.*[139])

Drug	Serious infections (%)	Cancer (%)	Tuberculosis (%)	Mortality (%)	Withdrawals due to AE (%)
Etanercept	0.6–13.2	1–5.7	0–1.4	0–3.1	0–13.6
Infliximab	0.8–13.8	0.16–5.1	0.06–4.6	0.06–2.0	6.4–12.8
Adalimumab	0.4–5.1	0.1–1.1	0–0.4	0.5–0.9	5.8–10.7

the rates of SAEs cover a broadly similar range across the three different biologic agents. However, all estimates were derived from a highly heterogeneous group of studies in terms of participants (e.g. inflammatory condition or disease severity), study design (e.g. length of follow-up) and treatment regimens (e.g. dose and frequency). Consequently, reliable estimates of the relative rate of SAEs for each drug could not be made.

Withdrawal rates due to AEs were typically < 10% for all drugs, with the highest reported single estimate being 13.6% for one etanercept study. This suggested that the majority of patients can tolerate biologic treatment in the medium term, although again the estimates were derived from a highly heterogeneous group of studies; therefore, the possibility of poorer tolerability in specific patient groups was not ruled out.

Open-label extensions of randomised trials

Of the longer-term follow-up studies included in our present review we evaluated those reporting AEs after 6 months (as the Cochrane review covered events occurring up to 6 months); 13 trial cohorts had studies which reported data after 6 months. Both the type of AEs assessed, and the periods over which they were assessed, varied across studies. *Table 26* compares results for studies with at least around 2 years of follow-up. The ATLAS[61] and GO-RAISE[90] trials both had extension study publications at the 2-year and 5-year time points.[140–143] Both cohorts were analysed using mITT data, in which patients had to have received at least one dose of treatment. This amounted to 99% of the randomised patients in both studies (311 out of 315 in ATLAS[61] and 353 out of 356 in GO-RAISE[90]). Davis[72,73] reported results for the 257 patients who enrolled in a 168-week open-label study following week 24 of the randomised phase; 277 patients had taken part in the earlier randomised study. All 257 patients in the open-label study had received at least one dose of etanercept.[144,145] The Calin trial[83–85] randomised 84 patients, with 81 patients enrolling in the open-label extension study. Results were presented separately for the 12-week to 2-year time points and the 2- to 5-year time points.[146,147] RAPID-axSpA[64] data at 96 weeks were reported in the manufacturer's submission. These data related to the mITT population: 315 (97%) of the 325 originally randomised patients.

The 2-year study of the ASSERT (Ankylosing Spondylitis Study for the Evaluation of Recombinant Infliximab Therapy)[102] (infliximab) cohort allowed dose escalation whereby, from week 36, patients with BASDAI scores of ≥ 3 could increase their dose to 7.5 mg/kg, which is a currently unlicensed dose. Results for the 5 mg/kg group (74 patients) between weeks 24 and 102 have therefore been presented in *Table 26*. The Braun cohort[148] was followed up for 8 years, but it was a small study which reported only SAEs and withdrawals due to SAEs.

Table 26 illustrates that rates of SAEs, cancer and serious infections were similar across all four anti-TNFs when using incidence per 100 patient-years as estimates. At 5 years, SAEs appeared more prevalent with adalimumab (45%) than golimumab (20%), although it is possible this difference is because of the way the data were reported; it was unclear whether the ATLAS[61] data related to the total number of SAEs or to the number of patients experiencing a SAE. At 2 years, the incidence of injection site reactions was higher in patients taking etanercept than in patients taking adalimumab, golimumab or certolizumab pegol. Withdrawal rates due to AEs were broadly similar across treatments. The reporting of tuberculosis and congestive heart failure was limited.

TABLE 26 Studies with adverse event data at around 2 years (or later)

	Adalimumab		Golimumab		Etanercept		Certolizumab		Infliximab
	ATLAS[61]		GO-RAISE[90]		Davis[72,73]	Calin[83–85]	RAPID-axSpA65		[a]ASSERT[102]
Event outcome	2 years ($n = 311$)	5 years ($n = 311$)	2 years ($n = 353$)	5 years ($n = 353$)	24–192 weeks ($n = 257$)[b]	12–108 weeks ($n = 81$)[b]	2–5 years ($n = 59$)	96 weeks ($n = 315$)	24–102 weeks ($n = 74$)
SAEs	48 events (15%) 10.5/100 PY	140 events (45%) 11.7/100 PY	40 events (11%)	72 events (20%)	33 events (13%) 8/100 PY	19 events (23%)	21/100 PY	AiC information has been removed	15 events (20%)
Withdrawals because of AEs	24 (8%) events 4.5/100 PY	–	19 events (5%)	32 events (9%) 2.13/100 PY	14 events (5%)	15 events (19%)	7 events (12%)	AiC information has been removed	–
Serious infections	6 (2%) events 1.1/100 PY	17 events (5%) 1.4/100 PY	11 events (3%)	21 events (6%) 2.1/100 PY	6 events (2%) 2/100 PY	5 events (6%)	3 events (5%) 3/100 PY	AiC information has been removed	3 events (4%)
Cancer	4 events (1%) 0.7/100 PY	3 events (1%) 0.2/100 PY	2 events (0.6%)	3 events (0.8%) 0.21/100 PY	–	4 events (5%)	3 events (5%)	–	1 events (1%)
NMSC	0.4/100 PY	–	–	–	–	–	–	–	–
Congestive heart failure	0 events	2 events (0.6%) 0.2/100 PY	–	–	–	–	–	–	–
Injection site reactions	42 events (14%) 17.6/100 PY	–	38 events (11%)	43 events (12%)	57 events (22%)	30 events (37%)	7 events (12%)	AiC information has been removed	9 events (12%)
Tuberculosis	0 events	0 events	–	–	–	0 events	0 events	AiC information has been removed	–

AiC, academic-in-confidence; ASSERT, Ankylosing Spondylitis Study for the Evaluation of Recombinant Infliximab Therapy; NMSC, non-melanoma skin cancer; PY, person-year.
a 5 mg/kg group.
b Weeks from randomisation.

Summary of adverse event data

Data from large systematic reviews, which included patients with a wide range of diseases, suggest that, in the short term, anti-TNFs as a group are associated with significantly higher rates of serious infections, tuberculosis reactivation, non-melanoma skin cancer, total AEs and withdrawals because of AEs than control treatments. Specifically, infliximab is associated with significantly higher rates of total AEs and withdrawals because of AEs, and certolizumab pegol is associated with significantly higher rates of serious infections and SAEs. Analyses from the present review showed etanercept to be statistically significantly more likely to result in an injection/infusion site reaction compared with infliximab, although analysable data on such reactions were not reported for the three other anti-TNFs. Evaluations of longer-term data are more scarce, although suggest similar safety profiles across anti-TNFs. Data from the open-label studies included in this review also do not suggest that there are important differences between treatments, other than a higher incidence of injection site reactions following treatment with etanercept. These open-label data are, however, limited by the small sample sizes and non-randomised study designs.

Review of natural history of ankylosing spondylitis and non-radiographic axial spondyloarthritis

In order to get some understanding of what happens to patients who, although eligible for anti-TNF therapy for their AS or nr-AxSpA, do not receive it, we conducted a rapid review of relevant literature. This was not a systematic review but one that started with the library of papers found by the main searches for RCTs of the anti-TNFs and then followed relevant citations to papers on AS and axSpA in patients not receiving an anti-TNF. Potentially relevant papers were those that reported on the pattern of disease, AS or nr-AxSpA or axSpA, without treatment with anti-TNFs over time. This process identified a number of relevant registries: OASIS,[118] Scotland and Ireland Registry for Ankylosing Spondylitis (SIRAS), Devenir des Spondylarthropathies Indifferenciées Récentes, Esperanza, Spanish Registry of spondyloarthritis, German Spondyloarthritis Inception Cohort (GESPIC) and St Mary RheumaToid Arthritis (SMART). Additional searches of MEDLINE were conducted using these specific registry names. All relevant studies identified through this process are presented in *Table 27*.

The studies collectively explore the associations between the various components of axSpA: disease activity, structural damage and spinal mobility. The exploration of the ASSERT trial baseline data[24,124] reveals that HRQoL as determined by SF-36 physical and mental components, is determined by BASFI and BASDAI; BASFI is determined by BASDAI, mSASSS and BASMI (spinal mobility); and BASMI is independently determined both by irreversible spinal damage (mSASSS) in late disease and reversible spinal damage (MRI) in early disease.

The studies identified that from a clinical practice and patients' point of view disease progression in terms of BASFI, a measure of the patient's functional ability, is very important. A number of studies on the disease progression of AS have been based on the European OASIS cohort[118] (a consecutive cohort, started in 1996, although there were no further specific eligibility criteria); the total cohort numbers 217 patients. One of these, a study by Landewe *et al.*,[10] demonstrated that physical function impairment (BASFI) is independently affected by both disease activity (BASDAI) and bony progression, usually assessed using mSASSS despite this being a measure of bony growth in the spine only (and not in the sacroiliac joints). Other studies by Ramiro[124,125] have demonstrated that radiographic progression, increases on average by around 2 mSASSS units every 2 years.[124,125] However, this progression is highly variable; the average patient with inactive disease [Ankylosing Spondylitis Disease Activity Score (ASDAS) 0] would progress by 5 mSASSS units over 12 years compared with a patient with 'very active disease' (ASDAS 4) who would have 19 units of progression.[124] In addition, of 68 patients who were followed for 12 years, 18% had no progression on mSASSS.[125] The variability is also demonstrated by the results based on a different cohort: a single German clinic ($n = 146$).[12] Baseline characteristics were similar to those in the OASIS cohort[118] (see *Table 27*). Mean follow-up was 3.8 years (SD 1.7 years) and mean mSASSS change was 1.3 units/year (SD 2.5 units/year) with a range of 0–22.8 mSASSS units. Thirty-four (23%) patients showed no progression.

TABLE 27 Natural history of axSpA: relevant outcomes and impact of anti-TNFs

Study	Description	Population characteristics	Summary of findings
Landewe 2009[10]	Examined the relationship between disease activity, radiographic damage and physical function in AS. Based on (European) OASIS cohort[118] baseline and 2-year data. $n = 217$ consecutive (from 1996) patients with AS (no specific criteria). BASDAI score mean 3.4 (SD 2.1), with 38% ≥ 4	BASFI score mean 3.4 (SD 2.6), 41% ≥ 4. mSASSS median 5, 69% > 0. Note: does mSASSS < 0 mean nr-AxSpA? None of the patients in the cohort had used anti-TNFs. Subgroup ($n = 188$) baseline BASDAI score of ≤ 6	Univariate correlation between baseline mSASSS and BASFI score = 0.45 (Spearman's rank correlation coefficient), but this was modified by baseline BASDAI: • BASDAI score of 0–2 ($n = 68$) = 0.68 • BASDAI score of > 2–4 ($n = 60$) = 0.58 • BASDAI score of > 4–6 ($n = 60$) = 0.43 • BASDAI score of > 6–8 ($n = 22$) = 0.40 • BASDAI score of > 8–10 ($n = 7$) = –0.20 Suggests a ceiling effect of BASFI. Owing to the high level of correlation between BASDAI and BASFI, a correlation between mSASSS and BASFI cannot be demonstrated at the highest level of BASDAI. There was a multivariate relationship between BASDAI and mSASSS with BASFI using baseline and 2-year data (but not longitudinal?) ($n = 188$, baseline BASDAI score of ≤ 6 only). Regression coefficients found that both BASDAI and mSASSS are statistically significant ($p < 0.001$) explanatory variables for BASFI (0.73 and 0.057 units respectively)
Ramiro 2014[124]	Analysed long-term relationship between disease activity (ASDAS, BASDAI) and radiographic damage (mSASSS) in AS. Used OASIS cohort[118] over 12 years	Subgroup used patients ($n = 184$) who had at least two sets of radiographs. Baseline characteristics of this subgroup: BASDAI score mean 3.4 (SD 2.0); mSASSS mean 10.8 (SD 15.2), 81% > 0. None of the patients had used anti-TNFs	On average patients had a progression of 1.9 mSASSS units/2 years. This varied with baseline ASDAS: • ASDAS < 1.3 progress = 0.7 mSASSS units/2 years • ASDAS > 3.5 progress = 3.1 mSASSS units/2 years The relationship with BASDAI was similar: • Baseline BASDAI score of < 4, 1.5 mSASSS units/2 years • BASDAI score of ≥ 4, 2.7 mSASSS units/2 years • BASDAI score of > 6 units, 2.0 mSASSS units/2 years The analysis found that the average patient with inactive disease (ASDAS score 1.0) would progress by 5 mSASSS units over 12 years compared with a patient with 'very active disease' (ASDAS score of 4) would have 19 units of progression
Ramiro 2013[125]	Earlier analysis of OASIS cohort[118] 12-year data to describe the evolution of radiographic abnormalities in AS patients	Subgroup used ($n = 186$) who had at least two sets of radiographs. Baseline characteristics of this subgroup: BASDAI score mean 3.4 (SD 2.0); mSASSS mean 11.6 (SD 16.2). None of the patients had used anti-TNFs	Long-term radiographic progression in AS highly variable at the patient level, but is more severe in men who are HLA-B27 positive. Over whole follow-up, 24% of patients (and 18% of the 68 patients who were followed for 12 years) had no progression on mSASSS. Duration of disease is not relevant. At the group level, progress is linear at 2 mSASSS units/2 years

TABLE 27 Natural history of axSpA: relevant outcomes and impact of anti-TNFs (*continued*)

Study	Description	Population characteristics	Summary of findings
Baraliakos 2009[12]	Natural course of radiographic progression in AS. Retrospective cohort, single clinic (Herne, Germany), 1993–2005. Mean follow-up 3.8 years (SD 1.7 years)	$n = 146$ anti-TNF naive patients. Baseline: • mSASSS mean 20.5 (SD 14.4) • BASDAI score mean 4.4 (SD 1.9) (range 0.5–7.3) • BASFI score mean 3.8 (SD 2.6) (range 1.0–8.4)	Mean mSASSS change was 1.3 units/year (SD 2.5 units/year). Note: range was mSASSS 0–22.8. Thirty-four (23%) patients showed no progression
Dean 2014,[149] poster at BSR meeting	SIRAS cohort. Study of BASDAI over time	BASDAI score at diagnosis data available for only 240 patients (out of the 1210 patient cohort). Baseline BASDAI score (at diagnosis) of 4.9 (SD 2.3). High disease activity group BASDAI score of 6.3 (SD 1.4) and low disease activity group BASDAI score of 2.5 (SD 1.3)	Baseline BASDAI remained fairly stable over time: across the whole cohort and in the high and low disease activity groups. The subgroup treated with anti-TNFs had higher mean BASDAI score [5.7 (SD 2.0)] than non-biologic patients [4.2 (SD 2.5)] and this remained so until around a year after treatment with anti-TNFs began, when mean BASDAI fell to the level of the non-biologic patients
Healey 2013[14]	Cohort study, single centre, England. Followed patients over 10 years [$n = 69$ who provided assessments at baseline (1998) and at 10 years (2008)]. Assessments using RLDQ, BASDAI, ASQoL and EQ-5D (and others)	At study entry patients were 84% male, mean age 49 years, disease duration 15.5 years, symptom duration 21.4 years. 1.5% on an anti-TNF at 10 years	Only RLDQ changed significantly over time for assessment 1 (1998) and 2 (2008): • RLDQ: mean 10.4 (SD 8.3); mean 13.6 (SD 10.9); $p = 0.002$ • BASDAI: mean 4.1 (SD 2.5); mean 4.4 (SD 2.7); $p = 0.36$ • ASQoL: mean 6.4 (SD 6.3); mean 7.5 (SD 6.4); $p = 0.15$ • EQ-5D: mean 0.64 (SD 0.28); mean 0.61 (SD 0.30); $p = 0.45$ However, as RLDQ (range 0–48) is a measure of function (comparable with BASFI) it does indicate progression with time even in these AS patients whose disease at study entry was already well established
Stone 2007[150]	Analysis of longitudinal data from SMART (Bath, UK) data set ($n = 224$). Regression analysis of BASDAI score on symptom duration and BASFI score adjusted for BASDAI score > 4 at baseline. Duration of follow-up was unclear	Overall, 68% had a baseline BASDAI score of ≥ 4. Mean symptom duration was 28.8 years	Only 20% experienced a significant change in BASDAI score over time (13% a decrease; 7% an increase). BASFI score increases over time by 0.035 units/symptom-year. In patients with baseline BASDAI score of ≥ 4, those who would be treated with anti-TNFs, the increase over time is 0.039 units/symptom-year
Machado 2010[151]	Baseline data from ASSERT.[102] Analysis of relation between mSASSS and MRI inflammation and BASMI	$n = 214$ AS patients (mNY criteria). Baseline median (IQR): • BASMI score of 4.6 (3.6–5.8) • BASDAI score of 6.5 (5.3–7.0) • CRP level 1.5 mg/dl (0.7–2.9 mg/dl) • mSASSS 13.8 (4.5–29.1)	Concluded that spinal mobility (BASMI) independently determined both by irreversible (mSASSS) and reversible spinal damage (MRI), the former in late disease and the latter in early disease

continued

TABLE 27 Natural history of axSpA: relevant outcomes and impact of anti-TNFs (*continued*)

Study	Description	Population characteristics	Summary of findings
Machado 2011[24]	Baseline data from ASSERT.[102] Analysis of relation between SF-36 and BASFI and BASDAI, ASDAS, CRP level, mSASSS, MRI inflammation and BASMI	$n = 214$ AS patients (mNY criteria)	Regression coefficients for associations reported in the publication. Briefly, SF-36 is determined by BASFI and BASDAI; and BASFI is determined by BASDAI, mSASSS and BASMI
Kobelt 2004[152]	A modelling study of infliximab but refers to large UK observational data set and generates an estimate for BASFI over time. Survey in 2002 ($n = 1413$). Value generated from patients who were captured in two surveys at two time points, 1992/1994 and November 2002, approximately 8 years apart ($n = 1100$). Data from a cohort of 493 patients who had been followed up for more than 3 years were used as a check for the result based on the survey	–	From the whole survey ($n = 1413$) mean BASDAI score 4.2 (SD 2.3) and mean BASFI score 4.4 (SD 2.8). The population was broader than that eligible for anti-TNFs, with 47% having a BASDAI score of < 4. It appears (but is unclear) that this is the BASDAI at the later time (2002) point not the earlier (1992/4). Estimate of annual BASFI progression was 0.07 points. Note: progression was faster (0.1 points) in patients with BASFI score of < 4 at baseline, but was stable (0?) in patients with BASFI score above 7. (Ceiling effect of BASFI?) When only patients with BASDAI score of ≥ 4 included BASFI progression was estimated as 0.054. Data from the cohort study generated similar findings; however, the number was not actually reported for whole survey. BASFI progression was 0.059 for patients with a BASDAI score of ≥ 4
Nr-axSpA			
Kiltz 2012[153]	Comparison of characteristics of patients with AS and nr-AxSpA. Cohort of 100 patients seen in 2010 in Herne clinic, Germany. Analysis tested if the proportion of patients reaching pre-specified cut-off criteria (markers of disease severity) differed between AS and nr-AxSpA	Consecutive, diagnosed with axSpA. None of the patients had used anti-TNFs. $n = 100$ AxSpA: $n = 44$ nr-AxSpA and $n = 56$ AS • Median BASDAI score, 4.3 (AS) and 3.6 (nr-AxSpA); $p = 0.2$ • Median BASFI score, 2.9 (AS) and 1.5 (nr-AxSpA); $p = 0.05$ • Median CRP level, 8.0 mg/l (AS) and 3.8 mg/l (nr-AxSpA); $p < 0.001$ • Median mSASSS 3.0 (AS) and 1.1 (nr-AxSpA); $p < 0.007$	Differences were statistically significant for ASDAS, CRP level, mSASSS and number of inflamed lesions. Proportion of males also significantly different. Results: • % male (p-value), 31.8% (nr-AxSpA) and 76.8% (AS); $p < 0.001$ • BASDAI score of ≥ 4: 43% (nr-AxSpA) and 53.5% (AS); $p = 0.1$ • BASFI score of ≥ 3: 34.1% (nr-AxSpA) and 46.4% (AS); $p = 0.08$ • ASDAS > 2: 54.5% (nr-AxSpA) and 78.6% (AS); $p = 0.01$ • CRP level > 5 mg/l: 29.5% (nr-AxSpA) and 69.1% (AS); $p < 0.001$ • mSASSS ≥ 3: 27.3%(nr-AxSpA) and 51.9% (AS); $p = 0.01$ • Number of inflamed lesions per patient ≥ 3: 9.1% (nr-AxSpA) and 46.4% (AS); $p = 0.01$

TABLE 27 Natural history of axSpA: relevant outcomes and impact of anti-TNFs (*continued*)

Study	Description	Population characteristics	Summary of findings
Rudwaleit 2009[18]	Cross-sectional study of GESPIC cohort ($n = 462$) patients with axSpA. Divided into AS ($n = 236$) and nr-AxSpA (with ≤ 5 years of symptoms, $n = 226$)	Baseline mean: • BASDAI score of 4.0 (SD 2.1) (AS) and 3.9 (SD 2.0) (nr-AxSpA) • BASDAI score of ≥ 4, 48.7% (AS) and 47.7% (nr-AxSpA) • BASFI score of 3.1 (SD 2.5) (AS) and 2.5 (SD 2.1) (nr-AxSpA) Note: mean BASFI score was the same for patients with AS more than or no more than 5 years of symptoms	When AS patients were divided into those with more than 5 years of symptoms and those no more than 5 years, there were no differences in characteristics at baseline. When AS (≤ 5 years) and nr-AxSpA were compared there were statistically significant differences (worse for patients with AS) in Physicians Global assessment, BASFI, BASMI, spinal mobility, lateral spinal flexion, CRP, ESR and all radiographic measures (mSASSS 4.9 in AS vs. 1.4 in nr-AxSpA). mSASSS was significantly worse in males vs. females and in CRP level > 6 vs. ≤ 6. Note: the AS patients had a very short symptom duration and must have progressed to AS rapidly. In addition, nr-AxSpA patients had only a short time from the start of symptoms and may therefore not reflect patients who remain nr-AxSpA for a longer time.
Poddubnyy 2011[154]	Study of radiographic progression of sacroiliitis in AS and nr-AxSpA. Radiographic evidence of sacroiliitis is a criterion in the mNY for AS; therefore, it is useful to see this analysis of progression rather than only the mSASSS	German cohort (GESPIC) $n = 210$ ($n = 115$ AS and $n = 95$ nr-AxSpA), 2 years' follow-up. (Baseline BASDAI score was 4 and BASFI score was 3 across AS and nr-AxSpA.) Overall the cohort had a short symptom duration of 4.2 years (5.2 years AS and 3.2 years nr-AxSpA). Only 3.5% had had treatment with anti-TNFs (3.5% AS and 1.1% nr-AxSpA)	After 2 years' follow-up, $n = 11$ of the 95 nr-AxSpA patients (11.6%, 95% CI 6.6% to 19.6%) fulfilled the mNY for AS. In addition, after 2 years approximately 10.5% of patients in the nr-AxSpA cohort had progressed by at least one mNY grade, compared with 8.7% of patients in the AS group (difference not statistically significant). Predictors of sacroiliitis progression were raised CRP level for both AS and nr-AxSpA. Male sex and HLA-B27 positive predicted lower progression in nr-AxSpA but higher progression in AS
Poddubnyy 2012[155]	GESPIC cohort. Radiographs of spine and SIJ at baseline and at 2 years	Baseline, all patients ($n = 210$): of the 2.4% patients treated with anti-TNFs, the BASDAI score was 4 and the BASFI score was 3. AS ($n = 115$): of the 3.5% patients treated with anti-TNFs the BASDAI score was 4 and the BASFI was score 3. nr-AxSpA ($n = 95$): of the 1.1% of patients treated with anti-TNFs, the BASDAI score was 4 and the BASFI score was 3	Regression analysis found syndemophytes at baseline, elevated ESR and CRP level and smoking were significantly associated with spinal progression (≥ mSASSS/2 years) in AS but only syndemophytes at baseline in axSpA. In AS patients mSASSS increased significantly from 5.86 (SD 10.30) to 6.81 (SD 11.71), mean difference 0.95 (SD 2.78). In nr-AxSpA patients mSASSS increased significantly from 2.30 (SD 4.24) to 2.76 (SD 5.26), mean difference 0.46 (SD 1.63). The difference between mean progression in AS and nr-AxSpA patients was not statistically significant, and neither was the difference between those with symptom duration of ≤ 5 years and > 5 years. Percentage that progressed by > 2 mSASSS units/2 years: all axSpA 14.3%; AS 20.0% (95% CI 13.7% to 28.2%); nr-AxSpA 7.4% (95% CI 3.6% to 14.4%). There was no difference in mSASSS change between patients not progressing to AS (0.49 units) and those who progressed to AS (0.27 units); $p = 0.53$

continued

TABLE 27 Natural history of axSpA: relevant outcomes and impact of anti-TNFs (*continued*)

Study	Description	Population characteristics	Summary of findings
Flares			
Cooksey 2010[28]	Cohort derived from full population of a trial comparing probiotic and placebo treatment in AS. Followed up for 1216 person-weeks and recorded localised/minor flares and generalised/major flares, plus BASDAI, BASFI and pain VAS	AS patients $n = 134$. Baseline mean BASDAI score of 3.7 (SD 2.1); mean BASFI score of 3.6 (SD 2.8). Mean duration of symptoms 21 years (SD 13 years, range 0–58 years)	The overall flare rate was 71.4 per 100 person-weeks: major flare rate of 12 per 100 person-weeks and minor flare rate of 59.4 per 100 person-weeks. Mean BASDAI scores were 5.5 (major flare), 3.1 (minor flare) and 2–2.5 (flare free). Mean BASFI scores were 5.5 (major flare), 3.1 (minor flare) and 2.5–3.5 (flare free). Note: these means are not from whole population but only from patients who experienced major flares plus flare-free periods ($n = 27$) and minor flares plus flare-free periods ($n = 77$)
Stone 2008[156]	A pilot study to investigate pattern of disease and impact of disease flares. It used the SMART cohort (Bath, UK). Patients were asked about four patterns of disease (see under Summary of findings in this table)	AS patients, although the diagnostic criteria was not stated, $n = 114$ (although not $n = 114$ for all of the percentage). Mean BASDAI score of 4.2 and BASFI score of 4.0	Overall, 96% of patients reported experiencing flares. The duration varied by patient: days (40%), weeks (30%) and months (30%). Of these, 83% reported experiencing symptoms between flares. The percentage of patients for the four patterns of the disease were: (a) relapsing/remitting (flares with no symptoms between): around 20% (b) flares on a background of symptoms: around 50% (c) gradually developing and resolving flare with periods of no symptoms: 7% (d) gradually developing and resolving flare after which symptoms worse than before start of flare: 26% (a) and (d) associated with higher BASFI score

IQR, interquartile range; mNY, modified New York criteria; RLDQ, Revised Leeds Disability Questionnaire; SIJ, sacroiliac joint.

There is evidence that BASDAI is relatively constant over time. An analysis of data from a UK registry, SIRAS, demonstrated that patients stratified into high or low disease activity (BASDAI) remain in their separate groups over many (12) years.[149] Data on the long-term pattern of patient function (BASFI) in patients not being treated with anti-TNFs are more scarce. A cohort study, from a single centre in England, provided data on 69 patients followed over 10 years [two data points: at baseline (1998) and at 10 years (2008)].[14] The assessment of BASDAI confirmed that it remains relatively constant [mean at baseline 4.1 units (SD 2.5 units) and after 10 years 4.4 units (SD 2.7 units) ($p = 0.36$)]. Patient function was assessed using the Revised Leeds Disability Questionnaire (RLDQ) rather than BASFI, but provided evidence of deteriorating function over time: mean RLDQ at baseline was 10.4 (SD 8.3), and after 10 years was 13.6 (SD 10.9) ($p = 0.002$). Analysis of longitudinal data from the SMART (Bath, UK) data set ($n = 223$) found that BASFI increased over time by 0.035 units/symptom year.[150] In patients with baseline BASDAI of ≥ 4 (those that would be treated with anti-TNFs and 68% of the total cohort) the rate of BASFI increase was 0.039 units/symptom year. Estimates of the rate of change in BASFI over time were also reported in a cost-effectiveness modelling study.[152] The data were from patients who were captured in two surveys at two time points 1992/1994 and November 2002 approximately 8 years apart ($n = 1100$). The estimate of annual BASFI progression was 0.07 points, but when only patients with BASDAI score of ≥ 4 were included in the analysis, BASFI progression was estimated as 0.054. It was reported that data from a cohort of 493 patients who had been followed up for more than 3 years generated similar findings; the number was not actually reported for the whole survey but was 0.059 for the BASDAI score of ≥ 4 subgroup.

Natural history data from patients with nr-AxSpA are even more scarce than those for AS patients, with no long-term data identified. A comparison of AS and nr-AxSpA patients from a cohort of 100 consecutive patients (Herne clinic, Germany) (AxSpA $n = 100$, nr-AxSpA $n = 44$, AS $n = 56$) found that slightly higher proportions of AS patients met pre-specified cut-off points of disease severity than did nr-AxSpA patients, but the differences were statistically significant only for ASDAS, CRP level, mSASSS and the number of inflamed lesions; the proportion of males was also statistically significantly different.[153] The results are given in *Table 27*. The difference for BASFI was very close to statistical significance.

A larger cross-sectional study of the GESPIC cohort [$n = 462$ patients with axSpA (AS or nr-AxSpA)] also found differences between AS and nr-AxSpA patients.[18] When AS (≤ 5 years) and nr-AxSpA were compared, there were statistically significant differences in Physician Global Assessment, BASFI (3.1 in AS vs. 2.5 in nr-AxSpA), BASMI (1.9 in AS vs. 1.1 in nr-AxSpA), spinal mobility and lateral spinal flexion, CRP level and ESR, and all radiographic measures (mSASSS 4.9 in AS vs. 1.4 in nr-AxSpA). mSASSS was statistically significantly worse in males versus females and between CRP level > 6 and < 6, although it is unclear whether or not this is a meaningful cut-off point for CRP level.

In two longitudinal studies of progression in nr-AxSpA,[154,155] also using the GESPIC cohort, progression in terms of sacroiliitis and in terms of radiographic progression in the spine (mSASSS) was slightly more rapid in AS than in nr-AxSpA but not statistically significantly so. Raised CRP level at baseline was a predictor of both measures of progression in AS but only for sacroiliitis in nr-AxSpA. The presence of syndesmophytes was predictive of higher progression rates as assessed by mSASSS in both AS and nr-AxSpA. Of the 95 patients with nr-AxSpA, 11 (11.6%) fulfilled the modified New York criteria for AS after 2 years of follow-up. A review of the burden of illness in nr-AxSpA[157] cited this (11.6%) progression rate along with a 10% rate over 2 years and a 24% rate over 10 years. However, the 10-year rate was derived from a broader, more heterogeneous population than the GESPIC cohort: patients had undifferentiated spondyloarthropathies with over half not having inflammatory low back pain.[156,158] The GESPIC study recruited only patients with axSpA (AS or nr-AxSpA).

Studies of disease progression in nr-AxSpA focus on aspects of the disease that can be assessed through imaging techniques: radiography or MRI. This may appear reasonable given the subjective, patient-questionnaire basis of the BASFI score.

Finally, there is evidence that as well as being progressive, the course of AS includes flares. A study based on the population of a trial comparing probiotic and placebo treatment in AS found that the overall flare rate was 71.4 per 100 person-weeks; the major flare rate was 12/100 person-weeks and the minor flare rate was 59.4/100 person-weeks.[28] BASDAI and BASFI varied with type of flare: mean BASDAI scores were 5.5 (major flare), 3.1 (minor flare) and 2–2.5 (flare free), and mean BASFI scores were 5.5 (major flare), 3.1 (minor flare) and 2.5–3.5 (flare free). A pilot study used the SMART cohort (Bath, UK) to investigate the pattern of disease and impact of disease flares.[159] Of the 114 patients, 96% reported experiencing flares. Flare duration varied by patient: days (40%), weeks (30%) and months (30%). Fifty per cent of patients reported flares on a background of symptoms, while 26% reported gradually developing and resolving flares after which symptoms were worse than before the start of the flare. These patterns were associated with higher BASFI scores. Around 20% reported flares with no symptoms between. A small proportion (7%) reported gradually developing and resolving flare with periods of no symptoms.

In summary, the available studies indicate that in AS and nr-AxSpA disease activity (BASDAI) is fairly stable over time and does not generally progress, although it can be at a high (severe) level early in the disease. Patients function (as assessed by BASFI) does deteriorate over time, although the course is not constant or predictable. BASFI is determined by both disease activity and bone neo-formation; progression of BASFI score over time is driven by progression of bony disease as assessed by imaging scores such as mSASSS,

or the presence of syndesmophytes. Best estimates of yearly disease progression rates without anti-TNF therapy are around 1.0 mSASSS units and 0.035–0.07 BASFI units. Information on the natural history of nr-AxSpA is relatively sparse. While disease progression appears to be faster in AS, patients with nr-AxSpA can have severe disease activity and hence poor function.

Clinical effectiveness summary and conclusions

Summary of randomised controlled trial results

The quality of the trial evidence was generally high; most studies were unlikely to have produced results which were biased. For both the AS and nr-AxSpA populations, the results of the meta-analyses demonstrated that anti-TNFs produce statistically significant and clinically relevant benefits to patients in terms of improving function and reducing disease activity. The common class-effect model used may have underestimated the uncertainty in the effect estimates. Although there is a possibility that infliximab is more effective than other TNF inhibitors at least at 12 weeks, there is no strong evidence to support this. For the disease activity, function and responder outcomes, the class-efficacy estimates were consistently slightly smaller for nr-AxSpA than for AS, most noticeably for BASFI and BASDAI 50. Statistical heterogeneity was more apparent in the nr-AxSpA analyses than in the AS analyses. This may be a result of both clinical heterogeneity in the nr-AxSpA trials (such as variation in CRP levels or the proportion of MRI positive patients) and the fact that fewer studies were available for analysis. In the light of the statistical heterogeneity across the nr-AxSpA trials, both the reliability of the nr-AxSpA-pooled estimates and their true relevance to patients seen in clinical practice is questionable.

The US FDA reanalyses of two key nr-AxSpA trials further emphasised the heterogeneity in the nr-AxSpA population. Results for an adalimumab trial in nr-AxSpA patients suggested reduced efficacy in a centrally diagnosed nr-AxSpA population than in a locally diagnosed population and that the treatment benefit in the whole trial population may have been driven by benefit in patients who actually had AS, not nr-AxSpA. Conversely, in a certolizumab pegol trial which recruited both populations, the efficacy findings were consistent across the AS and nr-AxSpA subpopulations, regardless of the discrepancy in local or central pelvic radiograph readings.

Long-term efficacy

The longest follow-up durations in patients with AS by anti-TNF were 5 years for adalimumab, 5 years for etanercept, 3 years for infliximab, around 5 years for golimumab and nearly 2 years for certolizumab pegol. The results showed that across all the anti-TNFs after approximately 2 years of treatment, around half of patients still achieved a good level of response to therapy. At 5 years around 60% of golimumab patients, 50% of etanercept patients and 30% of adalimumab patients still achieved a good treatment response. However, the long-term studies were not as well-reported as the RCTs, and their results were derived from less reliable data; it is therefore unknown if these are true treatment differences or a result of differences in follow-up protocols, and/or imputation and analysis methods.

The long-term follow-up for nr-AxSpA patients showed a continued high proportion of responders. At 1 year around half of patients on adalimumab, etanercept or certolizumab still achieved an ASAS 40 or BASDAI 50 level response. With certolizumab this is maintained at 2 years and with adalimumab at 3 years.

When the long-term data are presented as observed or as completer analyses, the long-term results are similarly good: withdrawal rates are not high and those patients who remain on treatment continue to achieve a good response.

For all anti-TNFs, at long-term follow-up mean final values or mean change from baseline for BASDAI, BASFI and BASMI, when reported, were generally maintained at levels indicative of clinically significant treatment benefit for those patients with AS and those with nr-AxSpA.

Four studies reported on radiographic disease progression over 2 years of follow-up in terms of mSASSS in patients taking adalimumab, infliximab, etanercept and golimumab. All four open-label, uncontrolled follow-up studies found that mSASSS increased by a mean of around 0.9 over 2 years. Three of these studies compared their rates with those from the OASIS cohort[118] (of patients not taking an anti-TNF) and found no difference. In conclusion, there is no real evidence for the impact of anti-TNF treatment on radiographic disease progression; a beneficial effect cannot be assumed, nor, given the short-term nature of the follow-up and the insensitivity of radiography as a tool for the evaluation of disease progression in AS, can one be discounted. There are some data to suggest an identifiable benefit from around 4 years but results from ongoing long-term studies should help to clarify this issue.

Registry data demonstrate that around 60% of patients with AS treated with a first anti-TNF will still be taking their therapy at 2 years, with median drug survival of 3.1 years (based on Danish registry $n = 1436$). Sequential treatment with anti-TNFs can be worthwhile but the drug survival response rates and benefits are reduced with second and third anti-TNFs, with the proportion of BASDAI 50 responders falling approximately 10% with each subsequent anti-TNF and the median BASDAI and BASFIs achieved increasing (worsening). The lower efficacy of a second anti-TNF relative to a first is reflected in lower median drug survival and proportion of patients remaining on therapy at 2 years. Interestingly, despite a further reduction in response and efficacy with a third anti-TNF, drug survival does not fall further, suggesting that patients may be allowed to, and be prepared to, continue with a less than optimally effective anti-TNF at this stage in their treatment history.

Adverse effects

Data from large systematic reviews, which included patients with a wide range of diseases, suggest that, in the short term, anti-TNFs as a group are associated with significantly higher rates of serious infections, tuberculosis reactivation, non-melanoma skin cancer, total AEs and withdrawals because of AEs, when compared with control treatments. Specifically, infliximab is associated with significantly higher rates of total AEs and withdrawals because of AEs, and that certolizumab pegol is associated with significantly higher rates of serious infections and SAEs. Analyses from the present review showed etanercept to be statistically significantly more likely to result in an injection/infusion site reaction compared with infliximab, although analysable data on such reactions were not reported for the other three anti-TNFs. Evaluations of longer-term data are more scarce, although they suggest similar safety profiles across anti-TNFs. Data from the open-label studies included in this review also do not suggest that there are important differences between treatments, other than a higher incidence of injection site reactions following treatment with etanercept. These open-label data are, however, limited by the small sample sizes and non-randomised study designs.

Natural history

The available studies indicate that in AS and nr-AxSpa disease activity (BASDAI) is fairly stable over time and does not generally progress, although it can be at a high (severe) level early in the disease. Patient function (as assessed by BASFI) does deteriorate over time, although the course is not constant or predictable. BASFI is determined by both disease activity and bony disease; progression of BASFI over time is driven by progression of bony disease as assessed by imaging scores such as mSASSS, or the presence of syndesmophytes. Best estimates of yearly disease progression rates without anti-TNF therapy are around 1.0 mSASSS units and 0.035 to 0.07 BASFI units. Information on the natural history of nr-AxSpA is relatively sparse. While disease progression appears to be faster in AS, patients with nr-AxSpA can have severe disease activity and hence poor function.

Overall conclusions

- For both the AS and nr-AxSpA populations the results of the meta-analyses demonstrated that anti-TNFs produce statistically significant and clinically important benefits to patients in terms of improving function and reducing disease activity. The efficacy estimates were consistently slightly smaller for nr-AxSpA than for AS.
- In AS, although there is a little variation in treatment effects and it is possible that infliximab may be more effective than other anti-TNFs at 12 weeks, the evidence for this is not strong and it is plausible that anti-TNFs may have a common class effect, with the treatments being equally effective.
- Statistical heterogeneity was more apparent in the nr-AxSpA analyses than in the AS analyses. This may be because of both clinical heterogeneity in the nr-AxSpA trials and the fact that fewer studies were available for analysis. In the light of this heterogeneity, both the reliability of the nr-AxSpA-pooled estimates and their true relevance to patients seen in clinical practice is questionable.
- Effectiveness was maintained over time. About 50% of patients maintained a benefit at 2 and 5 years.
- Evidence for an effect of anti-TNFs on radiographic disease progression was limited. The relatively short-term follow-up available to date and the insensitivity of radiography as an imaging tool precluded the drawing of firm conclusions regarding the role of anti-TNFs in preventing or delaying the progression of AS; there are some data to suggest an identifiable benefit from around 4 years, but results from ongoing long-term studies should help to clarify this issue.
- Sequential treatment with anti-TNFs can be worthwhile in patients with AS but the drug survival response rates and benefits are reduced with second and third anti-TNFs.

Chapter 4 Assessment of existing cost-effectiveness evidence

Systematic review of existing cost-effectiveness evidence

The following sections provide an overview of existing cost-effectiveness evidence and an assessment of the relevance of the data from the perspective of the UK NHS. The differences in the approaches and assumptions used across the studies are examined in order to explain any discrepancies in the findings and to identify key areas of remaining uncertainty. The findings from the review provide the basis for the development of a new decision-analytic model reported in *Chapter 6, Independent economic assessment: York model*.

Methods

An initial systematic search was undertaken in the NHS EED using a combination of technology names and disease terms. Further searches were undertaken in MEDLINE and EMBASE for modelling and utility studies using disease terms only (as known references were not identified from the initial search in NHS EED). Only full economic evaluations that compared two or more options and consider both costs and consequences (including cost-effectiveness, cost–utility and cost–benefit analyses) were included in the review of existing economic literature. No language and date limits were initially applied, although eligibility of studies was subsequently restricted to those reporting results which were specific to the UK. Full details of the search strategies used are reported in *Appendix 1*.

In addition, as part of the current multiple TAs process, each manufacturer submitted de novo evidence on the cost-effectiveness of the anti-TNFs in line with their indications for the treatment of AS and nr-AxSpA.[34–37] These submissions are reviewed and the findings compared with those found in the review of previously published studies.

Results of review of existing cost-effectiveness evidence

The combined searches retrieved 210 citations. A total of six UK studies reporting on the cost-effectiveness of anti-TNFs for the treatment of AS were identified. No previously published studies were identified for patients with nr-AxSpA.

Four of these studies were industry-funded assessments of the following anti-TNFs: infliximab (Kobelt *et al.* 2004[152] and Kobelt *et al.* 2007[160] both funded via an unrestricted grant by Schering-Plough), etanercept (Ara *et al.*[161] funded by Wyeth pharmaceutical P.C.) and adalimumab (Botteman *et al.*[162] funded by Abbott Laboratories). The three studies published in 2007 are largely based on the economic analyses originally submitted by the manufacturers to NICE as part of the previous multiple TA (TA143).[17] As the earlier publication by Kobelt has been superseded by the 2007 publication, only the latter publication is further considered in this review. The remaining two UK studies were publications of the assessments and/or critiques undertaken by the independent Assessment Group/Evidence Review Group (ERG) for infliximab, etanercept and adalimumab for TA143[38] and golimumab for TA233.[163] Therefore, a total of five studies met the inclusion criteria and are included in this review.

The following sections provide a narrative discussion of each publication. The quality assessment of these studies is shown in *Appendix 10*. A single critique section is used to highlight the key issues and potential limitations of existing published cost-effectiveness evidence. These issues are then revisited with respect to the de novo analyses submitted by the manufacturers considering how these key issues and potential limitations have been addressed in the two separate indications. The final section highlights the remaining issues and uncertainties and provides the basis for informing the development of a separate independent analysis of the cost-effectiveness of anti-TNFs for AS and nr-AxSpA relevant to informing decisions for the NHS.

Assessment of published cost-effectiveness studies

Kobelt *et al.*:[160] Comparison of the Cost-Effectiveness of Infliximab in the Treatment of Ankylosing Spondylitis in the UK Based on Two Different Clinical Trials

Kobelt *et al.*[160] estimated the cost-effectiveness of infliximab for the treatment of AS compared with standard care over a lifetime horizon (60 years). Results were presented from both societal and NHS/personal social services (PSS) perspectives, although only the latter are reported here in line with the current NICE reference case. Short-term effectiveness data were derived from two separate clinical trials (Braun *et al.*[98] and ASSERT[102,164]) to inform the proportion and magnitude of initial response to treatment expressed in terms of BASDAI 50 (or a BASDAI ≤ 4) response (12–24 weeks) and changes in BASDAI and BASFI scores. These were combined with longer-term observational evidence on disease progression (BASFI only) and other external sources on costs and utilities to estimate cost-effectiveness. Results were reported separately based on each trial. Costs and benefits were discounted at 3.5% and presented at 2005 prices.

Methods

The cost-effectiveness model was based on a short-term decision tree representing the double-blind periods of the trials (12–24 weeks) and a longer-term Markov model to estimate subsequent progression. The Markov model comprised three states: 'Off treatment', 'On treatment' and 'Dead'. Only patients responding to treatment as defined by the following criteria [BASDAI ≤ 4 (scale 0–10) or a $\geq 50\%$ improvement in BASDAI] remain on treatment at the end of the double-blind periods. Differential BASDAI and BASFI scores (scale 0–10) were derived from Braun ['Off treatment' BASDAI score = 6.3 and BASFI score = 5.4; 'On treatment' (responders) BASDAI score = 1.8 and BASFI score = 2.0] and ASSERT[102] ['Off treatment' BASDAI score = 6.4 and BASFI score = 5.8; 'On treatment' (responders) BASDAI score = 1.4 and BASFI score = 1.9]. Disease progression was expressed in terms of changes in BASFI and was estimated from two surveys conducted 10 years apart ($n = 1110$).[152] The mean absolute annual change in BASFI applied was + 0.07 (scale 0–10) and this was used to characterise the natural history of progression for patients with AS without infliximab. Three main scenarios were presented reflecting different assumptions concerning the impact of infliximab on disease progression: (1) no progression while on treatment; (2) 50% of natural history (0.035/year); and (3) same as natural history (0.07/year).

Fifteen per cent of patients were assumed to discontinue from infliximab annually based on data specific to responders from the open-label extension period in the Braun trial. Interestingly, the authors noted that the persistence rate was lower in responders compared with the entire sample in the Braun trial[98] and its extension (approximately 10% withdrawal rate per annum). The BASDAI and BASFI scores for patients who withdrew from infliximab were assumed to return to the mean score of the non-treated group. Mortality was modelled from general population life-tables applying a SMR of 1. Hence no additional mortality was assumed to be related to AS and no direct or indirect benefits for mortality were assumed for infliximab.

Disease costs and HRQoL were derived from a cross-sectional retrospective survey conducted at the University of Bath, with the sample covering the full range of BASDAI and BASFI (1–10). The annual cost of infliximab was based on 5 mg/kg body weight (weeks 0, 2 and 6 and then every 6 weeks). An initial cost was assigned to all patients starting treatment (£79.25) and an outpatient cost was applied to each infusion.

Results

From a NHS perspective, the cost per quality-adjusted life-years (QALYs) gained ranged from £28,332 and £26,751 (no progression while on treatment) to £49,417 and £46,167 (no effect of treatment on progression) as shown in *Table 28*. The model was also sensitive to the time horizon and the withdrawal rate. Using a 10-year horizon resulted in incremental cost-effectiveness ratios (ICERs) between 63% and 66% higher than the base-case lifetime horizon (60 years) and a withdrawal rate of 5% resulted in ICERs between 22% and 33% higher than the base case (15%).

TABLE 28 Lifetime cost per QALY estimates reported by Kobelt *et al.*[160] (NHS and PSS perspective)

Scenario	Incremental cost	QALY gain	Incremental cost-effectiveness ratio (£/QALY)
Braun[98]			
No progression on treatment	36,378	1.28	28,332
50% progression on treatment	35,756	1.01	35,332
Same progression on treatment	39,336	0.80	49,417
ASSERT			
No progression on treatment[102]	33,920	1.27	26,751
50% progression on treatment[102]	34,408	1.01	34,067
Same progression on treatment[102]	39,242	0.86	46,167

Ara *et al.*:[161] The Cost-Effectiveness of Etanercept in Patients with Severe Ankylosing Spondylitis in the UK

Ara *et al.*[161] estimated the cost-effectiveness of etanercept for the treatment of severe AS in the UK in accordance with BSR guidelines from a NHS/PSS perspective over a 25-year time horizon. Effectiveness data were derived from individual patient data from a large multicentre European RCT to inform the proportion and magnitude of initial response to treatment and associated changes in BASDAI and BASFI scores. These were combined with longer-term observational evidence on disease progression (BASFI) and other external sources on costs and utilities to estimate cost-effectiveness. Costs and benefits were discounted at 3.5%. The price year was not formally stated.

Methods

An individual patient model was used to estimate short-term and longer-term costs and outcomes. Patients in the model were assumed to have tried and failed at least two consecutive NSAIDs and have a BASDAI measurement ≥ 40 (scale 0–100). Response was defined as a ≥ 50% reduction in BASDAI (or all fall of ≥ 20 units) and a reduction of the spinal VAS by ≥ 2 units. Response rates at 12 and 24 weeks were derived from two RCTs (67% and 55% for etanercept and 24% and 16% for comparator arm at each respective time point). Individual patient data at 12 and 24 weeks were used to estimate the magnitude of change in BASDAI and BASFI for responders and non-responders. The mean BASDI and BASI scores at week 12 and 24 for responders and non-responders are reported in *Table 29* together with observed utility at week 12 and the predicted utility values mapped from BASDAI and BASFI at week 24.

TABLE 29 Bath Ankylosing Spondylitis Disease Activity Index, BASFI and EQ-5D measurements at weeks 12 and 24

	Week 12			Week 24		
Patient response	BASDAI	BASFI	EQ-5D[a]	BASDAI	BASFI	EQ-5D[b]
Treatment non-responder	53.02	54.86	0.48	56.87	56.87	0.46
Treatment responder	19.52	25.39	0.79	18.32	21.41	0.80
Comparator non-responder	55.60	57.55	0.46	47.67	47.78	0.42
Comparator responder	22.97	29.88	0.74	25.11	20.92	0.79

a Observed values.
b Predicted values using a mapping algorithm.

For patients who continued responding to treatment it was assumed that BASDAI and BASFI measures remained constant at the levels observed at week 24. For patients who withdrew after week 24, it was assumed patients would immediately revert back to their baseline values of BASDAI and BASFI. After 24 weeks in the model it was also assumed that patients with AS, not receiving anti-TNFs [conventional care (CC) and etanercept non-responders], would experience a worsening BASFI. A mean absolute change in BASFI of 0.7 (scale 0–100) was assumed based on a cross-sectional study of over 1000 UK patients.[152]

Quality-adjusted life-years were estimated using a relationship derived from BASDAI, BASFI and EQ-5D from a single European RCT (utility = 0.9235 – 0.004 × BASFI – 0.004 × BASDAI). Disease costs were derived from a separate costing study of 147 patients attending the Staffordshire Rheumatology Centre in Stoke-on-trent.[165] A relationship between BASDAI and BASFI measurements and costs was used to estimate the disease costs and impact of etanercept (annual costs = 5.862 + 0.006 × BASDAI + 0.016 × BASFI). An annual cost of £9372 was included to reflect the acquisition and monitoring costs associated with etanercept. An initial cost of £71 was also applied to the first 3-month period for etanercept, although no further details were provided by the authors concerning what this cost represented. The costs and/or HRQoL associated with AEs were not included.

The authors assumed that 10% of patients withdraw from etanercept every year. These data were derived from external sources and no explanation was provided concerning whether or not these data specifically applied to the post-24-week period or not and/or whether they were derived from responders to treatment or not. Mortality was modelled from general population life-tables applying a SMR of 1.50. No direct or indirect benefits for mortality were assumed for etanercept.

Separate scenarios were presented to explore alternative assumptions related to disease progression, long-term annual withdrawal and the model time horizon.

Results

The main results are summarised in *Table 30*. From a NHS perspective, the base-case cost per QALY gained was £22,704 for etanercept over a 25-year horizon. In contrast to the study by Kobelt *et al.*,[160] the impact of alternative progression assumptions appeared to have limited impact on the ICER, with alternative scenario results ranging from between £23,625 (50% progression on treatment) and £25,679 per QALY (same progression on treatment). The ICERs for alternative annual withdrawal rates ranged from £15,103 (5% withdrawal rate) to £29,428 per QALY (15% withdrawal rate). The ICERs for alternative time horizons ranged between £27,594 (2 years) and £22,704 (25 years).

TABLE 30 Twenty-five-year cost per QALY estimates reported by Ara *et al.*[161] (NHS and PSS perspective)

Scenario	Incremental cost	QALY gain	ICER (£/QALY)
Base case	35,978	1.59	22,704
No progression for any patient	36,825	1.43	25,679
50% progression on treatment (0.035 BASFI score)	36,032	1.56	23,155
Same progression on treatment (0.07 BASFI score)	36,088	1.53	23,625
Annual withdrawal rate, 5%	33,976	2.25	15,103
Annual withdrawal rate, 15%	36,968	1.26	29,428

Botteman *et al.*:[162] Cost-Effectiveness of Adalimumab for the Treatment of Ankylosing Spondylitis in the UK

Botteman *et al.*[162] evaluated the cost-effectiveness of adalimumab versus conventional therapy in patients with active AS from a NHS perspective over a 30-year time horizon. Effectiveness data were derived from pooled data from two Phase III studies in patients with an inadequate response to ≥ 1 NSAID. Micro-simulation methods were subsequently applied to these studies to simulate treatment decisions in accordance with BSR guidelines and associated outcomes. These were combined with author assumptions on disease progression (BASFI only), utility and cost data from the clinical trials and other external sources to estimate cost-effectiveness. Costs and benefits were discounted at 3.5% using a 2004 price year.

Methods

Micro-simulation methods were applied to patients ($n = 397$) recruited into two adalimumab RCTs: ATLAS[61] and M03-606. In the adalimumab clinical trials, patients were kept on active treatment even when response had not been achieved. Consequently, simulation methods were applied to the patients in the clinical trial to mimic treatment decisions which more closely reflected treatment guidelines and the requirements of the economic model. In accordance with BSR guidelines, a response in the model was defined as a reduction of BASDAI of 50% or a decrease of ≥ 2 cm (scale 0–10) accompanied by a reduction of spinal pain VAS of ≥ 2 cm. Assessment of initial response was assumed to take place 8 weeks after treatment initiation. If the response criteria were not met at 8 weeks, a second response assessment was assumed at 12 weeks. Failure to achieve response on both occasions was assumed to lead to withdrawal of adalimumab therapy. Therapeutic responses were then assumed to be reviewed every 3 months until the end of the simulation (year 30). Failure to maintain the original response led to repeat assessments after 6–12 weeks in the first 48 weeks. Failure to maintain response on both occasions led to withdrawal of adalimumab. After week 48, the simulation model defined inadequate response on the basis of BASDAI scores only. In the RCTs, patients were allowed to switch to open-label adalimumab at week 24; for these patients, LOCF at time of switch for BASDAI, BASFI and VAS values were used in the model.

The BASDAI, BASFI and spinal pain scores were based on directly observed trial scores (until week 48) and additional assumptions about disease progression (after week 48). The BASDAI, BASFI and spinal pain scores were adjusted at each time point by a fixed value equal to the average difference between adalimumab and CC patients observed at baseline. BASDAI scores after week 48 were assumed to remain constant at these levels for patients continuing to respond to adalimumab and for CC patients. BASFI was assumed, for CC patients, to worsen after week 48 by 0.05 units (scale 0–10) annually. The estimate applied to the increase in BASFI appears to be based on the authors' own assumption but is argued to be consistent with previous cost-effectiveness/epidemiological studies. In contrast, BASFI scores were assumed to remain stable for adalimumab while patients remained on therapy, which was argued to be consistent with the assumptions applied in previous published cost-effectiveness studies. It was assumed that patients who discontinued would revert back to the BASFI scores of CC patients within 12 weeks (i.e. any benefits in BASFI were not maintained over a longer period). This was argued by the authors to be a conservative assumption.

Utilities were derived from the Health Utilities Index 3 (HUI-3) from data at baseline and at 24 weeks from both adalimumab trials. A subsequent regression was estimated to predict utilities based on BASDAI, BASFI, sex and race [utility = 0.948857 − 0.041528 × BASDAI − 0.034481 × BASFI + 0.047080 × Gender (1 = male, 0 = female) − 0.063801 × Race(1 = white, 0 = other)].

Estimates of disease costs were based on 2-year data from 208 patients in the OASIS study, conducted in the Netherlands, Belgium and France.[118] An ordinary least squares regression was estimated using only BASDAI (and only BASFI in a sensitivity analysis). The regression utilised in the base-case was £708.45 + £750 × BASDAI. Hence each increase in BASDAI of 1 unit (scale 0–10) was assumed to be associated with an increase in costs of £750.

Additional acquisition costs were applied to adalimumab (£357.50 per injection). No additional administration costs were incorporated as patients were assumed to self-administer their injections. All patients, regardless of treatment, were assumed to require at least two rheumatologist visits per year. Routine safety monitoring costs were based on national guidance and included the cost of nursing and physician time. The cost of a routine tuberculosis screening test via chest radiography was assumed before and 6 months after initiation of therapy and tuberculosis skin testing before initiation of therapy. The cost of AEs was based on data collected from the two clinical trials. A cost of £5100 was applied to an active tuberculosis case.

An annual rate of withdrawal of 10% was applied based on a assumption by the authors. The estimate was argued to be consistent with estimates reported in previously published cost-effectiveness analyses.

Results

The main results are summarised in *Table 31*. From a NHS perspective, the base-case cost per QALY gained was £23,097 for adalimumab over a 30-year horizon. Similar to the study by Ara *et al.*,[161] the impact of alternative progression assumptions appeared to have limited impact on the ICER, with alternative scenario results ranging from between £23,802 (no BASFI progression on any treatment) and £23,812 per QALY (same BASFI progression on treatment). However, in contrast to Ara *et al.*,[161] the ICERs appeared more sensitive to the alternative time horizons with estimates ranging between £47,083 (48 weeks), £26,332 (5 years) and £23,097 (30 years).

McLeod *et al.*:[38] Adalimumab, Etanercept and Infliximab for the Treatment of Ankylosing Spondylitis: A Systematic Review and Economic Evaluation

McLeod *et al.*[38] evaluated the clinical effectiveness and cost-effectiveness of adalimumab, etanercept and infliximab compared with conventional treatment for AS. The publication is based on the independent assessment undertaken by Liverpool Reviews and Implementation Group (LRiG) for NICE TA143.[17] The cost-effectiveness of these interventions over the short term (1 year) and over alternative time horizons of up to 20 years was reported.

Methods

The authors assumed that all three interventions were of equal clinical effectiveness and analysed the anti-TNFs as a class versus placebo. Short-term effectiveness over 1 year was modelled using individual patient data from two RCTs (including an open-label extension from week 24) for adalimumab from 397 patients (246 adalimumab, 151 placebo). Of these, 315 of 397 patients were deemed to meet the BSR guidelines and were included within the Abbott economic model. There is a lack of transparency regarding the values used because of the commercial nature of the data. However, the estimates for response rates were reported to be similar to those reported by LRiG in a separate pooled analysis at weeks 12 and 24, which were 59% and 49.2%, respectively, for the TNF-α inhibitors (vs. 22.5% and 14%, respectively, for placebo). No information is reported on the magnitude of changes assumed. From week 30 onwards it was assumed that spontaneous recovery without treatment (for placebo patients) would occur at a rate of 17.1%, as estimated by LRiG from the patient-level analysis of the two adalimumab RCTs. This assumption was explored in a separate sensitivity analyses.

TABLE 31 Thirty-year cost per QALY estimates reported by Botteman *et al.*[162] (NHS and PSS perspective)

Scenario	Incremental cost	QALY gain	ICER (£/QALY)
Base case	23,857	1.03	23,097
No progression for any patient	NR	NR	23,802
Same progression on treatment (0.05 BASFI)	NR	NR	23,812

NR, not reported.

The assessment group model assumed that patients withdraw from TNF-α inhibitor treatment at a rate of 15% per year, which was considered to represent the 'central value' of the studies that were identified reporting longer-term discontinuation rates. This estimate is also the same as the annual rate reported in the open-label extension study for infliximab (Braun trial) also reported in the review undertaken by LRiG. Rates of 7% and 24% were also explored in separate sensitivity analyses, representing the range of values reported across the studies considered. The annual withdrawal rate (after the first 12 months, as observed rates are used in the first 12 months) was applied to the difference in response rate between the two arms of the evaluation, rather than the absolute number of responders. This was to account for a potential anomaly that could arise through the assumption of a constant level (17.1%) of spontaneous recovery without treatment.

The assessment group model took into account the cost of drug acquisition, administration, monitoring and AEs. No administration costs were assumed for etanercept and adalimumab, as it was assumed that both would be self-administered at home without supervision. The authors assumed an additional cost of £267 to administer infliximab infusions based on NHS Reference Costs estimates for the regular attender cost for chemotherapy with musculoskeletal primary diagnosis. Quarterly monitoring and testing was assumed for all patients receiving long-term treatment with TNF-α inhibitors. However, two of these assessments were assumed to take place at the patient's routine follow-up outpatient visit, so only the additional costs of tests for monitoring (£25) were applied to these assessments. The remaining two assessments were assumed to be undertaken at a general practitioner's surgery and an additional cost of £25 was assumed for nurse/general practitioner time in addition to the costs of tests for monitoring (£25). AEs costs were based on estimates reported by one manufacturer (Abbott) of £95.29 in the first year of treatment and £47.65 per patient-year thereafter.

Disease-related costs to the NHS were estimated by fitting an exponential cost model to the weighted aggregate data from the OASIS study; a 2-year prospective study of 208 AS patients from four centres in France, Belgium and the Netherlands ($n = 208$).[118] The exponential model estimated NHS cost = £1585.30 × exp(0.1832 × BASFI). The OASIS data were considered by the authors to provide a more reliable source than other published studies from Stoke and Bath, being prospective in design and over a longer period. BASFI was used by the authors as the major predictor of costs because it was considered to better reflect long-term disease progression compared with BASDAI.

Health-related quality of life was estimated using the utility model provided by Schering-Plough developed from the Bath Survey data set on the grounds that it used a comparatively larger sample of UK AS patients ($n = 1144$), and also because it incorporated age and sex variables: utility = 0.8772129 − 0.0384087 × BASDAI − 0.0322519 × BASFI − 0.0278913 × Male + 0.0016809 × Age.

The assessment group adopted a long-term increase in BASFI scores of 0.07 units per year for the conventional treatment comparator arm of the model. This progression rate is applied for all periods after week 20 in the model. In the base-case analysis, the same value was used in the intervention arm adjusted pro rata to the proportion remaining of the maximal excess response seen at 12 weeks. In effect, this assumes that patients withdrawn from anti-TNFs are assumed to return to the same trajectory as non-responders, such that there is no ensuring benefit associated with being an initial responder.

Results

Over a 1-year time horizon, base-case ICERs for adalimumab and etanercept versus CC were essentially the same (approximately £57,000 per QALY). In contrast, the ICER for infliximab was over £120,000 per QALY. With respect to modelling beyond 12 months, the results for adalimumab were considered as representative of etanercept, and only the former were provided. In contrast with other published models, the ICERs increased steadily from year 2 onwards. At a 20-year horizon the ICERs for adalimumab/etanercept increased to £98,910 per QALY and to £175,000 per QALY for infliximab.

Additional National Institute for Health and Care Excellence Decision Support Unit analyses

Given the discrepancy between the results reported by McLeod *et al.*[38] and the manufacturer's submissions[34–37] (largely reflected in the three industry-funded publications previously discussed in this section)[160–162] also submitted as part of TA143,[17] additional work was undertaken by NICE's Decision Support Unit (DSU) to reconcile the different models and to explore whether or not differences were a result of different parameter inputs or alternative structural assumptions.

A common set of parameter values were applied by the DSU to the three manufacturer models and the LRiG model. The purpose of this was to attempt to identify whether or not differences between the results of the models persisted once this common set of values were used. The specific parameter values which were implemented were:

1. no improvement in BASFI or BASDAI for patients not on anti-TNFs
2. BASFI progression prevented while on anti-TNFs
3. BASFI progresses at 0.07 per annum when patients are not on anti-TNFs
4. annual withdrawal rate of 7% from anti-TNFs
5. baseline BASDAI/BASFI score averages 6.5/5.6
6. utility model as in the Schering-Plough submission
7. assessment group parameters for cost parameters (drug costs only)
8. a 20-year time horizon.

These parameter values were reported to have been the values agreed at a separate NICE committee meeting and consequently the rationale for these values and assumptions is not formally stated by the NICE DSU.

The results of the DSU analysis found that the manufacturer models all gave relatively consistent results for each of the drugs. For Schering-Plough, the ICERs over 20 years for etanercept/adalimumab were £27,000 or £24,000 and for infliximab were £58,000 and £50,000. Two figures were presented because Schering-Plough presented two different versions of the model which reflected two different trials. The Wyeth model gave results of £20,000 for etanercept and £39,000 for infliximab. Abbott gave results of £17,000 for adalimumab and £43,000 for infliximab (over a 30-year time horizon). These ICERs were markedly different from those reported by the independent assessment group. Using a similar set of parameters the results for etanercept/adalimumab using the LRiG model were £42,000 and for infliximab £82,000.

Further work by the DSU revealed that the differences appeared largely driven by two key assumptions which differed between the LRiG and industry models relating to:

1. the modelling of a 'placebo' effect
2. the longer-term functions fitted to BASDAI and BASFI for responders to anti-TNFs.

The LRiG model applied a 17.1% rate of spontaneous recovery without treatment from week 30 onwards (i.e. akin to assuming a long-term 'placebo' response for CC) in contrast to the manufacturers who either assumed there would be no response with CC or that any response would be transient and dissipate quickly after the 12-week period.

The LRiG model also applied a quadratic function to the BASDAI and BASFI scores of responders over a longer-time horizon, compared with the linear functions used by the manufacturers. The use of a quadratic function assumes that that the difference compared with CC was decreasing (initially) with time; that is, over time, the differences in BASDAI/BASFI would slowly reduce in responders and eventually be the same as for CC. However, the logical problem of applying a quadratic function is clear. While the scores are reducing for a period, at longer-time periods the function starts to increase again. The issues were addressed by LRiG by using various assumptions and logical constraints (i.e. BASDAI/BASFI score not allowed to be higher than CC).

To further reconcile the models, the DSU incorporated a series of alternative structural assumptions within the LRiG model. These assumptions included removing the 17.1% rate of spontaneous improvement applied to CC and assuming constant BASDAI/BASFI scores after 1 year for responders. Applying these assumptions resulted in an ICER for etanercept/adalimumab of £30,100 per QALY (estimates for infliximab not reported) which were considered to be more consistent with the manufacturer results.

Importantly, the DSU highlighted that, although these analyses helped to reconcile the different model results, any progression in terms of BASDAI or BASFI over time while on treatment would cause the ICER to increase beyond £30,100. Similarly, the DSU concluded that the exclusion of the 17.1% spontaneous recovery, without a comparable adjustment made to the intervention group was favourable towards the cost-effectiveness of TNF-α inhibitors and any adjustment for this issue would similarly lead to a higher ICER.

Armstrong *et al.*:[163] Golimumab for the Treatment of Ankylosing Spondylitis: a National Institute for Health and Care Excellence Single Technology Appraisal

Armstrong *et al.*[163] summarises the report undertaken by the ERG on the clinical effectiveness and cost-effectiveness of golimumab for AS for a NICE single technology appraisal (TA233).[33] The ERG provided a critique of the manufacturer's submission (Merck Sharp & Dohme)[37] and undertook additional exploratory analyses. The manufacturer's model applied a 20-year time horizon in the base-case and a separate lifetime analysis (60.1 years) was presented in a separate sensitivity analysis. The discount rate applied was 3.5% for utilities and costs, and costs are considered from a NHS and PSS perspective.

Methods

The manufacturer's submission for golimumab[37] was based on a single trial versus placebo (GO-RAISE[90]). A total of seven additional placebo controlled trials were included of other anti-TNFs; five RCTs for etanercept and two for adalimumab. In the absence of head-to-head studies directly comparing the relative effectiveness of the alternative anti-TNFs, the manufacturer undertook a Bayesian random-effects mixed-treatment comparison (MTC) including BASDAI 50 response, discontinuations and SAEs. All treatments were reported by the manufacturer to be statistically significantly more effective than placebo in terms of BASDAI 50 response. No statistically significant differences were reported between each of the alternative anti-TNFs in terms of discontinuations and SAEs. When the alternative anti-TNFs were compared with each other, no significant differences between golimumab, adalimumab and etanercept were identified for BASDAI 50. A higher risk of discontinuation was reported for golimumab versus etanercept (relative risk 4.30, 95% CrI 1.01 to 18.50), although golimumab was associated with significant improvements in BASDAI versus etanercept (mean difference −0.88, 95% CrI −1.58 to −0.14) and BASMI versus adalimumab (mean difference 0.52, 95% CrI 0.23 to 0.80).

The manufacturer cost-effectiveness model[37] was based on a short-term decision tree (12 weeks) and a longer-term Markov model. The short-term tree was used to characterise response to each TNF-α inhibitor treatment based on the MTC results for BASDAI 50. After the short-term tree, patients entered a separate Markov model with a cycle length of 12 weeks and time horizon of 20 years. If patients were already receiving a TNF-α inhibitor, they either stayed on therapy ('on TNF inhibitor' state) or discontinued therapy because of lack of efficacy or AEs ('not on TNF-inhibitor' state). It was assumed that discontinuations occurred at a rate of 15% per year in line with NICE TA143.[17] To model the lower disease activity just after discontinuation of TNF-α inhibitor therapy, two 12-week tunnel states ('just discontinued' and 'discontinued') were also incorporated into the model. Patients who are in the health state 'on TNF-α inhibitor' are assumed to have at least a 50% improvement in BASDAI (BASDAI 50) during the first 12 weeks of treatment and do not discontinue. Treatment is discontinued in patients whose condition does not respond to treatment and they are switched to conventional therapy. Patients in the CC arm enter the Markov model in the 'not on TNF-α inhibitor' state. Patients could die at any point in the model.

Disease progression was incorporated in the model using BASDAI and BASFI scores. Data from the GO-RAISE[90] trial and the open-label extension period were used to develop predictive equations of mean change from baseline in BASDAI and BASFI scores over time. Two separate equations were developed based on the 24-week data for all patients and post-24-week data from GO-RAISE[90] for responders only. These equations were used for all anti-TNFs and the manufacturer assumed that the scores followed the GO-RAISE[90] data for 2 years before they either levelled off (BASDAI) or started to deteriorate [BASFI at 50% of the rate of CC, equivalent to an increase of 0.035 (scale 0–10) units per year].

Although the equations are critical to the model structure and parameter estimates, these are not reported in the paper by Armstrong *et al.*[163] A separate examination of the full ERG report[33] revealed that these were reported as CiC by the manufacturer and hence it is not possible to report the assumptions made in relation to the magnitude of change in BASDAI and BASFI over the initial 24-week period and subsequent post-24-week period for the anti-TNFs (responders, non-responders) and CC. BASFI scores for CC were reported to deteriorate according to the GO-RAISE trial (short-term equations were available only) after which they were assumed to deteriorate at a rate of 0.07 units per year. The assumptions related to the impact of discontinuation of anti-TNFs are not formally stated in the paper by Armstrong *et al.*[163] However, the structure of the model implies that patients will revert back to the subsequent trajectories of CC for both BASDAI and BASFI after 2 cycles (24 weeks).

Utilities were derived from the previous NICE TA (TA143[17]) and incorporated age, sex, BASFI and BASDAI. Costs included in the model comprised drug acquisition, short-term (12-week) costs, longer-term disease costs and AEs. Longer-term disease costs were based on BASFI scores from the GO-RAISE[90] trial using the same regression equation used for NICE TA guidance 143. Mortality was included in the model and was considered to be a constant across the comparator treatments at a relative risk of 1.47.

Results

The main base-case results from the manufacturer are summarised in *Table 32*. From a NHS perspective, the base-case cost per QALY gained was £26,597 for golimumab compared with CC over a 20-year horizon. Both etanercept and adalimumab were reported to be extendedly dominated by golimumab.

The ERG undertook a limited validation of the model and reported various errors which were corrected. However, they concluded that questions remained concerning the integrity of the manufacturer model. The ERG subsequently presented results based on an exploratory reanalysis of the manufacturer's submission, using results from a separate MTC analysis and employing a lifetime horizon. The results of the ERG reanalysis are reported in in *Table 33*. The results of this re-analysis resulted in golimumab being extendedly dominated by the other two anti-TNFs.

There is no discussion by Armstrong *et al.*[163] of the appropriateness of the assumptions applied to BASFI progression, despite this being a critical structural assumption. However, a separate sensitivity analysis was presented in the full ERG report which uses the same rate of disease progression for BASFI (0.07 units per year) for all patients after 2 years. As part of this analysis, the ERG corrected errors identified in the way the BASFI regression equations were incorporated by the manufacturer.

TABLE 32 Manufacturer cost-effectiveness results: 20-year horizon[37]

Technology	Costs (£)	QALYs	Incremental costs	Incremental QALYs	ICER (£)
CC	88,667	6.6581	–	–	–
Adalimumab	93,601	6.8426	4934	0.1845	N/A (extendedly dominated)
Etanercept	93,782	6.8504	5115	0.1923	N/A (extendedly dominated)
Golimumab	93,786	6.8506	5119	0.1925	26,597

N/A, not applicable.

TABLE 33 Evidence Review Group exploratory cost-effectiveness results: lifetime horizon

Technology	Costs (£)	QALYs	Incremental costs	Incremental QALYs	ICER (£)
CC	95,227	7.8762	–	–	–
Golimumab	99,361	8.0296	4134	0.1534	N/A (extendedly dominated)
Adalimumab	108,295	8.3683	8934	0.3387	N/A (extendedly dominated)
Etanercept	108,347	8.3712	52	0.0029	26,505

N/A, not applicable.

Table 34 reports the ERG results based only on correcting the error identified and *Table 35* reports the results of also applying a common rate of disease progression for all patients after 2 years as well as correcting for the error. Golimumab was reported to be extendedly dominated by the other two anti-TNFs in both scenarios. It is also worth noting that the ICER for etanercept versus CC exceeded £30,000 per QALY in both scenarios.

Summary and critique of published cost-effectiveness studies

No previously published studies were identified which assessed the cost-effectiveness of anti-TNFs for nr-AxSpA. Consequently, the de novo submissions provided by the manufacturers provide the only existing evidence which can be considered to inform decisions for the NHS. Of the previously published UK cost-effectiveness study identified, there appear marked differences between the results of the industry-funded assessments and the results from the independent assessment by LRiG. Importantly, the results of the independent critique and exploratory reanalysis by the ERG for TA233[33] also appear potentially less favourable than the industry-funded published assessments. Although the DSU review of models submitted as part of TA143[17] has reconciled many of the key differences and highlighted the key assumptions, a number of key

TABLE 34 Evidence Review Group exploratory cost-effectiveness results: correction for BASFI error (from NICE TA233[33])

Technology	Costs (£)	QALYs	Incremental costs	Incremental QALYs	ICER (£)
CC	77,505	6.7336	–	–	–
Golimumab	81,849	6.8746	4334	0.1410	N/A (extendedly dominated)
Adalimumab	91,340	7.1703	9491	0.2937	N/A (extendedly dominated)
Etanercept	91,408	7.1734	68	0.0031	31,612

N/A, not applicable.

TABLE 35 Evidence Review Group exploratory cost-effectiveness results: correction for BASFI error and common BASFI progression after 2 years (from NICE TA233[33])

Technology	Costs (£)	QALYs	Incremental costs	Incremental QALYs	ICER (£)
CC	74,980	6.8267	–	–	–
Golimumab	79,330	6.9675	4350	0.1408	N/A (extendedly dominated)
Adalimumab	88,994	7.2567	9664	0.2892	N/A (extendedly dominated)
Etanercept	89,055	7.2600	61	0.0033	32,483

N/A, not applicable.

uncertainties remain. The remainder of this section provides an overview of the issues and uncertainties identified based on existing published studies and the DSU reports. This summary provides an important basis for considering the extent to which the de novo submissions provided by the manufacturers for this appraisal have adequately addressed these.

All existing models are based on similar two part structures:

- initial-response period (short-term model used to determine initial response rate)
- post-response period (longer-term model used to characterise natural history of disease (i.e. without anti-TNFs) and impact of anti-TNFs (while on therapy and when therapy is stopped).

All models use changes in BASDAI and/or BASFI to quantitatively model the short- and longer-term costs and quality-of-life implications (using QALYs) of the use of anti-TNFs versus CC alone.

Although there are differences between the modelling of the initial response period, existing models are broadly comparable being based on an assessment around 12 weeks (and potentially at 24 weeks as well) using a particular variant of existing BSR guidelines. Patients receiving anti-TNFs who meet the response criteria at the 12-/24-week assessment are continued on anti-TNFs. Anti-TNFs are withdrawn in non-responders at the 12-/24-week assessment point and patients subsequently receive CC alone.

However, there are marked differences between existing studies in relation to the modelling of the post-response period and the assumptions used. This period is often separated into different time intervals allowing different assumptions to be made regarding the effect of anti-TNFs (i.e. initially improving with time in responders but then later 'levelling off' or even deteriorating over a longer-term time horizon relative to CC). An important difference between existing models is the timing of this 'levelling off' period and assumptions employed over a longer time horizon. The differences in approaches and the timing of this 'flattening off' period are also closely linked to the data used, that is whether or not the changes in BASDAI/BASFI used in the model are restricted to the 12- to 24-week data from RCT evidence reported during the double-blind phase (Kobelt *et al.* 2007[160] and Ara *et al.* 2007[161]) or also incorporate longer-term data from the open-label extensions. Studies which use change in BASDAI/BASFI data directly in the model, from the double-blind phase, appear to use shorter 'levelling off' periods than studies using data from the open-label extension phase (Botteman *et al.* 2007,[162] McLeod *et al.* 2007[38] and Armstrong *et al.* 2007[163]).

Those studies incorporating an open-label extension typically assume continuing changes in BASDAI/BASFI for responders to anti-TNFs versus non-responders/CC beyond the initial 12/24-week period. Importantly, none of the studies using open-label extension data appear to provide any discussion of the potential for selection bias (e.g. related to the initial consent for patients to participate and/or agree to switch treatments as well as ongoing selection issues concerning retention over a longer period) and how these should be considered and/or adjusted for in the economic model. However, the implication of this is important, as the assumption being made by several models appears to incorporate an assumption of an increasing effect of anti-TNFs in responders over time (i.e. in terms of continuing improvements in BASDAI/BASFI), which does not appear to be adequately justified or related to any underlying clinical/pharmacological mechanism. In the absence of the counter-factual (i.e. comparable data in patients who did not participate or were subsequently withdrawn from the open-label study) it is unclear whether the apparent increasing effect is simply a function of the selection issue or is a real effect of the anti-TNFs. Importantly, those studies which only use data from the double-blind periods of RCTs often cite the open-label data as providing supportive evidence regarding the maintenance of the effects observed at 12/24 weeks but do not use it to support an assumption of an increasing effect over time.

The longer-term impact on costs and utilities beyond the initial response period are subsequently quantified by estimating separate BASDAI/BASFI 'trajectories' for different patient categories. The three main categories are:

1. CC
2. non-responder to anti-TNFs at 12-/24-week assessment
3. initial responder to anti-TNFs at 12-/24-week assessment.

The 'trajectory' for patients who are responders to anti-TNFs at the initial 12-/24-week assessment are further separated into (1) the period up to the point that anti-TNFs are subsequently withdrawn (i.e. because of loss of efficacy or AEs) and (2) the period post TNF-α inhibitor withdrawal.

After the 'levelling off' period for BASDAI, the majority of existing studies assume BASDAI is constant over the longer term, that is the BASDAI of responders to anti-TNFs is assumed to be lower than the equivalent BASDAI value (lower disease activity) applied to CC/non-responders and a constant difference is assumed to be retained until patients discontinue. At the point of discontinuation of anti-TNFs, patients subsequently revert back to the same value assumed for CC/placebo and non-responders to anti-TNFs at 12/24 weeks. Hence, any improvement in BASDAI is assumed to dissipate immediately or within a short period (3–6 months) after discontinuation of anti-TNFs.

All existing studies model BASFI as a linearly increasing function over the longer term for non-responders/CC, that is a constant rate of change is subsequently applied which is used to characterise the impact of disease progression on functional ability, typically a worsening of 0.07 (0–10 scale) units per annum. Again, the same assumptions applied to BASDAI for non-responders to anti-TNFs are applied to BASFI, that is beyond 12/24 weeks non-responders are assumed to follow an identical BASFI 'trajectory' as that of CC/placebo patients. By contrast, patients who respond to anti-TNFs are typically assumed not to 'progress' further in terms of functional disability, or progress at a lower rate than CC patients, while continuing to receive anti-TNFs. Hence the difference in individual mean BASFI scores increases over time in existing economic models between patients who continue to receive anti-TNFs and non-responders/CC.

The only study which employs a markedly different approach to the modelling of BASDAI and BASFI for responders is the study undertaken by the previous independent assessment group (LRiG) for TA143.[17] Instead, LRiG applied a quadratic function to the BASDAI and BASFI scores of responders. This approach assumed that the difference compared with CC was decreasing (initially) with time, that is, over time, the differences in BASDAI/BASFI would slowly reduce in responders and eventually be the same as for CC. While the logical problems of applying a quadratic function over a longer period were recognised by the authors (i.e. function begins to increase after a particular period) and was addressed using a series of logical restrictions (i.e. BASDAI/BASFI score constrained to be the same or better than CC), the clinical 'face' validity of this approach also appears questionable in the context of longer-term projections which are required for appropriate assessments of cost-effectiveness.

Another key difference between existing studies relates to the assumptions made concerning the subsequent trajectory of BASFI for patients who withdraw from active treatment. Given that BASFI is linearly increasing with time for CC, the assumption of the subsequent BASFI trajectory is potentially an important driver of cost-effectiveness. This is often referred to as 'rebound'. Typically, two scenarios are used:

- Rebound equal to gain: when patients fail therapy (after initially responding), their BASFI deteriorates by the same amount by which it improves when they responded to therapy.
- Rebound back to natural history/CC: when patients fail therapy (after initially responding), their BASFI deteriorates to the level and subsequent trajectory it would have been had they not initially responded to therapy.

In the absence of evidence on the magnitude of any rebound, these alternative scenarios represent the 'best-case' and 'worst-case' scenarios possible. In other words, the reality regarding rebound is likely to be somewhere between these two scenarios which should, therefore, be seen as the limits.

The implications of the different rebound scenarios are clearly illustrated in *Figures 2* and *3*. Studies which are based on assumptions of rebound equal to gain incorporate an ongoing benefit of anti-TNFs in patients in whom therapy is subsequently withdrawn after an initial response. Hence, such an assumption is more optimistic than assuming no continuing benefit at the point treatment is withdrawn.

Although the impact of discontinuation in patients who initially respond is clearly an important issue, the assumptions underpinning the subsequent trajectories of patients who are non-responders at 12/24 weeks to anti-TNFs are rarely explicitly justified. The most common assumption applied is that non-responders during the initial period follow the same subsequent trajectory for BASDAI/BASFI as CC/placebo patients beyond the 12-/24-week assessment point. However, the appropriateness of this assumption does not appear to have been discussed in existing studies. Essentially, for this assumption to hold, the initial response to anti-TNFs has to be independent of baseline patient characteristics, such that response to treatment is effectively a random process. However, if response is not independent of patient characteristics, the implication is that responders/non-responders to TNF-α inhibitors may be systematically different from each other. This has implications for the appropriateness of current assumptions being applied to non-responders at 12/24 weeks and subsequent responders who later withdraw.

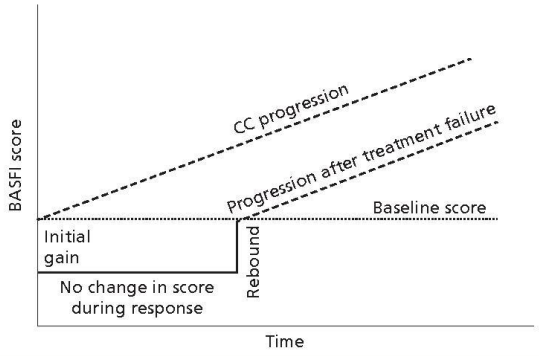


FIGURE 2 Illustration of the scenario of rebound equal to gain.

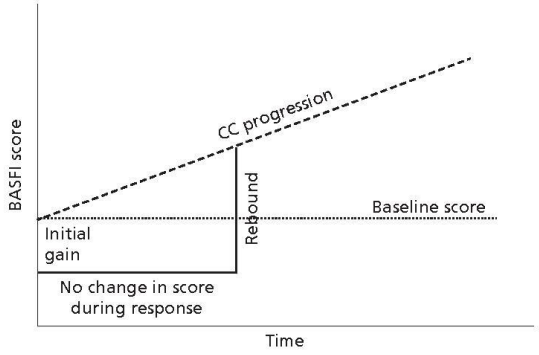


FIGURE 3 Illustration of the scenario of rebound to CC.

For example, all other things being equal, if patients with more severe disease (high BASDAI/high BASFI) were more likely not to respond, then assuming that the non-responders at 12/24 weeks follow the same trajectory as the 'average' CC/placebo patient is likely to be optimistic towards the anti-TNFs (and vice versa for if less severe patients are more likely to respond). Hence, rather than following the trajectory of an 'average' placebo/CC patient, a non-responder may actually follow a different trajectory, that is that of an equivalent more/less severe CC patient. Inevitably, the impact of different patient characteristics is likely to be more complex than the simplistic scenarios presented above.

As previously noted, all models use changes in BASDAI and/or BASFI to quantitatively model the short- and longer-term costs and quality-of-life implications (using QALYs) of the use of anti-TNFs versus CC. The justification for using these measures appears largely driven by the existence of external sources of costs and health utility estimates which can be directly linked to these measures and not to others (e.g. BASMI, ASDAS, mSASSS, etc.). Hence current models appear more of a function of the data which are available to link to costs and utilities rather than being based on a clear underlying biological or clinical process. This raises more general conceptual concerns regarding existing models and also regarding the generalisability of findings in an AS population to the separate nr-AxSpA population.

The use of BASDAI/BASFI per se is perhaps not the most significant issue, as in the absence of alternative mapping functions to costs and/or utilities it is unclear how to estimate longer-term costs and QALYs without ultimately linking to these measures. However, it is concerning that the majority of existing studies do not appear to link the data and assumptions applied to these measures to any coherent clinical underpinning regarding differences between population characteristics and the effect of anti-TNFs. Consequently, 'progression' over time is currently modelled entirely via changes in BASFI, as BASDAI is assumed to remain constant. However, no attempt is made to justify why BASFI increases, the rate at which it increases and how this rate might differ across different groups as well as the impact that anti-TNFs might have (i.e. any effect on BASFI which may be independent of the effect on BASDAI).

Modelling 'progression' implicitly (i.e. employing natural history estimates of the rate of change of BASFI from external studies) rather than explicitly (i.e. attempting to explain how BASFI evolves over time in relation to inflammatory and other processes and how these may differ within populations and across the AS and nr-AxSpA groups) has led to a series of implicit/evidence-free assumptions. These include:

- No change in BASFI while receiving anti-TNFs (i.e. assuming implicitly that these act as disease modifiers and that while patients respond and continue to receive them, further deterioration in functional progression is completely prevented).
- Lower BASFI changes while receiving anti-TNFs (i.e. assuming that anti-TNFs do not completely halt further deterioration in functional progression but that the rate of progression is reduced relative to progression on CC).
- Similar natural history rates of change in BASFI across different subgroups and populations (i.e. assuming that rate of change in BASFI is independent of time and/or patient characteristics).

Similar conceptual concerns were also highlighted by the NICE DSU in their work to support TA143, noting that in inflammatory arthritis a clearer conceptual relationship is assumed between disease activity, radiographic progression and physical functioning, such that changes in physical functioning can be more clearly related to different processes and evidence for the anti-TNFs on each separate process. In highlighting these issues, the DSU cited emerging longer-term data reported for anti-TNFs based on measures of radiographic progression (mSASSS) in AS. Although this evidence was not formally included in their analyses, the evidence was cited to indicate that an assumption of no further progression while on anti-TNFs for AS was potentially optimistic based on emerging longer-term radiographic progression data.

Importantly, the only UK study published since the NICE DSU review did subsequently use a less favourable assumption concerning the impact of anti-TNFs on functional progression (BASFI). The assumption used by the manufacturer for golimumab[37] assumed that the longer-term rate of change in BASFI for responders

who continued on treatment would be 50% of that assumed for CC/non-responders. Although this assumption is a significant departure from the base-case assumptions applied within previous industry-funded studies, no justification appeared to be identified to support this by Armstrong *et al.*[163] in the review of the manufacturer's submission.

In summary, there appear to be significant differences between the cost-effectiveness results reported in existing UK published studies. Many of these differences appear largely because of differences in data sources (i.e. double-blind period vs. open-label extensions), subsequent assumptions and estimates related to the magnitude and duration of the differences in BASDAI and BASFI measurements between responders and non-responders in the short to medium term (i.e. the 'levelling off' period) and then longer term in relation to assumptions concerning BASFI progression, and issues around 'placebo' effect and the withdrawal of anti-TNFs. Some of the main differences between existing studies have been highlighted in a separate review by the NICE DSU. However, while this review is helpful in identifying the impact of parameter and structural assumptions, it does not provide a basis for informing which assumptions appear most justified based on existing data and clinical understanding of the progression of AS and the impact of anti-TNFs. It is also concerning that many of the existing studies are based on CiC data and hence lack transparency regarding specific inputs and assumptions.

To date, only two UK studies have attempted to assess the cost-effectiveness of the alternative anti-TNFs. One of these studies, McLeod *et al.*,[38] assumed that the alternative treatments were identical in terms of clinical effectiveness and hence only considered differences in the acquisition, administration and monitoring costs. The justification provided by the authors was based on the lack of statistically significant differences across key outcome measures based on indirect comparisons. The other study, Armstrong *et al.*,[163] assumed differences in the clinical effectiveness of the alternative anti-TNFs based on a separate MTC. However, differences between the anti-TNFs appeared sensitive to the studies included and the specific outcomes considered. Hence different conclusions could be drawn concerning the most 'efficient' intervention depending on the analysis considered. However, the magnitude of differences in clinical effect and QALYs remained small and the clinical and economic value of this might appear questionable.

There are conceptual concerns surrounding all existing models relating to the subsequent projection of BASDAI and BASFI over a longer time horizon which are required in order to generate more appropriate lifetime estimates of costs and QALYs required for cost-effectiveness assessments. The speculative nature of these projections was highlighted as a significant concern by the previous independent assessment group (LRiG) and hence their longer-term results were presented as exploratory scenarios. However, it appears that all existing models are largely based on implicit approaches and assumptions, and lack a clearer conceptual basis which might help to more appropriately inform parameter estimates and structural assumptions, and facilitate a more evidence-based assessment of the potential longer-term impact of anti-TNFs.

The following sections present a summary of the de novo submissions provided by the manufacturers[34–37] for the separate AS and nr-AxSpA indications. Brief overviews of the manufacturers' submissions for AS and nr-AxSpA are provided alongside a summary of the base-case cost-effectiveness results. This is followed by a more in-depth comparison of key parameter and structural assumptions across the manufacturers and the separate indications. The issues and concerns regarding existing published studies are used as the basis for a more critical assessment of these submissions and the extent to which these concerns have been adequately addressed and key uncertainties which still remain have been highlighted is investigated.

It should be noted that although fully incremental results were routinely presented by each manufacturer, there were differences between manufacturers in terms of how the results were presented and also whether or not the correct calculations based on dominance and extended dominance were included. Consequently, the fully incremental ICER tables reported are based on our own calculations to ensure accuracy and consistency between the various manufacturer results tables.

Summary of manufacturers' de novo submissions

Manufacturers submitted de novo analyses for both AS (AbbVie,[34] UCB,[35] Pfizer[36] and Merck Sharp & Dohme[37]) and nr-AxSpA (AbbVie,[34] UCB[35] and Pfizer[36]) populations.

Overview of AbbVie (adalimumab) model

The economic model presented by AbbVie[34] compared the cost-effectiveness of adalimumab versus conventional therapy and other licensed anti-TNFs for nr-AxSpA and AS. Separate state-transition models were developed for the two indications separately based on the ASAS guidelines for the use of anti-TNFs. All patients were assumed to take conventional therapy/background therapy (e.g. NSAIDs) during the modelled horizon and also receive one of the licensed anti-TNFs or placebo (conventional therapy only). Specifically, patients were assumed to stay on therapy as long as they had an adequate therapeutic response (i.e. ASAS 40 for nr-AxSpA and ASAS 20 for AS) and patients were assumed to discontinue therapy when insufficient response occurred. Discontinuations due to AEs or reasons other than therapeutic failures were also included.

The model consists of a short-term component (first 12 weeks) and a longer-term component to estimate lifetime cost-effectiveness (40 years). In common with previously published models, the model was based on the estimation of BASDAI and BASFI scores over time. The model used the available long-term open-label extension data of trials of adalimumab (up to 156 weeks in ABILITY-1[58] for nr-AxSpA and 260 weeks in ATLAS[61] for AS, *Figures 4* and *5*) as well as including assumptions beyond these study durations to inform

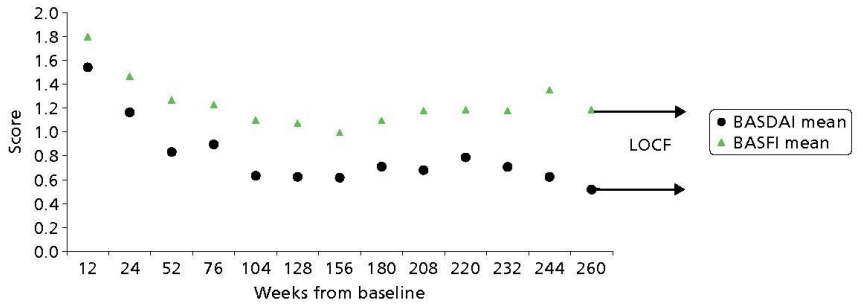


FIGURE 4 Observed mean BASDAI and BASFI scores for adalimumab ASAS 20 responders in the licensed population from ATLAS[61] (AS).

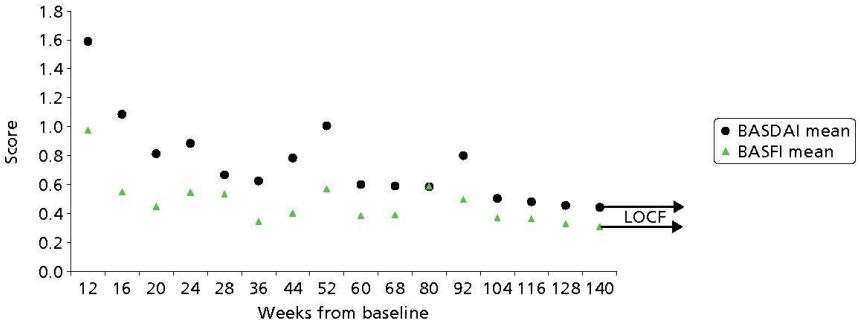


FIGURE 5 Observed mean BASDAI and BASFI scores for adalimumab ASAS 40 responders in the licensed population from ABILITY-1[58] (nr-AxSpA).

the life-time cost-effectiveness results. To avoid extrapolating life-time improvement by applying a functional form to the BASDAI/BASFI data, the manufacturer applied the mean observed BASDAI and BASFI scores until the last available data point and carried forward the last observed values to the end of horizon.

Response rates and other select treatment efficacy end points were based on a separate systematic review and network meta-analysis. In the base case, ASAS 40 for nr-AxSpA and ASAS 20 for AS were used to define clinical response at week 12, based on the primary outcome measures from the clinical trials of adalimumab. In the base-case analysis, placebo responders at week 12 were assumed to lose response and return to baseline disease severity. Patients who subsequently withdrew from TNF-α treatment at any time point were also assumed to return to baseline disease severity (rebound equal to gain). Longer-term discontinuation was assumed to be time-dependent and was based on a log-normal parametric distribution from the separate open-label RCTs adjusting for subsequent loss of response.

In the base-case model, the BASFI score for all patients not on TNF-α inhibitor treatment increases in a linear fashion by 0.084 (scale 0–10) per year in patients with nr-AxSpA, in line with the evidence from the ABILITY-1 trial,[58] in which each additional year of baseline symptom duration was reported to be associated with a significant (+0.084; $p = 0.0005$) increase in baseline BASFI score, adjusting for the age of onset (age at first reported axial SpA symptom) to control for the age effect on functional damage. An estimate of +0.056 was applied to patients with AS based on applying a similar approach to the ATLAS trial,[61] adjusting for age at disease diagnosis. Hence, a higher BASFI progression was applied to patients not on anti-TNFs in the nr-AxSpA population compared with the AS population.

The BASDAI and BASFI scores were used jointly to estimate quality of life associated with AS, using the relationship observed between the utility scores (measured in HUI-3) and the BASDAI and BASFI scores in the ATLAS trial.[61] Observed EQ-5D scores were mapped to BASDAI and BASFI for the relationship in the base case for nr-AxSpA from ABILITY-1.[58]

The relationship between BASDAI and costs, derived from a reanalysis of the OASIS data, was applied in the base case. Costs of drug, administration, initiation and monitoring, and AEs were also included. Discounting was applied at 3.5% for both costs and outcomes. SMRs of 1 and 1.5 were assumed for nr-AxSpA and AS, respectively. Uncertainty surrounding results was addressed using probabilistic sensitivity analyses (PSAs).

Base-case results from AbbVie (adalimumab) model

The main base-case ICER results from the manufacturer are summarised in *Table 36* for the AS population. From a NHS perspective, the base-case cost per QALY gained versus CC ranged from £16,391 per QALY (adalimumab) and £44,448 per QALY (infliximab).

TABLE 36 Tumour necrosis factor-α inhibitors compared with CC for AS: AbbVie[34] (base case)

Technology	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (£)
CC	112,762	8.62	–	–	–
Adalimumab	139,860	10.28	27,098	1.65	16,391
Certolizumab	133,273	9.82	20,511	1.20	17,067
Etanercept	139,574	10.21	26,812	1.59	16,897
Golimumab	138,385	10.17	25,624	1.55	16,535
Infliximab	197,100	10.52	84,339	1.90	44,448

Table 37 reports the results based on the fully incremental analysis. In the manufacturer base-case analysis, certolizumab and etanercept were ruled out by extended dominance. The ICER of adalimumab was £16,391 per QALY compared with CC. The ICER of the next more costly (and non-dominated) TNF-α inhibitor was £238,500 per QALY for the comparison between infliximab and adalimumab.

The main base-case ICER results from the manufacturer and fully incremental analysis are summarised in *Tables 38* and *39* for the nr-AxSpA population. The ICERs versus CC ranged from £12,866 (certolizumab) to £13,288 per QALY (adalimumab). In the fully incremental comparison, adalimumab was extendedly dominated and hence the ICER for certolizumab versus CC is the only ICER reported (£12,866).

The manufacturer reported more favourable ICERs versus CC in the nr-AxSpA population compared with the AS population. This appears largely driven by two inputs: (1) the lower BASDAI/BASFI scores assumed for responders based on ABILITY-1[58] (compared with ATLAS[61]) and (2) the higher annual BASFI progression rate assumed for non-responders/CC in the nr-AxSpA population (0.084 vs. 0.056).

TABLE 37 Fully incremental comparison of anti-TNFs for AS: assessment group analysis based on AbbVie[34] (base case)

Technology	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (£)
CC	112,762	8.62	–	–	–
Certolizumab	133,273	9.82	–	–	Extendedly dominated
Golimumab	138,385	10.17	–	–	Extendedly dominated
Etanercept	139,574	10.21	–	–	Extendedly dominated
Adalimumab	139,860	10.28	27,098	1.66	16,391
Infliximab	197,100	10.52	57,240	0.24	238,500

TABLE 38 Anti-TNFs compared with CC for nr-AxSpA: AbbVie[34] (base case)

Technology	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (£)
CC	126,075	8.88	–	–	–
Adalimumab	142,218	10.10	16,143	1.22	13,228
Certolizumab	142,608	10.16	16,532	1.28	12,866
Etanercept	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed

TABLE 39 Fully incremental comparison of anti-TNFs for nr-AxSpA: assessment group analysis based on AbbVie[34] (base case)

Technology	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (£)
CC	126,075	8.88	–	–	–
Adalimumab	142,218	10.10	–	–	Extendedly dominated
Certolizumab	142,608	10.16	390	0.06	12,866
Etanercept	Not assessed	Not assessed	Not assessed	Not assessed	Not assessed

Overview of UCB (certolizumab) model

The economic model presented by UCB[34] compared the cost-effectiveness of certolizumab with conventional therapy and other licensed anti-TNFs for nr-AxSpA and AS. Separate Markov cohort models were developed for the two indications separately based on the subpopulations of the RAPID-axSpA trial.[64] Separate analyses were argued to be necessary given that the comparators differed for each subpopulation. Analyses performed for the AS subpopulation consisted of all patients with AS from the RAPID-axSpA study,[64] including those who were anti-TNF therapy experienced or therapy naive. The nr-AxSpA subpopulation consisted of anti-TNF therapy-naive patients only, as there were no anti-TNF therapy-experienced patients in this subpopulation.

The analyses used a lifetime time horizon in the base case. An alternative time horizon of 20 years was tested in a scenario analysis. A NHS and PSS perspective was used and an annual discount rate of 3.5% was applied to costs and outcomes. All costs are reported at 2013 values.

The model consists of a short-term component and a longer-term component to estimate lifetime cost-effectiveness. The duration of the short-term component varied between the models used for the AS and the nr-AxSpA subpopulations based on the response end point assumed. Response was assessed at 24 weeks in the AS subpopulation which was argued by the manufacturer to be in accordance with clinical practice as indicated key British opinion leaders. For the nr-AxSpA subpopulation, response assessment was assumed at 12 weeks, as comparator data were available only at that time point. In their base case, the manufacturer used ASAS 20 to determine response in line with the primary outcome measure in the RAPID-axSpA.[64] However, it should be noted that ASAS 20 response at week 12 was the primary outcome in the RAPID-axSpA trial. Hence, although the measure of response used is in accordance with the primary outcome of the RAPID-axSpA trial, the differential timing of this applied across the separate populations clearly deviates from this. This has potential issues because at week 16 patients were allowed an 'early escape' from placebo and hence results at week 24 used for the AS subpopulation are no longer based on the original randomised population.

Assessment in Ankylosing Spondylitis 20 response rates for certolizumab and relative treatment effects for the other anti-TNFs were derived based on a separate systematic review and MTC. The base-case model inputs applied in the manufacturer's submission[34] are replicated (and associated footnotes) in *Tables 40* and *41*.

TABLE 40 Base-case model inputs: ASAS 20 response at week 24 (AS subpopulation, certolizumab pegol-pooled dosing)[34]

Treatment	ASAS 20 response (%)[a]	SE	Relative risk[b]	CI	Source
Certolizumab pegol	CiC information has been removed	CiC information has been removed	–	–	MTC
Adalimumab[b]	–	–	AiC information has been removed	AiC information has been removed	MTC
Etanercept[b]	–	–	AiC information has been removed	AiC information has been removed	MTC
Golimumab[b]	–	–	AiC information has been removed	AiC information has been removed	MTC
Infliximab[b]	–	–	AiC information has been removed	AiC information has been removed	MTC

AiC, academic in confidence.
a Proportion responding.
b Certolizumab pegol vs. comparator.
All footnotes supplied by the manufacturers are reported in their entirety for further clarification, although supporting references have been removed here. Based on pooled certolizumab pegol 200 mg Q2W and 400 mg Q4W arms from RAPID-axSpA.[64]

TABLE 41 Base-case model inputs: ASAS 20 response at week 12 (nr-AxSpA subpopulation, certolizumab pegol-pooled dosing)[34]

Treatment[a]	ASAS 20 response (%)[b]	SE	Relative risk[a]	CI	Source
Certolizumab pegol	AiC information has been removed	AiC information has been removed	–	–	MTC
Adalimumab[a]	–	–	AiC information has been removed	AiC information has been removed	MTC
Etanercept[a]	–	–	AiC information has been removed	AiC information has been removed	MTC

AiC, academic in confidence.
a Certolizumab pegol vs. comparator.
b Proportion responding.
All footnotes supplied by the manufacturers are reported in their entirety for further clarification, although supporting references have been removed here. Based on pooled certolizumab pegol 200 mg Q2W and 400 mg Q4W arms from RAPID-axSpA.[64]

The MTC was also used to determine change in baseline BASFI and BASDAI scores. The base-case inputs for change from baseline in BASFI and BASDAI at week 24 for the AS subpopulation reported by the manufacturer are replicated in *Tables 42* and *43*. The manufacturer noted that the mean change from baseline reported in the tables is that observed per trial arm, which includes both the ASAS 20 responders and non-responders in each arm. In order to determine the change in BASFI and BASDAI for responders alone, the manufacturer used the equation:

Mean change in BASFI score = (change in BASFI among ASAS 20 responders × proportion ASAS 20 responders) + (change in BASFI among ASAS 20 non-responders × proportion ASAS 20 non-responders). (1)

TABLE 42 Base-case model inputs: change from baseline in BASFI score at week 24 (AS subpopulation, certolizumab pegol-pooled dosing)[34]

	Change from baseline in BASFI score at Week 24: initial response assessment period		
Treatment	Mean	SD	Source
Certolizumab pegol	AiC information has been removed	AiC information has been removed	MTC
Adalimumab	AiC information has been removed	AiC information has been removed	MTC
Etanercept	AiC information has been removed	AiC information has been removed	MTC
Golimumab[a]	AiC information has been removed	AiC information has been removed	MTC[a]
Infliximab	AiC information has been removed	AiC information has been removed	MTC
CC[b]	AiC information has been removed	AiC information has been removed	Assumed zero in base case[b]

AiC, academic in confidence.
a Golimumab assumed same values as adalimumab, given that specific input values for BASFI at 24 weeks were not available from MTC.
b CC assumed to produce no change in BASFI score initially in the base case. As noted in main text, it is reasonable to assume patients receiving CC do not achieve a change from baseline (worsening or improvement) in BASDAI or BASFI as evidence from RAPID-axSpA,[64] ATLAS[61] and ABILITY-1[58] demonstrate that in the placebo arms of these studies where patients were essentially maintained on CC, patients did not achieve MCID for BASDAI or BASFI.[58,61,64] Furthermore, Dougados *et al.*[43] describe CC regimens as 'palliative at best, providing no alteration of the disease process'. This assumption is consistent with previous manufacturers' submissions to NICE in AS. However, a mean change in BASFI of (AiC information has been removed) estimated from the MTC was used in the sensitivity analysis.
All footnotes supplied by the manufacturers are reported in their entirety for further clarification, although supporting references have been removed here. Based on pooled certolizumab pegol 200 mg Q2W and 400 mg Q4W arms from RAPID-axSpA.[64]

TABLE 43 Base-case model inputs: change from baseline in BASDAI score at week 24 (AS subpopulation, certolizumab pegol-pooled dosing)[34]

	Change from baseline in BASDAI Score at week 24: initial response assessment period		
Treatment	Mean	SD	Source
Certolizumab pegol	AiC information has been removed	AiC information has been removed	MTC
Adalimumab	AiC information has been removed	AiC information has been removed	MTC
Etanercept	AiC information has been removed	AiC information has been removed	MTC
Golimumab	AiC information has been removed	AiC information has been removed	MTC
Infliximab	AiC information has been removed	AiC information has been removed	MTC
CC[a]	AiC information has been removed	AiC information has been removed	Assumed zero in base case[a]

AiC, academic in confidence; MTC, mixed-treatment comparison.
a CC assumed to produce no change in BASDAI score initially in the base case. As noted in main text, it is reasonable to assume patients receiving CC do not achieve a change from baseline (worsening or improvement) in BASDAI or BASFI as evidence from RAPID-axSpA,[64] ATLAS[61] and ABILITY-1[58] demonstrate that in the placebo arms of these studies where patients were essentially maintained on CC, patients did not achieve MCID for BASDAI or BASFI. Furthermore, Dougados *et al.*[43] describe CC regimens as 'palliative at best, providing no alteration of the disease process'. This assumption is consistent with previous manufacturers' submissions to NICE in AS.[33] However, a mean change in BASDAI of (AiC information has been removed) estimated from the MTC was used in the sensitivity analysis.
All footnotes supplied by the manufacturers are reported in their entirety for further clarification, although supporting references have been removed here. Based on pooled certolizumab pegol 200 mg Q2 W and 400 mg Q4 W arms from RAPID-axSpA.[64]

This approach assumed that the change in BASFI (and BASDAI) score among ASAS 20 non-responders is equal to that of the CC arm. Thus, the equation is used to algebraically solve for change in BASFI (and BASDAI) score among ASAS 20 responders. The manufacturer stated that:

> *As an example for the AS subpopulation base case, the change in BASFI among ASAS 20 responders for CZP [certolizumab pegol] is: [academic-in-confidence (AiC) information removed]. Thus, in this example, the actual change from baseline in AS responders to CZP is [AiC information removed]. The same approach was used for change from baseline for BASDAI. This approach, where the change from baseline for BASDAI and BASFI is calculated among responders only, is consistent with previous evaluations pharmacoeconomic evaluations conducted for AS*
>
> *Manufacturer's submission, pp. 69–70*[34]

The manufacturer base-case inputs for change from baseline in BASFI and BASDAI at week 12 for the nr-AxSpA subpopulation are replicated in *Tables 44* and *45*.

The manufacturer's submission[34] assumed no change in BASDAI and BASFI for CC during the response period. The manufacturer justified this assumption with reference to evidence from RAPID-axSpA,[64] ATLAS[61] and ABILITY-1[58] studies, although no specific data were reported to support this.

These change scores are assumed to be maintained for BASDAI as long as a patient continues to receive an anti-TNF. For AS patients on CC, an additional annual increase of 0.07 points (scale 0–10) in BASFI is assumed and justified by the manufacturer according to the assumptions deemed reasonable by a previous NICE committee. Hence, while the change scores are assumed constant, the absolute difference between patients receiving anti-TNFs and CC is increasing over time given the underlying progression assumed for BASFI for patients receiving CC. The assumption of no progression in BASFI for patients receiving anti-TNFs is not explicitly discussed within the manufacturer's submission, neither are separate results provided for alternative assumptions.

TABLE 44 Base-case model inputs: change from baseline in BASFI score at week 12 (nr-AxSpA subpopulation, certolizumab pegol-pooled dosing)[34]

	Change from baseline in BASFI score at week 12: initial response assessment period		
Treatment	Mean	SD	Source
Certolizumab pegol	AiC information has been removed	AiC information has been removed	MTC
Adalimumab	AiC information has been removed	AiC information has been removed	MTC
Etanercept	AiC information has been removed	AiC information has been removed	MTC
CC[a]	AiC information has been removed	AiC information has been removed	Assumed zero in base case[a]

a CC assumed to produce no change in BASFI score initially in the base case. As noted in main text, it is reasonable to assume patients receiving CC do not achieve a change from baseline (worsening or improvement) in BASDAI or BASFI as evidence from RAPID-axSpA,[64] ATLAS[61] and ABILITY-1[58] demonstrate that in the placebo arms of these studies in which patients were essentially maintained on CC, patients did not achieve MCID for BASDAI or BASFI. Furthermore, Dougados *et al.*[43] describe CC regimens as 'palliative at best, providing no alteration of the disease process'. This assumption is consistent with previous manufacturers' submissions to NICE in AS. However, a mean change in BASFI of (AiC information has been removed) estimated from the MTC was used in the sensitivity analysis.

All footnotes supplied by the manufacturers are reported in their entirety for further clarification, although supporting references have been removed here. Based on pooled certolizumab pegol 200 mg Q2 W and 400 mg Q4 W arms from RAPID-axSpA.[64]

TABLE 45 Base-case model inputs: change from baseline in BASDAI score at week 12 (nr-AxSpA subpopulation, certolizumab pegol-pooled dosing)[34]

	Change from baseline in BASDAI score at week 12: initial response assessment period		
Treatment	Mean	SD	Source
Certolizumab	AiC information has been removed	AiC information has been removed	MTC
Adalimumab	AiC information has been removed	AiC information has been removed	MTC
Etanercept	AiC information has been removed	AiC information has been removed	MTC
CC[a]	AiC information has been removed	AiC information has been removed	Assumed zero in base case[b]

a CC assumed to produce no change in BASDAI score initially in the base case. As noted in main text, it is reasonable to assume patients receiving CC do not achieve a change from baseline (worsening or improvement) in BASDAI or BASFI as evidence from RAPID-axSpA,[64] ATLAS[61] and ABILITY-1[58] demonstrate that in the placebo arms of these studies where patients were essentially maintained on CC, patients did not achieve MCID for BASDAI or BASFI. Furthermore, Dougados *et al.*[43] describe CC regimens as 'palliative at best, providing no alteration of the disease process'. This assumption is consistent with previous manufacturers' submissions to NICE in AS. However, a mean change in BASDAI of (AiC information has been removed) estimated from the MTC was used in the sensitivity analysis.

All footnotes supplied by the manufacturers are reported in their entirety for further clarification, although supporting references have been removed here. Based on pooled certolizumab pegol 200 mg Q2 W and 400 mg Q4 W arms from RAPID-axSpA.[64]

The same annual rate (0.07) in BASFI progression for CC is also applied to the nr-AxSpA subpopulation. In addition, it is assumed that some nr-AxSpA patients may progress to AS during their course of treatment. The manufacturer's model adopts an estimate for disease progression for the nr-AxSpA subpopulation based on a German cohort of axSpA patients, GESPIC. In this cohort, the rates and predictors of radiographic spinal progression over 2 years were estimated based on mSASSS. In total, 7.4% of the 95 nr-AxSpA patients were reported to show spinal radiographic progression, which was defined as a worsening of mSASSS by ≥ 2 units over 2 years. As this 7.4% progression represents a proportion, it was converted to a rate for use in the economic model, assuming an exponential distribution through the following formula:

$$1 - 0.074 = \exp(-\text{rate} \times 2 \text{ years}); \text{ rate} = 0.0384 \text{ or } 3.84 \text{ per } 100 \text{ patient-year.} \quad (2)$$

The manufacturer's submission[34] is not explicit about how this additional aspect of progression subsequently alters the BASDAI/BASFI trajectories within the nr-AxSpA model. However, examination of the electronic model submitted by the manufacturer reveals that once patients are assumed to show spinal radiographic progression, they effectively become AS patients by picking up the same absolute values of BASDAI and BASFI (on and off treatment) applied in their AS subpopulation model. The justification for this approach and the values subsequently assigned are not formally discussed by the manufacturer and the validity of the approach appears questionable (i.e. given other differences, e.g. disease duration, severity of radiographic disease etc., that may differ between the two populations even after radiographic progression has occurred in the nr-AxSpA subpopulation).

Patients who subsequently withdrew from TNF-α treatment at any time point were assumed to revert back to the same trajectory as CC over a 6-month period (i.e. rebound back to CC/natural history). A constant annual rate of discontinuation of 7% was assumed for all anti-TNFs over the longer-term period in both the AS and nr-AxSpA populations. The estimate of 7% applied to the AS subpopulation was justified by citing the rate apparently assumed by the NICE committee for TA143 and the lack of long-term evidence more generally. This estimate was referred to earlier in the review section of our report when the additional analyses undertaken by the NICE DSU were considered (see *Data extraction*). Identical assumptions for discontinuation rates were assumed for the nr-AxSpA subpopulations, although no justification was provided by the manufacturer.

The BASDAI and BASFI scores were used jointly to estimate quality of life in both subpopulations based on EQ-5D data collected in the RAPID-axSpA[64] study. Data from patients having EQ-5D, BASDAI and BASFI scores available at baseline and at weeks 12 and 24 were used to estimate a relationship between utility and the BASDAI and BASFI scores. Utilities were subsequently converted using a logistic transformation with the justification based on possible floor and ceiling effects, as they are bounded by 0 and 1. Without access to the original data, it is not possible to determine the impact of this transformation, although it should be noted that EQ-5D is not bounded by 0 (i.e. negative values are possible). The manufacturer used a repeated-measures logistic regression to model the relationship between utility and the BASDAI and BASFI scores.

The relationship between BASFI and costs, derived from the OASIS study and used by the previous independent assessment group in TA143,[17] was applied in the base case. Costs of drug, administration, initiation and monitoring were included. The costs and HRQoL of AEs were not included. Discounting was applied at 3.5% for both costs and outcomes. Uncertainty surrounding outcomes was addressed using PSA.

Base-case results from UCB (certolizumab) model

The main base-case ICER results from the manufacturer[35] are summarised in *Table 46* for the AS population, together with a fully incremental comparison of ICERs in *Table 47*. The ICERs versus CC ranged from £16,647 per QALY (certolizumab) and £42,671 per QALY (infliximab). In the fully incremental analysis, certolizumab dominated (i.e. less costly and more expensive) all other TNF-α treatments apart from infliximab. However, it should be noted that the costs of certolizumab are based on a patient access scheme (PAS) which has been proposed but is not yet formally agreed with the Department of Health and NICE. Results without the PAS were not reported by the manufacturer. UCB will make certolizumab pegol (Cimzia,® USB Pharma) available free of charge to all NHS patients for the first 3 months of therapy, at which point clinical response should be clear. Only after this 3-month stage will the NHS be charged for continuing to use this therapy.

The ICER of certolizumab was £16,647 per QALY versus CC and the ICER for infliximab was £113,871 (vs. certolizumab).

TABLE 46 Tumour necrosis factor-α inhibitors compared with CC for AS: UCB (base case)[35]

Technology	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (£)
CC	AiC information has been removed	AiC information has been removed	AiC information has been removed	AiC information has been removed	–
Adalimumab	AiC information has been removed	AiC information has been removed	AiC information has been removed	AiC information has been removed	19,932
Certolizumab	AiC information has been removed	AiC information has been removed	AiC information has been removed	AiC information has been removed	16,647
Etanercept	AiC information has been removed	AiC information has been removed	AiC information has been removed	AiC information has been removed	19,272
Golimumab	AiC information has been removed	AiC information has been removed	AiC information has been removed	AiC information has been removed	19,049
Infliximab	AiC information has been removed	AiC information has been removed	AiC information has been removed	AiC information has been removed	42,671

TABLE 47 Fully incremental comparison of anti-TNFs for AS: assessment group analysis based on UCB (base case)[35]

Technology	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (£)
CC	AiC information has been removed	AiC information has been removed	AiC information has been removed	AiC information has been removed	–
Certolizumab	AiC information has been removed	AiC information has been removed	AiC information has been removed	AiC information has been removed	16,647
Golimumab	AiC information has been removed	AiC information has been removed	AiC information has been removed	AiC information has been removed	Dominated
Adalimumab	AiC information has been removed	AiC information has been removed	AiC information has been removed	AiC information has been removed	Dominated
Etanercept	AiC information has been removed	AiC information has been removed	AiC information has been removed	AiC information has been removed	Dominated
Infliximab	AiC information has been removed	AiC information has been removed	AiC information has been removed	AiC information has been removed	113,871

The main base-case ICER results from the manufacturer[35] are summarised in *Table 48* for the nr-AxSpA population, together with a fully-incremental comparison of ICERs in *Table 49*. In contrast to the results for AS, there was a more marked difference between the ICERs of the alternative anti-TNFs and CC. The ICERs versus CC ranged from £15,615 (certolizumab) to £50,692 per QALY (etanercept). The higher differential ICERs appears to be largely a result of the more heterogeneous trials included in the MTC for the nr-AxSpA populations and a higher differential effect assumed for certolizumab vis-à-vis the other alternative anti-TNFs compared with the AS population. Importantly, other manufacturers (Pfizer) argue that the results for certolizumab in this population may be confounded by population characteristics which could invalidate the indirect comparison of certolizumab versus the other comparator treatments in the current nr-AxSpA MTC. In the fully incremental analysis, certolizumab dominated adalimumab and etanercept.

Overview of Pfizer (etanercept) model

The economic model submitted by Pfizer[36] compared the cost-effectiveness of etanercept versus conventional therapy and other licensed anti-TNFs for AS, nr-AxSpA and a combined population (axSpA). The results for the combined population are not summarised in this review but are reported separately in the manufacturer's submission. The model is based on a lifetime time horizon and costs and benefits are discounted at an annual rate of 3.5%. The reference year for costs was reported to be 2014.

The model was based on a patient-level simulation model based on a discrete event simulation (DES). The analysis was conducted from a NHS/PSS perspective. Data to populate the model were derived from key clinical trials for etanercept, the results of a clinical systematic review, MTC and, in a separate analysis presented for the nr-AxSpA population, a match-adjusted indirect comparison (MAIC). The model structure was reported to be developed in accordance with current OMERACT (Outcome Measures in Rheumatology) guidance and was constructed around the BASDAI and the BASFI in line with other published studies.

TABLE 48 Tumour necrosis factor-α inhibitors compared with CC for nr-AxSpA: UCB (base case)[35]

Technology	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (£)
CC	AiC information has been removed	AiC information has been removed	AiC information has been removed	AiC information has been removed	–
Adalimumab	AiC information has been removed	AiC information has been removed	AiC information has been removed	AiC information has been removed	30,370
Certolizumab	AiC information has been removed	AiC information has been removed	AiC information has been removed	AiC information has been removed	15,615
Etanercept	AiC information has been removed	AiC information has been removed	AiC information has been removed	AiC information has been removed	50,692

TABLE 49 Fully incremental comparison of anti-TNFs for nr-AxSpA: assessment group analysis based on UCB (base case)[35]

Technology	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (£)
CC	AiC information has been removed	AiC information has been removed	AiC information has been removed	AiC information has been removed	–
Certolizumab	AiC information has been removed	AiC information has been removed	AiC information has been removed	AiC information has been removed	15,615
Etanercept	AiC information has been removed	AiC information has been removed	AiC information has been removed	AiC information has been removed	Dominated
Adalimumab	AiC information has been removed	AiC information has been removed	AiC information has been removed	AiC information has been removed	Dominated

The AS population was defined based on current NICE guidance in TA143[17] and TA233.[33] The nr-AxSpA population was defined based on the scope issued by NICE and was defined by the manufacturer as people with severe axSpA without radiographic evidence of AS but with objective signs of inflammation, whose disease has responded inadequately to, or who are intolerant to, NSAIDs.

An important aspect of the submission for the nr-AxSpA population was an attempt to adjust analyses for differences in the baseline patient characteristics between the trials included. The manufacturer reported that:

> *The clinical systematic review identified that the baseline characteristics of nr-AxSpA patients within the randomised controlled trials of certolizumab pegol and adalimumab were heterogeneous, and potentially differed in characteristics that could act as treatment effect modifiers. Furthermore, the populations of these trials also included sizable proportions of AS patients who were originally classified as nr-AxSpA on the basis of a difference between centralised and localised readings of X-rays.*
>
> *To address the differences in the proportions of AS patents in the trials due to reclassification upon central assessment, analyses were conducted using match adjusted indirect comparison (MAIC) and simulated treatment comparison (STC) techniques that incorporated AS patients from the etanercept 314-EU trial. These analyses are referred to collectively as 'analyses adjusting for differences in study baseline characteristics'. A comparison of the results from the MAIC and STC approaches show that while the results of the two analyses are similar, when considering comparisons between etanercept and both adalimumab and certolizumab, the MAIC analysis provides a lower overall comparative estimate of the benefit of etanercept, and is therefore considered overall to be the more conservative of the two approaches. To maintain consistency in the analysis utilised in the economic section, the MAIC was used throughout as the adjusted comparative efficacy measure between etanercept versus adalimumab and etanercept versus certolizumab. For the analysis comparing etanercept against certolizumab pegol, it was possible to address the issue of patient reclassification and differences in baseline characteristics by utilising the RAPID-axSpA trial results that were also available at the level of AxSpA patients, an approach not possible in the comparison of etanercept versus adalimumab. We note that although not explicitly detailed within the scope, the AxSpA population encompasses both nr-AxSpA and AS patients, thus making it a relevant comparison to the decision problem outlined in the scope.*
>
> *Manufacturer's submission, pp. 226–7*[36]

The manufacturer argued that the use of DES conferred potential advantages in relation to modelling non-linearity because of heterogeneous patient characteristics and in relation to modelling time dependency. The latter was also argued as an advantage to considering the impact of sequential therapy which was argued to be complex within a more conventional Markov type structure. Pfizer's model was the only model which explicitly explored issues of treatment sequences. However, in the base case the use of second-line TNF-α inhibitor treatment was restricted to those patients who withdrew because of AEs and was assumed equal efficacy to first line usage. The schematic of the model provided by the manufacturer is replicated in *Figure 6*.

Etanercept RCT data were used to predict an initial 12-week response (in terms of reduction in BASDAI and BASFI) for etanercept for both nr-AxSpA and AS populations. Separate multivariate regressions were used to account for correlation between BASDAI and BASFI. A range of variables were initially included in the regression models based on potential predictors of response identified from their review of economic studies. The statistical significance and direction of effect were evaluated before final models were specified. The 12-week models of BASDAI and BASFI for the nr-AxSpA had R^2 values of (AiC information has been removed) respectively. For the AS population, the equivalent R^2 values were (AiC information has been removed). The regressions were used to estimate mean change in BASDAI and BASFI which through the patient level simulation were used to assign patients into BASDAI 50 responder/non-responder categories and to assess the associated magnitude of change at 12 weeks for these categories.

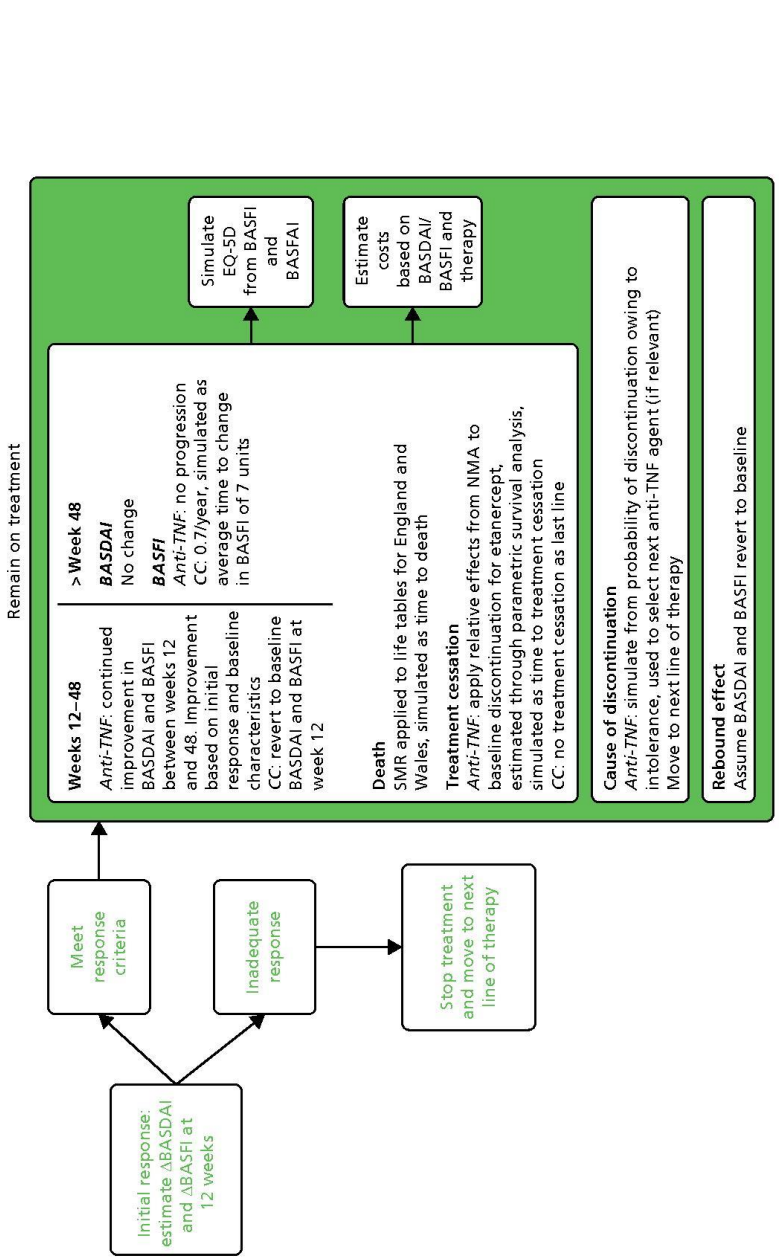


FIGURE 6 Pfizer DES model schematic. NMA, network meta-analysis.[36]

Relative effects from the MTC (or MAIC in the analyses adjusting for differences in study baseline characteristics), in terms of mean differences in BASDAI and BASFI, were applied in order to predict equivalent response and change scores for the other anti-TNF agents and CC at 12 weeks.

From week 12, the BASFI scores for CC were assumed to increase at a rate of 0.7 units per annum (0–100 scale). The modelling of change in BASDAI and BASFI at week 48 for responders to etanercept was conducted using the same approach used for the week 12 treatment response. However, change in BASDAI and BASFI from baseline at week 12 were included as additional covariates within the resulting models in order to ensure that an individual's response at week 48 was dependent on their response at week 12. The 48-week models of BASDAI and BASFI for the nr-AxSpA had R^2 values of (AiC information has been removed), respectively. For the AS population, the equivalent R^2 values were (AiC information has been removed). In the absence of relative effect estimates at week 48 for other therapies, it was assumed that patients who remained on TNF-α inhibitor treatment beyond week 12 (i.e. responders) would converge at the BASDAI and BASFI levels predicted for etanercept by week 48. Constant BASDAI and BASFI scores for TNF-α inhibitor responders were assumed at the level observed at week 48 for subsequent periods.

Treatment discontinuation was modelled by fitting separate parametric survival curves to long-term open-label study data from etanercept for the AS and nr-AxSpA populations. In order to predict treatment cessation in the population that was likely to continue treatment after 12 weeks, parametric curves were fitted only to subjects who achieved a BASDAI 50 response at week 12. Only patients who were randomised to etanercept at baseline were retained within these survival analyses and patients who began etanercept during open-label phases of studies were excluded. The distributions that provided the best fit were exponential (AiC information has been removed) and log-normal (AiC information has been removed), based on the minimisation of the Akaike information criterion and the Bayesian information criterion. The exponential model was chosen based, in part, on the goodness of fit but also because the use of hazard ratios, which were applied to estimate the effect of other anti-TNFs on the rate of discontinuation, required the use of a proportional hazard survival model [to avoid making further assumptions when applying the hazard ratio to the log-normal (accelerated failure time) model].

The same risk of discontinuation was applied to all individuals in the model. The models of discontinuation translate into annual probabilities of discontinuation for etanercept, for patients who achieve a BASDAI 50 response, of 5% and 11% for nr-AxSpA and AS populations, respectively. Based on data from the DANBIO registry, it was assumed that other anti-TNFs have an increased risk of discontinuation compared with etanercept: a hazard ratio of 1.3 is applied for infliximab and 1.12 for adalimumab. In the absence of evidence for golimumab and certolizumab, it was assumed that the relative effect is the same as for adalimumab on the basis that these have common molecular structure and belong to monoclonal antibodies.

After discontinuation of the first treatment, an alternative TNF-α inhibitor was modelled as second-line treatment for patients who discontinued due to AEs [(CiC information has been removed)% for AS and (CiC information has been removed)% for nr-AxSpA]. The same efficacy as applied for first-line treatments was assumed for second-line treatments for patients switching because of AEs. For patients who discontinued because of loss of efficacy, no further TNF-α treatment was modelled. These assumptions were considered by the manufacturer to be consistent with current NICE guidance. For the base-case model, it was assumed that following discontinuation from anti-TNFs, patients would rebound back to their baseline BASDAI and BASFI scores and that the rebound takes 6 months based on the approach used within the TA233[33] submission to NICE.

In the absence of previously published studies reporting the relationship between BASDAI/BASFI and EQ-5D utility scores in the nr-AxSpA population, a de novo relationship was estimated from the 1031 study;[166] variables included age, sex, baseline BASDAI and BASFI. Ordinary least squares regression models were used, with SEs clustered around each subject to account for repeated observations. For consistency, a

similar relationship was estimated for the AS population using the 314-EU study.[167] Alternative linear and non-linear relationships were evaluated and final model selection based on Akaike information criterion statistics. In the nr-AxSpA population, the final model included squared terms for BASDAI and BASFI and an interaction between BASDAI and BASFI, while in the AS population, the covariates for the interaction term, age and male were not included. Scenario analyses considered using alternative model specifications for mapping. The manufacturer reported that according to visual inspection, the estimated models were very similar between populations and reported a high degree of similarity between the results of the de novo estimated models and those published previously.

Figures 7 and *8* replicate the relationships reported by the manufacturer between EQ-5D, BASDAI and BASFI in the nr-AxSpA and AS populations, respectively. Additional figures were also presented by the manufacturer for predicted versus observed EQ-5D in each of the populations. The manufacturer concluded that the models overpredicted EQ-5D at low observed EQ-5D and underpredicted at higher observed EQ-5D values. The manufacturer argued that this was a common feature of mapping algorithms and argued that the approach would be conservative towards the use of anti-TNFs.

The manufacturer included the acquisition, administration and pre-treatment monitoring costs of TNF-α inhibitors. Subsequent monitoring costs were not included in order to avoid potential double counting of the costs which were estimated as a function of BASDAI and BASFI. In the base-case analysis the manufacturer used data from Rafia *et al.*[168] based on BASDAI scores only. A categorical approach was applied to BASDAI scores based on the following annual costs: BASDAI score of $< 40 =$ £151.96; $40 \leq$ BASDAI score $< 60 =$ £311.08; and BASDAI score of $\geq 60 =$ £1039.16. The manufacturer justified the use of this source as it provides the most recent UK specific data reported and permitted separation of particular cost items. The costs and HRQoL of AEs (serious infections only) were included in the base-case analysis (none observed in the nr-AxSpA trial 1031 study;[166] however, serious infections were observed in the AS trial 314-EU study[167]). A separate sensitivity analysis included the costs of serious infections.

A SMR of 1 for the nr-AxSpA population and 1.5 for the AS population were applied to general population life-tables.

Results of Pfizer (etanercept) model

The main base-case ICER results from the manufacturer[36] are summarised in *Table 50* for the AS population, together with a fully incremental comparison of ICERs in *Table 51*. The ICERs versus CC ranged from £19,586 per QALY (certolizumab) and £37,741 per QALY (infliximab). In common with the

FIGURE 7 AiC information has been removed.[36]

FIGURE 8 AiC information has been removed.[36]

TABLE 50 Tumour necrosis factor-α inhibitors compared with CC for AS: Pfizer (base case)[36]

Technology	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (£)
CC	18,122	7.318	–	–	–
Adalimumab	57,535	9.203	39,413	1.885	20,909
Certolizumab	51,843	9.040	33,721	1.722	19,586
Etanercept	60,338	9.334	42,216	2.016	20,938
Golimumab	62,698	9.412	44,576	2.094	21,288
Infliximab	98,340	9.443	80,218	2.125	37,741

TABLE 51 Fully incremental comparison of anti-TNFs for AS: assessment group analysis based on Pfizer (base case)[36]

Technology	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (£)
CC	18,122	7.318	–	–	–
Certolizumab	51,843	9.040	33,721	1.722	19,586
Adalimumab	57,535	9.203	–	–	Extendedly dominated
Etanercept	60,338	9.334	8495	0.294	28,834
Golimumab	62,698	9.412	2360	0.078	30,376
Infliximab	98,340	9.443	35,642	0.031	1,131,181

UCB model, it should be noted that the costs of certolizumab assumed within Pfizer's model were also based on the PAS for certolizumab which has been proposed but is not yet formally agreed with the Department of Health and NICE. Hence the ICER for certolizumab versus CC without the PAS will be higher than the estimates reported here.

In the fully incremental analysis, adalimumab was extendedly dominated. Of the remaining non-dominated treatments, the ICERs of the next most costly interventions compared with the previous non-dominated alternative were £19,586 (certolizumab vs. CC), £28,834 (etanercept vs. certolizumab), £30,376 (golimumab vs. etanercept) and £1,131,181 (infliximab vs. golimumab).

The main base-case ICER results from the manufacturer[36] are summarised in *Table 52* for the nr-AxSpA population together with a fully incremental comparison of ICERs (*Table 53*). The ICERs versus CC ranged from £23,195 (etanercept) and £23,575 (certolizumab). In contrast to the UCB analysis, the ICERs for the nr-AxSpA population were marginally less favourable than the results for the AS population. There was also less of a marked difference between the ICERs for each of the anti-TNFs and CC compared with the UCB results, although a large difference was evident relating to the magnitude of the incremental QALY estimates for certolizumab vis-à-vis the other anti-TNFs. *Table 54* reports the results of the fully incremental analysis. None of the anti-TNFs was ruled out via dominance or extended dominance, and the ICER of each comparison remained below £30,000 per QALY for each successively more expensive and effective treatment.

TABLE 52 Anti-TNFs compared with CC for nr-AxSpA: Pfizer (base case)[36]

Technology	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (£)
CC	20,609	10.221	–	–	–
Adalimumab	62,667	12.030	42,058	1.809	23,242
Certolizumab	74,282	12.497	53,673	2.276	23,575
Etanercept	59,635	11.903	39,026	1.682	23,195

TABLE 53 Fully incremental comparison of nr-AxSpA: assessment group analysis based on Pfizer (base case)[36]

Technology	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (£)
CC	20,609	10.221	–	–	–
Etanercept	59,635	11.903	39,026	1.683	23,195
Adalimumab	62,667	12.030	3033	0.127	23,871
Certolizumab	74,282	12.497	11,615	0.467	24,864

TABLE 54 Incremental results of etanercept vs. adalimumab in nr-AxSpA (using MAIC data): Pfizer[36]

Technology	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (£)
Adalimumab	48,494	11.473	–	–	–
Etanercept	60,404	11.928	11,910	0.455	26,176

To address the concerns noted by Pfizer relating to the heterogeneity across the different trials in the nr-AxSpA population, a separate matched indirect comparison was presented for etanercept versus adalimumab. A separate comparison was also presented versus certolizumab for the combined axSpA population in the manufacturer's submission. Using the MAIC approach, the ICER for etanercept versus adalimumab was £23,195 per QALY. Total cost and QALYs estimates were reversed in the MAIC approach compared with the base-case analysis (adalimumab generated greater QALYs at increased cost), demonstrating the potential impact of trying to minimise observable sources of possible confounding.

Overview of Merck Sharp & Dohme (golimumab, infliximab) model

The economic models submitted by Merck Sharp & Dohme[37] compared the cost-effectiveness of golimumab and infliximab with conventional therapy and other licensed anti-TNFs for AS. Although the manufacturer made separate submissions for golimumab and infliximab, the model structures and data sources used to inform the economic models are identical across the submissions. Hence this review focuses on the specific submission for golimumab but also considers key data sources and assumptions specific to infliximab. The model base case is based on a lifetime time horizon (approximately 60 years), and costs and benefits are discounted at an annual rate of 3.5%. A NHS and PSS perspective is used for costs. The reference year for costs was reported to be 2012/13.

The economic model submitted by the manufacturer for golimumab[37] is based on the same model structure submitted as part of NICE TA233[33] and summarised previously in *Data extraction* (Armstrong *et al.* 2013)[163] Hence, a description of the structure of the model is not repeated in this section. In summary, the manufacturer's cost-effectiveness model was based on a short-term decision tree (based on an assessment of BASDAI 50 response at 12 weeks in the base case) and a longer-term Markov model.

The proportion of patients achieving BASDAI 50 at week 12 (±2 weeks) for each TNF-α inhibitor was obtained from a systematic review and MTC undertaken by the manufacturer. The results are summarised in *Table 55*.

TABLE 55 Odds ratios and probability of BASDAI 50 score response to anti-TNFs and conventional therapy (Merck Sharp & Dohme)[37]

	BASDAI 50	
Treatment	OR (95% CrI)	Probability
Golimumab	5.54 (2.12 to 12.13)	0.49
Infliximab	22.44 (2.78 to 89.05)	0.79
Adalimumab	5.20 (2.14 to 10.62)	0.47
Etanercept	5.46 (2.03 to 11.74)	0.60
Certolizumab pegol	6.62 (1.66 to 17.59)	0.53
Conventional therapy	–	0.15

Data from the GO-RAISE[90] trial and the open-label extension period (up to week 108) were used to develop predictive equations of mean change from baseline in BASDAI and BASFI scores over time. Two separate equations were developed based on the 24-week data (0–24 weeks) for all patients and post-24-week (week 24–108) data from GO-RAISE[90] for patients who remained on treatment. The variables applied in each equation are summarised in *Tables 56* and *57*.

The treatment coefficient (and interaction term) in the short-term regression equation is used to estimate separate BASDAI/BASFI scores for anti-TNFs and CC. Hence, up to week 24, the same estimate of BASDAI/BASFI appears to be applied to all TNFs (i.e. regardless of the differential response rates assumed). Beyond week 24, the same BASDAI/BASFI score is applied to a responder to any of the TNFs, although a different response rate for each TNF-α inhibitor is assumed based on the MTC. The BASDAI/BASFI regressions are applied to responders who continue on TNF-α inhibitor therapy up to week 108 for BASDAI and up to week 108 for BASFI.

The BASDAI and BASFI scores beyond week 108 for responders who continue to receive anti-TNFs beyond this period in the model are assumed to remain constant (at the week 108 value). The BASFI scores beyond week 256 for responders who continue to receive anti-TNFs beyond this period in the model are assumed to remain constant (at the week 108 value) but are also subject to an annual progression rate of BASFI at this point which is set to half the rate of CC in the baseline (0.035 units per annum, 0–10 scale). The justification for this is not explicitly made by the manufacturer. For the base-case model, BASFI scores for CC patients on conventional therapy are assumed to progress at a rate of 0.07 units per year after week 24.

An annual discontinuation rate of 6.1% is applied for the entire time horizon after week 12 in the base-case analysis. This estimate is derived from data reported between weeks 24 and 256 in the 50 mg arm of golimumab from the GO-RAISE[90] extension period. The manufacturer does not formally state whether or

TABLE 56 Short-term regression equations used by Merck Sharp & Dohme[37] for BASDAI/BASFI score (0–24 weeks): all patients

Variable	Parameter	SE
BASFI		
Intercept	0.1008	0.557
Age	–0.0284	0.009874
Baseline BASFI	0.1780	0.05429
Treatment	1.8096	0.2551
Male	0.04156	0.2767
Week^{-2}	5.226	0.2767
Treatment × week^{-2}	–14.6396	2.2699
BASDAI		
Intercept	0.4685	0.8126
Age	–0.03399	0.0105
Baseline BASDAI	0.2212	0.08436
Treatment	2.0620	0.2742
Male	0.2652	0.2953
Week^{-2}	–3.4664	2.1365
Treatment × week^{-2}	–7.1029	2.6887

TABLE 57 Long-term regression equations used by Merck Sharp & Dohme[37] for BASDAI/BASFI score (24–108 weeks): responders only

Variable	Parameter	SE
BASFI		
Intercept	0.4933	0.7364
Age	–0.03915	0.01321
Baseline BASFI	0.5706	0.07292
Male	0.6523	0.4001
Log (week)	0.09524	0.04938
BASDAI		
Intercept	0.6277	1.0303
Age	–0.03531	0.01367
Baseline BASDAI	0.5762	0.1055
Male	0.2196	0.4094
Log (week)	0.2196	0.06908

not this is specific to those patients who were identified to be responders at 12 weeks. However, it appears to be based on all patients who continued to receive golimumab beyond 24 weeks regardless of their response status. The same discontinuation rate is applied to all TNF-α inhibitors. Following discontinuation from anti-TNFs, the BASFI and BASDAI scores are assumed to deteriorate/rebound over a 24-week period back to their baseline BASFI and BASDAI scores (i.e. rebound equal to gain). Therefore, in common with other models which apply this rebound assumption, patients are assumed to achieve a lifetime benefit from treatment with anti-TNFs for BASFI.

Utilities were derived from a NICE TA (TA143[17]) and incorporated age, sex, BASFI score and BASDAI score. Costs included in the model comprised drug acquisition, short-term (12-week) costs, longer-term disease costs and AEs. Longer-term disease costs were based on BASFI scores from the GO-RAISE[90] trial using the same regression equation used for NICE TA143.[17]

The proportion of males and females recruited in the GO-RAISE[90] trial is used to estimate a weighted average mortality risk by sex. The sex-specific SMR for AS from a study by Bakland *et al.*[16] is applied to the mortality rates from the general population to calculate adjusted mortality rates for AS patients in the model. The study by Bakland reported a SMR of 1.63 (95% CI 1.29 to 1.97) for males and 1.38 (95% CI 0.48 to 2.28) for females.

Results of the Merck Sharp & Dohme (golimumab, infliximab) model

The main base-case ICER results from the manufacturer[37] are summarised in *Table 58* for the AS population, together with a fully incremental comparison of ICERs in *Table 59*. The ICERs versus CC ranged from £19,070 (golimumab) to £42,532 (infliximab). In the fully incremental analysis, golimumab and certolizumab were the non-dominated anti-TNFs. The ICER for golimumab versus CC was £19,070 and for certolizumab versus golimumab was £21,441 per QALY.

Summary and critique of de novo cost-effectiveness submissions

In general, the manufacturer models[34–37] appeared to be constructed to a high standard and it is evident that significant work had been undertaken by each to identify and use previously published studies and to exploit existing individual patient data from their own RCTs and open-label extension periods to generate estimates that were appropriate for the requirements of the model.

TABLE 58 Anti-TNFs compared with CC for AS: Merck Sharp & Dohme (base case)[37]

Technology	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (£)
CC	160,837	10.5529	–	–	–
Adalimumab	181,589	11.6296	20,752	1.0766	19,275
Certolizumab	183,017	11.6962	22,180	1.1432	19,401
Etanercept	183,540	11.5862	22,703	1.0332	21,972
Golimumab	181,427	11.6326	20,590	1.0797	19,070
Infliximab	208,856	11.6819	48,019	1.1290	42,532

TABLE 59 Fully incremental comparison of anti-TNFs for AS: assessment group analysis based on Merck Sharp & Dohme (base case)

Technology	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (£)
CC	160,837	10.5529	–	–	–
Golimumab	181,427	11.6326	20,590	1.0797	19,070
Adalimumab	181,589	11.6296	–	–	Dominated
Certolizumab	183,017	11.6962	1590	0.0636	21,441
Etanercept	183,540	11.5862	–	–	Dominated
Infliximab	208,856	11.6819	–	–	Dominated

Despite the different model structures and assumptions applied across the various manufacturer's submissions,[34–37] the ICERs reported for the anti-TNFs versus CC were remarkably consistent in the AS population. *Table 60* presents a summary of the ICER reported by each manufacturer for each of the anti-TNFs versus CC. It is perhaps expected that the majority of manufacturer's reported the lowest ICER versus CC for their own products. The only exception to this is the Pfizer model, which estimated the lowest ICER versus CC for certolizumab in this population (etanercept was the next lowest), although it should be noted that Pfizer included the proposed PAS costs for certolizumab which was not universally applied across the different manufacturer's submissions. Hence, although differences between the ICER versus CC were quite similar, the variation in approaches used by each manufacturer appears partially driven by maximising any potential comparative advantage considered vis-à-vis other manufacturer

TABLE 60 Comparison of manufacturer ICER estimates vs. CC (AS population)

Technology	AbbVie[34] (adalimumab), ICER (£)	UCB[35] (certolizumab), ICER (£)	Pfizer[36] (etanercept), ICER (£)	Merck Sharp & Dohme[37] (golimumab, infliximab), ICER (£)
CC	–	–	–	–
Adalimumab	**16,391**	19,932	20,909	19,275
Certolizumab	17,067	**16,647**	**19,586**	19,401
Etanercept	16,897	19,272	20,938	21,972
Golimumab	16,535	19,049	21,288	**19,070**
Infliximab	44,448	42,671	37,741	42,532

The figures reported in bold indicate which specific TNF-α inhibitor has the lowest ICER vs. CC in each of the manufacturer's submissions.

products (i.e. in terms of assumptions made about similarities and differences for response rates, magnitude of changes in BASDAI and BASFI and withdrawal rates). However, it should be noted that no manufacturer makes a strong claim regarding differential efficacy between the alternative anti-TNFs which is borne out in the relatively small differentials reported between the different products in each of the submissions.

Table 61 presents a summary of the ICERs reported by each manufacturer for each of the anti-TNFs versus CC for the nr-AxSpA population. There appears to be much more heterogeneity across the manufacturer's submissions[34–36] compared with the AS population. There appears an almost twofold difference in the ICERs reported across the submissions for each of the anti-TNFs. Importantly, there also appears variation across the populations with more favourable ICERs reported than CC for the nr-AxSpA population vis-à-vis the estimates by AbbVie (both adalimumab and certolizumab) and UCB (certolizumab only). Hence, the differences in structural and parameter assumptions appear more evident in the results for the nr-AxSpA population compared with results for the AS population.

To assist in identifying possible reasons for the differences between populations, summaries of the key structural assumptions used by each manufacturer are provided in *Tables 62* and *63*. A more micro-level analysis of comparison of specific parameter estimates is reported separately in *Appendix 11*.

In general, it is difficult to identify the specific factors which can easily explain differences within and between the two populations across the manufacturer's submissions. In general, similar model structures were applied by each manufacturer across the separate populations. However, it is evident that there are important differences based on a number of key structural issues: (1) the response criteria and timings applied; (2) the magnitude of change scores and particularly the assumption concerning the time at which these were assumed to 'level off' (generally longer in the AS populations because of the longer open-label extension periods); (3) the underlying rate of progression of BASFI with CC and the impact of anti-TNFs on this rate; and (4) the rebound assumption and timing of this.

TABLE 61 Comparison of manufacturer ICER estimates vs. CC (nr-AxSpA population)

Technology	AbbVie[34] (adalimumab), ICER (£)	UCB[35] (certolizumab), ICER (£)	Pfizer[36] (etanercept), ICER (£)
CC	–	–	–
Adalimumab	13,228	30,370	23,242
Certolizumab	**12,866**	**15,615**	23,575
Etanercept	Not assessed	50,692	**23,195**

The figures reported in bold indicate which specific TNF-α inhibitor has the lowest ICER vs. CC in each of the manufacturer's submissions.

TABLE 62 Model structure and key structural assumptions: AS population

Parameter	Merck Sharp & Dohme economic model[27] (infliximab, golimumab)	AbbVie[24] economic model (adalimumab)	UCB[26] economic model (certolizumab)	Pfizer[26] economic model (etanercept)
Model type	Decision tree followed by Markov model	Markov model	Markov model	Patient-level simulation model (DES)
Time horizon	Lifetime	40 years	Lifetime	Lifetime
Response criterion	BASDAI 50 response at week 12	ASAS 20 response at week 12	ASAS 20 response at week 24	BASDAI 50 response at week 12
Response criterion justification	Efficacy outcome in GO-RAISE[90] study; recommended by the ASAS Working Group (Keat 2005)[169]	Primary end point of ATLAS[61] study	ASAS 20 is the primary end point of RAPID-axSpA[84] study	Based on the current NICE definition of treatment response (TA143[17])
Progression assumption BASDAI				
Anti-TNFs responders	Constant after week 108	Constant after week 260	Constant after week 24	Constant after week 48
Anti-TNFs non-responders	Constant	Constant	Constant	Constant
CC	Constant after week 24	Constant	Constant	Constant after week 12
Progression assumption BASFI				
Anti-TNFs responders	Constant after week 108; 0.035 after week 256	Constant after week 260	Constant after week 24	Constant after week 48
Anti-TNFs non-responders	0.07	0.056	0.07	0.07
CC	0.07 after week 24	0.056	0.07	0.07 after week 12
Rebound assumption	Rebound to baseline	Rebound to baseline	Rebound to conventional therapy	Rebound to baseline
Rebound assumption duration	Over a 6-month period	Immediately	Over a 6-month period	Over a 6-month period
Placebo response	14.5% at week 12; loss or maintenance of placebo response not reported	BASDAI and BASFI scores return to baseline at week 12	No placebo response	BASDAI and BASFI scores return to baseline at 12 weeks

TABLE 63 Model structure and key structural assumptions: nr-AxSpA population

Parameter	AbbVie[34] economic model (adalimumab)	UCB[35] economic model (certolizumab)	Pfizer[36] economic model (etanercept)
Model type	Markov model	Markov model	Patient-level simulation model (DES)
Time horizon	40 years	Lifetime	Lifetime
Response criterion	ASAS 40 response at week 12	ASAS 20 response at week 12	BASDAI 50 response at week 12
Response criterion justification	Primary end point of ABILITY-1[58] study	Primary end point of RAPID-axSpA[64] study	Based on the current NICE definition of treatment response (TA143[17])
Progression assumption BASDAI			
Anti-TNFs responders	Constant after week 140	Constant after week 12	Constant after week 48
Anti-TNFs non-responders	Constant	Constant	Constant
CC	Constant	Constant	Constant after week 12
Progression assumption BASFI			
Anti-TNFs responders	Constant after week 140	Constant after week 12	Constant after week 48
Anti-TNFs non-responders	0.084	0.07	Constant/0.07
CC	0.084	0.07	0.07 after week 12
Rebound assumption	Rebound to baseline	Rebound to conventional therapy	Rebound to baseline
Rebound assumption duration	Immediately	Over a 6-month period	Over a 6-month period
Placebo response	BASDAI and BASFI scores return to baseline at week 12	No placebo response	BASDAI and BASFI scores return to baseline at 12 weeks

Given the complex inter-relationship between these structural assumptions and subsequent parameter estimates, it is difficult to identify single specific reasons for differences. However, the structural differences clearly lead to marked differences in the BASDAI and BASFI scores estimated over time by each manufacturer for each population. *Figures 9–11* provide a graphical summary of the cohort BASDAI and BASFI scores, for the AS population, from three of the manufacturers. These highlight the significant differences in subsequent parameter estimates applied at a cohort level. Equivalent estimates are not presented for the Pfizer model because of the complexities of generating this data from the DES model. The BASDAI and BASFI scores are presented here only for the case made by each manufacturer for their own product.

Tables 64 and *65* summarise the mean difference in BASDAI and BASFI scores applied to responders to anti-TNFs and those applied to CC at various time points in each model. The tables clearly highlight the range of different values applied across the separate manufacturers. This further emphasises the variation in approaches, sources and assumptions.

The equivalent figures and tables are reported for the nr-AxSpA population (*Figures 12* and *13* and *Tables 66* and *67*).

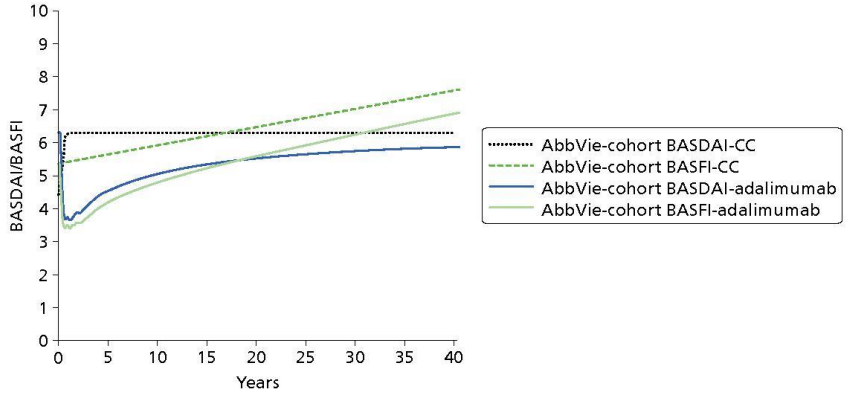


FIGURE 9 Comparison of cohort BASDAI/BASFI scores for AS population from AbbVie model (adalimumab).[34]

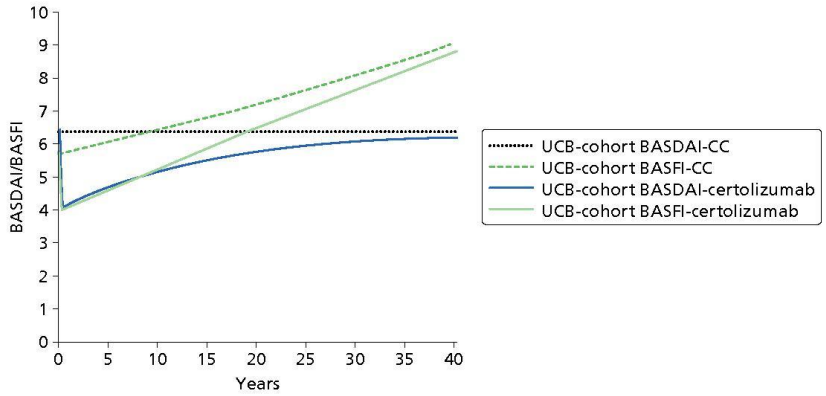


FIGURE 10 Comparison of cohort BASDAI/BASFI scores for AS population from UCB model (certolizumab).[35]

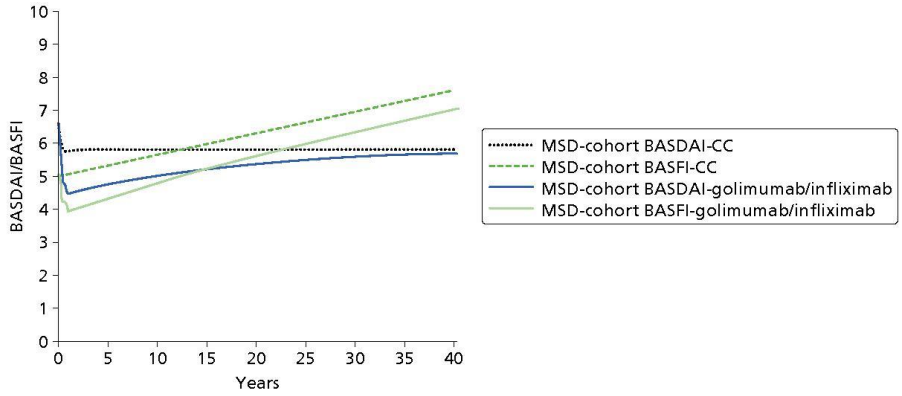


FIGURE 11 Comparison of cohort BASDAI/BASFI scores for AS population from Merck Sharp & Dohme models (golimumab/infliximab).[37] MSD, Merck Sharp & Dohme.

TABLE 64 The BASDAI score difference for treatment responders vs. CC: AS population

Time	Adalimumab vs. CC	Certolizumab vs. CC	Infliximab/golimumab vs. CC
12 weeks	–2.98	AiC information has been removed	–2.01
24 weeks	–4.42	AiC information has been removed	–2.05
1 year	–4.9	AiC information has been removed	–2.77
3 years	–5.23	AiC information has been removed	–2.83
5 years	–5.31	AiC information has been removed	–2.83
10 years	–5.31	AiC information has been removed	–2.83
20 years	–5.31	AiC information has been removed	–2.83
40 years	–5.31	CiC information has been removed	–2.83

TABLE 65 The BASFI score difference for responders vs. CC: AS population

Time	Adalimumab vs. CC	Certolizumab vs. CC	Infliximab/golimumab vs. CC
12 weeks	–2.03	AiC information has been removed	–1.68
24 weeks	–3.28	AiC information has been removed	–1.74
1 year	–3.71	AiC information has been removed	–2.49
3 years	–4.25	AiC information has been removed	–2.59
5 years	–4.25	AiC information has been removed	–2.66
10 years	–4.53	AiC information has been removed	–2.85
20 years	–5.09	AiC information has been removed	–3.18
40 years	–6.21	AiC information has been removed	–3.75

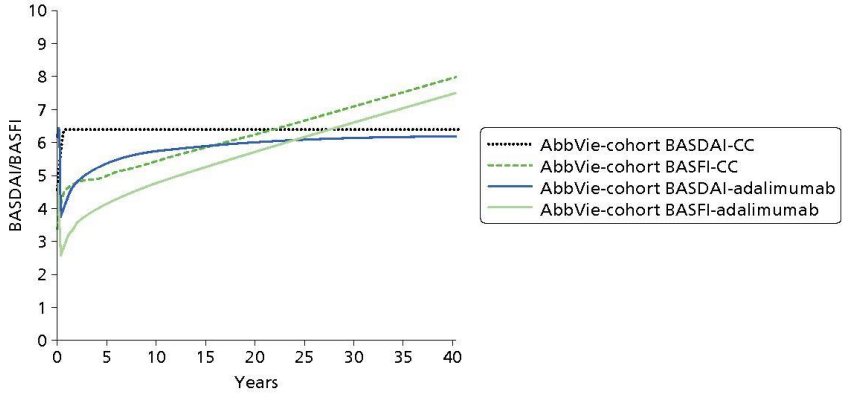


FIGURE 12 Comparison of cohort BASDAI/BASFI scores for nr-AxSpA population from AbbVie model (adalimumab).[34]

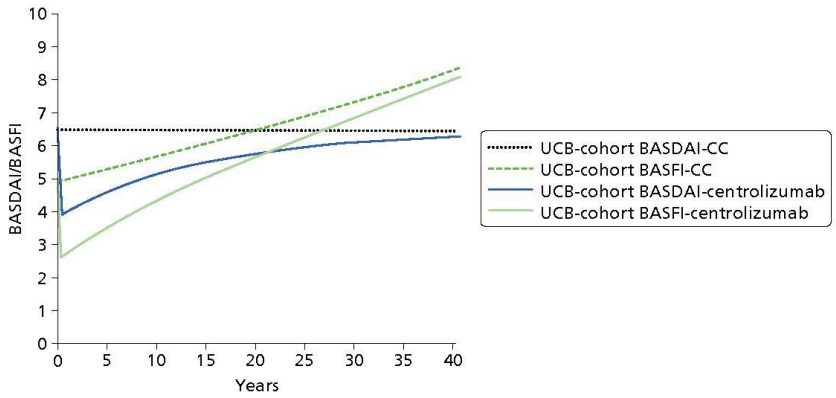


FIGURE 13 Comparison of cohort BASDAI/BASFI scores for nr-AxSpA population from UCB model (certolizumab).[35]

TABLE 66 The BASDAI score difference for responders vs. CC: nr-AxSpA population

Time	Adalimumab vs. CC	Certolizumab[a] vs. CC
12 weeks	–3.89	AiC information has been removed
24 weeks	–5.54	AiC information has been removed
1 year	–5.42	AiC information has been removed
3 years	–5.99	AiC information has been removed
5 years	–5.99	AiC information has been removed
10 years	–5.99	AiC information has been removed
20 years	–5.99	AiC information has been removed
40 years	–5.99	AiC information has been removed

a Certolizumab patients who remain nr-AxSpA and do not transition to AS.

TABLE 67 The BASFI score difference for responders vs. CC: nr-AxSpA population

Time	Adalimumab vs. CC	Certolizumab[a] vs. CC
12 weeks	–2.95	AiC information has been removed
24 weeks	–4.11	AiC information has been removed
1 year	–4.12	AiC information has been removed
3 years	–4.55	AiC information has been removed
5 years	–4.72	AiC information has been removed
10 years	–5.14	AiC information has been removed
20 years	–5.98	AiC information has been removed
40 years	–7.66	AiC information has been removed

a Certolizumab patients who remain nr-AxSpA and do not transition to AS.

The differences across manufacturers and between the populations are further illustrated by the summary of key parameter inputs reported in *Appendix 11*. As well as reporting the main parameter inputs, *Appendix 11* also explores differences in approaches at a parameter level for key inputs (e.g. withdrawal, costs, etc.).

It is evident from these comparisons that there are significant differences across the manufacturers in terms of key structural and parameter estimates. While it might appear reassuring that these differences do not appear to lead to significant differences across the ICER estimates reported for the AS population, the greater heterogeneity reported in the ICER estimates for the nr-AxSpA is clearly an issue. However, even within the AS population, any reassurance that one might have regarding the robustness and appropriateness that these estimates have for informing NHS practice needs to be carefully considered in relation to the key conceptual issues and concerns highlighted in *Summary and critique of published cost-effectiveness studies* for previously published cost-effectiveness studies. Despite significant work undertaken by each manufacturer in support of existing and new indications for their products, it is particularly concerning that many of the key conceptual issues and concerns appear to have not been fully addressed. Indeed, many of these issues seem to have not been addressed at all, such that many models still seem reliant on the use of open-label extension data (and even more so with the extended follow-up reported in the AS population) without any formal consideration of the potential issues with selection that the use of these studies inevitably are subject to. Consequently, the benefits of anti-TNFs are being projected over significant periods of time without any evidence on the counterfactual (i.e. what happens to patients who do not enter into the open-label extension periods? What happens to patients who subsequently withdraw from anti-TNFs? And what would have happened to patients over a longer-time horizon who did not receive anti-TNFs?).

It appears that much of the case being made concerning the cost-effectiveness of the anti-TNFs rests on comparison of single-arm studies (the subject of open-label data) and retrospective comparisons against historical cohorts (as the counterfactual, for patients not on treatment, is unknown). While such a comparison may be necessitated by the short-term nature of the double-blind periods, the lack of a more detailed consideration of the appropriateness of the comparisons being made in relation to sources of natural history data (and subsequent assumptions made concerning the BASDAI/BASFI trajectories of the different patient categories) is concerning and, hence, current ICER estimates reported by the manufacturers must be considered to be both speculative and highly uncertain.

Many of these problems can be associated with whether or not BASDAI and BASFI scores provide an appropriate conceptual basis for modelling the chronic and progressive nature of AS and nr-AxSpA. Hence current models appear largely driven by data availability (i.e. the extensive evidence which has been generated and continues to be generated investigating the relationship between BASDAI/BASFI and costs/utilities) rather than trying to develop a clearer underlying biological or clinical process which may better characterise the disease and subsequent progression across the separate populations.

Until such time that sufficient data-linking costs and utilities to other measures are reported, it seems inevitable that models will continue to be driven largely by BASDAI and BASFI scores over time together with assumptions concerning the longer-term effect of anti-TNFs. However, given the nature of existing models and the reliance on uncontrolled longer-term follow-up of anti-TNFs, and comparison with historical 'controls' (particularly in relation to BASFI progression over time and the assumptions being made concerning the potential disease modification properties of anti-TNFs in both AS and nr-AxSpA populations), it is surprising that greater efforts have not been made by the manufacturers to try to more formally link to the increasing evidence base being generated in relation to radiographic progression in the AS population.

It is also surprising that more thought has also not been given to characterising the potential difference in BASFI progression across the separate populations and how generalisable assumptions may be between these. The result is that many of the key assumptions concerning whether the anti-TNFs are primarily symptom control treatments or whether they are also potential disease modifiers remains implicitly dealt with within existing submissions. The result is that several manufacturers use identical assumptions across populations with respect to BASFI progression and the effect of the anti-TNFs. Interestingly, only one manufacturer appears to use differential rates of BASFI progression across the populations (AbbVie), although the same structural assumption concerning the effect of anti-TNFs is still made. Interestingly, this manufacturer applies a higher rate of change in BASFI for patients receiving CC in the nr-AxSpA population vis-à-vis the AS population. However, while such a difference is interesting, the basis of and implication for this differential is not fully explained or justified by the manufacturer.

The issue of intermittent and sequential use of anti-TNFs remain important clinical questions but the existing models do not provide a robust basis for informing these decisions. The cost-effectiveness of intermittent therapy versus continuous therapy was not formally considered in any model identified. However, it could be argued that such a comparison might be deemed outside the scope of a NICE appraisal. Although one manufacturer (Pfizer[36]) explored the potential cost-effectiveness of sequential therapy, much of this has been done via assumptions (e.g. assuming equal efficacy second line in patients who discontinue first line as a result of an AE) or via adjustments applied to first-line efficacy estimates based on 'real-world' evidence reported from large-scale registries (which typically show anti-TNFs to be clinically effective but with lower response rates than reported in naive patients). Consequently, existing attempts to model sequential therapy are largely based on applying adjustments to first-line efficacy data using observational evidence which are clearly subject to potential confounding. In large part, the limitations of existing cost-effectiveness models for informing these clinical questions appears less a function of the models themselves but rather that robust clinical data to date has not been generated to inform unbiased estimates of relative efficacy of alternative strategies for using the anti-TNFs.

The following sections report the development of a de novo model to address some of the key issues and uncertainties which have been identified in this review. *Chapter 5* reports the results of an extended synthesis which has been developed to provide a more generalisable framework for synthesising clinical efficacy data ensuring that appropriate estimates are generated for the model which make use of all relevant and available evidence. This is followed in *Chapter 6* by a description of the de novo model (York model) which attempts to link this framework to a more coherent conceptual model of the chronic and progressive nature of AS and nr-AxSpA.

Chapter 5 Independent economic assessment: extended synthesis

Existing evidence on the short-term clinical effectiveness of anti-TNF drugs has been presented and discussed in *Chapter 3*. The methods of evidence synthesis are extended in this section to more directly address the decision problem and the parameter inputs required for the economic model. There were two specific aims to these analyses. Firstly, we aimed to more formally explore the differences between individual anti-TNF treatments to inform the most appropriate assumption for the economic model (i.e. equivalence or drug-specific differences). Within *Chapter 3* of this report, the assumption of independent treatment effects was evaluated alongside the assumption of a common (equal) treatment effect across anti-TNFs, for every outcome of interest. While there is no evidence that supports differences in the effectiveness of these drugs, assuming equal effectiveness means that the trials are pooled as if the same drug had been trialled; this leads to an arguably overly precise estimate of effect for the class of drugs. For this reason, we explore an additional scenario in which treatments are assumed to have a similar, but not equal, effectiveness, that is there are differences between the effectiveness of treatments that we may not be able to explain but that we should consider.

The second aim was to generate appropriate effect-size estimates and their associated uncertainty to inform the main input parameters of the economic model by synthesising together evidence on BASDAI and BASFI outcomes jointly. Initially, we considered the two related BASDAI outcomes relevant to the decision model reported in the effectiveness evidence available: changes in BASDAI scores over a certain period of time and a probability of response to BASDAI 50 (that is, a 50% change in the BASDAI score in relation to baseline). The BASDAI 50 is important, as patients are expected to discontinue anti-TNFs if, at 12 weeks, they have not been able to achieve response to this criterion (according to NICE guidance).[17,33] Changes in BASDAI scores observed at this same time point determine the magnitude of initial response to treatment and have often been used in economic modelling as the basis for extrapolating treatment effects. Given that these outcomes are both central to informing effect parameters in the decision model, a synthesis model that considers the relation between these two outcomes provides a more consistent and coherent basis for informing these parameters.

We developed a synthesis model that pools evidence on the change in BASDAI score by considering both those studies that report this measure directly and also those that report the proportion of patients achieving a BASDAI 50 response. We expressed BASDAI 50 as a function of the absolute change in BASDAI and we use this relationship in the extended synthesis. We also aim to simultaneously synthesise information on BASFI score, a measure that is used together with BASDAI score to determine the long-term QALY and cost burden of the disease in the economic model. Treatments improving AS symptoms are expected to affect both disease activity and function, and thus we expect a reduction in both BASDAI and BASFI scores, this means that we expect the changes to these two measures to be correlated. Extending the synthesis modelling to consider BASFI scores not only allows all relevant evidence to contribute to the synthesis but also ensures that all measures are synthesised together to reflect the expected correlations between the two outcomes. Uncertainty is also more appropriately quantified than synthesising each outcome separately.

In the decision model, prognosis, costs and QALYs are determined by absolute BASDAI and BASFI scores. Given that treatment continuation is determined by response to BASDAI 50 at 12 weeks, it is important for the economic model to estimate the absolute change in BASDAI and BASFI separately for responders and non-responders, that is the conditional scores. However, the published clinical effectiveness evidence does not report the conditional scores. Consequently, we requested the conditional data from the pivotal trials in both the AS and nr-AxSpA indications from each manufacturer. These data were subsequently provided by the manufacturers for their four pivotal trials. However, this information was not available for the other trials in the evidence-base (there were 15 trials in the evidence base and conditional scores were only

available for four). In view of the limited data available on the conditional scores, another important extension of the synthesis approach was the evaluation of these. We used the results from the extended synthesis model to evaluate the conditional scores by simulating BASDAI and BASFI scores for two equivalent cohorts of patients the only difference being that one cohort was treated and the other was not.

This section provides only a summary of these analyses; full details are in *Appendices 5, 12* and *13*. We will describe first the approach for the synthesis of evidence on the AS population, followed by the approaches and results for the nr-AxSpA population.

Ankylosing spondylitis population

Brief description of the data

Based on study population and follow-up (i.e. around 12 weeks in duration), 16 of the RCTs are considered directly relevant to the decision problem for the AS population (studies 1–16 in *Table 68*). One of these studies did not report BASDAI or BASFI outcomes (study 3[57]) and thus could not be included in the analyses. The 15 remaining studies reported at least one outcome measure of BASDAI 50 score and/or change from baseline on BASDAI and BASFI scores.

TABLE 68 Evidence on BASDAI- and BASFI-related outcomes for the AS population

Study number	Trial name	Treatments	Number in treatment group	Number in placebo group	BASDAI 50 score	Change BASDAI score	Change BASFI score
1	Hu (2012)[55]	1	26	20		✗	✗
2	Huang (2014)[56]	1	229	115	✗	✗	✗
3	Lambert (2007)[57]	1	38	44			
4	ATLAS (2006)[61]	1	208	107	✗	✗	
5	RAPID-axSpA (2014)[64]	2	121	57	✗	✗	✗
6	Barkham (2010)[71]	3	20	20	✗	✗[a]	✗[a]
7	Davis (2003)[72]	3	138	139		✗	✗
8	Dougados (2011)[74]	3	39	43	✗	✗	✗
9	Gorman (2002)[79]	3	20	20			✗
10	Calin (2004)[83]	3	45	39		✗	✗
11	van der Heijde (2006)[86]	3	305	51	✗		
12	GO-RAISE (2008)[90]	4	138	78	✗		✗
13	Bao (2012)[96]	4	108	105	✗		✗
14	Braun (2002)[98]	5	34	35	✗	✗[a]	✗[a]
15	Marzo-Ortega (2005)[100]	5	28	14		✗	✗[a]
16	Van den Bosch (2002)[101]	5	9	12		✗[a]	✗[a]

a Do not report any measure of dispersion (such as SDs).
Treatment: 1, adalimumab; 2, certolizumab (certolizumab 200 mg and/or certolizumab 400 mg); 3, etanercept (etanercept 25 mg and/or etanercept 50 mg); 4, golimumab 50 mg; 5, infliximab. Note that some studies only report one of the BASDAI measures. For example, the golimumab trials (studies 12[90] and 13[96]) only report BASDAI 50 score and not the absolute change in this score.

General aspects of implementation and software

The synthesis was conducted from a Bayesian perspective, using WinBUGS [a Markov chain Monte Carlo (MCMC) simulation-based software for Bayesian inference]. For burn-in, we ran 100,000 simulations and another 100,000 were used in inferences. Convergence was assessed by running two chains and convergence was assumed if the Gelman–Rubin statistic was equal to 1. Goodness of fit was assessed using the DIC, a criterion developed by Spiegelhalter *et al.*[170] based on the trade-off between the fit of the data to the model and the complexity of the model. Fit is measured using the deviance, and complexity is included using a measure of the 'effective number of parameters' (i.e. posterior mean deviance minus deviance evaluated at the posterior mean of the parameters). Models with smaller DIC are better supported by the data: that is, the lower the DIC, the better the data fit the model. In the presence of autocorrelation, the MCMC simulation for inference was increased to 200,000 and a thin of 20 was applied (yielding a sample for inference of 10,000 for each chain).

The main synthesis models (approaches B and C described next) pooled differences between treatment and control in change in scores from baseline (BASDAI and BASFI). The treatment associated with the lowest (most negative) mean change score is expected to be best. However, it is important to quantify the uncertainty around the estimates and for this reason SDs were reported alongside expected values. When averaged ORs were presented median values instead of means were used, as ORs tend to follow a skewed distribution.

Relative effectiveness estimates for models assuming exchangeability across treatments are based on the predictive distribution, representing the distribution of the data averaged over all possible parameter values. This summary statistic best reflects the impact of uncertainty in the parameters of the model and is here judged as a more appropriate basis to be used in the decision model.[171]

When possible, meta-regression analyses were conducted to evaluate potential treatment effect modifiers. Meta-regression is a tool aimed at examining the impact of variables on effect size using regression-based techniques. In these explorations, the following baseline characteristics were considered: BASDAI score, BASFI score, age, sex, duration of symptoms (years) and CRP level.

Exploring assumptions for the relative effectiveness of individual anti-tumour necrosis factor treatments (modelling approach A)

In AS, pivotal trials for the licensed anti-TNFs do not perform head-to-head comparisons with other agents but instead compare the effect of treatments against standard care. These trials show anti-TNFs to be effective in relation to standard care. In view of the available evidence, previous NICE guidance (TA143[17] and TA233[33]) concluded that there was no compelling evidence on which it could reliably distinguish between the anti-TNFs on the basis of clinical effectiveness when making recommendations.

Our analysis, based on the most up-to-date evidence base, aimed to evaluate anti-TNF drugs using indirect comparisons across trials. Within this subsection, alternative assumptions of equivalence in the effectiveness of anti-TNF treatments will be more formally assessed. Note that at this stage each outcome was synthesised independently.

Brief description of synthesis methods

In brief, the synthesis model directly aggregates relative treatment effects, that is log OR for BASDAI 50 response and the difference between treatment and placebo in change in BASDAI from baseline (the data set analysed is shown in *Appendix 12*). In common with the approach implemented in *Chapter 3*, all outcomes are here assumed normally distributed. We implemented alternative models that differ in the way treatment effects are considered; a summary of each is presented below.

Model A1 (treatments: independent; studies: fixed effect): This model considers treatments to be independent: that is, it assumes the effects to differ between treatments. This is a fixed-effect model in that multiple studies evaluating the same treatment are considered to measure the same treatment effect.

Model A2 (treatments: independent; studies: random effects): This model differs from model A1 in that a random effect is assumed to describe the findings of multiple studies evaluating the same treatment.

Model A3 (treatments: equal; studies: fixed effect): This model differs from model A1 in that treatments are not assumed to differ. The model thus evaluates a common relative effectiveness for all anti-TNFs.

Model A4 (treatments: equal; studies: random effects): This model differs from model A3 in that a random effect is assumed to describe the findings of multiple studies evaluating the same treatment.

Model A5 (treatments: exchangeable; studies: fixed effect): This model differs from model A1 in that a random effect is used to describe any differences between treatments (exchangeability is assumed). This model thus assumes the treatments to have a similar, but not equal, effectiveness: there are differences between the effectiveness of treatments that we may not be able to explain but that we should consider.

There is some evidence that health outcomes may depend on patients' characteristics such as age, BASFI score, enthesitis, therapy, CRP level and HLA-B27 genotype.[172] There is, however, no evidence on which factors may modify the effects of treatment with anti-TNFs (note that Lord *et al.* 2010[173] studied predictors of BASDAI 50 response in patients receiving anti-TNFs, but by not including a placebo arm this study was not able to evaluate treatment effect modifiers). To our knowledge, previous meta-analyses of studies in AS have not explored how the effect of treatment may depend on characteristics of the patients or of their disease. Within this modelling approach we explored potential heterogeneity in treatment effects using meta-regression (i.e. potential treatment effect modifiers). We did so by extending the modelling approach in A1 to include treatment effect interactions with baseline characteristics (centred on their means where relevant). We have explored the inclusion of alternative covariates by evaluating the DIC associated with alternative models.

Results of modelling approach A

All models implemented synthesise results on each of the outcomes separately. The results of each modelling approach are shown in *Table 69*.

Models A1 and A2 consider that anti-TNF have distinct relative effects. Applying the assumptions of model A1, adalimumab is expected to be the least effective of the set of treatments analysed in terms of BASDAI 50 (the expected OR is 4.71), but in terms of the differences in change scores it is certolizumab that is expected to be the least effective, with differences of −1.45 and −1.10 in BASDAI and BASFI scores, respectively. It should be noted that studies on golimumab (studies 12[90] and 13[96] in *Table 68*) do not report absolute changes in BASDAI scores, and thus using this modelling approach we were unable to estimate a treatment effect for this outcome measure. Model A2 reports similar results to model A1, but the SE of the estimates is slightly higher, reflecting increased uncertainty because of the use of the random effects to characterise between study results. The DIC is lower in model A1 (52 vs. 57), indicating that model A1 is preferable to model A2.

Outcome 1 is OR for BASDAI 50; outcome 2 is the difference between treatment and placebo on change in BASDAI from baseline; and outcome 3 is the difference between treatment and placebo on change in BASFI from baseline.

Models A3 and A4 consider the treatments as equal in terms of their effectiveness in each of the three outcomes. This means drugs are assumed equally effective and results from trials are pooled together as if these trials evaluated the same drug, which will return more precise estimates (i.e. less uncertainty) and interpretations of this evidence may thus be overly confident. The DIC of these models is substantially lower than that of models A1 and A2, indicating that the data supports the assumption of equivalence, rather than one of independence. As with models A1 and A2, the random-effect assumption was also not deemed worthwhile.

TABLE 69 Assumptions over the relative effectiveness of anti-TNF treatments: results

	A1. Treatment: independent; studies: fixed effect	A2. Treatment: independent; studies: random effects	A3. Treatment: common; studies: fixed effect	A4. Treatment: common; studies: random effects	A5. Treatment: exchangeable; studies: fixed effect
Outcome 1: OR on BASDAI 50	***Median (SD)***	***Median (SD)***	***Median (SD)***	***Median (SD)***	***Median (SD)***
Adalimumab	4.71 (1.00)	4.69 (6.11)	5.21 (0.72)	5.30 (0.98)	5.34 (9.79)
Certolizumab	6.02 (3.33)	6.04 (22.87)			
Etanercept	4.73 (1.43)	4.72 (3.32)			
Golimumab	5.86 (1.81)	6.10 (7.45)			
Infliximab	11.9 (11.94)	12.10 (44.00)			
Outcome 2: change in BASDAI	***Mean (SD)***	***Mean (SD)***	***Mean (SD)***	***Mean (SD)***	***Mean (SD)***
Adalimumab	–1.56 (0.16)	–1.57 (0.27)	–1.66 (0.11)	–1.67 (0.15)	–1.70 (0.87)
Certolizumab	–1.45 (0.37)	–1.46 (0.51)			
Etanercept	–1.76 (0.20)	–1.73 (0.28)			
Golimumab	N/A	N/A			
Infliximab	–2.28 (0.46)	–2.27 (–2.28)			
Outcome 3: change in BASFI	***Mean (SD)***	***Mean (SD)***	***Mean (SD)***	***Mean (SD)***	***Mean (SD)***
Adalimumab	–1.22 (0.18)	–1.18 (0.29)	–1.38 (0.11)	–1.39 (0.13)	–1.41 (0.49)
Certolizumab	–1.10 (0.37)	–1.11 (0.47)			
Etanercept	–1.48 (0.19)	–1.50 (0.24)			
Golimumab	–1.45 (0.20)	–1.44 (0.29)			
Infliximab	–2.16 (0.53)	–2.17 (0.56)			
DIC	52.4	57.0	39.1	44.3	43.6

N/A, not applicable.

Model A5 assumes the treatments to have a similar, but not equal, effectiveness; this model introduces more flexibility than assuming treatment effects to be equal (model A3), but does not fully assume treatments to differ as in model A1. It does imply that there are differences between the effectiveness of treatments that we may not be able to explain but that we should consider. These may be a result of differences between the treatments themselves or because of differences in the design of the trials used to evaluate each treatment. Ignoring such difference could lead to misrepresenting uncertainty, with overprecise results. Model A5 shows a slightly higher DIC than model A3 but this difference is not significant (i.e. lower than 5 units), which means that both models represent equally well the existing data. Given the underlying assumptions, results differ to those of model A3 particularly in relation to the measures of uncertainty. As expected, results from model A3 are more precise than the results of model A5. The results from model A5 in *Table 69* relate to the predictive distribution which reflects uncertainty in all model parameters; in this case, such uncertainty explicitly accounts for the observed differences in the effects treatments.

Despite our preferred summary from model A5 in this evaluation being the common effect for the 'class of drugs' (see *Table 69*), the assumption of treatment effects being drug specific may still retain some plausibility. From model A5, drug-specific estimates can be retrieved (*Table 70*). Within this model drug-specific inferences will borrow strength from the common-class effect and estimates are thus shrunken towards the mean of this class effect (that is estimates are closer to the value reported for the class in *Table 2*).

Explorations of heterogeneity suggested only sex to potentially modify the effect of anti-TNF treatment, specifically for change in BASDAI as outcome; however, when sex is used together with all covariates, such evidence on effect modification disappears.

Interpretation/discussion

The models implemented above show that there is no significant heterogeneity across trials evaluating each treatment, that is the DIC of model A2 is higher than that of model A1, indicating the use of a random effect across studies to be unnecessary.

The statistical analysis has also shown the effectiveness of the different treatments to be similar. This is in line with the published evidence, in AS, that does not demonstrate one anti-TNF treatment to be significantly more effective than another. Specifically, we implemented a model considering a common effect for all anti-TNFs when compared with placebo (model A3). This model shows a better fit than the one estimating a different effect for each anti-TNF (model A1). However, unless we believe this assumption to hold AND the trials to be homogeneous in design and in the populations included, we believe adopting model A3 would misrepresent uncertainty in the estimates.

For this reason, we evaluated an alternative model (model A5) that assumes treatments to have a similar (but not equal) effect. In this model, the treatment effects for the anti-TNFs are assumed to come from a 'common' distribution, assumed Normal with a common mean, that is a 'class effect'. This is an assumption of exchangeability across treatments within the class, which we also refer to as a random-effect distribution. The DIC for this model is not significantly different from that of model A3, and it allows a more appropriate description of the uncertainty over the effects of anti-TNFs. However, it should be noted that this model is not explicit about the source of the differences in the effects of treatments.

The evidence available does not appear to suggest obvious treatment effect modifiers. However, because only aggregate data were available, the results may be prone to ecological fallacy in which statistical associations between variables present or absent at the group level may not be reflective of associations at the individual level.[174]

TABLE 70 Shrunken estimates of treatment effect from model A5

	Model A5		
Treatment	Outcome 1: OR on BASDAI 50 (median, SD)	Outcome 2: change in BASDAI (mean, SD)	Outcome 3: change in BASFI (mean, SD)
Adalimumab	5.05 (0.87)	−1.60 (0.15)	−1.31 (0.16)
Certolizumab	5.42 (1.71)	−1.59 (0.26)	−1.31 (0.23)
Etanercept	5.13 (1.08)	−1.72 (0.17)	−1.43 (0.15)
Golimumab	5.47 (1.25)	−1.69 (0.84)	−1.42 (0.16)
Infliximab	5.70 (3.30)	−1.88 (0.34)	−1.55 (0.33)

Extending the modelling approach to jointly relate outcomes (modelling approach B)

In the previous chapter the two outcomes based on BASDAI scores were synthesised separately; however, BASDAI 50 is the probability of having a reduction in BASDAI score of 50%, and thus it should be possible to relate the proportion of BASDAI 50 responders to the change in absolute BASDAI scores from baseline observed in each study. Such structural constraints should be incorporated into the synthesis, when possible, by expressing it algebraically.[174] Within this chapter, we use this structural relation within the synthesis, allowing change scores from baseline to be informed not only from direct data on this quantity but also from data on BASDAI 50 (*Joint synthesis of Bath Ankylosing Spondylitis Disease Activity Index outcomes*). We then extend the modelling framework further to consider BASFI outcomes [(*Extending the modelling framework to synthesise change in Bath Ankylosing Spondylitis Functional Index scores (modelling approach C)*].

Joint synthesis of Bath Ankylosing Spondylitis Disease Activity Index outcomes

The model implemented here pools the change in BASDAI score from baseline to evaluate the difference between treatment and placebo, using evidence reported in trials directly on the change scores for each arm and also data on BASDAI 50.

The following description briefly explains the approach used to model these data.

Brief description of synthesis methods

Data on the mean change in BASDAI score from baseline, alongside the SE for this measure, were assumed normally distributed (likelihood). The mean of this distribution was the treatment effect, defined as the sum of the change score for the placebo arm plus the difference in change score for the treatments. Some studies also reported the number of responders to BASDAI 50 (a 50% reduction in BASDAI score), out of the total individuals in the study. The likelihood for the BASDAI 50 data was expressed as a binomial distribution. The probability parameter of this distribution was then related to the change score as follows. The BASDAI score at baseline and the change score were assumed correlated using a bivariate normal distribution. To define the bivariate distribution a number of quantities were needed. First, the mean score at baseline was reported in the data and was thus assumed known. Second, the variability on BASDAI score at baseline was assumed equal to that of the change score. This was also reported in the data and was thus assumed known. Finally, the unknown correlation between baseline and change score was estimated within the model by assuming this quantity was independent of study. The correlation parameter was estimated separately for placebo and anti-TNF treatment. Under these assumptions, the probability parameter from BASDAI 50 data was expressed algebraically as a function of the change score. For treatment effects, our preferred approach was to assume a common class effect (i.e. exchangeable effects across treatments, analogous assumption to model A5). See *Appendix 12* for a fuller description of the methods used in analyses.

Results of modelling approach B

The summary results regarding relative treatment effects from this modelling approach are reported in *Table 71* for model B. The treatment effect reported here represents difference between treatment and placebo on BASDAI score changes from baseline.

TABLE 71 Modelling approach B: results

	Estimated	Assumed[a]	Predicted	
Treatment	Difference in change score from baseline, mean (SD)	Probability of having a BASDAI 50 response, placebo, mean (SD)	Probability of having a BASDAI 50 response, anti-TNF, mean (SD)	OR for BASDAI 50 response, anti-TNF vs. placebo, median (SD)
Anti-TNFs	−1.91 (0.48)	0.10 (–)	0.40 (0.08)	5.94 (4.06)

a This figure is based on a BASDAI baseline score of 6.11 (SD 1.56) and a placebo change score of −0.61 (SD 1.44), which represent the average across trials (weighted by number of patients)

With model B, we were now able to consider the evidence from trials only reporting information on BASDAI 50 to estimate the change in BASDAI score, an example being evidence on golimumab. The class effect of anti-TNFs is evaluated to be slightly higher −1.91 (SD 0.48) in comparison to model A5 [reporting a class effect on the change score of −1.70 (SD 0.87)], reflecting the inclusion of BASDAI 50 evidence. By using the indirect evidence on BASDAI 50, model B returns more precise estimates of the pooled change score than model A5 (SE of 0.48 in model B compared with 0.87 in model A5). This modelling approach, despite pooling absolute change scores, can be used to evaluate BASDAI 50 response for a specific baseline BASDAI score and change score in the placebo arm. We assumed a baseline BASDAI score of 6.11 (SD 1.56) and a change score for placebo of 0.61 (SD 1.44), which represent the average across trials (weighted by number of patients). According to these, the assumed probability of having a BASDAI 50 response to placebo is evaluated at 0.10. Based on the change score evaluated in the synthesis model, the probability of having a BASDAI 50 response when on anti-TNFs is evaluated at 0.40 (SD 0.08), which results in an OR for BASDAI 50 response of 5.94 (SD 4.06).

Drug-specific (shrunken) estimates from model B are shown in *Table 72*.

Interpretation/discussion

The current modelling approach, by synthesising together evidence on both BASDAI outcomes, is a theoretically coherent approach to the synthesis. Moreover, it allows using the whole of the evidence on this outcome. In addition, given these outcomes are to be both used in the decision model, the combined synthesis model will generate consistent estimates by considering their structural relation explicitly.

The results of modelling approach B show that using information on BASDAI 50 alongside direct evidence on change scores from baseline results in slightly higher estimates of effectiveness compared with approach A. There are two possible explanations for this. One is that higher treatment effects are observed in the trials only reporting BASDAI 50 compared with the remaining studies. The few studies that only report BASDAI 50 are studies 11,[86] 12[90] and 13;[96] these report ORs for BASDAI of, respectively, 5.9, 4.4 and 10.42. The second possible explanation relates to the assumptions used when defining the relation between the outcomes in the model. While we expected the model to use the BASDAI 50 evidence in such a way that would exactly predict the value of change score observed in the sample, we cannot guarantee this is the case as our analysis is based on assumptions over the distribution of BASDAI scores across patients. Given we did not have access to individual patient data when developing this relationship, and thus the validity of the assumptions of analysis cannot be established. The differences observed are, however, not significant and any misspecification of the model can be thus deemed irrelevant.

Extending the modelling framework to synthesise change in Bath Ankylosing Spondylitis Functional Index scores (modelling approach C)

The models implemented here extend those in *Joint synthesis of Bath Ankylosing Spondylitis Disease Activity Index outcomes* by adding the syntheses of changes in BASFI score. This is of particular relevance to the economic modelling because BASFI scores are used together with BASDAI scores. Given we expect that, within each trial, changes to BASDAI scores to be related to changes in BASFI scores, this section will model the trial evidence to reflect this correlation. *Figure 14* plots the BASDAI change scores against the BASFI change scores observed in the trials, showing support for the existence of correlation.

TABLE 72 Shrunken estimates of treatment effect from model B

Treatment	Change in BASDAI, mean (SD)
Adalimumab	−1.77 (0.25)
Certolizumab	−2.01 (0.37)
Etanercept	−1.88 (0.18)
Golimumab	−1.92 (0.30)
Infliximab	−2.02 (0.32)

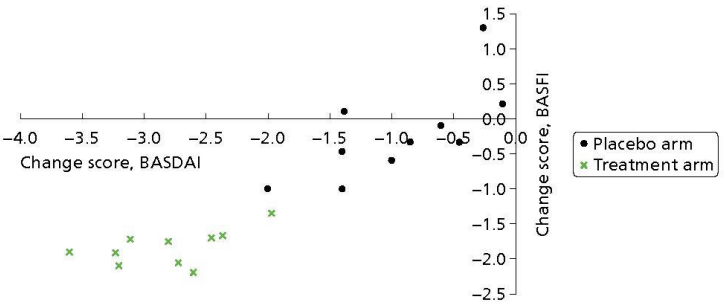


FIGURE 14 Scatterplot of BASDAI and BASFI change scores observed in the trials (AS).

Brief description of synthesis methods

The data on mean change in BASFI score reported in some of the studies was assumed normally distributed. The mean of this distribution was the treatment effect, defined as the sum of the change score for the placebo arm plus the difference in change score for the treatments (analogous to BASDAI). Treatment effects on BASFI were considered correlated to those on BASDAI across trials. The variation in treatment effects for both BASDAI and BASFI, and the correlation parameter between these were estimated from the data. As in *Joint synthesis of Bath Ankylosing Spondylitis Disease Activity Index outcomes*, we assumed again exchangeability across the effects of the different treatments (analogous to models '5' in modelling approach A).

Results of modelling approach C

The results on differences between treatment and placebo on change score form baseline are reported in *Table 73*, for both BASDAI and BASFI scores.

Based on the change score evaluated in the synthesis model, the probability of having a BASDAI 50 response when on anti-TNFs is evaluated at 0.41 (SD 0.05), which returns an OR for BASDAI 50 response of 6.3 (SD 1.56). Note that estimates on BASDAI treatment effects are more precise than in modelling approach B, reflecting the support to inferences from the data on BASFI; the correlation between outcomes observed in the data and allowed in the synthesis model allows inferences in BASDAI to borrow strength from those on BASFI. Drug-specific (shrunken) estimates from model C are shown in *Table 74*.

TABLE 73 Modelling approach C: results

	Estimated	Assumed[a]	Predicted	
Outcome	Difference in change score from baseline, mean (SD)	Probability of having a BASDAI 50 response, placebo, mean (SD)	Probability of having a BASDAI 50 response, anti-TNF, mean (SD)	OR for BASDAI 50 response, anti-TNF vs. placebo, median (SD)
Effect of anti-TNFs on BASDAI	−1.95 (0.30)	0.10 (–)	0.41 (0.05)	6.30 (1.56)
Effect of anti-TNFs on BASFI	−1.40 (0.28)	–	–	–

a Based on a BASDAI baseline score of 6.11 (SD 1.56) and a placebo change score of −0.61 (SD .44), which represent the average across trials (weighted by number of patients).

TABLE 74 Shrunken estimates of treatment effect from model C

Treatment	Change in BASDAI, mean (SD)	Change in BASFI, mean (SD)
Adalimumab	−1.89 (0.22)	−1.34 (0.17)
Certolizumab	−2.02 (0.28)	−1.36 (0.21)
Etanercept	−1.94 (0.18)	−1.43 (0.16)
Golimumab	−1.98 (0.25)	−1.42 (0.17)
Infliximab	−2.03 (0.27)	−1.49 (0.25)

Interpretation/discussion

We hypothesised that treatments improving AS symptoms are expected to affect both disease activity and function, and therefore we expected changes to these two measures to be correlated. We have thus extended the synthesis model to consider BASFI scores. This not only allows all relevant evidence to contribute to the synthesis but also ensures that all measures are synthesised together to reflect the expected correlations between the two outcomes.

The results obtained with this modelling approach for BASDAI outcomes are similar to those of modelling approach B, the difference being that estimates are now more precise because of the borrowing of strength between outcomes.

Non-radiographic axial spondyloarthritis population

This section examines the evidence on the effectiveness of anti-TNFs on the nr-AxSpA population.

Brief description of the data

On the nr-AxSpA population, five RCTs were considered directly relevant to the decision problem (studies 17–21[50,51,58,64,76] in *Table 75*). All studies reported BASDAI and BASFI outcomes and one study did not report BASDAI 50 (study 21[50]).

TABLE 75 Evidence on BASDAI and BASFI-related outcomes for the nr-AxSpA population

Study number	Trial name	Treatments	Number in treatment group	Number in placebo group	BASDAI 50 score	Change BASDAI score	Change BASFI score
17	Haibel 2008[52]	Adalimumab	22	24	✗	✗	✗
18	ABILITY-1 2013[58]	Adalimumab	69	73	✗	✗	✗
19	RAPID-axSpA 2014[64]	Certolizumab pegol	46 + 51	50	✗	✗	✗
20	Dougados 2014[76]	Etanercept 50	106	109	✗	✗	✗
21	Barkham 2009[50]	Infliximab	20	20		✗	✗

✗ denotes whether or not each outcome is measured in each of the studies.

Description of approaches to the synthesis

To synthesise these data we used the same implementation and software specifications as described in *Chapter 5, Ankylosing spondylitis population*. Analyses explored two different scenarios to consider these data:

- scenario 1 – data from nr-AxSpA trials were considered in isolation
- scenario 2 – data from AS population were also used, no difference between the populations was assumed.

All models implemented here jointly synthesise BASDAI and BASFI outcomes [for our preferred modelling approach, C, see description in *Extending the modelling framework to synthesise change in Bath Ankylosing Spondylitis Functional Index scores (modelling approach C)*].

Results of the synthesis

In what concerns scenario 1, in which only data from the nr-AxSpA trials has been considered, we implemented two models: one assuming an equal effect across treatments and another assuming exchangeable treatment effects. Both models represented the data equally well (DIC of 87.6 vs. 88.7), and thus we only present results in *Table 76* for the latter model [the preferred model; see *Exploring assumptions for the relative effectiveness of individual anti-tumour necrosis factor treatments (modelling approach A)*]. Results are qualitatively similar to those in AS but slightly lower estimates for both change scores: BASDAI –1.95 in AS and –1.86 in the nr-AxSpA population; and BASFI –1.40 in AS and –1.30 in the nr-AxSpA population. The uncertainty over these estimates is higher in the nr-AxSpA population, which was expected as the number of trials (and overall number of patients in the set of trials) is substantially lower.

When the data from the nr-AxSpA trials were considered together with data on AS (scenario 2), inferences were more precise. As treatment effects in AS trials are not significantly different from those observed in the nr-AxSpA population, pooled treatment effect estimates do not differ significantly from those reported in AS.

TABLE 76 Nr-axSpA population: results

	Estimated	Assumed[a]	Predicted	
	Difference in change score from baseline, mean (SD)	Probability of having a BASDAI 50 response, placebo, mean (SD)	Probability of having a BASDAI 50 response, anti-TNF, mean (SD)	OR for BASDAI 50 response, anti-TNF versus placebo, median (SD)
Scenario 1: data from nr-AxSpA trials				
Effect of anti-TNFs on BASDAI	–1.86 (0.79)	0.20 (–)	0.53 (0.13)	4.39 (6.59)
Effect of anti-TNFs on BASFI	–1.30 (0.84)	–	–	–
Scenario 2: data from AS and nr-AxSpA trials, no difference between the populations				
Effect of anti-TNFs on BASDAI	–1.97 (0.32)	0.20 (–)	0.55 (0.06)	4.94 (1.48)
Effect of anti-TNFs on BASFI	–1.37 (0.3)	–	–	–

a Based on a BASDAI baseline score of (AiC information has been removed) and a placebo change score of (AiC information has been removed), which represent the results seen in the certolizumab trial (RAPID-axSpA[64]).

Interpretation/discussion

The evidence base of the effect of anti-TNFs in the nr-AxSpA population consists of five trials that observed four treatments and CC in a total of 590 patients. The effect measures pooled across the five trials were not significantly different from the outcomes expected in the AS population. Thus, it may be reasonable to consider the evidence in nr-AxSpA and AS together for inferences over treatment effects.

Bath Ankylosing Spondylitis Disease Activity Index and Bath Ankylosing Spondylitis Functional Index scores conditional on Bath Ankylosing Spondylitis Disease Activity Index response

We previously highlighted that NICE guidance determines that BASDAI 50 at 12 weeks defines treatment continuation with anti-TNFs in clinical practice. Given much of the evidence on prognosis, costs and utility scores links to the absolute values of BASDAI and BASFI scores, it is important to consider absolute changes in BASDAI and BASFI separately for responders and non-responders, that is the conditional scores. However, the published clinical effectiveness evidence does not report the conditional scores. In this section we use the results from the extended synthesis model to evaluate the conditional scores by simulating BASDAI and BASFI scores for two equivalent cohorts of patients, one treated with an anti-TNF and the other with conventional therapy.

Brief description of methods

From the inferences obtained using the synthesis model above, it is possible to derive the conditional change score in responders and non-responders using simulation. While the synthesis focuses on the pooling of mean estimates of change scores and proportion of responders to BASDAI 50, to derive conditional mean scores there is the need to consider the distributions at the individual patient level. Hence, conditional scores could not directly be derived from the synthesis, but through a simulation procedure based on the assumptions and results of the synthesis model. The simulation procedure is described in detail in *Appendix 5*. Briefly, we used a simulation sample size of 10,000 patients. Given results depend on the baseline distributions of BASDAI and BASFI scores and on the change scores from baseline for placebo, we used the averages across trials (weighted by the number of patients in each trial) in AS. Baseline BASDAI scores were thus assumed normally distributed with mean 6.11 and SD 1.56; the change from baseline for placebo was simulated from a normal distribution with mean –0.61 and SD 1.44. For BASFI, the baseline was assumed to have a mean of 5.27 and a SD 1.79 and the change from baseline for placebo a mean of –0.19 and a SD 0.22. The correlation between baseline BASFI and BASDAI scores was valued at (AiC information has been removed). This value was based on the sample correlation on BASDAI and BASFI at baseline from etanercept studies [the individual patient data were available in the Microsoft Excel® (Microsoft Corporation, Redmond, WA, USA) file for the etanercept submission]; the Spearman's rank correlation coefficient was (AiC information has been removed) in 314-EU study[167] in AS and (AiC information has been removed) in 1031 study[166] in nr-AxSpA.

Results for ankylosing spondylitis

The conditional change scores derived from the synthesis model (and underlying assumptions) are reported in *Table 77*. While it is natural to consider that conditional change in BASDAI scores differ between respondents and non-respondents, differences in the baseline of respondents and non-respondents may be less intuitive. These are, however, natural. If we consider two patients that obtained the same change score in BASDAI from anti-TNF treatment, for example –2 units, but one started with a baseline of 4 and another with a baseline of 5, the first would be considered a responder and the second would not. For this reason, respondents are expected to have a lower BASDAI than non-responders. Results of the prediction of conditional scores using the synthesis model are presented in *Table 77*.

TABLE 77 Conditional scores predicted for the AS population using the synthesis model

	BASDAI		BASFI	
	Control	Treatment	Control	Treatment
Base case				
% responders to BASDAI 50	0.09	0.42		
Change in score				
Responders	–2.89	–3.86	–1.72	–3.08
Non-responders	–0.36	–1.64	–0.04	–0.44
All	–0.59	–2.57	–0.19	–1.55
Baseline				
Responders	4.01	4.80	3.52	4.20
Non-responders	6.33	7.08	5.46	6.07
All	6.12	6.12	5.28	5.28

Results show, as expected, that the change in BASDAI score for respondents is more negative than the mean change score (–3.86 for the 42% predicted anti-TNF responders vs. –2.63 for all anti-TNF users; in the control arm, responders were predicted to have a change score of –2.70 vs. –0.66 for all participants). Non-respondents were still expected to have a negative change score in both arms revealing some level of symptom control but this was lower than the mean (–1.73 vs. –2.63 for anti-TNF users and –0.45 vs. –0.66 in control arm). The baseline BASDAI and BASFI were predicted to be lower for respondents than non-respondents (e.g. the BASDAI baseline for responders to treatment was 4.76 in respondents when the group baseline was 6.08).

We requested the conditional data from the pivotal trials in AS from each manufacturer. These data were subsequently provided by all manufacturers for their pivotal trials. Conditional scores observed in the trials are summarised in *Table 78*. The results show that there are some differences between the conditional results predicted using the synthesis and the ones observed in trials. Differences are especially relevant for the conditional baseline scores; although the synthesis model predicts, for example, that treated patients that respond have a baseline BASDAI score of 4.76 and those that do not respond have a baseline score of 7.03, the trials show much smaller differences. Despite incorporating all evidence available at the aggregate level, the predictive ability of the conditional baseline score from the synthesis could only be improved if we had access to the individual patient data, as this methodology is strongly dependent on assumptions over the distribution of scores across patients.

TABLE 78 Conditional scores observed in trials in AS

	BASDAI		BASFI	
	Control	Treatment	Control	Treatment
ATLAS[61] trial (adalimumab, studies = 4)				
% responders to BASDAI 50	0.16	0.46		
Change in score				
Conditional on response	–4.5	–4.64	–2.74	–2.92
Conditional on non-response	–0.2	–0.82	–0.17	–0.72
Total	–0.90	–2.58	–0.59	–1.73
Baseline				
Conditional on response	6.31	6.14	4.50	4.53
Conditional on non-response	6.37	6.35	5.91	5.78
Total	6.36	6.25	5.68	5.21
GO-RAISE[90] (golimumab, studies = 12)				
% responders to BASDAI 50	0.15	0.46		
Change in score				
Conditional on response	–4.25	–4.74	–1.80	–3.03
Conditional on non-response	–0.18	–1.22	0.38	–0.53
Total	–0.81	–2.84	0.05	–1.68
Baseline				
Conditional on response	6.52	6.25	3.56	4.45
Conditional on non-response	6.63	6.69	5.39	5.48
Total	6.61	6.49	5.11	5.01
RAPID-axSpA[64] (certolizumab, studies = 5)				
% responders to BASDAI 50	CiC information has been removed	CiC information has been removed		
Change in score				
Conditional on response	CiC information has been removed	CiC information has been removed	CiC information has been removed	CiC information has been removed
Conditional on non-response	CiC information has been removed	CiC information has been removed	CiC information has been removed	CiC information has been removed
Total	CiC information has been removed	CiC information has been removed	CiC information has been removed	CiC information has been removed
Baseline				
Conditional on response	CiC information has been removed	CiC information has been removed	CiC information has been removed	CiC information has been removed
Conditional on non-response	CiC information has been removed	CiC information has been removed	CiC information has been removed	CiC information has been removed
Total	CiC information has been removed	CiC information has been removed	CiC information has been removed	CiC information has been removed

TABLE 78 Conditional scores observed in trials in AS (*continued*)

	BASDAI		BASFI	
	Control	Treatment	Control	Treatment
314-EU study[167] (etanercept, studies = 11)[a]				
% responders to BASDAI 50	CiC information has been removed	CiC information has been removed		
Change in score				
Conditional on response	CiC information has been removed	CiC information has been removed	CiC information has been removed	CiC information has been removed
Conditional on non-response	CiC information has been removed	CiC information has been removed	CiC information has been removed	CiC information has been removed
Total	CiC information has been removed	CiC information has been removed	CiC information has been removed	CiC information has been removed
Baseline				
Conditional on response	CiC information has been removed	CiC information has been removed	CiC information has been removed	CiC information has been removed
Conditional on non-response	CiC information has been removed	CiC information has been removed	CiC information has been removed	CiC information has been removed
Total	CiC information has been removed	CiC information has been removed	CiC information has been removed	CiC information has been removed

a Pooled results for etanercept arms (etanercept 25 mg twice weekly and etanercept 50 mg once weekly).
For the trials of adalimumab and etanercept were week 12 responders, for golimumab they were week 14 responders (week-12 data for week-14 responders is available but not reported in the table).

Results for non-radiographic axial spondyloarthritis

The conditional results were also predicted for the nr-AxSpA population using both scenarios implemented of the synthesis model.

For this population, conditional data were provided by only two manufacturers (Pfizer and AbbVie). Conditional scores observed are summarised in *Tables 79* and *80*.

Prediction results are consistent with those in AS, and the differences between the conditional results predicted using the synthesis and the ones observed in trials are also present in this analysis.

Interpretation/discussion

Conditional scores predicted using synthesis model C differ from those seen in the trials. Differences are probably because of distributional assumptions over the baseline and change scores. Only with access to the individual patient data could such predictions be improved. Note that the synthesis model itself does not rely as heavily on such assumptions, and thus any concerns should not be transposed to the results obtained in *Ankylosing spondylitis population* and *Non-radiographic axial spondyloarthritis population*.

TABLE 79 Conditional scores predicted for the nr-AxSpA population using results and assumptions of the synthesis model[a]

	BASDAI		BASFI	
	Control	Treatment	Control	Treatment
Scenario 1				
% responders to BASDAI 50	AiC information has been removed	AiC information has been removed		
Change in score	AiC information has been removed	AiC information has been removed		
Responders	AiC information has been removed	AiC information has been removed	AiC information has been removed	AiC information has been removed
Non-responders	AiC information has been removed	AiC information has been removed	AiC information has been removed	AiC information has been removed
All	AiC information has been removed	AiC information has been removed	AiC information has been removed	AiC information has been removed
Baseline	AiC information has been removed	AiC information has been removed		
Responders	AiC information has been removed	AiC information has been removed	AiC information has been removed	AiC information has been removed
Non-responders	AiC information has been removed	AiC information has been removed	AiC information has been removed	AiC information has been removed
All	AiC information has been removed	AiC information has been removed	AiC information has been removed	AiC information has been removed
Scenario 2				
% responders to BASDAI 50	AiC information has been removed	AiC information has been removed		
Change in score	AiC information has been removed	AiC information has been removed		
Responders	AiC information has been removed	AiC information has been removed	AiC information has been removed	AiC information has been removed
Non-responders	AiC information has been removed	AiC information has been removed	AiC information has been removed	AiC information has been removed
All	AiC information has been removed	AiC information has been removed	AiC information has been removed	AiC information has been removed
Baseline	AiC information has been removed	AiC information has been removed		
Responders	AiC information has been removed	AiC information has been removed	AiC information has been removed	AiC information has been removed
Non-responders	AiC information has been removed	AiC information has been removed	AiC information has been removed	AiC information has been removed
All	AiC information has been removed	AiC information has been removed	AiC information has been removed	AiC information has been removed

a Based on a BASDAI baseline score of (AiC information has been removed) a placebo change in BASDAI score of (AiC information has been removed), a BASFI baseline score of (AiC information has been removed) and a placebo change in BASFI score of (AiC information has been removed), which represent the results seen in the certolizumab trial (RAPID-axSpA[64]).

TABLE 80 Conditional scores observed in trials in nr-AxSpA

	BASDAI		BASFI	
	Control	Treatment	Control	Treatment
ABILITY-1[58] trial (adalimumab, studies = 18)				
% responders to BASDAI 50	0.14	0.40		
Change in score				
Conditional on response	–3.9	–4.79	–2.78	–2.75
Conditional on non-response	–0.69	–0.55	–0.40	–0.32
Total	–1.16	–2.23	–0.75	–1.29
Baseline				
Conditional on response	5.64	6.21	4.37	3.60
Conditional on non-response	6.46	6.53	4.91	4.97
Total	6.34	6.40	4.83	4.43
RAPID-axSpA[64] (certolizumab, study = 19)				
% responders to BASDAI 50	CiC information has been removed	CiC information has been removed		
Change in score				
Conditional on response	CiC information has been removed	CiC information has been removed	CiC information has been removed	CiC information has been removed
Conditional on non-response	CiC information has been removed	CiC information has been removed	CiC information has been removed	CiC information has been removed
Total	CiC information has been removed	CiC information has been removed	CiC information has been removed	CiC information has been removed
Baseline				
Conditional on response	CiC information has been removed	CiC information has been removed	CiC information has been removed	CiC information has been removed
Conditional on non-response	CiC information has been removed	CiC information has been removed	CiC information has been removed	CiC information has been removed
Total	CiC information has been removed	CiC information has been removed	CiC information has been removed	CiC information has been removed
EU-1031[166] (etanercept, studies = 20)				
% responders to BASDAI 50	CiC information has been removed	CiC information has been removed		
Change in score				
Conditional on response	CiC information has been removed	CiC information has been removed	CiC information has been removed	CiC information has been removed
Conditional on non-response	CiC information has been removed	CiC information has been removed	CiC information has been removed	CiC information has been removed
Total	CiC information has been removed	CiC information has been removed	CiC information has been removed	CiC information has been removed

continued

TABLE 80 Conditional scores observed in trials in nr-AxSpA (*continued*)

	BASDAI		BASFI	
	Control	Treatment	Control	Treatment
Baseline				
Conditional on response	CiC information has been removed	CiC information has been removed	CiC information has been removed	CiC information has been removed
Conditional on non-response	CiC information has been removed	CiC information has been removed	CiC information has been removed	CiC information has been removed
Total	CiC information has been removed	CiC information has been removed	CiC information has been removed	CiC information has been removed

Etanercept and adalimumab studies included week-12 responders. Pfizer reported results only for etanercept 50 mg.

Discussion/conclusion

The analyses developed in this section focused on extending the synthesis evidence on the short-term clinical effectiveness of anti-TNF drugs in *Chapter 3* that considered individually multiple outcomes of interest reported in the trials, namely the mean change in BASDAI scores at 12 weeks, the proportion of BASDAI 50 responders (i.e. those who had, at 12 weeks, a change in the baseline BASDAI score of 50% or more) and the mean change in BASFI scores at 12 weeks.

Initially, within such a univariate framework, we further explored assumptions over the relative effectiveness of anti-TNFs. We evaluated the possibility of the evidence suggesting treatment effects to be independent, equal or similar effects (treatment effects were assumed to come from a 'common' distribution, i.e. a 'class effect'). Independence was ruled out through statistical checks of goodness of fit; this is in line with the published evidence, in AS, that does not demonstrate one anti-TNF treatment to be significantly more effective than another. The data were as well represented by the other two models. However, unless we believe the equality assumption to hold AND the trials to be homogeneous in design and in the populations included, assuming equality in treatment effects will provide overprecise estimates. For this reason, our preferred assumption was that of similarity; however, it should be noted that this model is not explicit about the source of the differences in the effects of treatments. Whereas heterogeneity may be a plausible explanation, further research needs to examine data at the individual patient level to avoid the potential for ecological bias.

We also extended the synthesis in a way that allowed multiple outcomes to be jointly modelled. We did so by (1) structurally relating the BASDAI-based outcomes, allowing for trials reporting BASDAI 50 to inform BASDAI change scores, and (2) by concomitantly synthesising BASFI outcomes, allowing correlation between outcomes and the borrowing of strength between results to BASDAI and BASFI. For these reasons, the synthesis model developed here more directly addresses the decision problem. It also generates appropriate effect-size estimates and their associated uncertainty to inform the main input parameters of the economic model.

In the decision model, treatment continuation is determined by response to BASDAI 50 at 12 weeks. Given that prognosis, costs and QALYs are determined by absolute BASDAI and BASFI scores, it is important to evaluate the absolute change in BASDAI and BASFI separately for responders and non-responders: that is, the conditional scores. We used the results from the extended synthesis model to develop a simulation model that allowed prediction of the conditional scores. The results obtained differ from those seen in three pivotal trials (data provided by the manufacturers on request), probably because of distributional assumptions over the baseline and change scores. Only with access to the individual-patient data such predictions could be improved.

Chapter 6 Independent economic assessment: York model

Overview

Chapter 4 indicates that there are significant conceptual concerns and uncertainties arising from previously published studies and the submissions made by manufacturers. For this reason, it has been necessary to develop a de novo model (hereafter referred to as the 'York model'). Although it shares some of the assumptions and parameter estimates from the manufacturer models, it has a different conceptual structure and applies a more generalised framework for the synthesis of data from the double-blind periods of existing RCTs, combined with a more explicit approach to modelling the progressive nature of AS and nr-AxSpA and the potential impact of the anti-TNFs.

The aim of the York model is to assess the cost-effectiveness of adalimumab, certolizumab, etanercept, golimumab and infliximab, in accordance with their respective licences, for the treatment of AS and nr-AxSpA. The model uses short-term trial data, based on the extended evidence synthesis, to model the response of patients to TNF-α inhibitor therapy at 12 weeks based on BASDAI 50 measured in the trials. In contrast to the models submitted by the manufacturers, the York model is based on an assumption of similar (but not identical) effects for the alternative biologics based on the results of the extended synthesis reported in *Chapter 5*.

In common with all existing cost-effectiveness studies, measures of disease activity (BASDAI) and functioning (BASFI) are used to characterise the chronic, progressive nature of AS and nr-AxSpA and the effect of anti-TNFs. However, the York model uses an alternative conceptual model applied to estimate longer-term BASFI scores. The effect of response to TNF-α therapy is modelled in terms of the short- and longer-term impact on BASDAI and BASFI scores.

The NHS and PSS costs are based on the cost of the TNF-α therapies (acquisition, administration and monitoring) and disease costs linked to BASFI scores. HRQoL, in terms of utility, is based on both BASDAI and BASFI scores. Health effects are subsequently expressed in terms of QALYs. Both costs and QALYs are discounted at 3.5% per annum. Costs are presented based on current prices.

The model is developed in accordance with the NICE reference case. The model has a lifetime horizon (60 years) and considers costs from the perspective of the NHS and PSS.

Contribution of the York model

Although the York model shares some of the assumptions and parameters from existing studies and manufacturer's submissions, it also provides a number of significant developments to existing cost-effectiveness analyses. First, the short-term clinical effectiveness inputs are based on an evidence-synthesis approach which is based on all available trial data for each biological therapy and which jointly synthesises 'related' parameters ensuring uncertainty is more appropriately characterised. Second, the evidence-synthesis approach is more explicitly linked to the decision problem and the requirements of the economic model, that is the model requires estimates of response and the impact on BASDAI/BASFI conditional upon this. As the conditional response scores are not conventionally reported in existing publications, existing models have largely been based on selective approaches (i.e. using conditional scores from single studies or

assumptions) or appear to have ignored the conditional scores entirely and instead use estimates from longer-term follow-up and/or open-label sources (i.e. implicitly assuming that patients who continue to participate in longer-term follow-up and open-label sources are more likely to be responders than patients who do not). Neither approach appears satisfactory in terms of meeting the requirements of the economic model and ensuring that all relevant evidence is considered. The evidence synthesis approach which underpins the York model is based on a joint synthesis of related parameters which makes fuller use of existing evidence and which can more appropriately estimate the input parameters which are required to populate existing models and better characterise the uncertainty surrounding these.

Another important development of the York model is the approach to modelling longer-term BASFI changes over time to characterise the progressive nature of AS and nr-AxSpA. In previous sections we highlighted our concerns over how this has been previously modelled and the implicit assumptions underlying the effect of anti-TNFs (i.e. potential disease modification properties resulting in halting further 'progression', or reducing the rate of progression, while patients respond and continue to receive anti-TNFs). Within the York model, we attempt to model the impact of different processes on BASFI over time, relating the changes more explicitly to the existing clinical effectiveness data for anti-TNFs on these different processes. Specifically, we consider the independent effects on BASFI because of disease activity (BASDAI) and the extent and progression of radiographic disease (as measured by the mSASSS) for AS. For the nr-AxSpA population, we assume a similar underlying clinical process relating to BASFI.

This approach confers several advantages over current approaches by linking changes in BASFI to a more explicit clinical/biological process and facilitating a more formal consideration of the potential impact of anti-TNFs on BASFI, via the specific effects these drugs have on the different processes which independently relate to this parameter. This approach allows consideration of the impact on BASFI that might be achieved via symptomatic improvements (i.e. in terms of reductions in disease activity) and those which might be conferred by disease modification properties (i.e. the effect on the likelihood and/or rate of further radiographic progression). The latter aspect is particularly important given the increasing amount of published evidence reported on the potential impact of anti-TNFs on radiographic progression which has not been formally considered or incorporated within existing cost-effectiveness studies.

Comparators

Table 81 summarises the comparators included in each of the populations, in line with the relevant existing (or likely to be granted by the time of the NICE appraisal) marketing authorisations for each manufacturer.

TABLE 81 Comparators evaluated in the different indications

Comparator	Manufacturer	AS	nr-AxSpA
CC	–	Yes	Yes
Adalimumab	AbbVie	Yes	Yes
Certolizumab	UCB	Yes	Yes
Etanercept	Pfizer	Yes	Yes
Golimumab	Merck Sharp & Dohme	Yes	No
Infliximab	Merck Sharp & Dohme	Yes	No

Model structure

The York model is a cohort model and takes the form of a modified decision tree for AS and nr-AxSpA. A simplified version of the structure is shown in *Figure 15*. A similar structure has been previously been used to evaluate the cost-effectiveness of anti-TNFs in psoriatic arthritis.[175]

For the alternative TNF-α inihibitors, initial response is determined on the basis of a short-term BASDAI 50 response (12 weeks). For those who respond, there is then an ongoing risk of withdrawal of treatment at any time point in the model. Initial or later treatment failures are assumed to move on to CC. The use of BASDAI 50 is consistent with existing BSR guidelines and previous NICE appraisals for AS.[17,33,169] Ensuring consistency in the response measure between the various appraisals provides a more comparable basis for exploring any subsequent differences in results. In addition, using BASDAI 50 as a response measure for the economic model maximises the evidence base used to inform the various clinical-effectiveness parameters required and, as outlined in *Chapter 5*, uses the same clinical constructs to inform response and subsequent BASDAI score changes.

Those patients who receive anti-TNFs will experience an initial improvement which is based on results of the evidence synthesis (average of mean change in BASDAI and BASFI scores estimated for responders and non-responders). From week 12, patients who continue to receive anti-TNFs are assigned the conditional mean change in BASDAI and BASFI scores estimated from the evidence synthesis which is assumed to remain constant for the treatment duration period. In addition to this initial improvement in BASDAI and BASFI, patients continuing on anti-TNFs treatment are also assumed to experience a slower progression rate in BASFI as long as they are responding (see *Longer-term discontinuation*).

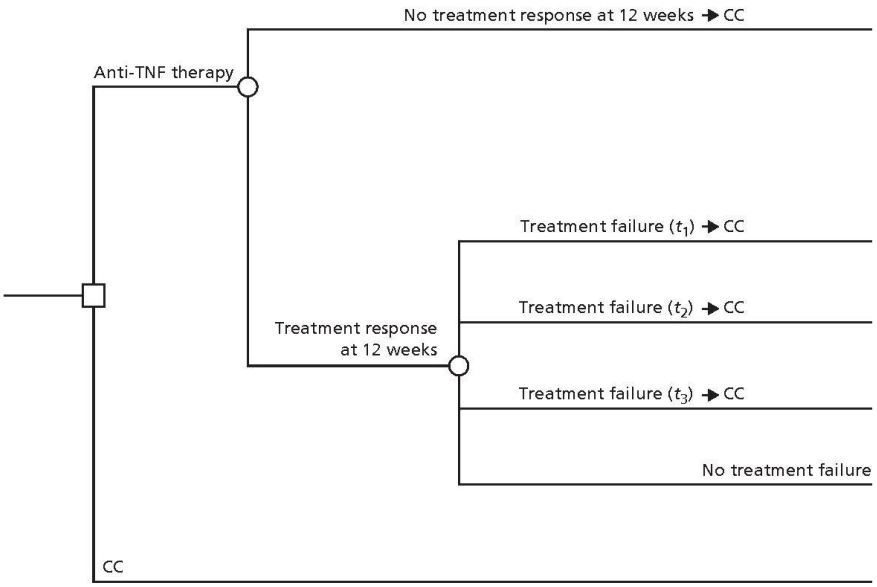


FIGURE 15 A simplified schematic of the York model structure.

Patients who fail on TNF-α inhibitor therapy after the initial (12-week) period will experience some form of rebound in terms of BASDAI/BASFI score, but the trial data are too short term to be able to characterise this accurately. The model, therefore, considers two rebound scenarios:

1. Rebound equal to gain (BASDAI and BASFI). When patients fail therapy (after initially responding), their BASDAI and BASFI deteriorates by the same amount by which it improves when they responded to therapy.
2. Rebound back to natural history/CC (BASFI only). When patients fail therapy (after initially responding), their BASFI deteriorates to the level and subsequent trajectory it would have been had they not initially responded to therapy. As BASDAI is not assumed to progress over time on CC, the same assumptions are applied to BASDAI in both scenarios.

Given the absence of evidence on rebound, both scenarios (rebound equal to gain and rebound back to natural history) are presented as the 'best-case' and 'worst-case' scenarios possible. In other words, the reality regarding rebound is likely to be somewhere between these two scenarios which should, therefore, be seen as the limits.

Importantly, the York model explores the impact of assuming different baseline BASDAI and BASFI scores for responders and non-responders. Hence, in contrast to existing models, the York model assumes that response is unlikely to be independent of baseline patient characteristics and hence the baseline characteristics of responders/non-responders to anti-TNFs may be systematically different from each other. Importantly, the results from the extended synthesis model estimated higher baseline BASDAI and BASFI scores for non-responders vis-à-vis responders and a similar relationship was also reported by those manufacturers who provided conditional response data requested by the assessment group. Consequently, assuming that non-responders revert back to the 'average' of the baseline BASDAI/BASFI score of all patients randomised to receive TNF-α inhibitor treatment, or the 'average' of patients receiving CC, is likely to be overly optimistic towards the subsequent cost-effectiveness of anti-TNFs. The model thus use different baselines for responders and non-responders (at 12 weeks) and at the point of discontinutation patients are assumed to revert to their respective baseline BASDAI and BASFI scores (i.e. at 12 weeks non-responders revert back to the non-responder baseline and after 12 weeks patients who subsequently discontinue from their TNF-α therapy revert back to their responder baseline). The impact of using these data is explored as part of the sensitivity analysis.

Patients are at risk of all-cause mortality at every time point in the model but no differential mortality risk between the therapies being evaluated. Aside from the cost of the TNF-α therapies themselves (i.e. acquisition, administration, monitoring and AEs), all other costs of AS and nr-AxSpA are assumed to vary according to BASFI score. Costs are presented based on current prices. HRQoL (in terms of utility) is implemented as a function of BASDAI and BASFI scores.

Model input parameters

The parameter estimates used in the York model, together with their sources, are detailed in *Tables 82* and *83*.

Baseline patient characteristics

Baseline characteristics applied to the AS and nr-AxSpA populations are summarised in *Tables 82* and *83*, respectively.

TABLE 82 List of parameter estimates used in the York model: AS population

Parameter	Mean value	SE	Distribution	Source
Annual discount rate costs/QALYs	3.5%	–	Fixed	–
Time horizon (years)	60	–	Fixed	–
Cycle length (years)	0.25	–	Fixed	–
Baseline patient characteristics				
Average age (years)	40	–	Fixed	Assumption
Proportion male	0.7	–	Fixed	Assumption
Average weight (kg)	73	–	Fixed	Assumption
Average baseline BASDAI	6.12	N/A	Derived from responder and non-responder baseline	Evidence synthesis (see *Chapter 5*)
Average baseline BASFI	5.28	N/A		
Baseline BASDAI CC responders	4.01	N/A	From evidence synthesis	Evidence synthesis (see *Chapter 5*)
Baseline BASDAI CC non-responders	6.33	N/A	From evidence synthesis	
Baseline BASFI CC responders	3.52	N/A	From evidence synthesis	
Baseline BASFI CC non-responders	5.46	N/A	From evidence synthesis	
Baseline BASDAI anti-TNF responders	4.80	N/A	From evidence synthesis	
Baseline BASDAI anti-TNF non-responders	7.08	N/A	From evidence synthesis	
Baseline BASFI anti-TNF responders	4.20	N/A	From evidence synthesis	
Baseline BASFI anti-TNF non-responders	6.07	N/A	From evidence synthesis	
Response (12-week BASDAI 50)				
Anti-TNF	42.0%	N/A	From evidence synthesis	Evidence synthesis (see *Chapter 5*)
Conventional therapy	9.1%	N/A	From evidence synthesis	
Treatment effect				
Initial BASDAI Changel Tx response: anti-TNF	−3.86	N/A	From evidence synthesis	Evidence synthesis (see *Chapter 5*)
Initial BASDAI Changel Tx response: CC	−2.89	N/A	From evidence synthesis	
Initial BASDAI Changel Tx no response: anti-TNF	−1.64	N/A	From evidence synthesis	
Initial BASDAI Changel Tx no response: CC	−0.36	N/A	From evidence synthesis	

continued

TABLE 82 List of parameter estimates used in the York model: AS population (*continued*)

Parameter	Mean value	SE	Distribution	Source
Initial BASFI Changel Tx response: anti-TNF	−3.08	N/A	From evidence synthesis	
Initial BASFI Changel Tx response: CC	−1.72	N/A	From evidence synthesis	
Initial BASFI Changel Tx no response: anti-TNF	−0.44	N/A	From evidence synthesis	
Initial BASFI Changel Tx no response: CC	−0.04	N/A	From evidence synthesis	
Long-term annual BASFI progression				
BASFI annual progression anti-TNF	0.034	–	Derived from probabilistic inputs (the following four rows)	–
BASFI annual progression CC	0.082	–		–
Annual rate of mSASSS change for mSASSS ≥ 10	1.44	0.133	Normal	Ramiro *et al.* (2013)[148]
BASFI change with 1 unit change in mSASSS	0.057	0.0049	Normal	Landewe *et al.* (2009)[10]
Treatment effect on progression (relative risk)	0.42	0.122	Normal	Haroon *et al.* (2013)[121]
Time to treatment effect (years)	4	–	–	Haroon *et al.* (2013),[121] Baraliakos *et al.* (2014)[122]
Long-term annual BASDAI progression				
BASDAI annual progression anti-TNF	0	N/A		
BASDAI annual progression CC	0	N/A		
Annual withdrawal probability				
Constant rate of annual withdrawal	0.11	AiC information has been removed	Log-normal; from exponential model [coefficient (AiC information has been removed); SE (AiC information has been removed)]	Pfizer submission[34]
Mortality				
SMR women	1.38	0.163	Normal	Bakland (2011)[16]
SMR men	1.63	0.163	Normal	Bakland (2011)[16]

TABLE 82 List of parameter estimates used in the York model: AS population (*continued*)

Parameter	Mean value	SE	Distribution	Source
Quality of life				
Intercept	(AiC information has been removed)	Uncertainty from reported variance–covariance matrix	Multivariate normal	Pfizer submission[34]
BASDAI coefficient	(AiC information has been removed)		Multivariate normal	
BASFI coefficient	(AiC information has been removed)		Multivariate normal	
BASDAI^2 coefficient	(AiC information has been removed)		Multivariate normal	
BASFI^2 coefficient	(AiC information has been removed)		Multivariate normal	
Initial 12-week period costs (drug + initiation + administration)				
Adalimumab	2422	–	Fixed	As discussed in *Resource use and costs*
Certolizumab pegol	3884	–	Fixed	
Etanercept	2454	–	Fixed	
Golimumab	2415	–	Fixed	
Infliximab	6878	–	Fixed	
Certolizumab pegol PAS	309	–	Fixed	
Subsequent 12-week costs (drug + monitoring + administration)				
Adalimumab	2171	–	Fixed	As discussed in *Resource use and costs*
Certolizumab pegol	2203	–	Fixed	
Etanercept	2203	–	Fixed	
Golimumab	2164	–	Fixed	
Infliximab	3435	–	Fixed	
Certolizumab pegol PAS	2203	–	Fixed	
Disease-related costs: annual				
Intercept	1284	0.165	Log-normal	OASIS data,[118] AbbVie submission[34]
BASFI coefficient	0.213	0.038	Normal	
AE costs (£ per patient)				
Year 1	18.2	–	Fixed	Excess rates for anti-TNFs from Cochrane review,[137] costs from NHS Reference Costs 2012/13[176]
Subsequent years	0	–	Fixed	

N/A, not applicable; Tx, treatment.

TABLE 83 List of parameter estimates used in the York model: nr-AxSpA population

Parameter	Mean value	SE	Distribution	Source
Annual discount rate costs/QALYs	3.5%	–	Fixed	–
Time horizon (years)	60	–	Fixed	–
Cycle length (years)	0.25	–	Fixed	–
Baseline patient characteristics				
Average age (years)	40	–	Fixed	Assumption
Proportion male	0.5	–	Fixed	Assumption
Average weight (kg)	73	–	Fixed	Assumption
Average baseline BASDAI	6.42	N/A	Derived from responder and non-responder baseline	Evidence synthesis (see *Chapter 5*)
Average baseline BASFI	4.92	N/A		
Baseline BASDAI CC responders	4.54	N/A	From evidence synthesis	Evidence synthesis (see *Chapter 5*)
Baseline BASDAI CC non-responders	6.86	N/A	From evidence synthesis	
Baseline BASFI CC responders	2.95	N/A	From evidence synthesis	
Baseline BASFI CC non-responders	5.38	N/A	From evidence synthesis	
Baseline BASDAI anti-TNF responders	5.45	N/A	From evidence synthesis	
Baseline BASDAI anti-TNF non-responders	7.51	N/A	From evidence synthesis	
Baseline BASFI anti-TNF responders	3.92	N/A	From evidence synthesis	
Baseline BASFI anti-TNF non-responders	6.04	N/A	From evidence synthesis	
Response (12-week BASDAI 50)				
Anti-TNF	52.9%	N/A	From evidence synthesis	Evidence synthesis (see *Chapter 5*)
Conventional therapy	18.9%	N/A	From evidence synthesis	
Treatment effect				
Initial BASDAI Changel Tx response: anti-TNF	–4.31	N/A	From evidence synthesis	Evidence synthesis (see *Chapter 5*)
Initial BASDAI Changel Tx response: CC	–3.34	N/A	From evidence synthesis	
Initial BASDAI Changel Tx no response: anti-TNF	–2.28	N/A	From evidence synthesis	
Initial BASDAI Changel Tx no response: CC	–1.06	N/A	From evidence synthesis	

TABLE 83 List of parameter estimates used in the York model: nr-AxSpA population (*continued*)

Parameter	Mean value	SE	Distribution	Source
Initial BASFI Changel Tx response: anti-TNF	−3.24	N/A	From evidence synthesis	
Initial BASFI Changel Tx response: CC	−1.88	N/A	From evidence synthesis	
Initial BASFI Changel Tx no response: anti-TNF	0.08	N/A	From evidence synthesis	
Initial BASFI Changel Tx no response: CC	−0.05	N/A	From evidence synthesis	
Long-term annual BASFI progression				
BASFI annual progression anti-TNF	0.017		Derived from probabilistic inputs (the following four rows)	–
BASFI annual progression CC	0.039			–
Annual rate of mSASSS change for mSASSS < 10	0.69	0.031	Normal	Ramiro *et al.* (2013)[148]
BASFI change with 1 unit change in mSASSS	0.057	0.0049	Normal	Landewe *et al.* (2009)[10]
Treatment effect on progression (relative risk)	0.42	0.122	Normal	Haroon *et al.* (2013)[121]
Time to treatment effect (years)	4	–	–	Haroon *et al.* (2013),[121] Baraliakos *et al.* (2014)[122]
Long-term annual BASDAI progression				
BASDAI annual progression anti-TNF	0	N/A	–	–
BASDAI annual progression CC	0	N/A	–	–
Annual withdrawal probability				
Constant rate of annual withdrawal	0.06	(AiC information has been removed)	Log-normal; from exponential model [coefficient (AiC information has been removed); SE (AiC information has been removed)]	Pfizer submission[36]
Mortality				
SMR women	1.38	0.163	Normal	Bakland (2011)[16]
SMR men	1.63	0.163	Normal	Bakland (2011)[16]

continued

TABLE 83 List of parameter estimates used in the York model: nr-AxSpA population (*continued*)

Parameter	Mean value	SE	Distribution	Source
Quality of life				
Intercept	AiC information has been removed	Uncertainty from reported variance–covariance matrix	Multivariate normal	Pfizer submission[36]
BASDAI coefficient	AiC information has been removed		Multivariate normal	
BASFI coefficient	AiC information has been removed		Multivariate normal	
Male coefficient	AiC information has been removed		Multivariate normal	
Age coefficient	AiC information has been removed		Multivariate normal	
BASDAI^2 coefficient	AiC information has been removed		Multivariate normal	
BASFI^2 coefficient	AiC information has been removed		Multivariate normal	
BASFI × BASDAI coefficient	AiC information has been removed		Multivariate normal	
Initial 12-week period costs (drug + initiation + administration)				
Adalimumab	2573	–	Fixed	As discussed in *Resource use and costs*
Certolizumab pegol	4035	–	Fixed	
Etanercept	2606	–	Fixed	
Golimumab	2566	–	Fixed	
Infliximab	7213	–	Fixed	
Certolizumab pegol PAS	460	–	Fixed	
Subsequent 12-week costs (drug + monitoring + administration)				
Adalimumab	2177	–	Fixed	As discussed in *Resource use and costs*
Certolizumab pegol	2210	–	Fixed	
Etanercept	2210	–	Fixed	
Golimumab	2170	–	Fixed	
Infliximab	3441	–	Fixed	
Certolizumab pegol PAS	2210	–	Fixed	
Disease-related costs: annual				
Intercept	1284	0.165	Log-normal	OASIS data,[118] AbbVie submission[34]
BASFI coefficient	0.213	0.038	Normal	
AE costs (£ per patient)				
Year 1	18.2	–	Fixed	Excess rates for anti-TNFs from Cochrane review,[137] costs from NHS Reference Costs 2012/13[176]
Subsequent years	0	–	Fixed	

N/A, not applicable; Tx, treatment.

Response, change in Bath Ankylosing Spondylitis Disease Activity Index/Bath Ankylosing Spondylitis Functional Index and conditional baselines

The BASDAI 50 response, the conditional change scores for BASDAI and BASFI at 12 weeks and the separate conditional baselines estimated for BASDAI and BASFI (responders vs. non-responders) were derived directly from the results of the extended synthesis model reported in *Chapter 5*. In the base case, it was assumed that the percentage of BASDAI 50 responders, change in BASDAI/BASFI and conditional baselines were the same for all anti-TNFs. The outputs [CODA (Convergence Diagnostic and Output Analysis) file format] from the simulations were incorporated directly into the model to maintain correlation and to avoid any additional distributional assumptions.

Longer-term Bath Ankylosing Spondylitis Functional Index progression

As previously highlighted in the overview section, the York model attempts to address some of the conceptual concerns outlined in *Chapter 4* surrounding the assumptions applied within existing models in relation to modelling BASFI progression over time. Specifically, we assume that BASFI is a function of separate processes which are independently related to disease severity/activity (BASDAI) and to the extent and subsequent progression of radiographic disease (mSASSS). The rationale for this is that the association between BASDAI and BASFI is already accounted for in the separate mean change scores applied to both BASDAI and BASFI for responders versus non-responders/CC patients. Differences in BASDAI are assumed to remain constant over the longer-term horizon (an assumption which is common across all models). Hence any additional changes which might affect BASFI need to be more explicitly related to a separate clinical process (or processes). Based on the studies included in the reviews reported in *Chapter 3* for natural history (see *Review of natural history of ankylosing spondylitis and non-radiographic axial spondyloarthritis*) and the effect on anti-TNFs on radiographic progression (see *Effect of anti-tumour necrosis factors on radiographic progression*), we modelled longer-term changes in BASFI (for CC and anti-TNFs) as a function of mSASSS scores.

The approach applied in the AS population is based on the following studies and assumptions:

1. The multivariate relationship reported in Landewe *et al.*,[10] based on longitudinal assessments of BASFI, BASDAI and mSASSS, was used to estimate the independent effect of a 1-unit change in mSASSS on BASFI scores (mean 0.057 units, SE 0.0049 units).
2. Data from a 12-year prospective follow-up of the OASIS study was used to estimate the annual rate of change in mSASSS. Although at the individual level progression of mSASSS is highly variable, the study by Ramiro *et al.*[124] demonstrated that at a group level (i.e. akin to the cohort approach applied in the York model) changes in mSASSS were stable, progressing at an annual rate of 0.98 mSASSS units per year.[125] Combining the estimates reported across the studies implies a change in BASFI of 0.056 units per annum (0–10 scale). However, as the population included in the study by Ramiro *et al.*[124] included patients who would not be eligible to receive anti-TNFs, we used data in the subgroup of patients with baseline mSASSS $\geq$ 10 units. The annual rate of mSASSS progression in this subgroup was 1.44 units (95% CI 1.18 to 1.70 units) per year with an implied annual BASFI score change of 0.082 units per year. This compares with an annual change of BASFI score of between 0.056 and 0.07 units assumed across the manufacturer's submissions. The specific subgroup (mSASSS $\geq$ 10 units) was chosen to reflect that AS patients eligible to receive anti-TNFs are likely to be more similar to this subgroup than the entire cohort reported by Ramiro *et al.*[124] This also provided a basis for differentiating between the AS and nr-AxSpA populations which is discussed in the following section.

   Given the uncertainties noted in *Chapter 3* (see *Effect of anti-tumour necrosis factors on radiographic progression*), surrounding the effect of anti-TNFs on radiographic progression, we explored alternative scenarios in the decision model. In the base case, we assumed that the effect was related to the duration of therapy which has been reported in recent studies by Haroon *et al.*[121] and Baraliakos *et al.*[122] Both studies consistently reported evidence that the difference in mSASSS between patients who received anti-TNFs and historical controls became different only in patients who had received

treatment for approximately 4 years or more. In the absence of any relative-effect measure reported by Baraliakos *et al.*,[122] we used results reported by Haroon *et al.*[121] applying a zero-inflated binomial model with a relative rate of mSASSS change of 0.42 units (95% CI 0.18 to 0.98 units). Hence, in the model, no effect on mSASSS was assumed until year 4 of the model and then only applied to patients who continued to receive therapy beyond this period.

3. Given the inherent uncertainties regarding the effect of anti-TNFs on radiographic progression we explored alternative scenarios based on (1) an assumption of no impact on radiographic progression; and (2) an immediate effect, applying the estimate of 0.42 from the outset.

For the nr-AxSpA population, we assume a similar underlying clinical process relating to BASFI but model separate BASFI processes for patients depending upon the probability of developing radiographic disease over time and thereafter modelling the extent and progression of radiographic disease via mSASSS changes. Hence, our intention in the nr-AxSpA model was to employ a constant BASFI score (on and off treatment) until a patient develops radiographic progression. At the time point of 'progression' an increasing BASFI would be assumed using a similar approach applied to the AS population. However, programming the additional transition to allow separate BASFI progression estimates based on the time of progression (and time since progression for patients who had previously progressed) proved more complex than anticipated. Consequently, a more simplified assumption was made such that all patients were assumed to incur progression in BASFI albeit at a lower rate relative to the AS population.

The approach we intended to apply in the nr-AxSpA population was based on the following studies and assumptions:

1. Poddubnyy *et al.*[155] is used to estimate the probability of nr-AxSpA patients progressing to radiographic disease based on the outcome '% progressed by ≥ 2 mSASSS over 2 years' (7.4%) reported. These estimates are converted into a rate to estimate the cycle-specific probability.
   Following progression, the mSASSS scores of patients are subsequently assumed to increase at a rate of 0.69 units (95% CI 0.63 to 0.75 units) per year, based on the subgroup of patients with baseline mSASSS < 10 reported by Ramiro *et al.*[148] BASFI is assumed to remain constant for patients who do not progress in each cycle of the model.
   The same results reported by Haroon *et al.*,[121] applying a zero-inflated binomial model with a relative rate of mSASSS change of 0.42 units (95% CI 0.18 to 0.98 units), were applied to the mSASSS scores for patients who progressed to estimate the treatment effect of anti-TNFs. Hence, in common with the AS model, no effect on mSASSS was assumed until year 4 and then it was only applied to patients who continued to receive therapy beyond this period.
2. Given the inherent uncertainties regarding the effect of anti-TNFs on radiographic progression we explored alternative scenarios based on (1) an assumption of no impact on radiographic progression and (2) an immediate effect – applying the estimate of 0.42 from the outset. We also considered an exploratory scenario where we assumed no radiographic progression for nr-AxSpA for patients receiving anti-TNFs, to investigate the untested hypothesis that early intervention in patients, prior to established radiographic disease, might halt subsequent progression.

Given the additional programming challenges that could not be overcome within the remaining time and funding constraints, the mSASSS scores of all nr-AxSpA patients were assumed to increase at the rate of 0.69 units per year. Hence the subsequent results reported for the nr-AxSpA population are potentially optimistic, as not all patients will develop radiographic progression. However, the use of mSASSS in this context inevitably represents an uncertain proxy process for BASFI changes. Further, it should also be noted that the BASFI trajectory of nr-AxSpA patients has been reported in publications to be similar to early AS patients.[177] Consequently, applying the change in mSASSS reported in the subgroup of patients with baseline mSASSS < 10 reported by Ramiro *et al.*[148] may not be an unreasonable proxy for the purposes of predicting future changes in BASFI over longer periods.

Longer-term discontinuation

Patients who achieve a response at 12 weeks are subsequently assumed to remain on that treatment until the treatment is discontinued (i.e. because of loss of efficacy or AEs) and hence the evidence required to inform the decision model is the post-12 week withdrawal data for responders. The rationales for this are (1) that discontinuation for lack of efficacy is higher during the first 3 months, and this has already been accounted for in the model using the probability of no BASDAI 50 response during the initial treatment period; and (2) that discontinuation rates in responders may differ from withdrawal rates in studies which potentially include both responders and non-responders. Although *Chapter 3, Drug survival and anti-tumour necrosis factor switching* identified 12 studies reporting on longer-term drug survival from registries, none of these appears to directly inform the model requirements (i.e. either including the initial 3-month period and/or not being specific to responders).

The most relevant estimates appeared to be those presented in previous and current submissions by the manufacturers. Three alternative approaches and sources were identified which appeared to meet the requirements of the economic model. These included:

1. A constant annual probability of 15% applied in the study by Kobelt *et al.*,[160] based on data from infliximab responders (BASDAI 50) reported as part of the 2nd year of the open-label extension period of the Braun trial ($n = 18$).
2. Separate time-dependent estimates of the probability for AS and nr-AxSpA reported in the AbbVie submission.[34] These estimates were based on a parametric function (log-normal distribution) estimated from responders (ASAS 20 for AS and ASAS 40 for nr-AxSpA at week 12) from the open-label extensions of ATLAS[61] (up to 260 weeks; $n =$ not stated) and ABILITY-1[58] (up to 156 weeks; $n = 28$).
3. A constant annual estimate (approximately 5% for nr-AxSpA and 11% for AS) reported in the Pfizer submission.[36] These estimates were based on a parametric function (exponential distribution) estimated from responders (BASDAI 50 at week 12) from the open-label extensions of studies 311-EU,[83–85] 312-EU[146] and 907-EU[147] (up to approximately 250 weeks for 311-EU; $n =$ not stated) for the AS population and 1031 study[166] (up to approximately 110 weeks; $n = 46$) for the nr-AxSpA population.

Figure 16 provides a comparison of the different estimates in terms of the subsequent drug survival over a longer-time horizon for AS.

Figure 17 provides a comparison of the different estimates in terms of the subsequent drug survival over a longer-time horizon for nr-AxSpA.

The base case of the York model is based on the estimates reported in the submission by Pfizer for both populations. The justification for this is that (1) the estimates relate to the response end point used in the York model (BASDAI 50); (2) full details were reported by Pfizer concerning the alternative parametric models and associated goodness-of-fit statistics and the exponential model appeared the most appropriate function;

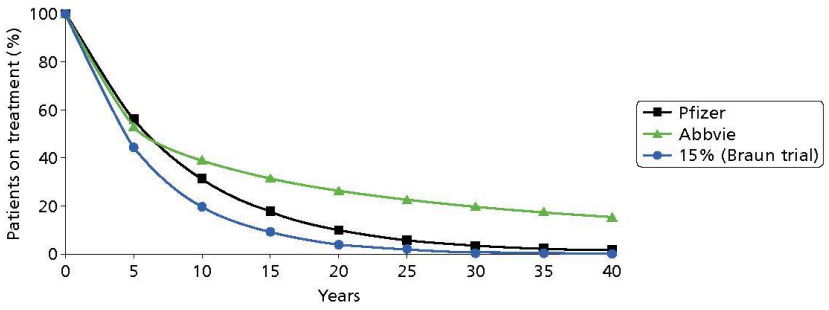


FIGURE 16 Comparison of withdrawal rates: AS population.

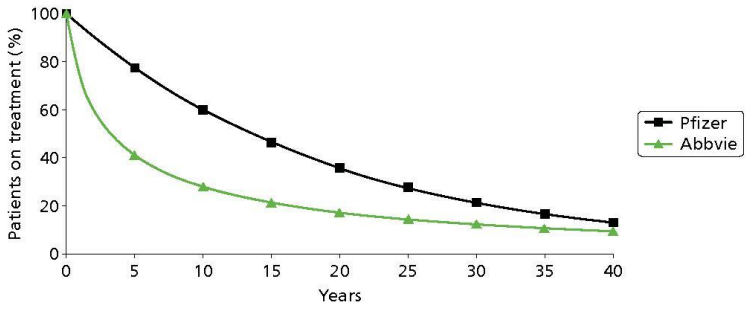


FIGURE 17 Comparison of withdrawal rates: nr-AxSpA population.

and (3) the continued use of a time-dependent function with long tails such as the log-normal distribution results in a significant proportion of patients who would still be assumed to be on TNF-α therapy even after 40 years. Although it is not possible to completely rule out this possibility, the approach by Pfizer was deemed to be a more appropriate basis for informing the York model based on a series of considerations.

Health-related quality of life

The current manufacturer's submissions are based on alternative mapping algorithms to link BASDAI and BASFI scores to a generic utility measure. The approach used by AbbVie in their base case was based on separate mapping algorithms for the AS and nr-AxSpA populations using data from the ATLAS[61] and ABILITY-1[58] trials, respectively. For the nr-AxSpA population, BASDAI and BASFI were mapped to EQ-5D, whereas the algorithm for the AS population mapped to HUI-3; reflecting the use of different generic utility measures used in the two trials. The approach employed by Pfizer in their base case was similarly based on separate algorithms for each population estimated using data from the 1031 study[166] (nr-AxSpA) and the 314-EU study[167] (AS) both mapped to EQ-5D. Both regressions were based on the relationships between BASDAI, BASFI and EQ-5D. The approach employed by UCB in their base case was based on the same single-mapping algorithm from the RAPID-axSpA[64] trial that included both patient populations. Merck Sharp & Dohme adopted the algorithm reported in McLeod *et al.*[38]

We undertook a separate search for other published utility algorithms and identified only the algorithm reported in Ara *et al.*,[161] which was based on the cost-effectiveness analysis submitted by Pfizer to NICE for TA143.[17] Full details of the search and an associated review of utility studies are reported in *Appendices 1* and *14*, respectively. A summary of the alternative algorithms based on EQ-5D is provided in *Tables 84* and *85*.

Figures 18–21 provide a comparison of the utility predictions for each algorithm in each population. For each population, two separate figures are presented. Each figure is based on the impact of holding either BASDAI or BASFI constant (at the mean value) and allowing the other measure to vary across the entire range. The baseline characteristics (BASDAI, BASFI and age) were derived from a weighted average of the baseline characteristics of the clinical trials for the AS population used in the manufacturer's economic models. For nr-AxSpA, baseline characteristics (BASDAI, BASFI and age) of the nr-AxSpA subpopulation from the RAPID-axSpA[64] study were used. Sex was assumed to be 65% male in AS and 35% male in nr-AxSpA.

It is evident that there is significant variation in the utility predictions arising from each separate algorithm. In particular, the non-linear function estimated by Pfizer results in important differences across several of the figures at the extremes of the BASDAI/BASFI ranges. However, limited details were provided in relation to goodness of fit and/or predictive performance for the majority of algorithms and hence a formal assessment of the validity of the different approaches is problematic. Only the submission by Pfizer[36] reported additional detail on these aspects and hence was subsequently used in the York model base case (separate algorithms for the different populations). The non-linear function for utilities was also considered to be more consistent with the non-linear approach applied to costs.

TABLE 84 Comparison of alternative EQ-5D utility regression models (AS)

AS	Ara 2007[161]	Merck Sharp & Dohme	UCB	Pfizer
BASDAI/BASFI scale	0–10	0–10	0–10	0–100
Regression model	Linear	Linear	Logistic	Non-linear
Intercept	0.92300000	0.877213	AiC information has been removed	AiC information has been removed
BASFI	–0.04318800	–0.032252	AiC information has been removed	AiC information has been removed
BASDAI	–0.04019000	–0.038409	AiC information has been removed	AiC information has been removed
Male	0.00000000	–0.027891	AiC information has been removed	AiC information has been removed
Age	0.00000000	0.001681	AiC information has been removed	AiC information has been removed
BASFI^2	0.00000000	0.000000	AiC information has been removed	AiC information has been removed
BASDAI^2	0.00000000	0.000000	AiC information has been removed	AiC information has been removed
BASFI × BASDAI	0.00000000	0.000000	AiC information has been removed	AiC information has been removed

TABLE 85 Comparison of alternative EQ-5D utility regression models (nr-AxSpA)

Nr-axSpA	UCB	AbbVie	Pfizer
BASDAI/BASFI scale	0–10	0–10	0–100
Regression model	Logistic	Linear	Non-linear
Intercept	AiC information has been removed	0.9220000	AiC information has been removed
BASFI	AiC information has been removed	–0.0411700	AiC information has been removed
BASDAI	AiC information has been removed	–0.0392400	AiC information has been removed
Male	AiC information has been removed	0.0000000	AiC information has been removed
Age	AiC information has been removed	0.0000000	AiC information has been removed
BASFI^2	AiC information has been removed	0.0000000	AiC information has been removed
BASDAI^2	AiC information has been removed	0.0000000	AiC information has been removed
BASFI × BASDAI	AiC information has been removed	0.0000000	AiC information has been removed

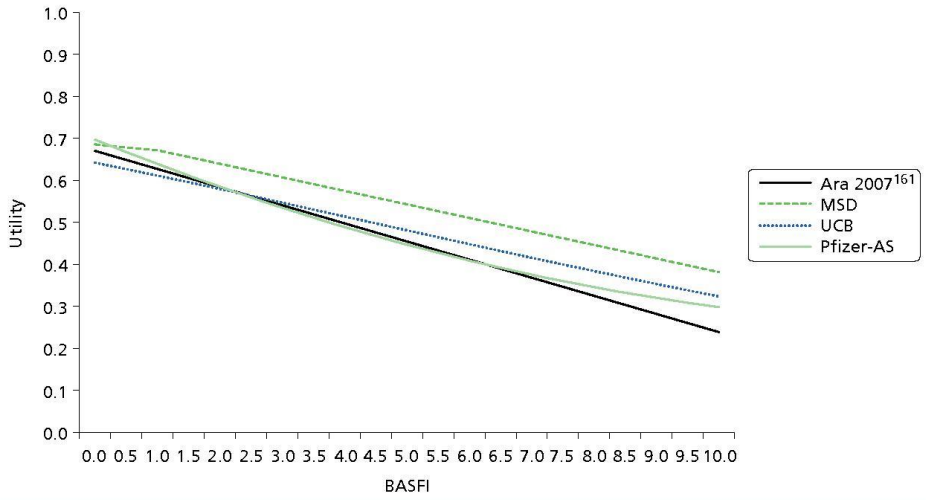


FIGURE 18 Illustration of predicted EQ-5D values using different mapping algorithms: constant BASDAI and varying BASFI (AS). MSD, Merck Sharp & Dohme.

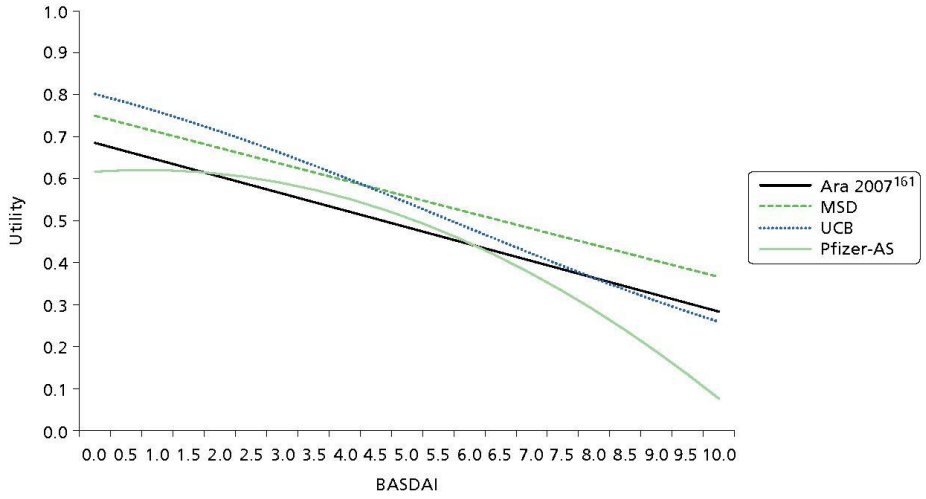


FIGURE 19 Illustration of predicted EQ-5D values using different mapping algorithms: constant BASFI and varying BASDAI (AS). MSD, Merck Sharp & Dohme.

FIGURE 20 AiC information has been removed.

FIGURE 21 AiC information has been removed.

A potential limitation of all the manufacturer analyses is that their algorithms are based on trial data. These data may represent a more limited range of BASDAI and BASFI values and hence there may be issues associated with their subsequent predictive performance in the context of the longer-term economic model, although, from the data reported by Pfizer[36] at least, it appeared as if the full range of BASDAI and BASFI scores were represented in the sample used. However, a separate sensitivity analysis was also undertaken based on the algorithm used by Merck Sharp & Dohme. This algorithm is based on a reanalysis of the Kobelt *et al.*[152] data from patients ($n = 1144$) who had BASDAI and BASFI scores across the whole 0–10 scale and was previously used by McLeod *et al.*[38] for the previous multiple TA. Hence this scenario also provides a more consistent basis for comparing the results from our new analysis.

Resource use and costs

Drug acquisition costs

The unit costs of anti-TNFs were sourced from the *British National Formulary*. Doses were calculated in accordance with their respective licences. *Tables 86* and *87* summarise the drug acquisition costs and the licensed dosage for AS and nr-AxSpA patients.

TABLE 86 Drug acquisition costs

Drug	Dose	Cost (£)	Source
Infliximab (Remicade®, Merck Sharp & Dohme)	i.v. infusion: 100-mg vial	419.62	BNF,[178] November 2014
Golimumab (Simponi®, Merck Sharp & Dohme)	Injection: 50-mg prefilled pen or prefilled syringe	762.97	BNF,[178] November 2014
	Injection: 100-mg prefilled pen	1525.94	
Adalimumab (Humira®, AbbVie)	Injection: 40-mg prefilled pen/prefilled syringe or 40-mg/0.8-ml vial	352.14	BNF,[178] November 2014
Certolizumab (Cimzia®, UCB)	Injection: 200-mg prefilled syringe	357.5	BNF,[178] November 2014
Etanercept (Enbrel®, Pfizer)	Injection: powder for reconstitution, 25-mg vial or 25-mg prefilled syringe	89.38	BNF,[178] November 2014
	Injection: 50-mg prefilled pen or prefilled syringe	178.75	BNF,[178] November 2014

BNF, *British National Formulary*; i.v. intravenous.

TABLE 87 Anti-TNFs licensed dosage in AS and nr-AxSpA

Drug	Licensed dosage in AS and nr-AxSpA
Infliximab (Remicade®, Merck Sharp & Dohme)	Dose of 5 mg/kg given as an intravenous infusion followed by additional 5-mg/kg infusion doses at 2 and 6 weeks after the first infusion, then every 6 to 8 weeks. If a patient does not respond by 6 weeks (i.e. after two doses), no additional treatment with infliximab should be given
Golimumab (Simponi®, Merck Sharp & Dohme)	Dose of 50 mg given once a month, on the same date each month. For patients with a body weight of more than 100 kg who do not achieve an adequate clinical response after three or four doses, increasing the dose of golimumab to 100 mg once a month may be considered
Adalimumab (Humira®, AbbVie)	Recommended dose for patients with AS and axSpA without radiographic evidence of AS is 40 mg of adalimumab administered every other week as a single dose via subcutaneous injection
Certolizumab (Cimzia®, UCB)	The recommended starting dose of Cimzia for adult patients is 400 mg (given as 2 subcutaneous injections of 200 mg each) at weeks 0, 2 and 4. After the starting dose, the recommended maintenance dose of Cimzia for adult patients with AS is 200 mg every 2 weeks or 400 mg every 4 weeks
Etanercept (Enbrel®, Pfizer)	The recommended dose is 25 mg of Enbrel administered twice weekly, or 50 mg administered once weekly

Patient access scheme details

Certolizumab with PAS: UCB will make Cimzia available free of charge to all NHS patients for the first 3 months of therapy, at which point clinical response should be clear. Only after this 3-month stage will the NHS be charged for continuing to use this therapy. However, it should be noted that the proposed PAS is not yet formally agreed with the Department of Health and NICE.

Golimumab PAS: the manufacturer provides the 100-mg dose of golimumab at the same cost as the 50-mg dose, agreed as part of the PAS.

Drug administration costs

Administration costs for intravenous therapies were based on a regular chemotherapy cost [Healthcare Resource Group (HRG) code SB15Z, Deliver subsequent elements of a chemotherapy cycle], similar to NICE TA143 (*Table 88*).[17] Therapies administered subcutaneously were assumed to be self-administered following instruction. The cost of instruction in the model was based on 1 hour of nurse time [Personal Social Services Research Unit (PSSRU) 2013].[179] Drug administration did not differ between the AS and nr-AxSpA indications.

Initiation and monitoring costs

The initiation and monitoring costs for anti-TNF therapies were restricted to the additional costs incurred compared with patients receiving CC alone, as these drugs are used in addition to current practice. The resource use assumptions for laboratory testing for anti-TNF initiation and monitoring have been sourced from the York model for psoriatic arthritis (TA199[175]) and conform to guidelines from the BSR[169] for the use of biologics.

Specifically, during the initial 12-week period AS patients on anti-TNF therapy are assumed to undertake a series of tests at treatment initiation and at week 12 when assessing treatment response (i.e. a full blood count, ESR, liver function test, and urea and electrolytes). Additional testing is conducted once during the initial period (i.e. chest radiography, tuberculosis Heaf test, antinuclear antibody and a double-stranded deoxyribonucleic acid test). AS patients on anti-TNF therapy are also assumed to visit a specialist twice during the initial 12-week period (at treatment initiation and when assessing 12-week response) and two times per year thereafter for monitoring. For quarterly monitoring, AS patients are assumed to receive a series of laboratory tests once every 3 months (i.e. a full blood count, ESR, liver function test, and urea and electrolytes).

Non-radiographic axial spondyloarthritis patients, in addition to the initiation and monitoring resource use assumed for AS patients on anti-TNF therapy, are also assumed to get a MRI test and a CRP test at treatment initiation, as well as a radiograph once per year after the initial period for monitoring, in order to assess radiographic progression.

Cost estimates for laboratory testing have been sourced from the York model for psoriatic arthritis (TA199[175]) and have been inflated to 2012/13 prices, using the Hospital and Community Health Services Pay and Prices Index.[175,179] The CRP test cost is derived from Henriksson 2010.[180] Specialist visits are costed at £100 (outpatient rheumatology follow-up attendance), using the NHS Reference Costs 2012/13.[176]

TABLE 88 Drug administration costs

Type of therapy	Cost (£)	Source
Subcutaneous therapies	49	Cost of nurse training for self-administration (PSSRU 2013)[179]
Intravenous therapies	291	HRG code SB15Z – Deliver Subsequent Elements of a Chemotherapy Cycle (NHS Reference Costs 2012/13)[176]

A summary of the initiation and monitoring resource use assumptions for anti-TNF therapies and the subsequent costs for the AS and nr-AxSpA populations is reported in *Tables 89* and *90*.

Summary of drug acquisition, administration and monitoring costs

Tables 91 and *92* summarise the drug acquisition, administration and monitoring costs applied in the economic model, for the initial 12-week period and on an annual basis thereafter.

TABLE 89 Initiation and monitoring resource use and costs: AS population

	Resource use		Cost	
Item	Initiation period (12 weeks)	Quarterly monitoring	Initiation period (12 weeks) (£)	Quarterly monitoring (£)
Full blood count	2	1	5.97	2.98
ESR	2	1	5.90	2.95
Liver function test	2	1	1.50	0.75
Urea and electrolytes	2	1	2.77	1.38
Chest radiography	1	0	26.19	0.00
Tuberculosis Heaf test	1	0	8.72	0.00
Antinuclear antibody	1	0	4.65	0.00
Double-stranded DNA test	1	0	4.65	0.00
Specialist visit	2	0.5	200.00	50.00
CRP level	0	0	0.00	0.00
Total	–	–	260	58

DNA, deoxyribonucleic acid.

TABLE 90 Initiation and monitoring resource use and costs: nr-AxSpA population

	Resource use		Cost	
Item	Initiation period (12 weeks)	Quarterly monitoring	Initiation period (12 weeks) (£)	Quarterly monitoring (£)
Full blood count	2	1	5.97	2.98
ESR	2	1	5.90	2.95
Liver function test	2	1	1.50	0.75
Urea and electrolytes	2	1	2.77	1.38
Chest radiography	1	0.25	26.19	6.55
Tuberculosis Heaf test	1	0	8.72	0.00
Antinuclear antibody	1	0	4.65	0.00
Double-stranded DNA test	1	0	4.65	0.00
Specialist visit	2	0.5	200.00	50.00
MRI	1	0	144.45	0.00
CRP level	1	0	6.62	0.00
Total	–	–	411	65

DNA, deoxyribonucleic acid.

TABLE 91 Summary of drug acquisition, administration and monitoring costs used in economic model: AS population

	Initial period (3 months)			Annual cost (after initial 3 months)			Total costs	
Treatment (dosage)	Acquisition cost (£)	Administration cost (£)	Monitoring costs (£)	Acquisition cost (£)	Administration cost (£)	Monitoring costs (£)	Initial period (3 months) (£)	Subsequent annual costs (£)
Adalimumab, 40 mg eow	2112.8	49.0	260.4	8451.4	0.0	232.3	2422.2	8683.6
Certolizumab, 200 mg/2 weeks	3575.0	49.0	260.4	8580.0	0.0	232.3	3884.4	8812.3
Certolizumab, 200 mg/2 weeks, with PAS	0.0	49.0	260.4	8580.0	0.0	232.3	309.4	8812.3
Etanercept, 25 mg twice/week	2145.1	49.0	260.4	8580.5	0.0	232.3	2454.5	8812.8
Etanercept, 50 mg once/week	2145.0	49.0	260.4	8580.0	0.0	232.3	2454.4	8812.3
Golimumab, 50 mg once monthly, with PAS	2105.6	49.0	260.4	8422.4	0.0	232.3	2415.0	8654.7
Infliximab, 5 mg/kg every 7 weeks, four vials	5639.7	978.8	260.4	11,509.6	1997.5	232.3	6878.8	13,739.3

eow, every other week.

TABLE 92 Summary of drug acquisition, administration and monitoring costs used in economic model: nr-AxSpA population

	Initial period (3 months)			Annual cost (after initial 3 months)			Total costs	
Treatment (dosage)	Acquisition drug cost (£)	Administration cost (£)	Monitoring costs (£)	Acquisition drug cost (£)	Administration cost (£)	Monitoring costs (£)	Initial period (3 months) (£)	Subsequent annual costs (£)
Adalimumab, 40 mg eow	2112.8	49.0	411.4	8451.4	0.0	258.5	2573.3	8709.8
Certolizumab, 200 mg/2 weeks	3575.0	49.0	411.4	8580.0	0.0	258.5	4035.4	8838.5
Certolizumab, 200 mg/2 weeks, with PAS	0.0	49.0	411.4	8580.0	0.0	258.5	460.4	8838.5
Etanercept, 25 mg twice/week	2145.1	49.0	411.4	8580.5	0.0	258.5	2605.5	8838.9
Etanercept, 50 mg once/week	2145.0	49.0	411.4	8580.0	0.0	258.5	2605.4	8838.5
Golimumab, 50 mg once monthly, with PAS	2105.6	49.0	411.4	8422.4	0.0	258.5	2566.0	8680.9
Infliximab, 5 mg/kg every 7 weeks, four vials	5796.08	1005.9	411.4	11,509.6	1997.5	258.5	7213.4	13,765.5

eow, every other week.

Long-term disease management costs

Patients who remain on anti-TNF treatment incur disease management costs. Previously published economic evaluations employed observational cohort studies to estimate disease management costs and modelled these according to BASDAI and/or BASFI (e.g. NICE TA143[17]). In addition, as discussed in *Chapter 4*, the majority of the manufacturer's submissions within this appraisal (and the LRiG model in TA143[17]) have analysed health-care resource use data from the OASIS[118] to estimate disease management costs. The submission by Pfizer[36] estimated disease-related costs using data from Rafia *et al.*[168] arguing that it is a more recent study and provides a UK-specific cost estimate. However, the comparative analysis of the different long-term cost models in *Appendix 11* showed that the Rafia model provided considerably lower cost estimates; the reasons for this discrepancy are not clear.

In NICE TA143[17] the committee judged that the OASIS data were the most reliable source, being a 2-year prospective study of 208 AS patients from four centres in France, Belgium and the Netherlands, and collecting clinical assessments and economic data including BASDAI and BASFI every 2 or 6 months. The NICE committee also decided that only BASFI should be employed as the major predictor of costs as it reflects long-term disease progression, while BASDAI appears to fluctuate but not increase over time.

The base case of the York model uses the exponential BASFI regression model from the AbbVie submission, which is a reanalysis of the OASIS resource utilisation data using up-to-date published tariffs (NHS Reference Costs 2012/13[176] and PSSRU 2013[179]) (*Table 93*).

Adverse events

Only serious infections and tuberculosis reactivation were included in the economic model. Anti-TNF excess rates versus CC for serious infections and tuberculosis reactivation for were sourced from the Cochrane review of AEs[137] which has been discussed in *Chapter 3, Cost-effectiveness results: adverse events*. The cost of a serious infection was sourced from the Pfizer submission[36] and was assumed to be £1457 based on a weighted average of relevant HRG costs from NHS Reference Costs 2012/13[176] (*Table 94*). The cost of tuberculosis was estimated to be £3204.50 per episode and was based on a weighted average of the relevant HRG codes with different levels of severity (codes DZ14C, DZ14D and DZ14E) from NHS Reference Costs 2012/13.

TABLE 93 Disease-related costs

Cost (£)	Source
£1284.186 × exp(0.213 × BASFI)	AbbVie submission;[34] reanalysis of OASIS[118] data

TABLE 94 Costs of serious infection (from Pfizer submission[36])

HRG code	HRG description	Activity	National average unit cost (£)
WA03C	Septicaemia, with CC score 0–1	44,956	1792
DZ23G	Bronchopneumonia with CC score 0–4	5231	1252
LA04M	Kidney or urinary tract infections, with interventions, with CC score 0–2	2587	2289
PA16B	Major infections with CC score 0	7859	1573
DZ22J	Unspecified acute lower respiratory infection with CC score 0–1	21,109	657
DZ21U	Chronic obstructive pulmonary disease or bronchitis, without NIV, without intubation, with CC Score 0–3	52,421	1453
Weighted average cost (£)			1457

CC, complications; NIV, non-invasive ventilation.
Activity refers to how many times each HRG has been used.
Source: NHS Reference Costs schedule 2012/13.[176]

Mortality

Sex-specific SMRs are applied to the mortality rates from the general population to calculate separate adjusted mortality rates for AS and nr-AxSpA populations in the model.[16]

Analytic methods

The expected costs and QALYs of the alternative anti-TNFs are estimated and cost-effectiveness assessed based on the incremental cost per additional QALY gained. As an assumption is made concerning the similarity in terms of clinical effect between the alternative anti-TNFs, the differences between each of the treatments are driven entirely by their respective acquisition, administration and monitoring costs. Under this assumption, inevitably the lowest cost TNF-α inhibitor would clearly dominate (i.e. lower cost and equal effect) in a fully incremental comparison of cost-effectiveness. Consequently, each TNF-α inhibitor is compared separately versus CC alone. This provides a more consistent basis for assessing the impact that the different drug costs have across each separate scenario.

Probabilistic sensitivity analysis is used to assess the implications of parameter uncertainty (the imprecision with which input parameters are estimated). The mean costs and QALY reported in the tables are derived from the PSA and the probabilities that each TNF-α inhibitor is more cost-effective than CC alone are reported at thresholds of £20,000 and £30,000 per QALY.

Sensitivity analyses

A number of separate scenarios are presented to assess the implications of key parameter assumptions and sources of structural uncertainty in the model. These include the following.

Scenario 1: no response to conventional care assumed at 12 weeks

The base-case model incorporates the probability of response to CC at 12 weeks and assigns separate baselines to responders and non-responders. Although the changes in BASDAI/BASFI estimated at 12 weeks for CC are assumed to disappear in the following 12-week cycle, the separate baselines estimated for responders and non-responders are retained for the remainder of the model horizon. Given uncertainties surrounding the nature of the 'placebo' response assumed to apply to CC and whether or not this would be evident in actual clinical practice, a separate scenario was modelled which assumed that no patients receiving CC would achieve a BASDAI 50 response. This scenario was based on a separate simulation using the extended synthesis model where the magnitude of 'placebo' effect was assumed to be 0. Hence, employing this scenario, the impact of the 'placebo' effect is effectively netted out of the model for both CC and the anti-TNFs. Hence, although the difference in response rates and BASDAI/BASFI scores for responders to anti-TNFs remains similar to the base-case model, the absolute response rate for anti-TNFs and the absolute BASDAI/BASFI scores are lower when the adjustment is applied. In addition, as no response is assumed for CC, a single baseline BASDAI and BASFI score is applied to CC patients.

Scenario 2: different baselines assumed for responders and non-responders

In the base-case analysis, the extended synthesis model is used to estimate both changes in BASDAI and BASFI conditional upon BASDAI 50 response as well as different baseline BASDAI/BASFI scores for responder and non-responders. It was noted in *Chapter 5* that there appeared a disparity in the magnitude of the difference in the conditional baseline scores estimated from the extended synthesis model compared with the differences reported by those manufacturers who provided additional data on request. Specifically, the difference between responders and non-responders appeared higher in our extended synthesis compared with the direct data reported by manufacturers. To explore the potential impact of this difference on the cost-effectiveness results, a separate scenario was undertaken wherein the difference in the conditional baselines was based on a pooled estimate of the differences across the trials provided by manufacturers rather than those estimated by the extended synthesis model.

In addition to exploring the impact of assuming different baselines, this scenario also included a pooled estimate of the change in BASDAI/BASFI scores for responders and non-responders reported by manufacturers. Hence, in this scenario, the extended synthesis model is used only to predict the response to BASDAI 50, the differences in the conditional baselines and change scores being derived from pooled estimates from the data reported by manufacturers.

Scenario 3: no effect of anti-tumour necrosis factors on Bath Ankylosing Spondylitis Functional Index progression

In the base-case model, a treatment effect is applied from year 4 of the model on the rate of further BASFI progression for patients who continue to receive TNF-α inhibitors beyond this time point. Given the uncertainty reported in *Chapter 3* surrounding existing evidence for anti-TNFs in relation to disease modification, a separate scenario was explored which assumed that the rate of BASFI progression would be the same for patients receiving anti-TNFs and CC alone.

Scenario 4: treatment effect of anti-tumour necrosis factors applied from start of model (Bath Ankylosing Spondylitis Functional Index progression)

A separate scenario was also undertaken assuming that the treatment effect on further BASFI progression would be incurred from the start of the model, as opposed to year 4. This scenario assumes that any disease modification would be achieved immediately compared with the delayed effect assumed in the base case.

Scenario 5: utilities – linear Bath Ankylosing Spondylitis Disease Activity Index/Bath Ankylosing Spondylitis Functional Index model

The base-case analysis in both the AS and nr-AxSpA populations are based on the non-linear mapping algorithms reported in the submission by Pfizer. A separate scenario was run in both populations using an alternative linear model which has been applied in previous NICE appraisals (referred to as the 'Merck Sharp & Dohme' algorithm in *Health-related quality of life*). This scenario was incorporated to explore the impact of using a linear model and to provide results which are more consistent with the utility approach applied in previous NICE appraisals (TA143[17] and TA233[33]).

Scenario 6 (non-radiographic axial spondyloarthritis only): trials in non-radiographic axial spondyloarthritis and ankylosing spondylitis populations combined

The base-case analysis for the nr-AxSpA population is based on the extended synthesis model using only the trials reporting in this population. A separate scenario was undertaken based on the results from the extended synthesis model which combined the AS and nr-AxSpA trials.

Following the consultation process to the NICE appraisal, additional analyses were undertaken to address comments received on *Chapters 5* and *6*. These focused on the conditional baseline BASDAI scores used in *Chapter 5* and on the existence of a biosimilar product for infliximab with a lower list price. Further details on these analyses and their results are shown in *Appendix 15*.

Model validation

The conceptualisation of the model and related structural assumptions were informed by the review of existing models and discussions with two clinical advisors. The face validity of the model structure, data sources and key assumptions were addressed using inputs based on systematic reviews, targeted searching and clinical input. Verification of the model and the associated inputs was undertaken using a staged process. One researcher developed the initial model structure and the preliminary coding. This was then checked and extended for the final model by a second researcher. Both researchers were subsequently involved in the subsequent quality-assurance process entailing detailed cross-checks of input data against their sources and undertook extensive logical checks and scenarios to assess the performance of the model. Two other researchers were involved in further checks of key aspects including the integration of the results from the extended synthesis within the Excel model. A fifth researcher was involved in all stages with preparing and checking parameter inputs for the model. Cross-validation was assessed by comparing the results with existing models and identifying differences and their causes.

Results of the independent economic assessment

Base-case results: ankylosing spondylitis population

The base-case results for the AS population, for the alternative rebound assumptions, are reported in *Tables 95* and *96*.

In the rebound equal to gain scenario, the ICER of the alternative anti-TNFs varied between £19,240 (certolizumab with the proposed PAS) to £40,576 per additional QALY (infliximab). Infliximab had the highest ICER (£40,576 per QALY) and the lowest probability of being cost-effective at a £20,000 and £30,000 per QALY threshold (0.001 and 0.089, respectively). Excluding infliximab, the ICERs of the other anti-TNFs were similar, ranging from £19,240 (certolizumab with the proposed PAS) to £23,133 (certolizumab without the proposed PAS).

TABLE 95 Base-case cost-effectiveness results: AS (rebound equal to gain)

Strategy	Total QALYs	Incremental QALYs	Total costs (£)	Incremental costs (£)	ICER (£)	Probability of CE £20,000	Probability of CE £30,000
Conventional therapy	7.245	–	110,821	–	–	–	–
Certolizumab with PAS	8.163	0.918	128,485	17,665	19,240	0.550	0.895
Golimumab	8.163	0.918	130,173	19,352	21,079	0.427	0.841
Adalimumab	8.163	0.918	130,257	19,436	21,170	0.423	0.839
Etanercept	8.163	0.918	130,630	19,810	21,577	0.402	0.826
Certolizumab	8.163	0.918	132,059	21,238	23,133	0.299	0.761
Infliximab	8.163	0.918	148,073	37,252	40,576	0.001	0.089

CE, cost-effectiveness.
The probability of CE £20,000/30,000 is the probability that the TNF-α inhibitor, compared with CC, is a cost-effective option at the stated threshold.

TABLE 96 Base-case cost-effectiveness results: AS (rebound to CC)

Strategy	Total QALYs	Incremental QALYs	Total costs (£)	Incremental costs (£)	ICER (£)	Probability of CE £20,000	Probability of CE £30,000
Conventional therapy	7.265	–	109,933	–	–	–	–
Certolizumab with PAS	7.867	0.603	130,277	20,344	33,762	0.035	0.399
Golimumab	7.867	0.603	131,960	22,027	36,554	0.019	0.299
Adalimumab	7.867	0.603	132,045	22,111	36,695	0.017	0.293
Etanercept	7.867	0.603	132,423	22,489	37,322	0.017	0.275
Certolizumab	7.867	0.603	133,851	23,918	39,693	0.011	0.203
Infliximab	7.867	0.603	150,022	40,088	66,529	0.000	0.001

CE, cost-effectiveness.
The probability of CE £20,000/30,000 is the probability that the TNF-α inhibitor, compared with CC, is a cost-effective option at the stated threshold.

As previously highlighted, the difference in the ICERs between the individual anti-TNFs is driven entirely by the different acquisition and administration costs associated with each. Excluding infliximab, the probability that each TNF-α inhibitor was more cost-effective than CC alone ranged between 0.299 and 0.550 at a £20,000 per QALY threshold and between 0.761 and 0.895 at a £30,000 threshold. There was less variation in these probabilities when the proposed PAS for certolizumab was included, ranging from 0.402 to 0.550 at a £20,000 per QALY threshold and from 0.826 to 0.895 at a £30,000 threshold.

In the rebound to CC scenario, the ICER of the alternative anti-TNFs varied between £33,762 (certolizumab with the proposed PAS) to £66,529 per additional QALY (infliximab). Infliximab had the highest ICER (£66,529 per QALY) and the lowest probability of being cost-effective at a £20,000 and £30,000 per QALY threshold (0.000 and 0.001, respectively). Excluding infliximab, the ICERs of the other anti-TNFs varied between £33,762 (certolizumab with the proposed PAS) to £39,693 (certolizumab without the proposed PAS) and the probability that each TNF-α inhibitor was more cost-effective than CC alone ranged between 0.011 and 0.035 at a £20,000 per QALY threshold and between 0.203 and 0.399 at a £30,000 threshold. There was less variation in these probabilities when the proposed PAS for certolizumab was included, ranging from 0.017 to 0.035 at a £20,000 per QALY threshold and from 0.275 to 0.399 at a £30,000 threshold.

Base-case results: non-radiographic axial spondyloarthritis population

The base-case results for the nr-AxSpA population, for the alternative rebound assumptions, are reported in *Tables 97* and *98*.

TABLE 97 Base-case cost-effectiveness results: nr-AxSpA (rebound equal to gain)

Strategy	Total QALYs	Incremental QALYs	Total costs (£)	Incremental costs (£)	ICER (£)	Probability of CE £20,000	Probability of CE £30,000
Conventional therapy	9.956	–	89,493	–	–	–	–
Certolizumab with PAS	11.351	1.395	128,911	39,418	28,247	0.139	0.591
Adalimumab	11.351	1.395	130,316	40,823	29,253	0.106	0.545
Etanercept	11.351	1.395	131,057	41,563	29,784	0.093	0.529
Certolizumab	11.351	1.395	132,484	42,991	30,807	0.066	0.482

CE, cost-effectiveness.
The probability of CE £20,000/30,000 is the probability that the TNF-α inhibitor, compared with CC, is a cost-effective option at the stated threshold.

TABLE 98 Base-case cost-effectiveness results: nr-AxSpA (rebound to CC)

Strategy	Total QALYs	Incremental QALYs	Total costs (£)	Incremental costs (£)	ICER (£)	Probability of CE £20,000	Probability of CE £30,000
Conventional therapy	9.963	–	90,219	–	–	–	–
Certolizumab with PAS	11.200	1.237	131,714	41,495	33,555	0.057	0.396
Adalimumab	11.200	1.237	133,109	42,890	34,684	0.038	0.343
Etanercept	11.200	1.237	133,859	43,640	35,290	0.035	0.318
Certolizumab	11.200	1.237	135,286	45,067	36,444	0.029	0.284

CE, cost-effectiveness.
The probability of CE £20,000/30,000 is the probability that the TNF-α inhibitor, compared with CC, is a cost-effective option at the stated threshold.

In the rebound equal to gain scenario, the ICER of the alternative anti-TNFs varied between £28,247 (certolizumab with the proposed PAS) and £30,807 per additional QALY (certolizumab without the proposed PAS). The probability that each TNF-α inhibitor was more cost-effective than CC alone ranged between 0.066 and 0.139 at a £20,000 per QALY threshold and between 0.482 and 0.591 at a £30,000 threshold. Again, there was less variation in these probabilities when only the proposed PAS for certolizumab was considered, ranging from 0.093 to 0.139 at a £20,000 per QALY threshold and from 0.529 to 0.591 at a £30,000 threshold.

In the rebound to CC scenario, the ICER of the alternative anti-TNFs varied between £32,528 (certolizumab with the proposed PAS) and £35,365 per additional QALY (certolizumab without the proposed PAS). The probability that each TNF-α inhibitor was more cost-effective than CC alone varied between 0.030 and 0.062 at a £20,000 per QALY threshold and between 0.312 and 0.429 at a £30,000 threshold. Again, there was less variation in these probabilities when only the proposed PAS for certolizumab was included, ranging from 0.039 to 0.062 at a £20,000 per QALY threshold and from 0.369 to 0.429 at a £30,000 threshold.

Sensitivity analyses results: ankylosing spondylitis population

Table 99 summarises the scenarios undertaken for the AS population.

Each of these scenarios was undertaken for the two alternative rebound assumptions. *Tables 100* and *101* summarise the ICER estimates for each scenario. Full ICER tables for each scenario are reported in *Appendix 16*.

TABLE 99 Summary of cost-effectiveness scenarios: AS population

Number	Parameter/structural	Approach in scenario	Approach in base case
1	CC ('placebo') response	No response to CC assumed at 12 weeks	Response to CC included at 12 weeks
2	Different baselines assumed for responders and non-responders and change in BASDAI/BASFI scores	Separate baselines based on pooled estimates provided by manufacturers. Changes in BASDAI/BASFI conditioned on response also based on pooled estimates provided by manufacturers	Separate baselines and changes in BASDAI/BASFI conditioned on responses estimated via extended synthesis model
3	BASFI progression	No effect of anti-TNFs on BASFI progression	Treatment effect applied from year 4 onwards
4	BASFI progression	Treatment effect of anti-TNFs applied from start of model	Treatment effect applied from year 4 onwards
5	Utilities	Linear BASDAI/BASFI model (based on Kobelt *et al.*[152])	Non-linear BASDAI/BASFI model (Pfizer submission[36])

TABLE 100 Summary of ICERs across scenarios (rebound equal to gain): AS population

		Scenario (£)				
Strategy	Base case (£)	1	2	3	4	5
Conventional therapy	–	–	–	–	–	–
Certolizumab (with PAS)	19,240	20,319	11,527	20,655	18,466	23,290
Golimumab	21,079	22,920	12,785	22,581	20,213	25,469
Adalimumab	21,170	23,013	12,851	22,677	20,301	25,579
Etanercept	21,577	23,425	13,143	23,106	20,695	26,073
Certolizumab	23,133	25,495	14,220	24,739	22,180	27,926
Infliximab	40,576	43,510	26,699	43,125	39,037	49,021

TABLE 101 Summary of ICERs across scenarios (rebound to CC): AS population

Strategy	Base case (£)	Scenario (£)				
		1	2	3	4	5
Conventional therapy	–	–	–	–	–	–
Certolizumab (with PAS)	33,762	34,229	25,530	36,518	32,222	29,414
Golimumab	36,554	38,068	27,986	39,483	34,910	31,827
Adalimumab	36,695	38,207	28,107	39,634	35,045	31,950
Etanercept	37,322	38,824	28,652	40,306	35,647	32,499
Certolizumab	39,693	41,885	30,731	42,828	37,928	34,554
Infliximab	66,529	68,815	54,045	71,565	63,684	58,022

The ICER estimates appeared to remain relatively stable across the majority of scenarios compared with the base-case ICER estimates. The exception to this appeared to be scenario 2 which used data submitted on request by several manufacturers which was used to inform the differences in the conditional baselines and the change scores assumed for responders versus non-responders. In summary, when the manufacturer's data were used, the ICER estimates became more favourable towards the anti-TNFs. The more favourable results are driven by smaller differences between responders and non-responders in terms of their conditional baselines and marginally higher differences in the conditional change scores. Both differences result in improvements in the ICER estimates compared with the base-case results derived from the extended synthesis model.

Sensitivity analyses results: non-radiographic axial spondyloarthritis population

Table 102 summarises the scenarios undertaken for the nr-AxSpA population.

Each of these scenarios was undertaken for the two alternative rebound assumptions. *Tables 103* and *104* summarise the ICER estimates for each scenario. Full ICER tables for each scenario are reported in *Appendix 16*.

TABLE 102 Summary of cost-effectiveness scenarios: nr-AxSpA population

Number	Parameter/structural	Approach in scenario	Approach in base case
1	CC ('placebo') response	No response to CC assumed at 12 weeks	Response to CC included at 12 weeks
2	Different baselines assumed for responders and non-responders and change in BASDAI/BASFI scores	Separate baselines based on pooled estimates provided by manufacturers. Changes in BASDAI/BASFI conditioned on response also based on pooled estimates provided by manufacturers	Separate baselines and changes in BASDAI/BASFI conditioned on responses estimated via extended synthesis model
3	BASFI progression	No effect of anti-TNFs on BASFI progression	Treatment effect applied from year 4 onwards
4	BASFI progression	Treatment effect of anti-TNFs applied from start of model	Treatment effect applied from year 4 onwards
5	Utilities	Linear BASDAI/BASFI model (based on Kobelt *et al.*[152])	Non-linear BASDAI/BASFI model (Pfizer submission[36])
6	Treatment effect of anti-TNFs	Trials in nr-AxSpA and AS populations combined	Only trials in nr-AxSpA included

TABLE 103 Summary of ICERs across scenarios (rebound equal to gain): nr-AxSpA population

		Scenario (£)					
Strategy	Base case (£)	1	2	3	4	5	6
Conventional therapy	–	–	–	–	–	–	–
Certolizumab (with PAS)	28,247	34,841	25,482	28,643	27,471	25,324	28,282
Adalimumab	29,988	37,884	27,302	29,670	28,466	29,228	29,512
Etanercept	29,253	38,507	27,821	30,208	28,988	29,753	30,041
Certolizumab	30,807	40,949	29,378	31,250	29,996	30,732	31,034

TABLE 104 Summary of ICERs across scenarios (rebound to CC): nr-AxSpA population

		Scenario (£)					
Strategy	Base case (£)	1	2	3	4	5	6
Conventional therapy	–	–	–	–	–	–	–
Certolizumab (with PAS)	32,528	40,928	29,884	34,416	31,841	26,900	33,184
Adalimumab	33,639	44,365	31,942	35,615	32,940	27,850	34,270
Etanercept	34,232	45,078	32,528	36,241	33,523	28,343	34,866
Certolizumab	35,365	47,842	34,288	37,456	34,642	29,303	35,985

In common with the AS scenarios, the ICER estimates appeared to remain relatively stable across the majority of scenarios compared with the base-case ICER estimates. However, the impact of applying adjustments to the conditional baseline estimates and BASDAI/BASFI scores provided by the manufacturers (scenario 2) had less of an impact in the nr-AxSpA population. The scenario which showed the largest variation compared with the base-case analysis was scenario 1. This scenario was based on results from the extended synthesis which excluded any placebo effect and resulted in a single baseline applied to all CC patients. The differences in the ICERs appear largely as a result of the impact of ignoring the non-linear relationship between baseline BASDAI/BASFI scores because of variation in the baseline of responders versus non-responders in scenario 1. Interestingly, the impact of this approach appears more marked in the nr-AxSpA population, compared with the AS population, which is likely to be driven by several inter-related factors including the magnitude of difference assumed between the conditional baseline scores and the absolute BASDAI and BASFI scores which differ across the populations.

Discussion and comparison with manufacturer models

Based on an underlying assumption of similarity in the clinical effectiveness of each of the anti-TNFs, the York model demonstrates that the cost-effectiveness results are dependent on several factors, including (1) the different acquisition and administration costs; (2) the rebound assumption applied to patients who discontinue therapy; (3) the magnitude of the change in BASDAI/BASFI scores assumed for responders versus non-responders; (4) the different baseline BASDAI/BASFI scores assumed for responders versus non-responders; and (5) the impact of anti-TNFs on the rate of longer-term BASFI progression.

Interestingly, the importance of specific factors also appears to vary across the separate indications. For example, the impact of the alternative rebound assumptions appears more marked in the AS population compared with the nr-AxSpA population. This appears largely driven by the smaller rate of BASFI progression applied in the York model to the nr-AxSpA population, such that the impact of alternative

assumptions regarding possible rebound effects has a less significant impact within this population. This difference also has an important bearing on the subsequent interpretation of the base-case ICERs estimated by the York model in the separate populations. Our findings suggest that the ICER estimates for anti-TNFs appear more favourable for the AS population, relative to those estimated for the nr-AxSpA population, based on the rebound equal to gain scenario. The more favourable results in the AS population based on the rebound equal to gain scenario appear to be driven by two main factors: (1) the smaller conditional change in BASDAI/BASFI scores estimated for the nr-AxSpA population and (2) the lower rate of BASFI progression assumed for the nr-AxSpA population. However, this finding appears reversed in the rebound to CC scenario. Interestingly, within this scenario, the lower conditional change in BASDAI/BASFI scores appears offset by the less significant influence of BASFI progression in the nr-AxSpA model, that is the impact on the ICERs of the two rebound assumptions is closely related to the underlying rate of BASFI progression assumed and the contribution that this makes to the ICER estimates under the separate scenarios. However, it should also be noted that, although the ICERs for the nr-AxSpA population appear more favourable in this scenario compared with those estimated for the AS population, all of the ICER estimates exceeded £30,000 per QALY in the York base case across both populations.

Tables 105 and *106* compare the results of the York model with the base-case results reported by each manufacturer for the alternative populations. In contrast to the manufacturer models which reported a single base case based on an assumption of either rebound equal to gain (AbbVie, Pfizer, Merck Sharp & Dohme) or rebound to CC (UCB), the York model presents both rebound scenarios in order to represent the potential limits to the ICER; recognising that the reality lies somewhere between these scenarios.

TABLE 105 Comparison of cost-effectiveness results from York model vs. manufacturers (AS population)

Strategy	AbbVie, ICER (£)	UCB, ICER (£)	Pfizer, ICER (£)	Merck Sharp & Dohme, ICER (£)	York (rebound equal to gain), ICER (£)	York (rebound to CC), ICER (£)
CC	–	–	–	–	–	–
Adalimumab	16,391	19,932	20,909	19,275	21,170	36,695
Certolizumab	17,067	16,647[a]	19,586[a]	19,401[a]	19,240[a]	33,762[a]
Etanercept	16,897	19,272	20,938	21,972	21,577	37,322
Golimumab	16,535	19,049	21,288	19,070	21,079	36,554
Infliximab	44,448	42,671	37,741	42,532	40,576	66,529

a PAS costs assumed for certolizumab.

TABLE 106 Comparison of cost-effectiveness results from York model vs. manufacturers (nr-AxSpA population)

Strategy	AbbVie (adalimumab), ICER (£)	UCB (certolizumab), ICER (£)	Pfizer (etanercept), ICER (£)	York (rebound equal to gain), ICER (£)	York (rebound to CC), ICER (£)
CC	–	–	–	–	–
Adalimumab	13,228	30,370	23,242	29,988	33,639
Certolizumab	12,866	15,615[a]	23,575[a]	28,247[a]	32,528[a]
Etanercept	Not assessed	50,692	23,195	29,253	34,232

a PAS costs assumed for certolizumab.

Although there are a number of important differences in approaches both among the different manufacturer models and compared with the York model, the comparison of ICERs based on the York rebound equal to gain scenario appear broadly consistent in the AS population. This might appear surprising given that the York model is based on two key assumptions that appear less favourable than those used by manufacturers, specifically: (1) incorporating separate baseline BASDAI/BASFI scores for responders and non-responders which assume that responders are likely to be less severe in terms of their baseline BASDAI and BASFI scores than non-responders; and (2) only incorporating an effect of anti-TNFs on disease progression for patients remaining on therapy for at least 4-years. However, these appear counterbalanced by the higher rate of BASFI progression applied to AS patients [0.082 (0–10 scale) units per annum compared with estimates between 0.056 and 0.07 assumed by the manufacturers]. As we highlighted at the start of this section, it is our view that the York model has a more coherent basis for modelling longer-term BASFI progression.

Another important counterbalancing effect is the use of the conditional scores for responders and non-responders obtained via the extended synthesis within the York model. This contrasts with the selective approaches (i.e. using conditional scores from single studies or assumptions) or use of longer-term follow-up and/or open-label sources (i.e. implicitly assuming that patients who continue to participate in longer-term follow-up and open label sources are more likely to be responders than patients who do not). Consequently, the change scores assumed in the York model for BASDAI 50 responders appear higher than those assumed by several of the manufacturers. The approach applied within the York model is based on a more generalised framework for synthesis and hence utilises more evidence than considered by the manufacturers. This approach directly informs the conditional change scores which are fundamental to an appropriate assessment of the cost-effectiveness when a response-based assessment is incorporated to determine eligibility for continued treatment.

In *Chapter 4* it was noted that there appeared more variation in the ICER estimates reported across the manufacturer's submissions in the nr-AxSpA population compared with those reported in the AS population. Again, the ICER estimates reported by the York model in the nr-AxSpA population do not appear inconsistent with the range of ICERs reported across the separate manufacturers. However, any attempt to formally cross-validate the results from the York model with those reported by the manufacturers is difficult given the contrasting approaches and assumptions employed. As the York model uses several of the key parameter inputs reported in the submission by Pfizer,[36] a comparison may be more usefully made by comparing the results of the York model and those reported by Pfizer. In general, the ICER estimates appear less favourable in the York model compared with those reported by Pfizer. One possible explanation for these differences is that the York model uses a lower rate of BASFI progression and only assumes that anti-TNFs affect this rate after at least 4 years of treatment. However, our results have also shown that the impact of progression appears less of a driver of cost-effectiveness in the nr-AxSpA model. Another possible explanation is the use of different baselines assumed for responders and non-responders assumed in the York model, that is the York model assumes that responders and typically less severe in terms of baseline BASDAI/BASFI scores compared with non-responders. Consequently, an additional scenario was undertaken using the York model to further assist in cross-validation. For this scenario, an assumption was made that the responders and non-responders did not differ in terms of baseline BASDAI/BASFI scores.

The results of the additional validation scenario are reported in *Table 107*. The ICERs in this scenario appeared closer to those reported by Pfizer.[36] Hence this additional validation scenario is important in helping to identify potential drivers of difference between the results of the York model and those reported by the manufacturers. The scenario also demonstrates that the assumption made concerning potential differences (and the magnitude of any difference) between the baseline BASDAI/BASFI scores of responders and non-responders has an important impact on the cost-effectiveness results. Hence, studies which are based on similar baselines are likely to be potentially overly optimistic in the subsequent ICER estimates reported for anti-TNFs. Equally, it might be argued that the results from the York base-case model may be conservative towards the anti-TNFs because the magnitude of differences in the baseline

TABLE 107 Non-radiographic axial spondyloarthritis: additional validation scenario (rebound equal to gain and responders/non-responders do not differ in terms of baseline BASDAI/BASFI scores)

Strategy	Total QALYs	Incremental QALYs	Total costs (£)	Incremental costs (£)	ICER (£)	Probability of CE £20,000	Probability of CE £30,000
Conventional therapy	9.977	–	88,692	–	–	–	–
Certolizumab (with PAS)	11.551	1.574	125,205	36,513	23,199	0.390	0.759
Adalimumab	11.551	1.574	126,606	37,914	24,089	0.341	0.733
Etanercept	11.551	1.574	127,350	38,658	24,562	0.319	0.720
Certolizumab	11.551	1.574	128,777	40,085	25,469	0.272	0.702

CE, cost-effectiveness.
The probability of CE £20,000/30,000 is the probability that the TNF-α inhibitor, compared with CC, is a cost-effective option at the stated threshold.

scores estimated from the extended synthesis model appeared higher than those obtained on request from manufacturers (although the direction of the difference was consistent). Hence, in a similar manner to which the different rebound assumptions represent the potential limits on the ICER given uncertainties surrounding rebound, the differences in the ICERs based on assuming no difference in baselines and the magnitude of differences employed in the York base case may also represent the limits of the ICER based on uncertainty surrounding the magnitude of this difference. Given the potential importance of this assumption, *Appendix 16* reports the full ICER results for each population (and under each rebound assumption) assuming identical baselines for responders and non-responders.

Although the York model provides a number of significant developments to existing cost-effectiveness analyses, there are still several potential limitations. First, in common with all existing models, subsequent linkages to costs and QALYs are related to BASDAI and BASFI, largely because of the existence of data. Second, the cost-effectiveness estimates are based on uncertain projections of BASDAI and BASFI over a longer time horizon in order to generate more appropriate lifetime estimates of costs and QALYs required for cost-effectiveness assessments. Although extensive efforts have been made to identify a more appropriate basis for informing these longer-term estimates (particularly for BASFI), inevitably, significant uncertainty remains. Third, it should be noted that there are potential benefits which have not been formally captured and quantified within the current model, specifically any potential impact on productivity costs and any additional benefits that anti-TNFs may confer for other comorbidities (e.g. inflammatory bowel disease, psoriasis, etc.). A final limitation is that it was not possible to include the biosimilar version of infliximab (CT-P13) within the analysis as a formal list price was not available at the time of the assessment.

In addition, the York model has not specifically addressed important clinical questions concerning the issue of intermittent and sequential use of anti-TNFs. However, in the absence of robust clinical evidence from RCTs, existing evidence is clearly subject to potential confounding. Consequently, existing attempts to model sequential therapy within the current manufacturer's submissions (Pfizer[36] only) are largely based on applying simple adjustments to first-line efficacy but which are unlikely to provide a robust basis for informing these decisions. Clearly, until such time that more robust data are available, a rough rule of thumb could similarly be applied to the results presented from the York model, such that the ICERs of a second-line TNF-α inhibitor in a patient who had previously responded but subsequently lost response, might be in the order of one-third higher than the results presented here.

Finally, it is important to appreciate that the assessments of cost-effectiveness reported in the York model are based on a normative approach, that is they are based on the assumption that 12-week continuation rules (and ongoing monitoring of response) would be fully adhered to in clinical practice. Hence, they do not necessarily reflect the cost-effectiveness of how anti-TNFs are currently used in the management of AS within the NHS or how they might be used, in the event of positive guidance from the NICE in nr-AxSpA. The findings of West Midland Rheumatology Audit from 2010 give some grounds for potential concern.[181] This regional audit was undertaken to assess compliance with the NICE guidelines (TA143[17]) in 17 rheumatology centres across the East and West Midlands. The findings from this audit revealed that (1) the proportion of patients being assessed at 12 weeks after treatment initiation was suboptimal; (2) fewer than 20% of patients with an inadequate response at 12 weeks had their treatment discontinued; and (3) fewer than half of the patients received regular 12-weekly assessments. During the course of our assessment we contacted the British Society for Rheumatology Biologics Register (BSRBR) Ankylosing Spondylitis Register to assess the feasibility of obtaining access to data which has been collected since the register was set up in 2012. Although our request was positively received, it was clear during ongoing discussions that the data and analyses requested could not be undertaken within the time frame of our assessment.

Chapter 7 Assessment of factors relevant to the NHS and other parties

The results of this technology assessment have some implications for clinical practice. Existing NICE guidance recommends adalimumab, certolizumab, etanercept and golimumab for the treatment of AS and therefore the use of these drugs is already widespread in the NHS. However, in the light of the additional evidence presented here, the use of these agents in AS may increase further.

Furthermore, the available clinical evidence indicates that adalimumab, certolizumab and etanercept are effective in patients with nr-AxSpA, although there is some uncertainty regarding the definition of the nr-AxSpA patient population who would benefit most from these anti-TNFs. The effectiveness demonstrated in the nr-AxSpA population suggests that early treatment of AS/nr-AxSpA patients is warranted. A key study on flares in AS suggested that the 12-week period required to confirm sustained active spinal disease in AS patients commencing an anti-TNF may be too long. The findings suggest that shorter time periods may therefore be considered in future guidance, which would minimise the delay in starting treatment and the discomfort experienced by patients.

The potential extra cost to the NHS of providing anti-TNFs for patients with nr-AxSpA in addition to AS patients is unclear because the prevalence of nr-AxSpA in the UK is somewhat uncertain. The potentially large volume of new patients to be assessed for eligibility for anti-TNF treatment could add a large burden to existing services. NICE guidance recommending the use of adalimumab, certolizumab and etanercept in nr-AxSpA would further increase the impact of these agents on the NHS budget.

Chapter 8 Discussion

Statement of principal findings

The systematic review of clinical efficacy identified a substantial and, generally, high-quality evidence base on the efficacy and safety of anti-TNFs in patients with AS, either as individual treatments or as a common class; there was limited evidence to suggest meaningful differences between the therapies in terms of efficacy, other than infliximab providing more rapid improvements during the first few months of treatment. The results of our meta-analyses demonstrated that anti-TNFs (when compared with placebo) produce statistically significant and clinically important benefits in patients with AS in terms of improving function and reducing disease activity over a 3- to 6-month period (none of the trials maintained randomised treatment allocations across groups beyond 6 months). Of the limited number of trials which reported quality-of-life outcomes, significant improvements were found following anti-TNF therapy but very few data were available on efficacy relating to any peripheral symptoms (other than enthesitis) or other possible symptoms such as uveitis, inflammatory bowel disease and psoriasis.

Although far fewer trials have been performed in the nr-AxSpA population, similar, although slightly smaller, benefits were achieved. The smaller benefit was most noticeable for the function (BASFI) and disease-activity (BASDAI 50) outcomes. However, in the nr-AxSpA trials, both clinical and statistical heterogeneity were evident, bringing into question both the reliability of the nr-AxSpA meta-analysis results and their true relevance to patients seen in clinical practice. This heterogeneity may have been compounded by the inclusion criteria applied in previous nr-AxSpA trials. For example, ABILITY-1[58] recruited patients who fulfilled the ASAS classification criteria and relied on the expertise of the local clinicians and/or radiologists to read sacroiliac joint radiographs and MRI images, as happens in real clinical practice. RAPID-axSpA[64] selected its population carefully by requiring objective evidence of disease activity at study entry by either a positive MRI showing signs of sacroiliac joint inflammation according to the ASAS/OMERACT definition or an elevated than normal CRP level. The difficulty of identifying which nr-AxSpA patients should receive anti-TNFs remains.

Results from open-label trial extension studies suggested that across all the anti-TNFs around half of patients still achieve a good level of response after around 2 years of treatment. The data also suggest that at 5 years around 60% of golimumab patients, 50% of etanercept patients and 30% of adalimumab patients still achieve a good treatment response. However, these longer-term studies were not as well-reported as the RCTs, and their results were derived from less reliable data; it is therefore unknown if these are true treatment differences or if they are a result of differences in the follow-up protocols (e.g. stopping rules) and/or methods used to impute missing data.

Evidence for an effect of anti-TNFs on radiographic disease progression was limited; the relatively short-term follow-up available to date and the insensitivity of radiography as an imaging tool precluded the drawing of firm conclusions regarding the role of anti-TNFs in preventing or delaying the progression of AS. There are some data to suggest an identifiable benefit from around 4 years but results from ongoing long-term studies should help to clarify this issue.

The results from studies based on registry data demonstrated that sequential treatment with anti-TNFs can be worthwhile in patients with AS. However, the drug survival, response rates and benefits were reduced with second and third anti-TNFs, with the proportion of BASDAI 50 responders falling approximately 10% with each subsequent anti-TNF and the median BASDAIs and BASFIs achieved increasing (worsening).

Data from large systematic reviews, which included patients with a wide range of diseases, suggest that, in the short-term, anti-TNFs as a group are associated with significantly higher rates of serious infections, tuberculosis reactivation, non-melanoma skin cancer, total AEs and withdrawals because of AEs than control treatments. Specifically, infliximab is associated with significantly higher rates of total AEs and withdrawals because of AEs, and certolizumab pegol is associated with significantly higher rates of serious infections and SAEs. Evaluations of longer-term data are more scarce and are limited by small sample sizes and uncontrolled designs. They suggest similar safety profiles across anti-TNFs, other than a higher incidence of injection site reactions following treatment with etanercept.

The systematic review of cost-effectiveness studies revealed significant conceptual issues and uncertainties arising from previously published studies and the submissions made by manufacturers. For this reason, a de novo model ('York model') was developed. Although it shared some of the assumptions and parameter estimates from the manufacturer models, it was based on a different conceptual structure and applies a more generalised framework for the synthesis of data from the double-blind periods of existing RCTs, combined with a more explicit approach to modelling the progressive nature of AS and nr-AxSpA and the potential impact of the TNF-α inhibitors.

Based on an underlying assumption of similarity in the clinical effectiveness of each of the TNF-α inhibitors, the York model demonstrates that the cost-effectiveness results are dependent on several factors, including (1) the different acquisition and administration costs; (2) the rebound assumption applied to patients who discontinue therapy; (3) the magnitude of the change in BASDAI/BASFI scores assumed for responders versus non-responders; (4) the different baseline BASDAI/BASFI scores assumed for responders versus non-responders; and (5) the impact of TNF-α inhibitors on the rate of longer-term BASFI progression.

Although there are a number of important differences in approaches both among the different manufacturer models and compared with the York model, the comparison of ICERs based on the York rebound equal to gain scenario appear broadly consistent with those reported by the manufacturers in both populations.

Strengths and limitations of the assessment

Strengths

Through our comprehensive searches we sought to identify all relevant published and unpublished trials, which minimised the possibility of publication or language biases affecting the review results. A full evaluation of the risk of bias in each RCT was performed, which incorporated an additional assessment of key baseline characteristics to allow firmer judgements to be made on the risk of selection bias. The use of multiple-treatment meta-analyses allowed for greater precision in random-effect models, and the calculation of relative risks was based on the population risk across all the trials. A key further strength of our review lies in the extensive breadth of other types of study we included, such as non-randomised trial extension studies, registry studies of patients taking anti-TNFs, systematic reviews and other large studies of adverse effects of anti-TNFs and a review of the natural history of AS and nr-AxSpA. Our review of AEs incorporated a wealth of data from RCTs in patients on anti-TNFs with diseases other than AS and nr-AxSpA, although the results only relate to short-term use. Our review was performed according to Centre for Reviews and Dissemination guidance, so the potential for reviewer errors and biases was minimised. Our review was reported according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement.

The York model confers several advantages over current cost-effectiveness studies by linking changes in BASFI to a more explicit clinical/biological process and facilitating a more formal consideration of the potential impact of TNF-α inhibitors on BASFI, via the specific effects these drugs have on the different processes which independently relate to this parameter. This approach allows consideration of the impact on BASFI that might be achieved via symptomatic improvements (i.e. in terms of reductions in disease

activity) and those which might be conferred by disease modification properties (i.e. the effect on the likelihood and/or rate of further radiographic progression). The latter aspect is particularly important given the increasing amount of published evidence reported on the potential impact of TNF-α inhibitors on radiographic progression that has not been formally considered or incorporated within existing cost-effectiveness studies. In addition, the evidence synthesis approach which underpins the York model is based on a joint synthesis of related parameters which makes fuller use of existing evidence and which can more appropriately estimate the input parameters and better characterise the uncertainty surrounding these.

Limitations

A key limitation of the systematic review was the variation in the reporting of outcomes across trials. ASAS 20 was the most commonly reported responder outcome, but its value in determining efficacy was somewhat limited by the relatively high rates of 'placebo' response associated with the 20% threshold. Results for 40%, 50% and 70% improvements (i.e. ASAS 40, ASAS 50 and ASAS 70) were reported less frequently, despite the fact that trial investigators would have had the data available to do so. Many trials did not report HRQoL outcomes and most trials were also limited in their assessment (or reporting) of improvement in any peripheral symptoms or symptoms of extra-articular manifestations. Although largely free of important biases, most RCTs had quite short durations (generally around 3 months) and several were limited by their small sample sizes (increasing the possibility of chance results for some outcomes).

Although we sought data beyond those available from RCTs, much of the data reported in studies using other designs may have been affected by biases or confounding; furthermore, key method details (e.g. imputation methods, or anti-TNF stopping rules) were often absent from publications. Much less reliability and certainty could therefore be ascribed to the results obtained from these other studies.

The York model did not directly address important clinical questions concerning the issue of intermittent and sequential use of anti-TNFs because of the lack of robust clinical evidence from RCTs.

Uncertainties

- The magnitude of treatment effect of anti-TNFs in patients with nr-AxSpA remains uncertain because of the heterogeneous nature of the trials performed to date.
- The limited design and reporting of the studies looking at the long-term use of anti-TNFs means there is uncertainty whether or not there are differences in efficacy between the different anti-TNFs in the long term.
- The evidence on the long-term risk of AEs is uncertain because of small study sample sizes and the study designs used.
- The long-term impact of anti-TNFs on other important outcomes in AS and nr-AxSpA remain uncertain, such as AS-related causes of death (cardiac valvular disease, amyloidosis and fractures), and extra-articular symptoms such as uveitis, inflammatory bowel disease and psoriasis. Studies based on ongoing anti-TNF registries (e.g. BSRBR) that record such data should inform this.
- With the patents of some anti-TNFs studied in this assessment due to expire shortly, biosimilars are likely to become available in the next few years (CT-P13 became available early in 2015). As they are difficult to produce, the number of biosimilars which will become available, and their price, is uncertain.

Chapter 9 Conclusions

Meta-analysis results derived from a substantial and, generally, high-quality evidence base on the efficacy of anti-TNFs in patients with AS (considered either as individual treatments or as a common class) show statistically significant and clinically important benefits in terms of improved function and reduced disease activity following around 3 months of treatment with an anti-TNF. Smaller benefits were seen across outcomes in patients with nr-AxSpA, being most noticeably smaller for the function and disease-activity outcomes. However, in the light of the clinical and statistical heterogeneity seen across the nr-AxSpA trials, both the reliability of the nr-AxSpA-pooled estimates and their true relevance to patients seen in clinical practice are questionable. Data from (less robust) observational studies suggest that good levels of treatment response are maintain in around 50% of patients after around 2 years of treatment. Evidence for an effect of anti-TNFs on radiographic disease progression is limited, although results from ongoing studies should clarify whether or not progression rates are reduced in the longer term. The results from studies based on registry data demonstrated that sequential treatment with anti-TNFs can be worthwhile in patients with AS, although drug survival, response rates and benefits were reduced with second and third anti-TNFs. Data from large systematic reviews, which included patients with a wide range of diseases, suggested that, in the short term, anti-TNFs as a group were associated with significantly higher rates of serious infections, tuberculosis reactivation, non-melanoma skin cancer, total AEs and withdrawals because of AEs than control treatments. Longer-term data on AEs were limited.

Implications for service provision

- From our review of natural history a key study on flares suggested that the 12-week period required to confirm sustained active spinal disease in AS patients commencing an anti-TNF may be too long. The findings suggest that shorter time periods might therefore be considered in future guidance, which would minimise the delay in starting treatment and the discomfort experienced by patients.

Suggested research priorities

- Randomised trials are needed to identify the nr-AxSpA population that will benefit the most from TNF-inhibitors; trials using stratified randomisation and pre-planned analyses by stratified group should inform this issue. Groups could be stratified according to their imaging status (i.e. MRI positive or not) and their CRP level; both the cut-off points to be used for CRP level elevation, and the eligibility criteria used for CRP level elevation, should be given careful consideration, given the variation evident in previous trials. These studies should help to inform clearer guidance as to what ASAS and the anti-TNF licences mean when referring to 'elevated CRP level' in patients with nr-AxSpA. There is also a clear need for more accurate biomarkers, or other measures of disease activity, to be developed. In the previous nr-AxSpA trials the placebo-controlled phases typically lasted around 3 months; a placebo-controlled follow-up period of at least 6 months in future trials would therefore be useful for studying persistence of response.
- Long-term longitudinal studies are needed on the natural history of nr-AxSpA to help clarify the characteristics of patients who do (or do not) eventually develop AS. Similar to the RCT recommendations, these studies should include analyses stratified by how patients were diagnosed; a comparison of patients with imaging (MRI) evidence of nr-AxSpA versus patients who are diagnosed with only clinical criteria evidence, would be particularly useful, albeit difficult to perform.

- Large, long-term longitudinal, cohort studies are needed to clarify the effect of anti-TNFs on the progression of structural damage in AS. In the absence of a gold standard imaging tool across the spectrum from nr-AxSpA to AS, sequential MRI and radiography assessment should be used at pre-defined end points to ascertain the true sensitivity and specificity of these tools in the diagnosis and assessment of neo-formation, and ankyloses characteristic of structural progression in the spine and sacroiliac joints of these patients.

Studies are also needed to better inform the efficacy estimates relating to sequential use of anti-TNFs. An ongoing study is looking at comparing the effect of intermittent versus standard use of anti-TNFs in patients with stable (low-active) disease.[182]

Acknowledgements

We acknowledge Nicky Welton, Reader in Evidence Synthesis, School of Social and Community Medicine, University of Bristol, for the guidance provided in the development of the synthesis model.

Rider on responsibility for report

The views expressed in this report are those of the authors and not necessarily those of the NIHR HTA programme. Any errors are the responsibility of the authors.

Contributions of authors

Mark Corbett contributed to the protocol, study selection, data extraction, validity assessment, synthesis of the included studies, interpretation of the results and the writing of the report.

Marta Soares developed the extended synthesis model and undertook the evidence synthesis for the purposes of the economic model. She also contributed to the development of the economic model and the writing of the report.

Gurleen Jhuti contributed to the protocol, the development of the economic model, the review of economic analyses, the interpretation of the results and the writing of the report.

Stephen Rice contributed to data extraction, validity assessment, undertook synthesis of the included studies for the clinical effectiveness section, contributed to the interpretation of the results and the writing of the report.

Eldon Spackman contributed to the development of the economic model, performed the economic analyses and contributed to the interpretation of the results.

Eleftherios Sideris contributed to the review of economic evaluations, the development of the economic model and the writing of the report.

Thirimon Moe-Byrne contributed to the protocol, study selection, data extraction and validity assessment of the included studies.

Dave Fox contributed to the protocol development, developed the search strategies, conducted a range of searches to locate studies and wrote the sections of the report relating to the searches.

Helena Marzo-Ortega provided expert clinical advice, contributed to the protocol, interpretation of the results and commented on drafts of the report.

Lesley Kay provided expert clinical advice, contributed to the protocol, interpretation of the results and commented on drafts of the report.

Nerys Woolacott contributed to the protocol, study selection, data extraction, validity assessment, synthesis of the included studies, interpretation of the results and the writing of the report, and took overall responsibility for the clinical effectiveness section.

Stephen Palmer had overall responsibility for the cost-effectiveness sections. He contributed to the protocol development and to all aspects of the cost-effectiveness work including the writing of the report.

Data sharing statement

All available data are either included within the main report or available in the appendices.

References

1. Lipton S, Deodhar A. The new ASAS classification criteria for axial and peripheral spondyloarthritis. *Int J Clin Rheumatol* 2012;**7**:675–82. http://dx.doi.org/10.2217/ijr.12.61
2. Vanderlinden S, Valkenburg HA, Cats A. Evaluation of diagnostic-criteria for ankylosing-spondylitis – a proposal for modification of the New York criteria. *Arthritis Rheum* 1984;**27**:361–8. http://dx.doi.org/10.1002/art.1780270401
3. Sieper J, Rudwaleit M, Baraliakos X, Brandt J, Braun J, Burgos-Vargas R, *et al.* The Assessment of SpondyloArthritis international Society (ASAS) handbook: a guide to assess spondyloarthritis. *Ann Rheum Dis* 2009;**68**:1–44. http://dx.doi.org/10.1136/ard.2008.104018
4. Rudwaleit M, van der Heijde D, Landewe R, Listing J, Akkoc N, Brandt J, *et al.* The development of Assessment of SpondyloArthritis International Society classification criteria for axial spondyloarthritis (part II): validation and final selection. *Ann Rheum Dis* 2009;**68**:777–83. http://dx.doi.org/10.1136/ard.2009.108233
5. Goh L, Samanta A. Update on biologic therapies in ankylosing spondylitis: a literature review. *Int J Rheum Dis* 2012;**15**:445–54. http://dx.doi.org/10.1111/j.1756-185X.2012.01765.x
6. Robinson PC, Wordsworth BP, Reveille JD, Brown MA. Viewpoint axial spondyloarthritis: a new disease entity, not necessarily early ankylosing spondylitis. *Ann Rheum Dis* 2013;**72**:162–4. http://dx.doi.org/10.1136/annrheumdis-2012-202073
7. van Tubergen A, Weber U. Diagnosis and classification in spondyloarthritis: identifying a chameleon. *Nat Rev Rheumatol* 2012;**8**:253–61. http://dx.doi.org/10.1038/nrrheum.2012.33
8. Deodhar A. Rheumatologists Make Progress Defining Spectrum of Axial Spondyloarthritis. *The Rheumatologist*, May 2014. URL: www.the-rheumatologist.org/issue/may-2014 (accessed 1 December 2015).
9. National Ankylosing Spondylitis Society. *National Ankylosing Spondylitis Society (NASS) Website*. URL: http://nass.co.uk/ (accessed 11 December 2014).
10. 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;**68**:863–7. http://dx.doi.org/10.1136/ard.2008.091793
11. Braun J, Sieper J. Ankylosing spondylitis. *Lancet* 2007;**369**:1379–90. http://dx.doi.org/10.1016/S0140-6736(07)60635-7
12. Baraliakos X, Listing J, von der Recke A, 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;**36**:997–1002. http://dx.doi.org/10.3899/jrheum.080871
13. Healey EL, Haywood KL, Jordan KP, Garratt AM, Packham JC. Disease severity in ankylosing spondylitis: variation by region and local area deprivation. *J Rheumatol* 2010;**37**:633–8. http://dx.doi.org/10.3899/jrheum.090899
14. Healey EL, Haywood KL, Jordan KP, Garratt AM, Packham JC. Patients with well-established ankylosing spondylitis show limited deterioration in a ten-year prospective cohort study. *Clin Rheumatol* 2013;**32**:67–72. http://dx.doi.org/10.1007/s10067-012-2092-3
15. Callhoff J, Sieper J, Weiss A, Zink A, Listing J. Efficacy of TNFα blockers in patients with ankylosing spondylitis and non-radiographic axial spondyloarthritis: a meta-analysis. *Ann Rheum Dis* 2015;**74**:1241–8. http://dx.doi.org/10.1136/annrheumdis-2014-205322

16. Bakland G, Gran JT, Nossent JC. Increased mortality in ankylosing spondylitis is related to disease activity. *Ann Rheum Dis* 2011;**70**:1921–5. http://dx.doi.org/10.1136/ard.2011.151191

17. National Institute for Health and Care Excellence. *Technology Appraisal TA143: Ankylosing Spondylitis – Adalimumab, Etanercept and Infliximab*. London: NICE; 2008.

18. Rudwaleit M, Haibel H, Baraliakos X, Listing J, Marker-Hermann E, Zeidler H, *et al.* The early disease stage in axial spondylarthritis results from the German Spondyloarthritis Inception Cohort. *Arthritis Rheum* 2009;**60**:717–27. http://dx.doi.org/10.1002/art.24483

19. Reveille JD, Sims A-M, Danoy P, Evans DM, Leo P, Pointon JJ, *et al.* Genome-wide association study of ankylosing spondylitis identifies non-MHC susceptibility loci. *Nat Genet* 2010;**42**:123–7. http://dx.doi.org/10.1038/ng.513

20. NHS Choices. *Ankylosing Spondylitis*. NHS Choices. URL: www.nhs.uk/conditions/Ankylosing-spondylitis/Pages/Introduction.aspx (accessed 29 April 2014).

21. Reveille JD, Weisman MH. The epidemiology of back pain, axial spondyloarthritis and HLA-B27 in the United States. *Am J Med Sci* 2013;**345**:431–6. http://dx.doi.org/10.1097/MAJ.0b013e318294457f

22. Sieper J, van der Heijde D. Review: nonradiographic axial spondyloarthritis: new definition of an old disease? *Arthritis Rheum* 2013;**65**:543–51. http://dx.doi.org/10.1002/art.37803

23. National Institute for Health and Care Excellence. *Ankylosing Spondylitis: Prevalence*. London: NICE; 2013. URL: http://cks.nice.org.uk/ankylosing-spondylitis#!backgroundsub:2 (accessed 29 April 2014).

24. Machado P, Landewe R, Braun J, Hermann KG, Baraliakos X, Baker D, *et al.* A stratified model for health outcomes in ankylosing spondylitis. *Ann Rheum Dis* 2011;**70**:1758–64. http://dx.doi.org/10.1136/ard.2011.150037

25. Wanders AJB, Landewe RBM, Spoorenberg A, Dougados M, van der Linden S, Mielants H, *et al.* What is the most appropriate radiologic scoring method for ankylosing spondylitis? *Arthritis Rheum* 2004;**50**:2622–32. http://dx.doi.org/10.1002/art.20446

26. Hróbjartsson A. What are the main methodological problems in the estimation of placebo effects? *J Clin Epidemiol* 2002;**55**:430–5. http://dx.doi.org/10.1016/S0895-4356(01)00496-6

27. Hawkins N, Scott DA. Cost-effectiveness analysis: discount the placebo at your peril. *Med Decis Making* 2010;**30**:536–43. http://dx.doi.org/10.1177/0272989X10362106

28. Cooksey R, Brophy S, Gravenor MB, Brooks CJ, Burrows CL, Siebert S. Frequency and characteristics of disease flares in ankylosing spondylitis. *Rheumatology* 2010;**49**:929–32. http://dx.doi.org/10.1093/rheumatology/kep435

29. Hróbjartsson A, Gøtzsche PC. Placebo interventions for all clinical conditions. *Cochrane Database Sys Rev* 2010;**1**:CD003974. http://dx.doi.org/10.1002/14651858.cd003974.pub3

30. Braun J, van den Berg R, Baraliakos X, Boehm H, Burgos-Vargas R, Collantes-Estevez E, *et al.* 2010 update of the ASAS/EULAR recommendations for the management of ankylosing spondylitis. *Ann Rheum Dis* 2011;**70**:896–904. http://dx.doi.org/10.1136/ard.2011.151027

31. Dagfinrud H, Hagen KB, Kvien TK. Physiotherapy interventions for ankylosing spondylitis. *Cochrane Database Sys Rev* 2008;**1**:CD002822. http://dx.doi.org/10.1002/14651858.cd002822.pub3

32. National Ankylosing Spondylitis Society (NASS). *Looking Ahead: Best Practice for the Care of People with Ankylosing Spondylitis (AS)*. Richmond: NASS; 2013.

33. National Institute for Health and Care Excellence. *Technology Appraisal TA233: Ankylosing Spondylitis – Golimumab*. London: NICE; 2011.

34. AbbVie. *National Institute for Health and Care Excellence (NICE) Multiple Technology Appraisal: TNF-alpha Inhibitors for Severe Active Ankylosing Spondylitis and Severe Axial Spondyloarthritis without Radiographic Evidence of Ankylosing Spondylitis*. Submission by AbbVie, manufacturer of adalimumab (HUMIRA®), September 2014.

35. UCB. *National Institute for Health and Care Excellence (NICE) Multiple Technology Appraisal: Certolizumab Pegol for the Treatment of Ankylosing Spondylitis and Axial Spondyloarthritis without Radiographic Evidence of Ankylosing Spondylitis*. Submission by UCB, manufacturer of certolizumab pegol, September 2014.

36. Pfizer. *National Institute for Health and Care Excellence (NICE) Multiple Technology Appraisal: Ankylosing Spondylitis and Axial Spondyloarthritis (Non-radiographic) – Adalimumab, Etanercept Infliximab and Golimumab*. Submission by Pfizer, manufacturer of etanercept (Enbrel®), September 2014.

37. Merck Sharp & Dohme. *National Institute for Health and Care Excellence (NICE) Multiple Technology Appraisal: TNF-alpha Inhibitors for Ankylosing Spondylitis and Axial Spondyloarthritis without Radiographic Evidence of Ankylosing Spondylitis (Including a Review of TA143 and TA233)*. Submission by Merck Sharp & Dohme, manufacturer of golimumab (Simponi®), September 2014.

38. McLeod C, Bagust A, Boland A, Dagenais P, Dickson R, Dundar Y, *et al.* Adalimumab, etanercept and infliximab for the treatment of ankylosing spondylitis: a systematic review and economic evaluation. *Health Technol Assess* 2007;**11**(28). http://dx.doi.org/10.3310/hta11280

39. Higgins JPT, Altman DG, Gotzsche PC, Jueni P, Moher D, Oxman AD, *et al.* The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;**343**:d5928. http://dx.doi.org/10.1136/bmj.d5928

40. Corbett MS, Higgins JPT, Woolacott NF. Assessing baseline imbalance in randomised trials: implications for the Cochrane risk of bias tool. *JRSM* 2014;**5**:79–85. http://dx.doi.org/10.1002/jrsm.1090

41. Pedersen SJ, Sorensen IJ, Loft AG, Hindrup JS, Thamsborg G, Asmussen K, *et al. Efficacy of Adalimumab in Patients with Axial Spondyloarthritis: Results of an Investigator-Initiated 12-Weeks Randomized Double-Blind Placebo Controlled Trial with a 12 Weeks Open-Label Extension Phase*. Annual European Congress of Rheumatology of the European League Against Rheumatism, EULAR, Madrid, 12–15 June 2013.

42. Sieper J, Lenaerts J, Wollenhaupt J, Rudwaleit M, Mazurov VI, Myasoutova L, *et al.* Efficacy and safety of infliximab plus naproxen versus naproxen alone in patients with early, active axial spondyloarthritis: results from the double-blind, placebo-controlled INFAST study, Part 1. *Ann Rheum Dis* 2014;**73**:101–7. http://dx.doi.org/10.1136/annrheumdis-2012-203201

43. Dougados M, Wood E, Combe B, Schaeverbeke T, Miceli-Richard C, Berenbaum F, *et al.* Evaluation of the nonsteroidal anti-inflammatory drug-sparing effect of etanercept in axial spondyloarthritis: results of the multicenter, randomized, double-blind, placebo-controlled SPARSE study. *Arthritis Res Ther* 2014;**16**:481. http://dx.doi.org/10.1186/s13075-014-0481-5

44. ClinicalTrials.gov. *Danish Multicenter Study of Adalimumab in Spondyloarthritis*. URL: http://ClinicalTrials.gov/show/NCT00477893 (accessed 1 December 2014).

45. ClinicalTrials.gov. *Effect of Golimumab in Participants with Active Axial Spondyloarthritis (P07642 AM2)*. URL: http://ClinicalTrials.gov/show/NCT01453725 (accessed 1 December 2014).

46. 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;**48**:1667–75. http://dx.doi.org/10.1002/art.11017

47. Zhang J, Huang F, Chen SL, Gu JR, Li ZG, Wu DH, *et al.* A multicenter, double-blind, placebo-controlled, randomized III clinical study of etanercept in the treatment of Chinese subjects with active ankylosing spondylitis. *Int J Rheum Dis* 2010;**13**:151.

48. Zhang J, Zhang YM, Zhang JL, Deng XH, Huang F. [Efficacy of etanercept in patients with ankylosing spondylitis: a double-blind, randomized, placebo controlled trial.] *Chin J New Drugs* 2009;**18**:1846–9.

49. Inman RD, Maksymowych WP, Candle Study Group. A double-blind, placebo-controlled trial of low dose infliximab in ankylosing spondylitis. *J Rheumatol* 2010;**37**:1203–10. http://dx.doi.org/10.3899/jrheum.091042

50. Barkham N, Keen HI, Coates LC, O'Connor P, Hensor E, Fraser AD, *et al.* Clinical and imaging efficacy of infliximab in HLA-B27-positive patients with magnetic resonance imaging-determined early sacroiliitis. *Arthritis Rheum* 2009;**60**:946–54. [Erratum published in *Arthritis Rheum* 2010;**62**:3005.] http://dx.doi.org/10.1002/art.24408

51. Barkham N, Keen HI, Coates LC, O'Connor P, Hensor EMA, Fraser A, *et al.* A randomized controlled trial of infliximab shows clinical and MRI efficacy in HLA B27 positive very early ankylosing spondylitis. *Clin Exp Rheumatol* 2008;**26**:745.

52. 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 Rheum* 2008;**58**:1981–91. http://dx.doi.org/10.1002/art.23606

53. Haibel H, Rudwaleit M, Heldmann F, Listing J, Wong R, Kupper H, *et al.* Adalimumab is effective in patients with preradiographic axial spondyloarthritis: results of a 12-week, randomized controlled trial. *Rheumatology* 2007;**46**(Suppl. 1):i52.

54. Haibel H, Rudwaleit M, Heldmann F, Listing J, Wong R, Kupper H, *et al.* Efficacy of adalimumab in the treatment of preradiographic axial spondyloarthritis: results of a 12-week, randomized controlled trial. *Arthritis Rheum* 2006;**54**:4044–5. http://dx.doi.org/10.1002/art.23606

55. Hu Z, Xu M, Li Q, Lin Z, Liao Z, Cao S, *et al.* Adalimumab significantly reduces inflammation and serum DKK-1 level but increases fatty deposition in lumbar spine in active ankylosing spondylitis. *Int J Rheum Dis* 2012;**15**:358–65. [Erratum published in *Int J Rheum Dis* 2012;**15**:564.] http://dx.doi.org/10.1111/j.1756-185X.2012.01734.x

56. Huang F, Gu J, Zhu P, Bao C, Xu J, Xu H, *et al.* Efficacy and safety of adalimumab in Chinese adults with active ankylosing spondylitis: results of a randomised, controlled trial. *Ann Rheum Dis* 2014;**73**:587–94. http://dx.doi.org/10.1136/annrheumdis-2012-202533

57. 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;**56**:4005–14. http://dx.doi.org/10.1002/art.23044

58. Sieper J, van der Heijde D, Dougados M, Mease PJ, Maksymowych WP, Brown MA, *et al.* Efficacy and safety of adalimumab in patients with non-radiographic axial spondyloarthritis: results of a randomised placebo-controlled trial (ABILITY-1). *Ann Rheum Dis* 2013;**72**:815–22. http://dx.doi.org/10.1136/annrheumdis-2012-201766

59. Sieper J, Van Der Heijde D, Dougados M, Mease PJ, Brown LS, Pangan A. Efficacy and safety of adalimumab in patients with non-radiographic axial spondyloarthritis results from a phase 3 study. *Arthritis Rheum* 2011;**63**(Suppl. 10):2486A. http://dx.doi.org/10.1136/annrheumdis-2012-201766

60. Maksymowych WP, Mease PJ, Rao S, Pangan A, Brown LS, Arora V, *et al.* Effect of adalimumab on function, health-related quality of life, work productivity, and daily activities in patients with non-radiographic axial spondyloarthritis. *Arthritis Rheum* 2011;**63**(Suppl. 10):1312.

61. 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;**54**:2136–46. http://dx.doi.org/10.1002/art.21913

62. Davis JC Jr, Revicki D, van der Heijde DM, 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 Rheum* 2007;**57**:1050–7. http://dx.doi.org/10.1002/art.22887

63. Davis J, Kivitz A, Schiff M, Sieper J, Dijkmans B, Braun J, *et al.* Major clinical response and partial remission in ankylosing spondylitis subjects treated with adalimumab: the ATLAS trial. *Arthritis Rheum* 2005;**52**:S208–9.

64. Landewe R, Braun J, Deodhar A, Dougados M, Maksymowych WP, Mease PJ, *et al.* Efficacy of certolizumab pegol on signs and symptoms of axial spondyloarthritis including ankylosing spondylitis: 24-week results of a double-blind randomised placebo-controlled Phase 3 study. *Ann Rheum Dis* 2014;**73**:39–47. http://dx.doi.org/10.1136/annrheumdis-2013-204231

65. van der Heijde D, Maksymowych WP, Landewe R, Stach C, Hoepken B, Fichtner A, *et al.* Effect of certolizumab pegol on inflammation of spine and sacroiliac joints in patients with axial spondyloarthritis: 12-week magnetic resonance imaging results of rapid-axSpA study. *Ann Rheum Dis* 2013;**72**:A515–16. http://dx.doi.org/10.1136/annrheumdis-2013-eular.1546

66. Sieper J, Kivitz A, Van Tubergen A, Deodhar A, Coteur G, Woltering F, *et al.* Rapid improvements in patient-reported outcomes with certolizumab pegol in patients with axial spondyloarthritis, including ankylosing spondylitis: 24-week results of rapid-axSpA study. *Ann Rheum Dis* 2013;**72**(Suppl. 3):A287. http://dx.doi.org/10.1136/annrheumdis-2013-eular.888

67. Landewe R, Rudwaleit M, van der Heijde D, Dougados M, Maksymowych WP, Braun J, *et al.* Effect of certolizumab pegol on signs and symptoms of axial spondyloarthritis, including ankylosing spondylitis and non-radiographic axial spondyloarthritis: 24-week results of rapid-axSpA study. *Ann Rheum Dis* 2013;**72**:A86–7. http://dx.doi.org/10.1136/annrheumdis-2013-eular.311

68. Sieper J, Kivitz A, Van Tubergen A, Deodhar A, Coteur G, Woltering F, *et al.* Rapid improvements in patient-reported outcomes with certolizumab pegol in patients with axial spondyloarthritis, including ankylosing spondylitis and non-radiographic axial spondyloarthritis: 24-week results of a phase 3 double blind randomized placebo-controlled study. *Value Health* 2013;**16**:A227. http://dx.doi.org/10.1016/j.jval.2013.03.1150

69. Landewe RBM, Rudwaleit M, van der Heijde D, Dougados M, Maksymowych WP, Braun J, *et al.* Effect of certolizumab pegol on signs and symptoms of ankylosing spondylitis and non-radiographic axial spondyloarthritis: 24 week results of a double blind randomized placebo-controlled phase 3 axial spondyloarthritis study. *Arthritis Rheum* 2012;**64**:S336–7.

70. Sieper J, Kivitz AJ, Van Tubergen AM, Deodhar AA, Coteur G, Woltering F, *et al.* Rapid improvements in patient reported outcomes with certolizumab pegol in patients with axial spondyloarthritis, including ankylosing spondylitis and non-radiographic axial spondyloarthritis: 24 week results of a phase 3 double blind randomized placebo-controlled study. *Arthritis Rheum* 2012;**64**:S243.

71. Barkham N, Coates LC, Keen H, Hensor E, Fraser A, Redmond A, *et al.* Double-blind placebo-controlled trial of etanercept in the prevention of work disability in ankylosing spondylitis. *Ann Rheum Dis* 2010;**69**:1926–8. [Erratum published in *Ann Rheum Dis*2011;**70**:151.] http://dx.doi.org/10.1136/ard.2009.121327

72. 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;**48**:3230–6. http://dx.doi.org/10.1002/art.11325

73. Davis JC, van der Heijde D, Braun J, Dougados M, Cush JJ, Clegg DO, *et al.* Etanercept (Enbrel®) improves signs and symptoms of ankylosing spondylitis: Results of a phase 3 multicenter clinical trial. *Ann Rheum Dis* 2003;**62**(Suppl. 1):65.

74. Dougados M, Braun J, Szanto S, Combe B, Elbaz M, Geher P, *et al.* Efficacy of etanercept on rheumatic signs and pulmonary function tests in advanced ankylosing spondylitis: results of a randomised double-blind placebo-controlled study (SPINE). *Ann Rheum Dis* 2011;**70**:799–804. [Erratum appears in *Ann Rheum Dis* 2011;**70**:1349]. http://dx.doi.org/10.1136/ard.2010.139261

75. Dougados M, Braun J, Szanto S, Combe B, Elbaz M, Geher P, *et al.* Efficacy of etanercept on rheumatic signs and pulmonary function tests in advanced ankylosing spondylitis. Results of a randomized double blind placebo-controlled study (SPINE). *Arthritis Rheum* 2009;**60**:524. http://dx.doi.org/10.1136/ard.2010.139261

76. Dougados M. Symptomatic efficacy of etanercept and its effects on objective signs of inflammation in early nonradiographic axial spondyloarthritis: a multicenter, randomized, double-blind, placebo-controlled trial. *Arthritis Rheum* 2014;**66**:2091–102. http://dx.doi.org/10.1002/art.38721

77. Dougados M, van der Heijde D, Sieper J, Braun J, Maksymowych WP, Citera G, *et al.* Clinical and imaging efficacy of etanercept in early non-radiographic axial spondyloarthritis: a 12-week, randomized, double-blind, placebo-controlled trial. *Ann Rheum Dis* 2013;**72**:A87–8. http://dx.doi.org/10.1136/annrheumdis-2013-eular.313

78. Dougados M, Tsai WC, Saaibi DL, Bonin R, Bukowski J, Pedersen R, *et al.* Patient-reported outcomes of etanercept in early non-radiographic axial spondyloarthritis: a 12-week, randomized, double-blind, placebo-controlled trial. *Arthritis Rheum* 2013;**65**:S656–7.

79. Gorman JD, Sack KE, Davis JC Jr. Treatment of ankylosing spondylitis by inhibition of tumor necrosis factor alpha. *N Engl J Med* 2002;**346**:1349–56. http://dx.doi.org/10.1056/NEJMoa012664

80. Gorman JD, Sack KE, Davis JC. Efficacy of etanercept (Enbrel®) in the treatment of ankylosing spondylitis: a randomized, placebo-controlled, double-blind study. *Arthritis Rheum* 2001;**44**:2947.

81. Gorman JD, Sack KE, Davis JC. A randomized, double-blind, placebo-controlled trial of (Enbrel®) in the treatment of ankylosing spondylitis. *Arthritis Rheum* 2001;**44**:S90.

82. Gorman JD, Sack KE, Davis JC. Etanercept in the treatment of ankylosing spondylitis: a randomized, double-blind, placebo-controlled study. *Arthritis Rheum* 2000;**43**:S403.

83. 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;**63**:1594–600. http://dx.doi.org/10.1136/ard.2004.020875

84. Calin A, Dijkmans B, Emery P, Hakala M, Kalden J, Leirisalo-Repo M, *et al.* Assessments of disease activity and functionality by Enbrel-treated ankylosing spondylitis patients in a multicenter, placebo-controlled trial. *Arthritis Rheum* 2003;**48**:S172.

85. Calin A, Dijkmans B, Emery P, Hakala M, Kalden J, Leirisalo-Repo M, *et al.* A multicentre, placebo-controlled trial of enbrel in ankylosing spondylitis. *Ann Rheum Dis* 2003;**62**:95.

86. van der Heijde D, Da Silva JC, Dougados M, Geher P, van der Horst-Bruinsma 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;**65**:1572–7. http://dx.doi.org/10.1136/ard.2006.056747

87. 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**:999–1004. http://dx.doi.org/10.1093/rheumatology/kem069

88. 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* 2010;**30**:1437–40. http://dx.doi.org/10.1007/s00296-009-1157-3

89. 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. *Ann Rheum Dis* 2007;**66**:399. http://dx.doi.org/10.1007/s00296-009-1157-3

90. Inman RD, Davis JC Jr, van der Heijde D, Diekman L, Sieper J, Kim SI, *et al.* Efficacy and safety of golimumab in patients with ankylosing spondylitis: results of a randomized, double-blind, placebo-controlled, phase III trial. *Arthritis Rheum* 2008;**58**:3402–12. http://dx.doi.org/10.1002/art.23969

91. Braun J, Davis J, Van Der Heijde D, Deodhar A, Diekman L, Sieper J, *et al.* Golimumab, a new, human, TNFA antibody, in ankylosing spondylitis: 24-week efficacy and safety results of the golimumab, a new, human, a antibody, in ankylosing spondylitis: 24-week efficacy and safety results of the GO-RAISE study. *Intern Med J* 2009;**39**:A56.

92. Braun J, Davis J, Van Der Heijde D, Deodhar A, Diekman L, Sieper J, *et al.* Golimumab, a new, human, TNFalpha antibody, in ankylosing spondylitis (AS): 24-week efficacy and safety results of the go-raise study. *Rheumatology* 2009;**48**:i57.

93. Braun J, Davis JC, van der Heijde D, Diekman L, Sieper J, Kim SI, *et al.* Golimumab, a new, human, TNF-alpha antibody administered subcutaneously every 4 weeks, in ankylosing spondylitis: 24-week efficacy and safety results of the randomized, placebo-controlled GO-RAISE study. *Clin Exp Rheumatol* 2008;**26**:747.

94. Braun J, Davis JC, Van der Heijde D, Diekman L, Sieper J, Kim SI, *et al.* Golimumab, a new, human, TNF-alpha antibody administered as a monthly subcutaneous injection in ankylosing spondylitis (AS): 24-week efficacy and safety results of the randomized, placebo-controlled GO-RAISE study. *Arthritis Rheum* 2007;**56**:4236–7.

95. Bao C, Huang F, Khan MA, Fei K, Wu Z, Han C, *et al.* Safety and efficacy of golimumab in Chinese patients with active ankylosing spondylitis: 1-year results of a multicentre, randomized, double-blind, placebo-controlled phase III trial. *Rheumatology* 2014;**53**:1654–63. http://dx.doi.org/10.1093/rheumatology/keu132

96. Bao C, Huang F, Khan MA, Fei K, Wu Z, Hsia EC. Golimumab administered subcutaneously every 4 weeks in Chinese patients with active ankylosing spondylitis: week 24 safety and efficacy results from a randomized, placebo-controlled study. *Arthritis Rheum* 2012;**64**:S589.

97. Tam L-S, Shang Q, Kun EW, Lee K-L, Yip M-L, Li M, *et al.* The effects of golimumab on subclinical atherosclerosis and arterial stiffness in ankylosing spondylitis – a randomized, placebo-controlled pilot trial. *Rheumatology* 2014;**53**:1065–74. http://dx.doi.org/10.1093/rheumatology/ket469

98. 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;**359**:1187–93. http://dx.doi.org/10.1016/S0140-6736(02)08215-6

99. Brandt J, Listing I, Alten R, Krause A, Gromnica-Ihle E, Kellner H, *et al.* Quality of life improvement in patients with severe ankylosing spondylitis upon treatment with the anti-TNF alpha antibody infliximab in a placebo controlled multicenter trial. *Arthritis Rheum* 2001;**44**:S89.

100. 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;**64**:1568–75. http://dx.doi.org/10.1136/ard.2004.022582

101. 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) versus placebo in active spondylarthropathy. *Arthritis Rheum* 2002;**46**:755–65. http://dx.doi.org/10.1002/art.511

102. 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;**52**:582–91. http://dx.doi.org/10.1002/art.20852

103. van der Heijde D, Landewe R, Baraliakos X, Houben H, van Tubergen A, Williamson P, *et al.* Radiographic findings following two years of infliximab therapy in patients with ankylosing spondylitis. *Arthritis Rheum* 2008;**58**:3063–70. http://dx.doi.org/10.1002/art.23901

104. van der Heijde D, Han C, DeVlam K, Burmester G, van den Bosch F, Williamson P, *et al.* Infliximab improves productivity and reduces workday loss in patients with ankylosing spondylitis: results from a randomized, placebo-controlled trial. *Arthritis Rheum* 2006;**55**:569–74. http://dx.doi.org/10.1002/art.22097

105. van der Heijde D, Landewe R, Hermann K, Han J, Williamson P, Braun J. The effect of infliximab therapy on spinal inflammation assessed by magnetic resonance imaging in a randomized, placebo-controlled trial of 279 patients with ankylosing spondylitis. *Ann Rheum Dis* 2005;**64**:317.

106. 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). *Ann Rheum Dis* 2005;**64**:318–19.

107. Braun J, Maksymowych WP, Dougados M, Steinfeld S, Dewoody K, Williamson P, *et al.* Efficacy of infliximab in subgroups of patients with ankylosing spondylitis: results from the ankylosing spondylitis study for the evaluation of recombinant infliximab therapy (ASSERT). *Ann Rheum Dis* 2004;**63**:402.

108. Braun J, Kellner H, Deodhar A, Inman R, Han C, DeWoody K, *et al.* Infliximab improves quality of life in patients with ankylosing spondylitis: results of a randomized, placebo-controlled trial (ASSERT). *Ann Rheum Dis* 2004;**63**:402.

109. van der Heijde D, Dijkmans B, Geusens P, Sieper J, DeWoody K, Williamson R, *et al.* Efficacy and safety of infliximab in patients with ankylosing spondylitis: results of a 24-week randomized, placebo-controlled trial (ASSERT). *Arthritis Rheum* 2005;**52**:582–91.

110. Park W, Hrycaj P, Jeka S, Kovalenko V, Lysenko G, Miranda P, *et al.* A randomised, double-blind, multicentre, parallel-group, prospective study comparing the pharmacokinetics, safety, and efficacy of CT-P13 and innovator infliximab in patients with ankylosing spondylitis: the PLANETAS study. *Ann Rheum Dis* 2013;**72**:1605–12. http://dx.doi.org/10.1136/annrheumdis-2012-203091

111. Park W, Hrycaj P, Kovalenko V, Miranda P, Gutierrez-Urena S, Lee Y, *et al.* A randomized, double-blind, phase 1 study demonstrates equivalence in pharmacokinetics, safety, and efficacy of CT-P13 and infliximab in patients with ankylosing spondylitis. *Ann Rheum Dis* 2013;**71**:111.

112. Glintborg B, Ostergaard M, Krogh NS, Dreyer L, Kristensen HL, Hetland ML. Predictors of treatment response and drug continuation in 842 patients with ankylosing spondylitis treated with anti-tumor necrosis factor: results from 8 years' surveillance in the Danish nationwide DANBIO registry. *Ann Rheum Dis* 2010;**69**:2002–8. http://dx.doi.org/10.1136/ard.2009.124446

113. 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;**32**:80–5.

114. US Food and Drug Administration (FDA). *Summary Minutes of the Arthritis Advisory Committee Meeting July 23, 2013*. Silver Spring, MD: FDA; 2013.

115. US Food and Drug Administration (FDA). *FDA Briefing Package: sBLA 125160/215: Cimzia® (Certolizumab) for the Treatment of Active Axial Spondyloarthritis, Including Patients with Ankylosing Spondylitis*. Silver Spring, MD: FDA; 2013.

116. US Food and Drug Administration (FDA). *FDA Briefing Package: sBLA 125057/323: Adalimumab for the Treatment of Active Non-Radiographic Axial Spondyloarthritis in Adults with Objective Signs of Inflammation by Elevated C-reactive Protein (CRP) or Magnetic Resonance Imaging (MRI), Who Have Had an Inadequate Response to, or are Intolerant to, a Nonsteroidal Anti-Inflammatory Drug*. Silver Spring, MD: FDA; 2013.

117. van der Heijde D, Salonen D, Weissman BN, Landewe R, Maksymowych WP, Kupper H, *et al.* Assessment of radiographic progression in the spines of patients with ankylosing spondylitis treated with adalimumab for up to 2 years. *Arthritis Res Ther* 2009;**11**:R127. http://dx.doi.org/10.1186/ar2794

118. Boonen A, van der Heijde D, Landewe R, Guillemin F, Rutten-van Molken M, Dougados M, *et al.* Direct costs of ankylosing spondylitis and its determinants: an analysis among three European countries. *Ann Rheum Dis* 2003;**62**:732–40. http://dx.doi.org/10.1136/ard.62.8.732

119. van der Heijde D, Landewe R, Einstein S, Ory P, Vosse D, Ni L, *et al.* Radiographic progression of ankylosing spondylitis after up to two years of treatment with etanercept. *Arthritis Rheum* 2008;**58**:1324–31. http://dx.doi.org/10.1002/art.23471

120. Braun J, Baraliakos X, Hermann KG, Deodhar A, van der Heijde D, Inman R, *et al.* The effect of two golimumab doses on radiographic progression in ankylosing spondylitis: results through 4 years of the GO-RAISE trial. *Ann Rheum Dis* 2014;**73**:1107–13. http://dx.doi.org/10.1136/annrheumdis-2012-203075

121. Haroon N, Inman RD, Learch TJ, Weisman MH, Lee M, Rahbar MH, *et al.* The impact of tumor necrosis factor alpha inhibitors on radiographic progression in ankylosing spondylitis. *Arthritis Rheum* 2013;**65**:2645–54. http://dx.doi.org/10.1002/art.38070

122. Baraliakos X, Haibel H, Listing J, Sieper J, Braun J. Continuous long-term anti-TNF therapy does not lead to an increase in the rate of new bone formation over 8 years in patients with ankylosing spondylitis. *Ann Rheum Dis* 2014;**73**:710–15. http://dx.doi.org/10.1136/annrheumdis-2012-202698

123. Baraliakos X, Listing J, Brandt J, Haibel H, Rudwaleit M, Sieper J, *et al.* Radiographic progression in patients with ankylosing spondylitis after 4 yrs of treatment with the anti-TNF-alpha antibody infliximab. *Rheumatology* 2007;**46**:1450–3. http://dx.doi.org/10.1093/rheumatology/kem166

124. Ramiro S, van der Heijde D, van Tubergen A, Stolwijk C, Dougados M, van den Bosch F, *et al.* Higher disease activity leads to more structural damage in the spine in ankylosing spondylitis: 12-year longitudinal data from the OASIS cohort. *Ann Rheum Dis* 2014;**73**:1455–61. http://dx.doi.org/10.1136/annrheumdis-2014-205178

125. Ramiro S, Stolwijk C, van Tubergen A, van der Heijde D, Dougados M, van den Bosch F, *et al.* Evolution of radiographic damage in ankylosing spondylitis over 12 years of follow-up. *Ann Rheum Dis* 2013;**72**:665–6. http://dx.doi.org/10.1136/annrheumdis-2013-eular.1973

126. Bonafede M, Fox KM, Watson C, Princic N, Gandra SR. Treatment patterns in the first year after initiating tumor necrosis factor blockers in real-world settings. *Adv Ther* 2012;**29**:664–74. http://dx.doi.org/10.1007/s12325-012-0037-5

127. Choquette D, Sauvageau D, Raynauld JP. Retention rate of adalimumab, etanercept and infliximab at 5 years in patients with ankylosing spondylitis: report from the rhumadata computerized database. *Ann Rheum Dis* 2013;**71**:647.

128. Gulfe A, Kapetanovic ML, Kristensen E. Efficacy and drug survival of anti-tumor necrosis factor-alpha therapies in patients with non-radiographic axial spondyloarthritis: an observational cohort study from Southern Sweden. *Scand J Rheumatol* 2014;**43**:493–7. http://dx.doi.org/10.3109/03009742.2014.918173

129. Nell-Duxneuner V, Schroeder Y, Reichardt B, Bucsics A. The use of TNF-inhibitors in ankylosing spondylitis in Austria from 2007 to 2009 – a retrospective analysis. *Int J Clin Pharmacol Ther* 2012;**50**:867–72. http://dx.doi.org/10.5414/CP201707

130. Yeaw J, Watson C, Fox KM, Schabert VF, Goodman S, Gandra SR. Treatment patterns following discontinuation of adalimumab, etanercept, and infliximab in a US managed care sample. *Adv Ther* 2014;**31**:410–25. http://dx.doi.org/10.1007/s12325-014-0110-3

131. Scire CA, Caporali R, Sarzi-Puttini P, Frediani B, Di Franco M, Tincani A, *et al.* Drug survival of the first course of anti-TNF agents in patients with rheumatoid arthritis and seronegative spondyloarthritis: analysis from the MonitorNet database. *Clin Exp Rheumatol* 2013;**31**:857–63.

132. Zufferey P, Ghosn J, Becce F, Ciurea A, Aubry-Rozier B, Finckh A, *et al.* Anti-tumor necrosis factor drug survival in axial spondyloarthritis is independent of the classification criteria. *Rheumatol Int* 2014;**35**:295–302. http://dx.doi.org/10.1007/s00296-014-3094-z

133. Pavelka K, Forejtova S, Stolfa J, Chroust K, Buresova L, Mann H, *et al.* Anti-TNF therapy of ankylosing spondylitis in clinical practice. Results from the Czech national registry ATTRA. *Clin Exp Rheumatol* 2009;**27**:958–63.

134. Lie E, van der Heijde D, Uhlig T, Mikkelsen K, Rødevand E, Koldingsnes W, *et al.* Effectiveness of switching between TNF inhibitors in ankylosing spondylitis: data from the NOR-DMARD register. *Ann Rheum Dis* 2011;**70**:157–63. http://dx.doi.org/10.1136/ard.2010.131797

135. Glintborg B, Ostergaard M, Krogh NS, Tarp U, Manilo N, Loft AGR, *et al.* Clinical response, drug survival and predictors thereof in 432 ankylosing spondylitis patients after switching tumour necrosis factor alpha inhibitor therapy: results from the Danish nationwide DANBIO registry. *Ann Rheum Dis* 2013;**72**:1149–55. http://dx.doi.org/10.1136/annrheumdis-2012-201933

136. Rudwaleit M, Claudepierre P, Wordsworth P, Cortina EL, Sieper J, Kron M, *et al.* Effectiveness, safety, and predictors of good clinical response in 1250 patients treated with adalimumab for active ankylosing spondylitis. *J Rheumatol* 2009;**36**:801–8. http://dx.doi.org/10.3899/jrheum.081048

137. Singh JA, Wells GA, Christensen R, Ghogomu TE, Maxwell L, MacDonald JK, *et al.* Adverse effects of biologics: a network meta-analysis and Cochrane overview. *Cochrane Database Syst Rev* 2011;**2**:CD008794. http://dx.doi.org/10.1002/14651858.cd008794.pub2

138. Askling J, Fahrbach K, Nordstrom B, Ross S, Schmid CH, Symmons D. Cancer risk with tumor necrosis factor alpha (TNF) inhibitors: meta-analysis of randomized controlled trials of adalimumab, etanercept, and infliximab using patient level data. *Pharmacoepidemiol Drug Saf* 2011;**20**:119–30. http://dx.doi.org/10.1002/pds.2046

139. Rodgers M, Epstein D, Bojke L, Yang H, Craig D, Fonseca T, *et al.* Etanercept, infliximab and adalimumab for the treatment of psoriatic arthritis: a systematic review and economic evaluation. *Health Technol Assess* 2011;**15**(10). http://dx.doi.org/10.3310/hta15100

140. 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;**68**:922–9. http://dx.doi.org/10.1136/ard.2007.087270

141. Sieper J, van der Heijde D, Dougados M, Brown LS, Lavie F, Pangan AL. Early response to adalimumab predicts long-term remission through 5 years of treatment in patients with ankylosing spondylitis. *Ann Rheum Dis* 2012;**71**:700–6. http://dx.doi.org/10.1136/annrheumdis-2011-200358

142. Braun J, Deodhar A, Inman RD, van der Heijde D, Mack M, Xu S, *et al.* Golimumab administered subcutaneously every 4 weeks in ankylosing spondylitis: 104-week results of the GO-RAISE study. *Ann Rheum Dis* 2012;**71**:661–7. http://dx.doi.org/10.1136/ard.2011.154799

143. Deodhar A, Braun J, Inman RD, van der Heijde D, Zhou Y, Xu S, *et al.* Golimumab administered subcutaneously every 4 weeks in ankylosing spondylitis: 5-year results of the GO-RAISE study. *Ann Rheum Dis* 2015;**74**:757–61. http://dx.doi.org/10.1136/annrheumdis-2014-205862

144. Davis JC Jr, van der Heijde DM, Braun J, Dougados M, Clegg DO, Kivitz AJ, *et al.* Efficacy and safety of up to 192 weeks of etanercept therapy in patients with ankylosing spondylitis. *Ann Rheum Dis* 2008;**67**:346–52. http://dx.doi.org/10.1136/ard.2007.078139

145. Davis JC, van der Heijde DM, Braun J, Dougados M, Cush J, Clegg D, *et al.* Sustained durability and tolerability of etanercept in ankylosing spondylitis for 96 weeks. *Ann Rheum Dis* 2005;**64**:1557–62. http://dx.doi.org/10.1136/ard.2004.035105

146. Dijkmans B, Emery P, Hakala M, Leirisalo-Repo M, Mola EM, Paolozzi L, *et al.* Etanercept in the longterm treatment of patients with ankylosing spondylitis. *J Rheumatol* 2009;**36**:1256–64. [Erratum published in *J Rheumatol* 2010;**37**:2198.] http://dx.doi.org/10.3899/jrheum.081033

147. Martin-Mola E, Sieper J, Leirisalo-Repo M, Dijkmans BA, Vlahos B, Pedersen R, *et al.* Sustained efficacy and safety, including patient-reported outcomes, with etanercept treatment over 5 years in patients with ankylosing spondylitis. *Clin Exp Rheumatol* 2010;**28**:238–45.

148. Baraliakos X, Listing J, Fritz C, Haibel H, Alten R, Burmester GR, *et al.* Persistent clinical efficacy and safety of infliximab in ankylosing spondylitis after 8 years – early clinical response predicts long-term outcome. *Rheumatology* 2011;**50**:1690–9. http://dx.doi.org/10.1093/rheumatology/ker194

149. Dean LE, Macfarlane GJ, MacDonald AG. *The Natural History of Ankylosing Spondylitis – Results from the Scotland and Ireland Registry for Ankylosing Spondylitis (SIRAS).* British Society for Rheumatology Annual Scientific Meeting. Liverpool: University of Aberdeen; 2013.

150. Stone MA, Sengupta R, Gordon D, Pomeroy E, Mogg R, Keat A. Longitudinal analyses of disease outcomes in ankylosing spondylitis yield insight into the natural history of AS. *Ann Rheum Dis* 2007;**66**:410.

151. Machado P, Landewe R, Braun J, Hermann KG, Baker D, van der Heijde D. Both structural damage and inflammation of the spine contribute to impairment of spinal mobility in patients with ankylosing spondylitis. *Ann Rheum Dis* 2010;**69**:1465–70. http://dx.doi.org/10.1136/ard.2009.124206

152. Kobelt G, Andlin-Sobocki P, Brophy S, Jonsson L, Calin A, Braun J. The burden of ankylosing spondylitis and the cost-effectiveness of treatment with infliximab (Remicade®). *Rheumatology* 2004;**43**:1158–66. http://dx.doi.org/10.1093/rheumatology/keh271

153. Kiltz U, Baraliakos X, Karakostas P, Igelmann M, Kalthoff L, Klink C, *et al.* Do patients with non-radiographic axial spondylarthritis differ from patients with ankylosing spondylitis? *Arthritis Care Res* 2012;**64**:1415–22. http://dx.doi.org/10.1002/acr.21688

154. Poddubnyy D, Rudwaleit M, Haibel H, Listing J, Maerker-Hermann E, Zeidler H, *et al.* Rates and predictors of radiographic sacroiliitis progression over 2 years in patients with axial spondyloarthritis. *Ann Rheum Dis* 2011;**70**:1369–74. http://dx.doi.org/10.1136/ard.2010.145995

155. Poddubnyy D, Haibel H, Listing J, Marker-Hermann E, Zeidler H, Braun J, *et al.* Baseline radiographic damage, elevated acute-phase reactant levels, and cigarette smoking status predict spinal radiographic progression in early axial spondylarthritis. *Arthritis Rheum* 2012;**64**:1388–98. http://dx.doi.org/10.1002/art.33465

156. 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* 2008;**47**:1213–18. http://dx.doi.org/10.1093/rheumatology/ken176

157. Sampaio-Barros PD, Bortoluzzo AB, Conde RA, Costallat LTL, Samara AM, Bertolo MB. Undifferentiated spondyloarthritis: a longterm followup. *J Rheumatol* 2010;**37**:1195–9. http://dx.doi.org/10.3899/jrheum.090625

158. Sampaio-Barros PD, Bertolo MB, Kraemer MHS, Marques-Neto JF, Samara AM. Undifferentiated spondyloarthropathies: a 2-year follow-up study. *Clin Rheumatol* 2001;**20**:201–6. http://dx.doi.org/10.1007/s100670170066

159. Boonen A, Sieper J, van der Heijde D, Dougados M, Bukowski JF, Valluri S, *et al.* The burden of non-radiographic axial spondyloarthritis. *Semin Arthritis Rheum* 2015;**44**:556–62. http://dx.doi.org/10.1016/j.semarthrit.2014.10.009

160. Kobelt G, Sobocki P, Sieper J, Braun J. Comparison of the cost-effectiveness of infliximab in the treatment of ankylosing spondylitis in the United Kingdom based on two different clinical trials. *Int J Technol Assess Health Care* 2007;**23**:368–75. http://dx.doi.org/10.1017/S0266462307070432

161. Ara RM, Reynolds AV, Conway P. The cost-effectiveness of etanercept in patients with severe ankylosing spondylitis in the UK. *Rheumatology* 2007;**46**:1338–44. http://dx.doi.org/10.1093/rheumatology/kem133

162. Botteman MF, Hay JW, Luo MP, Curry AS, Wong RL, Hout BA. Cost effectiveness of adalimumab for the treatment of ankylosing spondylitis in the United Kingdom. *Rheumatology* 2007;**46**:1320–8. http://dx.doi.org/10.1093/rheumatology/kem031

163. Armstrong N, Joore M, van Asselt T, Misso K, Manning N, Tomini F, *et al.* Golimumab for the treatment of ankylosing spondylitis: a NICE single technology appraisal. *Pharmacoeconomics* 2013;**31**:415–25. http://dx.doi.org/10.1007/s40273-013-0049-2

164. ClinicalTrials.gov. *Study Comparing Etanercept (ETN) Against a Placebo for Etanercept on a Background Nonsteroidal Anti Inflammatory Drug (NSAIDs) in the Treatment of Early Spondyloarthritis (SpA) Patients who do not Have X-ray Structural Changes (AS EARLY).* ClinicalTrials.gov Identifier: NCT01258738.

165. ClinicalTrials.gov. *Study Comparing Etanercept 50 mg Once Weekly to 25 mg Twice Weekly in Patients With Ankylosing Spondylitis.* ClinicalTrials.gov Identifier: NCT00418548.

166. Braun J, Brandt J, Listing J, Zink A, Alten R, Burmester G, *et al.* Two year maintenance of efficacy and safety of infliximab in the treatment of ankylosing spondylitis. *Ann Rheum Dis* 2005;**64**:229–34. http://dx.doi.org/10.1136/ard.2004.025130

167. Ara RM, Packham J, Haywood K. The direct healthcare resource costs associated with ankylosing spondylitis patients attending a UK secondary care rheumatology unit. *Ann Rheum Dis* 2006;**65**:273.

168. Rafia R, Ara R, Packham J, Haywood K, Healey E. Healthcare costs and productivity losses directly attributable to ankylosing spondylitis. *Clin Exp Rheumatol* 2012;**30**:246–53.

169. Keat A, Barkham N, Bhalla A, Gaffney K, Marzo-Ortega H, Paul S, *et al.* BSR guidelines for prescribing TNF-alpha blockers in adults with ankylosing spondylitis. Report of a working party of the British Society for Rheumatology. *Rheumatology* 2005;**44**:939–47. http://dx.doi.org/10.1093/rheumatology/keh669

170. Spiegelhalter DJ, Best NG, Carlin BR, van der Linde A. Bayesian measures of model complexity and fit. *J R Stat Soc Ser B* 2002;**64**:583–616. http://dx.doi.org/10.1111/1467-9868.00353

171. Welton N, Ades AE. Research decisions in the face of heterogeneity: what can a new study tell us? *Health Econ* 2012;**21**:1196–200. http://dx.doi.org/10.1002/hec.1797

172. Vastesaeger N, van der Heijde D, Inman RD, Wang YX, Deodhar A, Hsu B, *et al.* Predicting the outcome of ankylosing spondylitis therapy. *Ann Rheum Dis* 2011;**70**:973–81. http://dx.doi.org/10.1136/ard.2010.147744

173. Lord PA, Farragher TM, Lunt M, Watson KD, Symmons DP, Hyrich KL, *et al.* Predictors of response to anti-TNF therapy in ankylosing spondylitis: results from the British Society for Rheumatology Biologics Register. *Rheumatology* 2010;**49**:563–70. http://dx.doi.org/10.1093/rheumatology/kep422

174. Dias S, Sutton AJ, Welton NJ, Ades AE. Evidence synthesis for decision making 3: heterogeneity – subgroups, meta-regression, bias, and bias-adjustment. *Med Decis Making* 2013;**33**:618–40. http://dx.doi.org/10.1177/0272989X13485157

175. National Institute for Health and Care Excellence. *Technology Appraisal TA199: Etanercept, Infliximab and Adalimumab for the Treatment of Psoriatic Arthritis*. London: NICE; 2010.

176. Department of Health. *NHS Reference Costs 2012/13*. London: Department of Health; 2013.

177. Poddubnyy D, Haibel H, Braun J, Rudwaleit M, Sieper J. Patients with non-radiographic axial spondyloarthritis and ankylosing spondylitis demonstrate the same clinical disease course over two years: results from the GESPIC cohort. *Ann Rheum Dis* 2014;**73**(Suppl. 2):201. http://dx.doi.org/10.1136/annrheumdis-2014-eular.3599

178. Joint Formulary Committee. *British National Formulary (Online)*. London: BMJ Group and Pharmaceutical Press. URL: www.medicinescomplete.com (accessed 8 December 2014).

179. Curtis L. *Unit Costs of Health and Social Care 2013*. Canterbury: PSSRU, University of Kent; 2013.

180. Henriksson M, Palmer S, Chen R, Damant J, Fitzpatrick NK, Abrams K, *et al.* Assessing the cost effectiveness of using prognostic biomarkers with decision models: case study in prioritising patients waiting for coronary artery surgery. *BMJ* 2010;**340**:b5606. http://dx.doi.org/10.1136/bmj.b5606

181. Rees F, Peffers G, Bell C, Obrenovic K, Sandhu R, Packham J, *et al.* Compliance with NICE guidance on the use of anti-TNF alpha agents in ankylosing spondylitis: an east and west Midlands regional audit. *Clin Med* 2012;**12**:324–7. http://dx.doi.org/10.7861/clinmedicine.12-4-324

182. University Hospital Montpellier. *Effect of Spacing of Anti-TNF Drugs in Ankylosing Spondylitis with Low Disease Activity (SPACING)*. ClinicalTrials.gov; 2012. URL: http://ClinicalTrials.gov/show/NCT01610947 (accessed 11 December 2014).

183. Higgins JPT, Green S. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.1.0 (updated March 2011). The Cochrane Collaboration; 2011.

184. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002;**21**:1539–58. http://dx.doi.org/10.1002/sim.1186

185. Haibel H, Rios V, Hermann KG, Althoff C, Rudwaleit M, Kupper H, *et al.* No progression of chronic sacroiliac changes in patients with active, non-radiographic axial spondyloarthritis treated with adalimumab over 52 weeks. *Arthritis Rheum* 2009;**60**:1798.

186. Haibel H, Rudwaleit M, Amtenbrink A, Heldmann F, Listing J, Wong R, *et al.* Adalimumab is effective in patients with preradiographic axial spondyloarthritis (PRSPA): results of a 12-week, randomised controlled trial with a 40-week open-label extension (OLE). *Rheumatology* 2008;**47**:il80.

187. Haibel H, Rudwaleit M, Amtenbrink A, Heldmann F, Listing J, Kupper H, *et al.* Efficacy of adalimumab in the treatment of preradiographic axial spondyloarthritis: results of a 20-week, randomized controlled trial with an open-label extension phase. *Z Rheumatol* 2007;**66**:14.

188. Haibel H, Rudwaleit M, Amtenbrink A, Heldmann F, Listing J, Wong R, *et al.* Efficacy of adalimumab in the treatment of preiradiographic axial spondyloarthritis: results of a 20-week, randomized controlled trial with an open-label extension phase. *Ann Rheum Dis* 2007;**66**:64. http://dx.doi.org/10.1002/art.23606

189. van der Heijde D, Maksymowych W, Sieper J, Lambert R, Brown MA, Rathmann S, *et al.* Relationship between MRI and clinical remission in patients with non-radiographic axial spondyloarthritis after two years of adalimumab therapy. *Ann Rheum Dis* 2013;**72**:A287. http://dx.doi.org/10.1136/annrheumdis-2013-eular.889

190. Sieper J, Baeten DL, Van Den Bosch F, Rathmann SS, Anderson J, Pangan AL. Sustained clinical remission in patients with non-radiographic axial spondyloarthritis after two years of adalimumab treatment. *Ann Rheum Dis* 2013;**72**:i140. http://dx.doi.org/10.1136/annrheumdis-2013-eular.314

191. Sieper J, Van Der Heijde D, Dougados M, Van Den Bosch F, Goupille P, Sarkar S, *et al.* Sustained efficacy of adalimumab in patients with non-radiographic axial spondyloarthritis: week 68 results from ability 1. *Ann Rheum Dis* 2013;**71**(Suppl. 3):248.

192. van der Heijde DM, Mease PJ, Pangan AL, Rao SA, Chen N, Cifaldi M. Improvement in physical function, health-related quality of life, and work productivity with adalimumab treatment in non-radiographic axial spondyloarthritis. *Arthritis Rheum* 2013;**65**:S1052.

193. Sieper J, Van Der Heijde D, Dougados M, Van Den Bosch F, Goupille P, Rathmann SS, *et al.* Sustained efficacy of adalimumab in patients with non-radiographic axial spondyloarthritis with positive MRI of the sacroiliac joints or spine or elevated C-reactive protein at baseline. *Rheumatology* 2013;**52**:i52–3.

194. van der Heijde D, Mease P, Pangan AL, Rao S, Chen N, Cifaldi MA. Improvement in physical function, health-related quality of life, and work productivity with adalimumab treatment in nonradiographic axial SPA: WK-52 results from ability-1. *Arthritis Rheum* 2012;**64**:S583.

195. Kimel M, Revicki D, Rao S, Fryback D, Feeny D, Harnam N, *et al.* Norms-based assessment of patient-reported outcomes associated with adalimumab monotherapy in patients with ankylosing spondylitis. *Clin Exp Rheumatol* 2011;**29**:624–32.

196. van der Heijde DM, Revicki DA, Gooch KL, Wong RL, Kupper H, Harnam N, *et al.* Physical function, disease activity, and health-related quality-of-life outcomes after 3 years of adalimumab treatment in patients with ankylosing spondylitis. *Arthritis Res Ther* 2009;**11**:R124. http://dx.doi.org/10.1186/ar2790

197. Revicki DA, Luo MP, Wordsworth P, Wong RL, Chen N, Davis JC Jr, *et al.* 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). *J Rheumatol* 2008;**35**:1346–53. [Erratum published in *J Rheumatol* 2011;**38**:788.]

198. van der Heijde D, Breban MA, Halter DG, DiVittorio G, Bratt J, Cantini F, *et al.* Sustained improvement of spinal mobility, physical function, and quality of life in patients with ankylosing spondylitis: 5-year results. *Arthritis Rheum* 2011;**63**(Suppl. 10):535. http://dx.doi.org/10.1093/rheumatology/keu438

199. Maksymowych W, Van Der Heijde D, Landewe R, Pangan A, Brown S, Lavie F. Predictors of radiographic progression in adalimumab-treated patients with Ankylosing Spondylitis. *Arthritis Rheum* 2010;**62**:1949. http://dx.doi.org/10.1186/ar2794

200. van der Heijde D, Emery P, Patra K, Lavie F. Relationship between bone formation in patients with ankylosing spondylitis and clinical signs and symptoms. *Rheumatology* 2010;**49**:i18.

201. Scott DG, Van Der Heijde D, Schiff MH, Sieper J, Kivitz A, Dijkmans BA, *et al.* Improvement in long-term spinal mobility in patients with ankylosing spondylitis (AS) is sustained during up to 3 years of adalimumab treatment: 3-year atlas results. *Rheumatology* 2009;**48**:i56. http://dx.doi.org/10.1136/ard.2007.087270

202. Landewe RBM, Rudwaleit M, Van Der Heijde DM, Dougados M, Mease PJ, Reveille JD, *et al.* Effect of certolizumab pegol over 48 weeks in patients with axial spondyloarthritis, including ankylosing spondylitis and non-radiographic axial spondyloarthritis. *Arthritis Rheum* 2013;**65**:S767. http://dx.doi.org/10.1136/annrheumdis-2013-204231

203. Sieper J, Kivitz A, Van Tubergen A, Deodhar A, Coteur G, Singh P, *et al.* Long-term maintenance of improvements in patient-reported outcomes with certolizumab pegol in patients with axial spondyloarthritis, including ankylosing spondylitis and non-radiographic axial spondyloarthritis: 48-week results of the rapid-axspa study. *Value Health* 2013;**16**:A569. http://dx.doi.org/10.1016/j.jval.2013.08.1523

204. Maksymowych W, Landewe R, Rudwaleit M, van der Heijde D, Dougados M, Mease P, *et al.* Effect of certolizumab pegol over 48 weeks in patients with axial spondyloarthritis, including ankylosing spondylitis and non-radiographic axial spondyloarthritis. *J Rheumatol* 2014;**41**:1496–7.

205. Boonen A, Patel V, Traina S, Chiou CF, Maetzel A, Tsuji W. Rapid and sustained improvement in health-related quality of life and utility for 72 weeks in patients with ankylosing spondylitis receiving etanercept. *J Rheumatol* 2008;**35**:662–7.

206. Davis JC, van der Heijde D, Dougados M, Woolley JM. Reductions in health-related quality of life in patients with ankylosing spondylitis and improvements with etanercept therapy. *Arthritis Rheum* 2005;**53**:494–501. http://dx.doi.org/10.1002/art.21330

207. Dougados M, Van Der Heijde D, Sieper J, Braun J, Maksymowych WP, Citera G, *et al.* Safety and efficacy of etanercept in early non-radiographic axial spondyloarthritis: a randomized, double-blind, placebo-controlled trial at 24 weeks. *Arthritis Rheum* 2013;**65**:S642–3.

208. Davis J Jr, Webb A, Lund S, Sack K. Results from an open-label extension study of etanercept in ankylosing spondylitis. *Arthritis Rheum* 2004;**51**:302–4. http://dx.doi.org/10.1002/art.20241

209. Bao C, Huang F, Khan MA, Fei K, Wu Z, Zhuang Y, *et al.* Safety and efficacy of golimumab, a human anti-TNF monoclonal antibody injected subcutaneously every 4 weeks, in Chinese patients with active ankylosing spondylitis: one-year results of a phase 3, randomized, placebo-controlled study. *Ann Rheum Dis* 2013;**72**(Suppl. 3):945. http://dx.doi.org/10.1136/annrheumdis-2013-eular.2835

210. Bao C, Huang F, Khan A, Fei K, Wu Z, Zhuang Y, *et al.* Safety and efficacy of golimumab, a human anti-tumor necrosis factor monoclonal antibody injected subcutaneously every 4 weeks, in Chinese patients with active ankylosing spondylitis: 1-year results of a phase 3, randomized, placebo-controlled study. *Arthritis Rheum* 2013;**65**:S651.

211. van der Heijde D, Braun J, Deodhar A, Inman RD, Xu S, Mack ME, *et al.* Comparison of three enthesitis indices in a multicentre, randomized, placebo-controlled trial of golimumab in ankylosing spondylitis (GO-RAISE). *Rheumatology* 2013;**52**:321–5. http://dx.doi.org/10.1093/rheumatology/kes251

212. Deodhar A, Braun J, Inman R, Van Der Heijde D, Zhou Y, Hsu B. Long-term safety and efficacy of golimumab in the treatment of ankylosing spondylitis: results through 5 years of the go-raise trial. *Ann Rheum Dis* 2013;**72**:283. http://dx.doi.org/10.1136/annrheumdis-2013-eular.880

213. Braun J, Van Der Heijde D, Hermann KG, Baraliakos X, Deodhar A, Beutler A, *et al.* The effect of anti-tumor necrosis factor therapy with golimumab on radiographic progression in definite ankylosing spondylitis: 4-year results. *Ann Rheum Dis* 2013;**71**(Suppl. 3):249. http://dx.doi.org/10.1136/annrheumdis-2012-eular.2243

214. van der Heijde D, Deodhar A, Braun J, Mack M, Hsu B, Gathany T, *et al.* Achieving ASDAS-CRP major improvement and inactive disease in patients with ankylosing spondylitis after treatment with golimumab is associated with normalized health related quality of life: two-year results from the go-raise trial. *Ann Rheum Dis* 2013;**71**(Suppl. 3):112. http://dx.doi.org/10.1136/annrheumdis-2012-eular.1853

215. van der Heijde D, Deodhar A, Braun J, Mack M, Hsu B, Gathany T, *et al.* Achieving ankylosing spondylitis disease activity score c-reactive protein major improvement and inactive disease in patients with ankylosing spondylitis after treatment with golimumab is associated with normalized health-related quality of life: 2-year results from go-raise. *Rheumatology* 2013;**52**:i161–2.

216. Braun J, Beutler A, Deodhar A, Hsu B, Van Der Heijde D, Hermann KG, *et al.* The effect of anti-tumor necrosis factor therapy with golimumab on radiographic progression in definite ankylosing spondylitis: 4-year results. *Int J Rheum Dis* 2012;**15**:91. http://dx.doi.org/10.1136/annrheumdis-2012-eular.2243

217. Braun J, Van Der Heijde D, Hermann KG, Baraliakos X, Deodhar A, Beutler A, *et al.* The effect of anti-tumor necrosis factor therapy with two different doses of golimumab on radiographic progression in definite ankylosing spondylitis: 4-year results. *Arthritis Rheum* 2011;**1**:423. http://dx.doi.org/10.1136/annrheumdis-2012-203075

218. Braun J, Van Der Heijde D, Deodhar A, Beutler A, Mack M, Hsu B, *et al.* Sustained clinical response with golimumab administered subcutaneously every 4 weeks in ankylosing spondylitis: 104-week results of a randomized, placebo-controlled study. *Arthritis Rheum* 2011;**1**:1227.

219. Braun J, Van Der Heijde D, Deodhar A, Diekman L, Sieper J, Kim SI, *et al.* Golimumab, a new, human, TNF-alpha antibody administered subcutaneously every 4 weeks, in ankylosing spondylitis: 104-week efficacy and safety results of the randomized, placebo-controlled go-raise study. *Scand J Rheumatol* 2010;**39**:24.

220. Kim SI, Braun J, Van Der Heijde D, Deodhar A, Dieckman L, Sieper J, *et al.* Golimumab, a new, human, TNF-alpha antibody administered subcutaneously every 4 weeks, in ankylosing spondylitis: 104 week efficacy and safety results of the randomized, placebo-controlled GO-RAISE study. *Int J Rheum Dis* 2010;**13**:152–3.

221. Braun J, Van Der Heijde DMFM, Deodhar AA, Diekman L, Sieper J, Kim SI, *et al.* Golimumab, a new, human, TNF-antibody administered subcutaneously every 4 weeks, in ankylosing spondylitis (AS): 104-week efficacy and safety results of the randomized, placebo-controlled GO-RAISE study. *Arthritis Rheum* 2009;**60**:1259.

222. Braun J, van der Heijde D, Deodhar A, Diekman L, Sieper J, Kim SI, *et al.* Golimumab, a new, human, TNF-alpha antibody administered subcutaneously every 4 weeks, in ankylosing spondylitis: 104-week efficacy and safety results of the randomized, placebo controlled GO-RAISE study. *Clin Exp Rheumatol* 2010;**28**:618.

223. Park W, Miranda P, Brzosko M, Wiland P, Gutierrez-Urena S, Mikazane H, *et al.* Efficacy and safety of CT-P13 (infliximab biosimilar) over two years in patients with ankylosing spondylitis: comparison between continuing with CT-P13 and switching from infliximab to CT-P13. *Arthritis Rheum* 2013;**65**:3326.

224. Braun J, Baraliakos X, Listing J, Fritz C, Alten R, Burmester G, *et al.* Persistent clinical efficacy and safety of anti-tumour necrosis factor alpha therapy with infliximab in patients with ankylosing spondylitis over 5 years: evidence for different types of response. *Ann Rheum Dis* 2008;**67**:340–5. http://dx.doi.org/10.1136/ard.2007.075879

225. Braun J, Baraliakos X, Brandt J, Listing J, Zink A, Alten R, *et al.* Persistent clinical response to the anti-TNF-alpha antibody infliximab in patients with ankylosing spondylitis over 3 years. *Rheumatology* 2005;**44**:670–6. http://dx.doi.org/10.1093/rheumatology/keh584

226. Braun J, Brandt J, Listing J, Zink A, Alten R, Burmester G, *et al.* Long-term efficacy and safety of infliximab in the treatment of ankylosing spondylitis: an open, observational, extension study of a three-month, randomized, placebo-controlled trial. *Arthritis Rheum* 2003;**48**:2224–33. http://dx.doi.org/10.1002/art.11104

227. Baraliakos X, Brandt J, Listing J, Rudwaleit M, Sieper J, Braun J. Clinical response to long-term therapy with infliximab in patients with ankylosing spondylitis – results after 3 years. *Ann Rheum Dis* 2004;**63**:412.

228. Baraliakos X, Brandt J, Listing J, Rudwaleit M, Sieper J, Braun J. Clinical response to withdrawal of anti-TNF therapy in patients with ankylosing spondylitis (AS) after 3 years of continuous treatment with infliximab. *Ann Rheum Dis* 2004;**63**:413.

229. Brandt J, Listing J, Alten R, Bumester G, Gromnica-Ihle E, Kellner H, *et al.* Two year follow up results of a controlled trial of the anti-TNF alpha antibody infliximab in active ankylosing spondylitis. *Arthritis Rheum* 2003;**48**:S172–3.

230. Brandt J, Listing J, Alten R, Burmester G, Gromnica-Ihle E, Schneider M, *et al.* One year results of a double-blind placebo controlled, phase-III clinical trial of infliximab in active ankylosing spondylitis. *Arthritis Rheum* 2002;**46**:S429.

231. Heldmann F, Brandt J, van der Horst-Bruinsma IE, Landewe R, Sieper J, Burmester GR, *et al.* The European ankylosing spondylitis infliximab cohort (EASIC): a European multicentre study of long term outcomes in patients with ankylosing spondylitis treated with infliximab. *Clin Exp Rheumatol* 2011;**29**:672–80.

232. 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;**59**:1270–8. http://dx.doi.org/10.1002/art.24001

233. Machado P, Landewe R, Braun J, Baraliakos X, Hermann KGA, Hsu B, *et al.* MRI inflammation and its relation with measures of clinical disease activity and different treatment responses in patients with ankylosing spondylitis treated with a TNF inhibitor. *Ann Rheum Dis* 2013;**71**.

234. Heldmann F, Brandt J, Braun J. European ankylosing spondylitis infliximab cohort (EASIC): long-term (5 years) efficacy of infliximab on disease activity and function: a real life experience after the end of ASSERT. *Arthritis Rheum* 2009;**60**:520.

235. Braun J, Han C, Kellner H, Deodhar A, Inman R, Xu W, *et al.* Infliximab improves spinal mobility in patients with ankylosing spondylitis: two-year findings from the assert clinical trial. *Rheumatology* 2008;**47**(Suppl. 1):i172.

236. Braun J, Han C, Kellner H, Deodhar A, Inman R, Xu W, *et al.* Infliximab improves spinal mobility in patients with ankylosing spondylitis: two-year findings from the assert clinical trial. *Ann Rheum Dis* 2007;**66**:390–1.

237. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. *Methods for the Economic Evaluation of Health Care Programmes*. 3rd edn. Oxford: Oxford University Press; 2005.

238. National Institute for Health and Care Excellence. *Guide to the Methods of Technology Appraisal*. London: NICE; 2013.

239. Pumford N, Wade A, Crawford G, McConnachie A. Sensitivity of PRO's to detect changes in quality of life in patients treated with a biologic agent. *Value Health* 2011;**14**:A314. http://dx.doi.org/10.1016/j.jval.2011.08.450

240. Wade AG, Crawford GM, Pumford N, Koscielny V, Maycock S, McConnachie A. Baseline characteristics and patient reported outcome data of patients prescribed etanercept: web-based and telephone evaluation. *BMC Med Res Methodol* 2011;**11**:91. http://dx.doi.org/10.1186/1471-2288-11-91

241. Lee TJ, Park BH, Son HK, Kim JW, Shin K, Lee EB, *et al.* Differences in cost-of-illness and quality of life between rheumatoid arthritis and ankylosing spondylitis in South Korea. *Value Health* 2011;**14**:A306. http://dx.doi.org/10.1016/j.jval.2011.08.407

242. Lee T-J, Park B-H, Kim JW, Shin K, Lee EB, Song Y-W. Cost-of-illness and quality of life in patients with ankylosing spondylitis at a tertiary hospital in Korea. *J Korean Med Sci* 2014;**29**:190–7. http://dx.doi.org/10.3346/jkms.2014.29.2.190

243. Joore M, Brunenberg D, Nelemans P, Wouters E, Kuijpers P, Honig A, *et al.* The impact of differences in EQ-5D and SF-6D utility scores on the acceptability of cost–utility ratios: results across five trial-based cost-utility studies. *Value Health* 2010;**13**:222–9. http://dx.doi.org/10.1111/j.1524-4733.2009.00669.x

244. Van Tubergen A, Boonen A, Landewe R, Rutten-Van Molken M, Van der Heijde D, Hidding A, *et al.* Cost effectiveness of combined spa-exercise therapy in ankylosing spondylitis: a randomized controlled trial. *Arthritis Rheum-Arthritis Care Res* 2002;**47**:459–67. http://dx.doi.org/10.1002/art.10658

245. Kobelt G, Andlin-Sobocki P, Collantes-Estevez E, Gratacos-Masmitja J, Pocovi A, Mulero-Mendoza J. The cost effectiveness of infliximab in the treatment of ankylosing spondylitis in Spain. Comparison of clinical trial and clinical practice data. *Ann Rheum Dis* 2006;**65**:282–3. http://dx.doi.org/10.1080/03009740701607224

246. Kobelt G, Sobocki P, Mulero J, Gratacos J, Collantes-Estevez E, Braun J. The cost-effectiveness of infliximab in the treatment of ankylosing spondylitis in Spain: comparison of clinical trial and clinical practice data. *Scand J Rheumatol* 2008;**37**:62–71. http://dx.doi.org/10.1080/03009740701607224

247. Boonen A, van der Heijde D, Landewe R, Spoorenberg A, Schouten H, Rutten-van Molken M, *et al.* Work status and productivity costs due to ankylosing spondylitis: comparison of three European countries. *Ann Rheum Dis* 2002;**61**:429–37. http://dx.doi.org/10.1136/ard.61.5.429

248. Boonen A, van der Heijde D, Landewe R, Guillemin F, Spoorenberg A, Schouten H, *et al.* Costs of ankylosing spondylitis in three European countries: the patient's perspective. *Ann Rheum Dis* 2003;**62**:741–7. http://dx.doi.org/10.1136/ard.62.8.741

249. Dolan P. Modeling valuations for EuroQol health states. *Med Care* 1997;**35**:1095–108. http://dx.doi.org/10.1097/00005650-199711000-00002

250. Boonen A, van der Heijde D, Landewe R, van Tubergen A, Mielants H, Dougados M, *et al.* How do the EQ-5D, SF-6D and the well-being rating scale compare in patients with ankylosing spondylitis? *Ann Rheum Dis* 2007;**66**:771–7. http://dx.doi.org/10.1136/ard.2006.060384

251. Gordeev VS, Maksymowych WP, Evers SM, Ament A, Schachna L, Boonen A. Role of contextual factors in health-related quality of life in ankylosing spondylitis. *Ann Rheum Dis* 2010;**69**:108–12. http://dx.doi.org/10.1136/ard.2008.100164

252. Haywood KL, Garratt AM, Dziedzic K, Dawes PT. Generic measures of health-related quality of life in ankylosing spondylitis: reliability, validity and responsiveness. *Rheumatology* 2002;**41**:1380–7. http://dx.doi.org/10.1093/rheumatology/41.12.1380

253. Kind P, Dolan P, Gudex C, Williams A. Variations in population health status: results from a United Kingdom national questionnaire survey. *BMJ* 1998;**316**:736–41. http://dx.doi.org/10.1136/bmj.316.7133.736

254. Dolan P, Gudex C, Kind P, Williams A. *A Social Tariff for EuroQol: Results from a UK General Population Survey*. York: University of York Centre for Health Economics; 1995.

255. Dolan P, Gudex C, Kind P, Williams A. The time trade-off method: results from a general population study. *Health Econ* 1996;**5**:141–54. http://dx.doi.org/10.1002/(SICI)1099-1050(199603)5:2<141::AID-HEC189>3.0.CO;2-N

256. Haywood KL, Garratt AM, Jordan KP, Healey EL, Packham JC. Evaluation of ankylosing spondylitis quality of life (EASi-QoL): reliability and validity of a new patient-reported outcome measure. *J Rheumatol* 2010;**37**:2100–9. http://dx.doi.org/10.3899/jrheum.091359

257. Kvamme MK, Kristiansen IS, Lie E, Kvien TK. Identification of cutpoints for acceptable health status and important improvement in patient-reported outcomes, in rheumatoid arthritis, psoriatic arthritis, and ankylosing spondylitis. *J Rheumatol* 2010;**37**:26–31. http://dx.doi.org/10.3899/jrheum.090449

258. Osnes-Ringen H, Kvamme MK, Kristiansen IS, Thingstad M, Henriksen JE, Kvien TK, *et al.* Cost-effectiveness analyses of elective orthopaedic surgical procedures in patients with inflammatory arthropathies. *Scand J Rheumatol* 2011;**40**:108–15. http://dx.doi.org/10.3109/03009742.2010.503661

259. Lindstrom U, Bremander A, Bergman S, Haglund E, Petersson IF, Jacobsson LT. Patients with non-as axial SPA have similar prevalence compared to as, but worse perceived health. Results from a population based study. *Ann Rheum Dis* 2013;**72**(Suppl. 3):A667–8. http://dx.doi.org/10.1136/annrheumdis-2013-eular.1978

Appendix 1 Search strategies for clinical and economic reviews

MEDLINE and MEDLINE In-Process & Other Non-Indexed Citations (Ovid)

Date range searched: 1946 to present.

Searched: 5 June 2014 via OVID interface.

Search strategy

1. spondylarthritis/ or spondylitis, ankylosing/ (12,386)
2. ((ankyl$ or axial) adj2 spondyl$).ti,ab. (10,322)
3. (ankyl$ adj2 (spine$ or spinal or vertebra$)).ti,ab. (402)
4. ((Bechtere$ or Bekhtere$ or "Marie Strumpell$" or "Marie Struempell$") adj2 (disease or syndrome)).ti,ab. (451)
5. 1 or 2 or 3 or 4 (14,886)
6. (adalimumab or humira or 331731-18-1).af. (3751)
7. (certolizumab or CDP870 or cimzia or 428863-50-7).af. (497)
8. (etanercept or enbrel or altebrel or 185243-69-0).af. (5540)
9. (golimumab or CNTO 148 or simponi or 476181-74-5).af. (328)
10. (infliximab or remicade or 170277-31-3 or inflectra or remsima or CT-P13).af. (9166)
11. 6 or 7 or 8 or 9 or 10 (13,950)
12. randomized controlled trial.pt. (375,396)
13. controlled clinical trial.pt. (88,473)
14. randomized.ab. (295,232)
15. placebo.ab. (154,473)
16. drug therapy.fs. (1,704,080)
17. randomly.ab. (213,686)
18. trial.ab. (306,623)
19. groups.ab. (1,359,351)
20. 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 (3,348,700)
21. animals/ not (animals/ and humans/) (3,855,883)
22. 20 not 21 (2,872,482)
23. 5 and 11 and 22 (1008)

EMBASE

Date range searched: 1974 to week 22, 2014.

Searched: 5 June 2014 via OVID interface.

Search strategy

1. exp spondylarthritis/ or exp ankylosing spondylitis/ (20,531)
2. ((ankyl$ or axial) adj2 spondyl$).ti,ab. (14,760)
3. (ankyl$ adj2 (spine$ or spinal or vertebra$)).ti,ab. (542)
4. ((Bechtere$ or Bekhtere$ or "Marie Strumpell$" or "Marie Struempell$") adj2 (disease or syndrome)).ti,ab. (551)
5. 1 or 2 or 3 or 4 (22,426)
6. (adalimumab or humira or 331731-18-1).af. (15,439)
7. (certolizumab or CDP870 or cimzia or 428863-50-7).af. (3097)
8. (etanercept or enbrel or altebrel or 185243-69-0).af. (19,368)
9. (golimumab or CNTO 148 or simponi or 476181-74-5).af. (2124)
10. (infliximab or remicade or 170277-31-3 or inflectra or remsima or CT-P13).af. (29,667)
11. 6 or 7 or 8 or 9 or 10 (41,065)
12. (random$ or factorial$ or crossover$ or cross over$ or cross-over$ or placebo$ or (doubl$ adj blind) or (singl$ adj blind$) or assign$ or allocat$ or volunteer$).ti,ab,kw. (1,351,644)
13. crossover-procedure/ or double-blind procedure/ or randomized controlled trial/ or single-blind procedure/ (390984)
14. 12 or 13 (1,428,385)
15. 5 and 11 and 14 (603)
16. limit 15 to embase (581)
17. animals/ not (animals/ and humans/) (1,188,711)
18. 16 not 17 (581)

Cumulative Index to Nursing and Allied Health Literature Plus

Date range searched: inception to 5 June 2014.

Searched: 5 June 2014 via EBSCO*host* interface.

Search strategy

S19 S6 AND S12 AND S18 (87)
S18 S13 OR S14 OR S15 OR S16 OR S17 (148,267)
S17 singl* N blind* or doubl* N blind* or singl* N mask* or doubl* N mask (285)
S16 (ZT "randomized controlled trial") (38,240)
S15 (allocate* or assign* or divid*) N5 (condition* or experiment* or treatment* or control* or group*) (26,737)
S14 crossover or "cross over" or "latin square" or placebo* (41,898)
S13 randomi* or random N allocate* or random N assign* or random N divid* or random N trial* or random N study or random N studies (108,710)
S12 S7 OR S8 OR S9 OR S10 OR S11 (3091)
S11 TX (infliximab or remicade or 170277-31-3 or inflectra or remsima or CT-P13) (1792)
S10 TX (golimumab or CNTO 148 or simponi or 476181-74-5) (119)
S9 TX (etanercept or enbrel or altebrel or 185243-69-0) (1298)
S8 TX (certolizumab or CDP870 or cimzia or 428863-50-7 (91)
S7 TX (adalimumab or humira or 331731-18-1) (647)
S6 S1 OR S2 OR S3 OR S4 OR S5 (2566)
S5 TX ((Bechtere* or Bekhtere* or "Marie Strumpell*" or "Marie Struempell*") N2 (disease or syndrome)) (3)

S4 TX (ankyl* N2 (spine* or spinal or vertebra*)) (91)
S3 TX ((ankyl* or axial) N2 spondyl*) (2277)
S2 MH spondylitis, ankylosing (1803)
S1 MH spondylarthritis (500)

Science Citation Index

Date range searched: 1900 to 2014.

Searched: 16 June 2014 via Web of Science.

Indexes=SCI-EXPANDED Timespan=1900-2014.

Search strategy

#13 #12 AND #11 AND #5 (1001)
#12 TS= clinical trial* OR TS=research design OR TS=comparative stud* OR TS=evaluation stud* OR TS=controlled trial* OR TS=follow-up stud* OR TS=prospective stud* OR TS=random* OR TS=placebo* OR TS=(single blind*) OR TS=(double blind*) (2,435,907)
#11 #10 OR #9 OR #8 OR #7 OR #6 (20,446)
#10 TOPIC: ((infliximab or remicade or 170277-31-3 or inflectra or remsima or CT-P13)) (13,285)
#9 TOPIC: ((golimumab or CNTO 148 or simponi or 476181-74-5)) (494)
#8 TOPIC: ((etanercept or enbrel or altebrel or 185243-69-0)) (7138)
#7 TOPIC: ((certolizumab or CDP870 or cimzia or 428863-50-7)) (916)
#6 TOPIC: ((adalimumab or humira or 331731-18-1)) (4754)
#5 #4 OR #3 OR #2 OR #1 (14,918)
#4 TOPIC: (((Bechtere* or Bekhtere* or "Marie Strumpell*" or "Marie Struempell*") NEAR/2 (disease or syndrome))) (191)
#3 TOPIC: ((ankyl* NEAR/2 (spine* or spinal or vertebra*))) (644)
#2 TOPIC: (((ankyl* or axial) NEAR/2 spondyl*)) (13,854)
#1 TOPIC: (spondylarthritis OR spondyloarthritis) (2394)

National Institutes of Health ClinicalTrials.gov register

Searched: 23 July 2014 online at http://clinicaltrials.gov/ct2/search.

Search strategy

((spondylarthritis OR spondyloarthritis OR spondylitis) AND (infliximab OR remicade OR inflectra OR remsima OR golimumab OR simponi OR etanercept OR enbrel OR altebrel OR certolizumab OR cimzia OR adalimumab OR humira))

160 results.

Cochrane Library (includes Cochrane Database of Systematic Reviews, Database of Abstracts of Reviews of Effects, Health Technology Assessment Database, Cochrane Central Register of Controlled Trials and NHS Economic Evaluation Database)

Searched: 5 June 2014 online at http://onlinelibrary.wiley.com/o/cochrane/cochrane_search_fs.html.

Search strategy

#1 MeSH descriptor: [Spondylitis, Ankylosing] explode all trees
#2 MeSH descriptor: [Spondylarthritis] explode all trees
#3 ((ankyl* or axial) near/2 spondyl*):ti,ab,kw (Word variations have been searched)
#4 (ankyl* near/2 (spine* or spinal or vertebra*)):ti,ab,kw (Word variations have been searched)
#5 ((Bechtere* or Bekhtere* or "Marie Strumpell*" or "Marie Struempell*") near/2 (disease or syndrome)):ti,ab,kw
#6 #1 or #2 or #3 or #4 or #5
#7 (adalimumab or humira or 331731-18-1):ti,ab,kw
#8 (certolizumab or CDP870 or cimzia or 428863-50-7):ti,ab,kw
#9 (etanercept or enbrel or altebrel or 185243-69-0):ti,ab,kw
#10 (golimumab or CNTO 148 or simponi or 476181-74-5):ti,ab,kw
#11 (infliximab or remicade or 170277-31-3 or inflectra or remsima or CT-P13):ti,ab,kw
#12 #7 or #8 or #9 or #10 or #11
#13 #6 and #12

284 total results comprised two Cochrane Database of Systematic Reviews, five Database of Abstracts of Reviews of Effects, 21 HTA, 233 Cochrane Central Register of Controlled Trials and 14 NHS Economic Evaluation Database.

Conference Proceedings Citation Index – Science

Searched: 2 September 2014 via Wiley Web of Science interface.

Indexes=CPCI-S Timespan=1900-2014.

Search strategy

#12 #11 AND #5 (341)
#11 #10 OR #9 OR #8 OR #7 OR #6 (4745)
#10 TOPIC: ((infliximab or remicade or 170277-31-3 or inflectra or remsima or CT-P13)) (2537)
#9 TOPIC: ((golimumab or CNTO 148 or simponi or 476181-74-5)) (141)
#8 TOPIC: ((etanercept or enbrel or altebrel or 185243-69-0)) (1221)
#7 TOPIC: ((certolizumab or CDP870 or cimzia or 428863-50-7)) (291)
#6 TOPIC: ((adalimumab or humira or 331731-18-1)) (1140)
#5 #4 OR #3 OR #2 OR #1 (2117)
#4 TOPIC: (((Bechtere* or Bekhtere* or "Marie Strumpell*" or "Marie Struempell*") NEAR/2 (disease or syndrome))) (4)
#3 TOPIC: ((ankyl* NEAR/2 (spine* or spinal or vertebra*))) (55)
#2 TOPIC: (((ankyl* or axial) NEAR/2 spondyl*)) (1906)
#1 TS=(spondylarthritis OR spondyloarthritis) (393)

International Prospective Register of Systematic Reviews (PROSPERO)

Searched: 7 October 2014 online at www.crd.york.ac.uk/prospero/search.asp.

Search strategy

spondylitis [In All Fields]

OR

spondylarthritis [In All Fields]

OR

spondyloarthritis [In All Fields]

6 results.

National Guideline Clearinghouse

Searched: 7 October 2014 online at www.guideline.gov.

Search strategy

spondylitis OR spondylarthritis OR spondyloarthritis

15 results.

NHS Evidence

Searched 27 October 2014 online at www.evidence.nhs.uk.

Search strategy

((((ankyl* or axial) near/2 spondyl*) OR (ankyl* near/2 (spine* or spinal or vertebra*)) AND (adalimumab or humira or certolizumab or CDP870 or cimzia or etanercept or enbrel or altebrel or golimumab or CNTO 148 or simponi or infliximab or remicade or inflectra or remsima or CT-P13))

350 results.

NHS Clinical Knowledge Summaries

Searched: 27 October 2014 online at http://cks.nice.org.uk/#?char=A.

1 result for ankylosing spondylitis.

Searches for economic review

NHS Economic Evaluation Database

Searched: 5 June 2014 online at http://onlinelibrary.wiley.com/o/cochrane/cochrane_search_fs.html.

Search strategy

#1 MeSH descriptor: [Spondylitis, Ankylosing] explode all trees
#2 MeSH descriptor: [Spondylarthritis] explode all trees
#3 ((ankyl* or axial) near/2 spondyl*):ti,ab,kw (Word variations have been searched)
#4 (ankyl* near/2 (spine* or spinal or vertebra*)):ti,ab,kw (Word variations have been searched)
#5 ((Bechtere* or Bekhtere* or "Marie Strumpell*" or "Marie Struempell*") near/2 (disease or syndrome)):ti,ab,kw
#6 #1 or #2 or #3 or #4 or #5
#7 (adalimumab or humira or 331731-18-1):ti,ab,kw
#8 (certolizumab or CDP870 or cimzia or 428863-50-7):ti,ab,kw
#9 (etanercept or enbrel or altebrel or 185243-69-0):ti,ab,kw
#10 (golimumab or CNTO 148 or simponi or 476181-74-5):ti,ab,kw
#11 (infliximab or remicade or 170277-31-3 or inflectra or remsima or CT-P13):ti,ab,kw
#12 #7 or #8 or #9 or #10 or #11
#13 #6 and #12
14 results.

Searches for European Quality of Life-5 Dimensions

MEDLINE and MEDLINE In-Process & Other Non-Indexed Citations (Ovid)

Date range searched: 1946 to present.

Searched: 16 June 2014 via OVID interface.

Search strategy

1. spondylarthritis/ or spondylitis, ankylosing/ (12,394)
2. ((ankyl$ or axial) adj2 spondyl$).ti,ab. (10,334)
3. (ankyl$ adj2 (spine$ or spinal or vertebra$)).ti,ab. (402)
4. ((Bechtere$ or Bekhtere$ or "Marie Strumpell$" or "Marie Struempell$") adj2 (disease or syndrome)).ti,ab. (451)
5. 1 or 2 or 3 or 4 (14,899)
6. (5d or 5-d or 5 dimension or eq-5d or eq5d or eq 5d).ti,ab. (13,976)
7. 5 and 6 (27)

EMBASE (June 2014)

Searched: 16 June 2014 via OVID interface.

Search strategy

1. exp spondylarthritis/ or exp ankylosing spondylitis/ (20,653)
2. ((ankyl$ or axial) adj2 spondyl$).ti,ab. (14,855)
3. (ankyl$ adj2 (spine$ or spinal or vertebra$)).ti,ab. (545)
4. ((Bechtere$ or Bekhtere$ or "Marie Strumpell$" or "Marie Struempell$") adj2 (disease or syndrome)).ti, ab. (552)
5. 1 or 2 or 3 or 4 (22,550)
6. (5d or 5-d or 5 dimension or eq-5d or eq5d or eq 5d).ti,ab. (17,019)
7. 5 and 6 (60)
8. limit 7 to embase (55)

Searches for economic models

MEDLINE and MEDLINE In-Process & Other Non-Indexed Citations (Ovid)

Date range searched: 1946 to present.

Searched: 25 July 2014 via OVID interface.

Search strategy

1. spondylarthritis/ or spondylitis, ankylosing/ (12,505)
2. ((ankyl$ or axial) adj2 spondyl$).ti,ab. (10,436)
3. (ankyl$ adj2 (spine$ or spinal or vertebra$)).ti,ab. (407)
4. ((Bechtere$ or Bekhtere$ or "Marie Strumpell$" or "Marie Struempell$") adj2 (disease or syndrome)).ti, ab. (455)
5. 1 or 2 or 3 or 4 (15,038)
6. exp models, economic/ (10,268)
7. ((economic$ or cost$ or pric$ or value or statistic$) and model$).ti,ab. (245,686)
8. 6 or 7 (250,668)
9. 5 and 8 (107)

EMBASE

Date range searched: 1974 to 24 July 2014.

Searched: 25 July 2014 via OVID interface.

Search strategy

1. exp spondylarthritis/ or exp ankylosing spondylitis/ (20,858)
2. ((ankyl$ or axial) adj2 spondyl$).ti,ab. (14,996)
3. (ankyl$ adj2 (spine$ or spinal or vertebra$)).ti,ab. (553)
4. ((Bechtere$ or Bekhtere$ or "Marie Strumpell$" or "Marie Struempell$") adj2 (disease or syndrome)).ti,ab. (553)
5. 1 or 2 or 3 or 4 (22,760)
6. statistical model/ (102,203)
7. ((economic$ or cost$ or pric$ or value or statistic$) adj2 model$).ti,ab. (24,642)
8. 6 or 7 (119,366)
9. 5 and 8 (63)
10. limit 9 to embase (55)

Appendix 2 Synthesis methods for clinical efficacy network meta-analyses

Estimating standard deviations from interquartile ranges

Where Q_1 is the lower quartile, Q_3 is the upper quartile and σ is the SD, then the SD was estimated as:

$$\sigma = \frac{(Q_3 + Q_1)}{2 \times 0.67}. \quad (3)$$

Calculating change from baseline outcomes and standard deviations

Given baseline and final values and their SDs, the change from baseline values and SDs can be calculated if the within-study correlation between baseline and final values is known. Similarly, the final values can be computed.

The within-study correlation ρ between baseline and final values can be calculated as follows, as stated in the Cochrane Handbook,[183] where

$$\rho = \frac{SD^2_{baseline} + SD^2_{final} - SD^2_{change}}{2 \times SD_{final} \times SD_{baseline}}. \quad (4)$$

The SD of the change from baseline can be found by rearranging the above equation. The SD of the final value can be found by rearranging the above equation which produces a quadratic. As a range of correlation estimates were obtained from the studies available, we tested 0.3 and 0.7 correlation estimates in our analyses. In calculating the SD of final values, this sometimes resulted in complex roots. In these cases, the lowest correlation estimate that allowed a real root was used in the calculation.

Prior distribution for the between-study standard deviation for the placebo absolute risk

In running fixed-effect and random-effect models to estimate the placebo absolute risk, the random-effect models had better fit. For ASAS 70 response, there were insufficient trials to run a random-effects model, so a prior distribution for the between-study SD was specified. This was derived from the between-study SD from the ASAS 40 analysis. The prior distribution was specified as a log-normal distribution, and the log-normal distribution parameters μ and σ^2 were derived from the following equations:

$$Mean = e^{\mu + \sigma^2/2}. \quad (5)$$

$$Median = e^{\mu}. \quad (6)$$

I^2

As noted in Higgins *et al.*,[184] I^2 was calculated as

$$I^2 = \frac{\tau^2}{\tau^2 + s^2}, \tag{7}$$

where τ^2 is the between-study variance estimated in the multiple-treatment meta-analysis,

$$s^2 = \frac{\Sigma w_i(k-1)}{(\Sigma w_i)^2 - \Sigma w_i^2}, \tag{8}$$

which was calculated in Excel, and w_i is the precision of study i.

Correlation

Table 108 presents the results for BASDAI change from baseline assuming a class effect and independent treatment effects, and assuming 0.3 and 0.7 within-study correlation. It is clear that the different correlation assumptions make no difference in this case. This is perhaps because the studies affected by the correlation assumption were small studies.

TABLE 108 The class and independent BASDAI change from baseline of the anti-TNFs vs. placebo assuming 0.3 and 0.7 within-study correlation

	0.3 correlation		0.7 correlation	
Intervention	Mean	95% CrI	Mean	95% CrI
Class	–1.66	–1.89 to –1.43	–1.66	–1.88 to –1.43
Adalimumab	–1.55	–1.88 to –1.23	–1.56	–1.88 to –1.24
Centrolizumab	–1.46	–2.16 to –0.74	–1.46	–2.16 to –0.74
Etanercept	–1.76	–2.15 to –1.37	–1.76	–2.15 to –1.37
Infliximab	–2.28	–3.18 to –1.38	–2.28	–3.18 to –1.38

Appendix 3 Risk-of-bias data

TABLE 109 Full risk-of-bias results

Trial	Sequence generation	Allocation concealment	Important baseline imbalance	Blinding of participants and researchers	Blinding of outcome assessment	Incomplete outcome data	Selective reporting	Exclude in sensitivity analysis?
Haibel 2008[52]	Unclear	Unclear	Unclear	Low	Low	Low	Low	No
	–	–	Imbalance for HLA-B27 positive and MRI positive	–	–	Number of withdrawals and dropouts: 0 Imputation used for continuous outcomes: N/A	All main relevant outcomes reported	
Hu 2012[55]	Unclear	Unclear	Low	Unclear	Low	Unclear	Unclear	Yes
	There is no description of the randomisation procedure and no explanation for the imbalance in number of patients in treatment arms, 26 vs. 20	No details reported	Main prognostic indicators were similar across trial arms at baseline	No details reported	–	Number of withdrawals and dropouts: NR Imputation used for continuous outcomes: NR	No reporting of AEs	
Huang 2014[56]	Low	Low	Low	Low	Low	Low	Low	No
	Centralised computer-based system	Centralised computer-based system	Groups comparable for all important factors	Matching placebo and all study personnel and patients stated to be blinded	–	Number of withdrawals and dropouts: 12 Imputation used for continuous outcomes: LOCF	All main outcomes reported	

Trial	Sequence generation	Allocation concealment	Important baseline imbalance	Blinding of participants and researchers	Blinding of outcome assessment	Incomplete outcome data	Selective reporting	Exclude in sensitivity analysis?
Lambert 2007[57]	Unclear	Unclear	Low	Unclear	Low	Low	Low	No
	–	–	No imbalances in possible prognostic factors	Stated to be double blind	–	Number of withdrawals and dropouts: 0 at week 12; two from placebo arm at week 52 Imputation used for continuous outcomes: no imputation for missing SPARCC enthesitis index score	BASDAI score not reported at follow-up	
ABILITY-1 2013[58]	Low	Low	Low	Low	Low	Low	Low	No
	Centralised randomisation with interactive voice response system	–	–	Matching placebo	–	Number of withdrawals and dropouts: $n = 6$ Imputation used for continuous outcomes: LOCF imputed values	–	

continued

TABLE 109 Full risk-of-bias results (*continued*)

Trial	Sequence generation	Allocation concealment	Important baseline imbalance	Blinding of participants and researchers	Blinding of outcome assessment	Incomplete outcome data	Selective reporting	Exclude in sensitivity analysis?
ATLAS 2006[61]	Unclear	Unclear	Low	Low	Low	Low	Low	No
	–	–	Balanced across treatment arms	Matching placebo	–	Number of withdrawals and dropouts: by week 12, $n=4$ placebo and $n=4$ active; by week 24, $n=6$ placebo and $n=13$ active (Note: week 24 was still RCT although no responders permitted early escape after week 12). Imputation used for continuous outcomes: LOCF	Primary and all main outcomes reported	
RAPID-axSpA (Landewe 2014)[64]	Low	Low	Low	Unclear	Low	Low	Low	No
	Central randomisation	Central randomisation	Small difference in baseline CRP level and HLA-B27 positive, making the placebo group have slightly increased risk (but unclear possible impact)	Administration of treatment was by unblinded trained personnel; their role in assessment is unclear and so the impact of their unblinded status is unclear	–	Number of withdrawals and dropouts: unclear at 12 weeks but at 24 weeks, $n=10$ placebo, $n=6$ 200 mg and $n=9$ 400 mg Imputation used for continuous outcomes: LOCF	Hierarchical analysis plan adhered to	

Trial	Sequence generation	Allocation concealment	Important baseline imbalance	Blinding of participants and researchers	Blinding of outcome assessment	Incomplete outcome data	Selective reporting	Exclude in sensitivity analysis?
Barkham 2010[71]	Unclear	Unclear	Unclear	Low	Low	Unclear	Low	Yes
	–	–	–	–	–	Number of withdrawals and dropouts: unclear, although it appears to be $n = 9$ for etanercept and $n = 8$ for placebo (the number for which ASAS 40 data were available)	–	
						Imputation used for continuous outcomes: ITT LOCF		
Davis 2003[72]	Unclear	Unclear	Low	Low	Low	Low	Low	No
	–	–	–	–	–	Number of withdrawals and dropouts: at 12 weeks, $n = 6$ etanercept and $n = 5$ placebo; at 24 weeks, $n = 12$ etanercept, $n = 19$ placebo	–	
						Imputation used for continuous outcomes: LOCF using etanercept $n = 138$, and using placebo $n = 139$		

continued

TABLE 109 Full risk-of-bias results (*continued*)

Trial	Sequence generation	Allocation concealment	Important baseline imbalance	Blinding of participants and researchers	Blinding of outcome assessment	Incomplete outcome data	Selective reporting	Exclude in sensitivity analysis?
Dougados 2011[74]	Unclear	Unclear	Low	Low	Low	Low	Low	No
	–	–	–	–	–	Number of withdrawals and dropouts: $n = 1$ etanercept and $n = 4$ placebo. Imputation used for continuous outcomes: mITT (at least one dose) with LOCF	–	
Dougados 2014[75]	Low	Low	Low	Low	Low	Low	Low	No
	–	–	–	–	–	Number of withdrawals and dropouts: $n = 6$ etanercept and $n = 3$ placebo. In addition to this, five patients in each group were excluded from analyses because of misdiagnosis Imputation used for continuous outcomes: LOCF in mITT population, $n = 106$ etanercept and $n = 109$ placebo	–	

Trial	Sequence generation	Allocation concealment	Important baseline imbalance	Blinding of participants and researchers	Blinding of outcome assessment	Incomplete outcome data	Selective reporting	Exclude in sensitivity analysis?
Gorman 2002[79]	Low	Low	High (chance imbalance)	Low	Low	Low	Low	Yes
	–	–	BASFI	–	–	Number of withdrawals and dropouts: $n = 3$	–	
						Imputation used for continuous outcomes: not totally clear, but appears to be proper ITT with LOCF		
Calin 2004[83]	Unclear	Unclear	High	Low	Low	Low	Low	Yes
	–	–	Important difference in CRP level, borderline important difference in age	–	–	Number of withdrawals and dropouts: $n = 2$. etanercept	–	
						Imputation used for continuous outcomes: LOCF for mITT population (placebo $n = 39$ and etanercept $n = 45$)		

continued

TABLE 109 Full risk-of-bias results (*continued*)

Trial	Sequence generation	Allocation concealment	Important baseline imbalance	Blinding of participants and researchers	Blinding of outcome assessment	Incomplete outcome data	Selective reporting	Exclude in sensitivity analysis?
van der Heijde 2006[86]	Unclear	Unclear	Unclear	Low	Low	Low	Low	No
	–	–	No data for HLA-B27	–	–	Number of withdrawals and dropouts: $n = 14$ in 50-mg group, $n = 14$ in 25-mg group and $n = 7$ in placebo group. In addition to this, five patients did not receive one dose of treatment (no further details) Imputation used for continuous outcomes: mITT population analysed (had at least one dose) using the $n = 155$, $n = 150$, $n = 51$ group sizes and LOCF was used to impute missing data	–	
Giardina 2010[88]	High	High	Low	High	High	Low	Low	Yes
	–	–	–	–	–	Number of withdrawals and dropouts: $n = 0$ Imputation used for continuous outcomes: none needed	–	

Trial	Sequence generation	Allocation concealment	Important baseline imbalance	Blinding of participants and researchers	Blinding of outcome assessment	Incomplete outcome data	Selective reporting	Exclude in sensitivity analysis?
GO-RAISE 2008[90]	Low	Low	Low	Low	Low	Low	Low	No
	Random assignment using voice response system	Random assignment using voice response system	No important imbalance in key prognostic variables	Matching placebo used	–	Number of withdrawals and dropouts: $n = 17$ to week 24 ($n = 2$ placebo, $n = 9$ 50 mg and $n = 6$ 100 mg). Not clear how many at week 14 (primary time point)	Primary end point and all other main outcomes (BASDAI, BASFI, BASMI and SF-36) reported	
						Imputation used for continuous outcomes: LOCF (ITT population)		
Bao 2012[96]	Unclear	Unclear	Unclear	Low	Low	Unclear	Low	No
	–	–	No HLA-B27 data	–	–	Number of withdrawals and dropouts: not reported for week 14	Primary outcome and other main outcomes reported	
						Imputation used for continuous outcomes: NR		
Tam 2014[97]	Unclear	Unclear	Unclear	Unclear	Low	Unclear	Unclear	No
	Abstract only, very small study (Chinese)	Abstract only, very small study (Chinese)	Abstract only, very small study (Chinese)	Abstract only, very small study (Chinese)	–	Number of withdrawals and dropouts: NR	Abstract only, very small study (Chinese)	
						Imputation used for continuous outcomes: NR		

continued

TABLE 109 Full risk-of-bias results (*continued*)

Trial	Sequence generation	Allocation concealment	Important baseline imbalance	Blinding of participants and researchers	Blinding of outcome assessment	Incomplete outcome data	Selective reporting	Exclude in sensitivity analysis?
Barkham 2009[50]	Unclear	Unclear	High	Low	Low	Low	Low	Yes
	–	–	Median CRP level 11.5 placebo vs. 5 infliximab; likely a result of chance, as there was higher CRP level in placebo group (and higher CRP level was associated with better responses)	–	–	Number of withdrawals and dropouts: $n = 1$ in the placebo group (at 12 weeks) Imputation used for continuous outcomes: NR but ITT population analysed	–	
Braun 2002[98]	Low	Low	Low	Low	Low	Low	Low	No
	–	–	–	–	–	Number of withdrawals and dropouts: $n = 0$ Imputation used for continuous outcomes: none required	–	
Marzo-Ortega 2005[100]	Low	Low	Unclear	Low	Low	Unclear	Low	No
	–	–	The only issue is with age, and the difference of 2 years could be caused by rounding	–	–	Number of withdrawals and dropouts: $n = 5$ out of 14 for placebo and $n = 2$ out of 28 for infliximab Imputation used for continuous outcomes: ITT with LOCF	–	

Trial	Sequence generation	Allocation concealment	Important baseline imbalance	Blinding of participants and researchers	Blinding of outcome assessment	Incomplete outcome data	Selective reporting	Exclude in sensitivity analysis?
Van den Borsch 2002[101]	Unclear	Unclear	High;	Low	Low	Low	Low	Yes
	–	–	BASFI > 1 point	–	–	Number of withdrawals and dropouts: $n = 0$	–	
						Imputation used for continuous outcomes: N/A		
ASSERT 2005[102]	Unclear	Unclear	Low	Low	Low	Low	Low	No
	–	–	–	–	–	Number of withdrawals and dropouts: $n = 8$ ($n = 4$ in each group)	–	
						Imputation used for continuous outcomes: not reported, but ITT population analysed. LOCF was used for ASAS 20		
PLANETAS 2013[110]	Low	Low	Unclear	Low	Low	Low	Low	No
	–	–	HLA-B27 was not reported	–	–	Number of withdrawals and dropouts: $n = 21$ ($n = 12$ CT-P13 *vs.* $n = 9$ infliximab)	–	
						Imputation used for continuous outcomes: NR, although ITT population was analysed		

N/A, not applicable; NR, not reported; SPARCC, Spondyloarthritis Research Consortium of Canada.
High, low and unclear relate to the trial's risk-of-bias judgement based on the Cochrane risk-of-bias tool. See *Chapter 4, Summary and critique of published cost-effectiveness studies.*

TABLE 110 Prognostic indicators of important baseline imbalance used in risk-of-bias assessment

Possible prognostic indicator	Results of association			Implications for baseline imbalance across groups within a trial (and variation in efficacy across trials)
	Glintborg 2010 (DANBIO registry)[112] ($n = 842$); 6-month time point; adalimumab, etanercept, infliximab	Vastesaeger 2011[172] (ASSERT[102] and GO-RAISE[90] trial data)[a] ($n = 635$); 3-month time point; infliximab, golimumab	Lord 2009 (BSRBR)[173] ($n = 261$); 6-month time point; adalimumab, etanercept, infliximab	
HLA-B27 status	No data	Moderate association	No data	HLA-B27 positive patients have a better outcome. Use 20% group difference as an important imbalance?
CRP level	≤ 14 mg/l vs. > 14 mg/l (OR 0.45, $p < 0.001$)	≤ 6 mg/l vs. > 6 mg/l to 20 mg/l: moderate; ≤ 6 mg/l vs. > 20 mg/l: moderate to strong	'Raised inflammatory markers' an important predictor, but result only available for CRP level or ESR (not CRP level alone)	Higher CRP levels are associated with a better outcome. Use Glintborg and Vastesaeger cut-offs (providing there's at least a 2 mg/l difference between groups)
Age	OR 0.98 per year; $p = 0.03$	< 40 years vs. > 40 years: weak to moderate	No significant association	Younger age associated with a better outcome. Use Vastesaeger cut-off point providing at least a 2-year difference between groups
BASFI score	OR 0.87 per cm increase; $p = 0.008$	< 6.5 vs. > 6.5 moderate to strong	Per unit increase, OR 0.78 (95% CI 0.64 to 0.99)	Lower BASFI scores associated with a better outcome. Use a 1-point difference as an indication of important imbalance? Use < 6.5 vs. > 6.5 providing there's at least a 0.5-point difference between groups
BASDAI	Not analysed	No significant association	Per unit increase, OR 1.30 (95% CI 1.04 to 1.62)	Higher BASDAI scores associated with a better outcome. Use a 1-point difference as an indication of important imbalance?
Disease duration	No significant association	No significant association	No significant association	Do not assess
Sex	No significant association	No significant association	No significant association	Do not assess

a OR results categorised association using weak, moderate, strong or very strong terminology (based on size of OR).

Appendix 4 Trial results

TABLE 111 Continuous outcomes: final values results

Trial	Population	Treatment arm	Dose	Time point (weeks)	Patients	Mean final values					
						BASDAI	BASFI	BASMI	MASES	SF-36 PCS	SF-36 MCS
Haibel 2008[52]	Nr-axSpA	Adalimumab	40 mg every 2 weeks	12	22	3.8 (SD 2.5)	3 (SD 2.4)	1.3 (SD 1.4)	2.5 (SD 3.5)	38.8 (SD 11.8)	44.6 (SD 12.7)
	Nr-axSpA	Placebo	–	12	24	5 (SD 2.4)	4.1 (SD 2.6)	1.7 (SD 1.5)	2.8 (SD 3.4)	34.9 (SD 9.6)	43.9 (SD 11.8)
Hu 2012[55]	AS	Adalimumab	40 mg every 2 weeks	12	NR	2.3 (SD 1.8)	1.8 (SD 1.6)	–	–	–	–
	AS	Placebo	–	12	NR	4.2 (SD 2.6)	2.9 (SD 1.9)	–	–	–	–
Huang 2014[56]	AS	Adalimumab	40 mg every 2 weeks	12	229	–	–	–	–	–	–
	AS	Placebo	–	12	115	–	–	–	–	–	–
Lambert 2007[57]	AS	Adalimumab	40 mg every 2 weeks	12	38	–	–	–	–	–	–
	AS	Placebo	–	12	44	–	–	–	–	–	–
[a]ABILITY-1 2013[58]	Nr-axSpA	Adalimumab	40 mg every 2 weeks	12	69	–	–	–	–	–	–
	Nr-axSpA	Placebo	–	12	73	–	–	–	–	–	–
ATLAS 2006[61]	AS	Adalimumab	40 mg every 2 weeks	12	208	–	1.414	–	–	–	–
	AS	Placebo	–	12	107	–	–	–	–	–	–
RAPID-axSpA 2014[64]	AS	Certolizumab pegol	200 mg every 2 weeks	12	65	–	–	–	–	–	–
	AS	Certolizumab pegol	400 mg every 4 weeks	12	56	–	–	–	–	–	–
	AS	Placebo	–	12	57	–	–	–	–	–	–

Trial	Population	Treatment arm	Dose	Time point (weeks)	Patients	Mean final values						
						BASDAI	BASFI	BASMI	MASES	SF-36 PCS	SF-36 MCS
RAPID-axSpA 2014[64]	Nr-axSpA	Certolizumab pegol	200 mg every 2 weeks	12	46	–	–	–	–	–	–
	Nr-axSpA	Certolizumab pegol	400 mg every 4 weeks	12	51	–	–	–	–	–	–
	Nr-axSpA	Placebo	–	12	50	–	–	–	–	–	–
Barkham 2010[71]	AS	Etanercept	25 mg twice weekly	12	20	–	–	–	–	–	–
	AS	Placebo	–	12	20	–	–	–	–	–	–
Davis 2003[72]	AS	Etanercept	25 mg twice weekly	12	138	–	–	–	–	–	–
	AS	Placebo	–	12	139	–	–	–	–	–	–
	AS	Etanercept	25 mg twice weekly	24	138	3.45 (SE 0.21)	3.6 (SE 0.22)	–	–	–	–
	AS	Placebo	–	24	139	5.51 (SE 0.2)	5.47 (SE 0.22)	–	–	–	–
Dougados 2011[74]	AS	Etanercept	50 mg weekly	12	39	3.7 (SD 2.6)	4.1 (SD 2.9)	5.1 (SD 1.7)	–	–	–
	AS	Placebo	–	12	43	4.5 (SD 1.9)	4.8 (SD 2.1)	5.6 (SD 1.3)	–	–	–
Dougados 2014[76]	AS	Etanercept	50 mg weekly	12	106	–	–	–	–	43.7 (SD 8.9)	–
	AS	Placebo	–	12	109	–	–	–	–	41 (SD 7.8)	–
	Nr-axSpA	Etanercept	50 mg weekly	12	94	–	–	–	–	–	–
	Nr-axSpA	Placebo	–	12	95	–	–	–	–	–	–
Gorman 2002[79]	AS	Etanercept	25 mg twice a week	16	20	–	2.2 (SD 2.1)	–	–	–	–
	AS	Placebo	–	16	20	–	3.1 (SD 3)	–	–	–	–
Calin 2004[83]	AS	Etanercept	25 mg twice weekly	12	45	3.38	3.96	–	–	–	–
	AS	Placebo	–	12	39	5.01	5.39	–	–	–	–

continued

TABLE 111 Continuous outcomes: final values results (*continued*)

Trial	Population	Treatment arm	Dose	Time point (weeks)	Patients	Mean final values					
						BASDAI	BASFI	BASMI	MASES	SF-36 PCS	SF-36 MCS
van der Heijde 2006[86]	AS	Etanercept	25 mg twice weekly	12	150	–	–	–	–	–	–
	AS	Etanercept	50 mg weekly	12	155	–	–	–	–	–	–
	AS	Placebo	–	12	51	–	–	–	–	–	–
Giardina 2010[88]	AS	Etanercept	50 mg weekly	12	25	–	5	–	–	–	–
	AS	Infliximab	5 mg/kg at 0, 2 and ≥ 6 weeks	12	25	–	3.5	–	–	–	–
GO-RAISE 2008[90]	AS	Golimumab	50 mg (two every 4 weeks)	14	138	–	–	–	–	–	–
	AS	Golimumab	100 mg (two every 4 weeks)	14	140	–	–	–	–	–	–
	AS	Placebo	–	14	78	–	–	–	–	–	–
Bao 2012[96]	AS	Golimumab	50 mg every 4 weeks	14	–	–	–	–	–	–	–
	AS	Placebo	–	14	–	–	–	–	–	–	–
Tam 2014[97]	AS	Golimumab	50 mg monthly	26	NR	–	–	–	–	–	–
	AS	Placebo	–	26	NR	–	–	–	–	–	–
Barkham 2009[50]	Nr-axSpA	Infliximab	5 mg/kg (0, 2 and ≥ 6 weeks)	16	20	–	–	–	–	–	–
	Nr-axSpA	Placebo	–	16	20	–	–	–	–	–	–
Braun 2002[98]	AS	Infliximab	5 mg/kg (0, 2 and 6 weeks)	12	34	3.3	–	–	–	–	–
	AS	Placebo	–	12	35	5.7	–	–	–	–	–

Trial	Population	Treatment arm	Dose	Time point (weeks)	Patients	Mean final values					
						BASDAI	BASFI	BASMI	MASES	SF-36 PCS	SF-36 MCS
Marzo-Ortega 2005[100]	AS	Infliximab + methotrexate	5 mg/kg[b]	10	28	3.34 (SD 2.56)	4.96	–	–	–	–
	AS	Placebo + methotrexate	–[c]	10	14	5.19 (SD 2.52)	6.1	–	–	–	–
	AS	Infliximab + methotrexate	5 mg/kg[b]	30	28	4.6 (SD 2.85)	5.04	–	–	–	–
	AS	Placebo + methotrexate	–[c]	30	14	5.74 (SD 2.34)	5.68	–	–	–	–
Van den Bosch 2002[101]	AS	Infliximab	5 mg/kg (0, 2 and 6 weeks)	12	9	2.66	2.74	4	–	–	–
	AS	Placebo	–	12	12	5.01	7.19	4	–	–	–
ASSERT 2005[102]	AS	Infliximab	5 mg/kg (0, 2 and ≥ 6 weeks)	24	201	–	–	–	–	–	–
	AS	Placebo	–	24	78	–	–	–	–	–	–
PLANETAS 2013[110]	AS	CT-P13	5 mg/kg	14	125	–	–	–	–	–	–
	AS	Infliximab	5 mg/kg	14	125	–	–	–	–	–	–
	AS	CT-P13	5 mg/kg	30	125	–	–	–	–	–	–
	AS	Infliximab	5 mg/kg	30	125	–	–	–	–	–	–
	AS	CT-P13	5 mg/kg	54	125	–	–	–	–	–	–
	AS	Infliximab	5 mg/kg	54	125	–	–	–	–	–	–

NR, not reported.
a Licensed population.
b Infliximab 5 mg/kg (infusion at weeks 0, 2, 6, 14 and 22) + methotrexate oral 7.5 mg with folic acid (5 mg twice a week which increased to 10 mg a week).
c Methotrexate oral 7.5 mg with folic acid (5 mg twice a week which increased to 10 mg a week).

TABLE 112 Continuous outcomes: change from baseline results

Trial	Population	Treatment arm	Dose	Time point (weeks)	Patients	Mean change from baseline					
						BASDAI	BASFI	BASMI	MASES	SF-36 PCS	SF-36 MCS
Haibel 2008[52]	Nr-axSpA	Adalimumab	40 mg every 2 weeks	12	22	–	–	–	–	–	–
	Nr-axSpA	Placebo	–	12	24	–	–	–	–	–	–
Hu 2012[55]	AS	Adalimumab	40 mg every 2 weeks	12	NR	–	–	–	–	–	–
	AS	Placebo	–	12	NR	–	–	–	–	–	–
Huang 2014[56]	AS	Adalimumab	40 mg every 2 weeks	12	229	−2.8 (SD 1.9)	−1.75 (SD 2.02)	−0.5 (SD 0.6)	−1.2 (SD 2.1)	6.6 (SD 6.4)	5.1 (SD 9.9)
	AS	Placebo	–	12	115	−1.4 (SD 1.9)	−0.47 (SD 1.64)	−0.2 (SD 0.7)	−0.8 (SD 1.7)	4 (SD 6.3)	2.8 (SD 9.4)
Lambert 2007[57]	AS	Adalimumab	40 mg every 2 weeks	12	38	–	–	–	–	–	–
	AS	Placebo	–	12	44	–	–	–	–	–	–
[a]ABILITY-1 2013[58]	Nr-axSpA	Adalimumab	40 mg every 2 weeks	12	69	−2.2 (SD 2.5)	−1.28 (SD 2.02)	−0.2 (SD 0.73)	−0.7 (SD 2.78)	6.9 (SD 9.32)	1.4 (SD 8.63)
	Nr-axSpA	Placebo	–	12	73	−1.1 (SD 1.96)	−0.63 (SD 1.79)	−0.2 (SD 0.64)	−1 (SD 2.71)	2.3 (SD 6.81)	0.7 (SD 11.38)
ATLAS 2006[61]	AS	Adalimumab	40 mg every 2 weeks	12	208	−2.6 (SE 0.2)	–	−0.5 (SE 0.1)	−2.7 (SE 0.4)	6.9 (SE 0.6)	2.7 (SE 0.7)
	AS	Placebo	–	12	107	−0.8 (SE 0.2)	–	0.1 (SE 0.1)	−1.3 (SE 0.5)	1.6 (SE 0.8)	2.4 (SE 1)
RAPID-axSpA 2014[64]	AS	Certolizumab	200 mg every 2 weeks	12	65	−2.5 (SE 0.3)	−1.7 (SE 0.3)	−0.6 (SE 0.1)	–	8.73 (SD 7.63)	2.42 (SD 9.08)
	AS	Certolizumab	400 mg every 4 weeks	12	56	−2.4 (SE 0.3)	−1.7 (SE 0.3)	−0.3 (SE 0.2)	–	7.6 (SD 7.65)	2.22 (SD 10.44)
	AS	Placebo	–	12	57	−1 (SE 0.3)	−0.6 (SE 0.3)	−0.2 (SE 0.1)	–	2.56 (SD 5.67)	1.07 (SD 10.92)

Trial	Population	Treatment arm	Dose	Time point (weeks)	Patients	Mean change from baseline					
						BASDAI	BASFI	BASMI	MASES	SF-36 PCS	SF-36 MCS
RAPID-axSpA 2014[64]	Nr-axSpA	Certolizumab	200 mg	12	46	−3.3 (SE 0.4)	−2.3 (SE 0.4)	−0.6 (SE 0.2)	–	9.56 (SD 9.46)	4.59 (SD 9.7)
	Nr-axSpA	Certolizumab	400 mg	12	51	−3.4 (SE 0.4)	−2.3 (SE 0.4)	−0.5 (SE 0.2)	–	8.72 (SD 8.84)	6.12 (SD 10.94)
	Nr-axSpA	Placebo	–	12	50	−1.5 (SE 0.4)	−0.4 (SE 0.4)	0 (SE 0.1)	–	2.13 (SD 7.47)	1.39 (SD 10.24)
Barkham 2010[71]	AS	Etanercept	25 mg twice weekly	12	20	−1.97	−1.35	–	–	–	–
	AS	Placebo	–	12	20	−0.1	0.21	–	–	–	–
Davis 2003[72]	AS	Etanercept	25 mg twice weekly	12	138	−2.36 (SE 0.19)	−1.67 (SE 0.2)	–	–	–	–
	AS	Placebo	–	12	139	−0.45 (SE 0.18)	−0.33 (SE 0.21)	–	–	–	–
	AS	Etanercept	25 mg twice weekly	24	138	–	–	–	–	–	–
	AS	Placebo	–	24	139	–	–	–	–	–	–
Dougados 2011[74]	AS	Etanercept	50 mg weekly	12	39	−2.6 (SD 2)	−2.2 (SD 1.8)	−0.57 (SD 0.65)	–	–	–
	AS	Placebo	–	12	43	−1.4 (SD 2)	−1 (SD 1.8)	−0.2 (SD 0.65)	–	–	–
Dougados 2014[75]	AS	Etanercept	50 mg weekly	12	106	−2 (SE 0.3)	−1.4 (SE 0.2)	−0.3 (SE 0.2)	–	–	–
	AS	Placebo	–	12	109	−1.3 (SE 0.3)	−0.8 (SE 0.2)	−0.3 (SE 0.1)	–	–	–
	Nr-axSpA	Etanercept	50 mg weekly	12	94	–	–	–	–	–	–
	Nr-axSpA	Placebo	–	12	95	–	–	–	–	–	–

continued

TABLE 112 Continuous outcomes: change from baseline results (*continued*)

Trial	Population	Treatment arm	Dose	Time point (weeks)	Patients	Mean change from baseline					
						BASDAI	BASFI	BASMI	MASES	SF-36 PCS	SF-36 MCS
Gorman 2002[79]	AS	Etanercept	25 mg twice a week	16	20	–	–	–	–	–	–
	AS	Placebo	–	16	20	–	–	–	–	–	–
Calin 2004[83]	AS	Etanercept	25 mg twice weekly	12	45	–2.72 (SE 0.34)	–2.06 (SE 0.33)	–	–	–	–
	AS	Placebo	–	12	39	–0.85 (SE 0.35)	–0.33 (SE 0.31)	–	–	–	–
van der Heijde 2006[86]	AS	Etanercept	25 mg twice weekly	12	150	–	–	–	–	–	–
	AS	Etanercept	50 mg weekly	12	155	–	–	–	–	–	–
	AS	Placebo	–	12	51	–	–	–	–	–	–
Giardina 2010[88]	AS	Etanercept	50 mg weekly	12	25	–	–	–	–	–	–
	AS	Infliximab	5 mg/kg (0, 2 and ≥ 6 weeks)	12	25	–	–	–	–	–	–
GO-RAISE 2008[90]	AS	Golimumab	50 mg (two every 4 weeks)	14	138	–	–1.4 (IQR –3.1 to –0.1)	0 (IQR –1 to 0)	–0.5 (SD 2.6)	7.3 (IQR 1.5 to 15.3)	1.5 (IQR –2.2 to 7.8)
	AS	Golimumab	100 mg (two every 4 weeks)	14	140	–	–1.5 (IQR –3.0 to –0.1)	0 (IQR –1 to 0)	–1.3 (SD 3.11)	8.4 (IQR 2.3 to 14.1)	3.7 (IQR –3.2 to 12.1)
	AS	Placebo	–	14	78	–	0.1 (IQR –1.1 to 1.1)	0 (SD –1 to 0)	–0.2 (SD 2.99)	2.4 (IQR –1.4 to 7.8)	0.1 (IQR –4.3 to 5.3)
Bao 2012[96]	AS	Golimumab	50 mg every 4 weeks	14		–	–1.26 (SD 2.57)	–0.42 (SD 0.91)	–	6.25 (SD 7.95)	3.86 (SD 8.92)
	AS	Placebo	–	14		–	0.11 (SD 2.1)	–0.19 (SD 0.72)	–	1.59 (SD 6.12)	0.82 (SD 9.44)

Trial	Population	Treatment arm	Dose	Time point (weeks)	Patients	Mean change from baseline					
						BASDAI	BASFI	BASMI	MASES	SF-36 PCS	SF-36 MCS
Tam 2014[97]	AS	Golimumab	50 mg monthly	26	NR	−1.82 (SD 1.64)	−0.13 (SD 0.25)[b]	−1 (IQR −2 to 0)	–	–	–
	AS	Placebo	–	26	NR	−0.66 (SD 1.24)	0.17 (SD 0.72)[b]	0 (IQR −1 to 0)	–	–	–
Barkham 2009[50]	Nr-axSpA	Infliximab	5 mg/kg (0, 2 and ≥ 6 weeks)	16	20	−3.41 (SD 2.53)	−2.7 (SD 2.36)	–	–	–	–
	Nr-axSpA	Placebo	–	16	20	−0.75 (SD 2.42)	−0.47 (SD 2.25)	–	–	–	–
Braun 2002[98]	AS	Infliximab	5 mg/kg (0, 2 and ≥ 6 weeks)	12	34	−3.2	−2.1	–	–	–	–
	AS	Placebo	–	12	35	−0.6	−0.1	–	–	–	–
Marzo-Ortega 2005[100]	AS	Infliximab + methotrexate	5 mg/kg[c]	10	28	−3.11 (SD 2.23)	–	–	–	–	–
	AS	Placebo + methotrexate	–[d]	10	14	−1.38 (SD 2.11)	–	–	–	–	–
	AS	Infliximab + methotrexate	5 mg/kg[c]	30	28	−1.85 (SD 2.84)	–	–	–	–	–
	AS	Placebo + methotrexate	–[d]	30	14	−0.84 (SD 1.8)	–	–	–	–	–
Van den Bosch 2002[101]	AS	Infliximab	5 mg/kg (0, 2 and 6 weeks)	12	9	–	–	–	–	–	–
	AS	Placebo	–	12	12	–	–	–	–	–	–
ASSERT 2005[102]	AS	Infliximab	5 mg/kg (0, 2 and ≥ 6 weeks)	24	201	−2.9 (IQR −4.9 to −0.9)	−1.7 (IQR −3.6 to −0.6)	−1 (IQR −1 to 0)	–	10.2 (IQR 3.9 to 17.1)	2.7 (IQR −2.9 to 8.8)
	AS	Placebo	–	24	78	−0.4 (IQR −1.4 to 0.7)	0 (IQR −1 to 1)	0 (IQR −1 to 0)	–	0.8 (IQR −1.9 to 6)	2 (IQR −2.6 to 7.5)

continued

TABLE 112 Continuous outcomes: change from baseline results (*continued*)

Trial	Population	Treatment arm	Dose	Time point (weeks)	Patients	Mean change from baseline					
						BASDAI	BASFI	BASMI	MASES	SF-36 PCS	SF-36 MCS
PLANETAS 2013[110]	AS	CT-P13	5 mg/kg	14	125	–2.91 (SD 2.17)	–2.51 (SD 2.14)	–0.7 (SD 1.2)	–	–	–
	AS	Infliximab	5 mg/kg	14	125	–2.77 (SD 2.08)	–2.47 (SD 2.18)	–0.7 (SD 1.4)	–	–	–
	AS	CT-P13	5 mg/kg	30	125	–3.04 (SD 2.23)	–2.6 (SD 2.19)	–1 (SD 1.4)	–	7.6	6.5
	AS	Infliximab	5 mg/kg	30	125	–2.71 (SD 2.24)	–2.54 (SD 2.17)	–0.9 (SD 1.4)	–	8.5	5.2
	AS	CT-P13	5 mg/kg	54	125	–	–	–	–	–	–
	AS	Infliximab	5 mg/kg	54	125	–	–	–	–	–	–

IQR, interquartile range; NR, not reported.
a Licensed population.
b These values are uncertain because of poor reporting.
c Infliximab 5 mg/kg (infusion at weeks 0, 2, 6, 14 and 22 weeks) + methotrexate oral 7.5 mg with folic acid (5 mg twice a week which increased to 10 mg a week).
d Methotrexate oral 7.5 mg with folic acid (5 mg twice a week which increased to 10 mg a week).

TABLE 113 Binary response outcomes results

Trial	Population	Treatment arm	Dose	Time point (weeks)	Patients	Number of responders (%)				
						ASAS 20	ASAS 40	ASAS 50	ASAS 70	BASDAI 50
Haibel 2008[52]	Nr-axSpA	Adalimumab	40 mg every 2 weeks	12	22	15 (68)	12 (55)	–	–	11 (50)
	Nr-axSpA	placebo	–	12	24	6 (25)	3 (13)	–	–	5 (21)
Hu 2012[55]	AS	Adalimumab	40 mg every 2 weeks	12	26	–	–	–	–	–
	AS	Placebo	–	12	20	–	–	–	–	–
Huang 2014[56]	AS	Adalimumab	40 mg every 2 weeks	12	229	154 (67)	102 (45)	–	–	114 (50)
	AS	Placebo	–	12	115	35 (30)	11 (10)	–	–	19 (17)
Lambert 2007[57]	AS	Adalimumab	40 mg every 2 weeks	12	38	18 (47)	–	12 (32)	–	–
	AS	Placebo	–	12	44	12 (27)	–	5 (11)	–	–
[a]ABILITY-1 2013[58]	Nr-axSpA	Adalimumab	40 mg every 2 weeks	12	69	41 (59)	28 (41)	24 (35)	13 (19)	27 (39)
	Nr-axSpA	Placebo	–	12	73	23 (32)	10 (14)	6 (8)	3 (4)	10 (14)
ATLAS 2006[61]	AS	Adalimumab	40 mg every 2 weeks	12	208	121 (58)	83 (40)	–	–	94 (45)
	AS	Placebo	–	12	107	22 (21)	14 (13)	–	–	17 (16)
RAPID-axSpA 2014[64]	AS	Certolizumab pegol	200 mg every 2 weeks	12	65	37 (57)	26 (40)	–	–	27 (42)
	AS	Certolizumab pegol	400 mg every 4 weeks	12	56	36 (64)	28 (50)	–	–	23 (41)
	AS	Placebo	–	12	57	21 (37)	11 (19)	–	–	6 (11)
	Nr-axSpA	Certolizumab pegol	200 mg	12	46	27 (59)	22 (48)	–	–	23 (50)
	Nr-axSpA	Certolizumab pegol	400 mg	12	51	32 (63)	24 (47)	–	–	24 (47)
	Nr-axSpA	Placebo	–	12	50	20 (40)	8 (16)	–	–	8 (16)
Barkham 2010[71]	AS	Etanercept	25 mg twice weekly	12	20	–	4 (20)	–	–	7 (35)
	AS	Placebo	–	12	20	–	0 (0)	–	–	1 (5)

continued

TABLE 113 Binary response outcomes results (*continued*)

Trial	Population	Treatment arm	Dose	Time point (weeks)	Patients	Number of responders (%)				
						ASAS 20	ASAS 40	ASAS 50	ASAS 70	BASDAI 50
Davis 2003[72]	AS	Etanercept	25 mg twice weekly	12	138	82 (59)	–	62 (45)	40 (29)	–
	AS	Placebo	–	12	139	39 (28)	–	18 (13)	10 (7)	–
	AS	Etanercept	25 mg twice weekly	24	138	78 (57)	–	58 (42)	–	–
	AS	Placebo	–	24	139	31 (22)	–	14 (10)	–	–
Dougados 2011[74]	AS	Etanercept	50 mg weekly	12	39	25 (64)	17 (44)	15 (38)	10 (26)	18 (46)
	AS	Placebo	–	12	43	14 (33)	10 (23)	6 (14)	4 (9)	10 (23)
Dougados 2014[76]	AS	Etanercept	50 mg weekly	12	106	55 (52)	34 (32)	–	–	46 (43)
	AS	Placebo	–	12	109	39 (36)	17 (16)	–	–	26 (24)
	Nr-axSpA	Etanercept	50 mg weekly	12	94	–	33 (35)	–	–	–
	Nr-axSpA	Placebo	–	12	95	–	16 (17)	–	–	–
Gorman 2002[79]	AS	Etanercept	25 mg twice a week	16	20	16 (80)	–	–	–	–
	AS	Placebo	–	16	20	6 (30)	–	–	–	–
Calin 2004[83]	AS	Etanercept	25 mg twice weekly	12	45	26 (58)	–	–	–	–
	AS	Placebo	–	12	39	9 (23)	–	–	–	–
van der Heijde 2006[86]	AS	Etanercept	25 mg twice weekly	12	150	107 (71)	80 (53)	–	–	87 (58)
	AS	Etanercept	50 mg weekly	12	155	115 (74)	90 (58)	–	–	93 (60)
	AS	Placebo	–	12	51	19 (37)	11 (22)	–	–	10 (20)
Giardina 2010[88]	AS	Etanercept	50 mg weekly	12	25	15 (60)	11 (44)	–	–	–
	AS	Infliximab	5 mg/kg (0, 2 and ≥ 6 weeks)	12	25	19 (76)	14 (56)	–	–	–

Trial	Population	Treatment arm	Dose	Time point (weeks)	Patients	Number of responders (%)				
						ASAS 20	ASAS 40	ASAS 50	ASAS 70	BASDAI 50
GO-RAISE 2008[90]	AS	Golimumab	50 mg (two every 4 weeks)	14	138	82 (59)	62 (45)	–	–	61 (44)
	AS	Golimumab	100 mg (two every 4 weeks)	14	140	84 (60)	69 (49)	–	–	56 (40)
	AS	Placebo	–	14	78	17 (22)	12 (15)	–	–	12 (15)
Bao 2012[98]	AS	Golimumab	50 mg every 4 weeks	14	108	53 (49)	38 (35)	–	–	37 (34)
	AS	Placebo	–	14	105	26 (25)	10 (10)	–	–	5 (5)
Tam 2014[97]	AS	Golimumab	50 mg monthly	26	20	11 (55)	–	–	–	–
	AS	Placebo	–	26	21	3 (14)	–	–	–	–
Barkham 2009[50]	Nr-axSpA	Infliximab	5 mg/kg (0, 2 and ≥ 6 weeks)	16	20	–	11 (55)	–	–	–
	Nr-axSpA	Placebo	–	16	20	–	3 (15)	–	–	–
Braun 2002[96]	AS	Infliximab	5 mg/kg (0, 2 and 6 weeks)	12	34	23 (68)	–	16 (47)	–	18 (53)
	AS	Placebo	–	12	35	10 (29)	–	2 (6)	–	3 (9)
Marzo-Ortega 2005[100]	AS	Infliximab + methotrexate	5 mg/kg[b]	10	28	20 (71)	–	–	–	–
	AS	Placebo + methotrexate	–[c]	10	14	4 (29)	–	–	–	–
	AS	Infliximab + methotrexate	5 mg/kg[b]	30	28	14 (50)	–	–	–	–
	AS	Placebo + methotrexate	–[c]	30	14	3 (21)	–	–	–	–

continued

TABLE 113 Binary response outcomes results (*continued*)

Trial	Population	Treatment arm	Dose	Time point (weeks)	Patients	Number of responders (%)				
						ASAS 20	ASAS 40	ASAS 50	ASAS 70	BASDAI 50
Van den Bosch 2002[101]	AS	Infliximab	5 mg/kg (0, 2 and 6 weeks)	12	9	–	–	–	–	–
	AS	Placebo	–	12	12	–	–	–	–	–
ASSERT 2005[102]	AS	Infliximab	5 mg/kg (0, 2 and ≥ 6 weeks)	24	201	123 (61)	93 (46)	–	–	–
	AS	Placebo	–	24	78	15 (19)	9 (12)	–	–	–
PLANETAS 2013[110]	AS	CT-P13	5 mg/kg	14	125	72 (58)	48 (38)	–	–	–
	AS	Infliximab	5 mg/kg	14	125	79 (63)	56 (45)	–	–	–
	AS	CT-P13	5 mg/kg	30	125	79 (63)	58 (46)	–	–	–
	AS	Infliximab	5 mg/kg	30	125	84 (67)	55 (44)	–	–	–
	AS	CT-P13	5 mg/kg	54	125	71 (57)	51 (41)	–	–	–
	AS	Infliximab	5 mg/kg	54	125	75 (60)	46 (37)	–	–	–

NR, not reported.
a Licensed population.
b Infliximab 5 mg/kg (infusion at 0, 2, 6, 14 and 22 weeks) + methotrexate oral 7.5 mg with folic acid (5 mg twice a week which increased to 10 mg a week).
c Methotrexate oral 7.5 mg with folic acid (5 mg twice a week which increased to 10 mg a week).

Appendix 5 Bath Ankylosing Spondylitis Disease Activity Index and Bath Ankylosing Spondylitis Functional Index scores conditional on Bath Ankylosing Spondylitis Disease Activity Index response

In this appendix we use the results from the extended synthesis model (*Appendices 12* and *13*) to evaluate the conditional scores by simulating BASDAI and BASFI scores for two equivalent cohorts of patients one treated with an anti-TNF and the other with conventional therapy.

Description of methods

From the inferences obtained using the synthesis model in *Appendix 13* it is possible to derive the conditional change score in responders and non-responders using simulation. Whereas the synthesis focuses on the pooling of mean estimates of change scores and proportion of responders to BASDAI 50, to derive conditional mean scores there is the need to consider the distributions at the individual patient level. Hence, conditional scores could not directly be derived from the synthesis but instead were derived through a simulation procedure based on the assumptions and results of the synthesis model.

The steps undertaken within the simulation procedure were:

1. Simulate baseline BASDAI scores, $x^{BASDAI*}$, from beliefs over its distribution, $X \sim N(\nu, \sigma)$
2. Simulate $y^*_{k=1}$ from beliefs over the mean (μ) of this quantity considering correlation with $x^{BASDAI*}$:

$$Y_{k=1|X=x} \sim N(\mu + \rho(x-\mu), (1-\rho_{pla}{}^2)\sigma^2). \quad (9)$$

3. Simulate $y^*_{k=2}$ (where $k=2$ represents treatment with anti-TNF) by considering:

$$Y_{k\neq 1|X=x} \sim N(\mu + d + \rho(x-\mu), (1-\rho_{anti-TNF}{}^2)\sigma^2). \quad (10)$$

4. Calculate final score for placebo and treatment separately, by summing $x^{BASDAI*}_{final} = y^{BASDAI*}_k + x^{BASDAI*}$
5. Compute response variables for both groups as $y^{BASDAI*}_k + x^*/2 < 0$

Repeat steps 1 to 4 until the desired sample size is achieved, and calculate conditional scores based on response variable and change in scores.

To evaluate BASFI conditional on BASDAI scores one needs to firstly consider we have available information on the BASFI scores at baseline: $X^{BASFI} \sim N(\nu^{BASFI}, (se^{BASFI})^2)$, and also on correlation with BASDAI scores, φ (at individual level). By considering $x^{BASDAI*}$, one can:

6. Simulate from the distribution of the baseline BASFI score conditional on the baseline BASDAI score being x^*:

$$x^{BASFI}_{|x^{BASDAI}=x} \sim N(\nu^{BASFI} + \frac{\sigma_{BASFI}}{\sigma_{BASDAI}}\varphi(x^{BASDAI*} - \nu^{BASDAI}), (1-\varphi^2)\sigma^2_{BASFI}). \quad (11)$$

Note the correlation parameter, φ which represents the individual-level correlation between baseline BASFI and BASDAI scores.

7. Simulate the change from baseline on BASFI for placebo $y_{k=1}^{BASFI*}$ from belief over this quantity, consider this to be correlated with the $y_{k=1}^{*}$ simulated for BASDAI (use correlation parameter estimated within the synthesis):

$$y_{k=1}^{BASFI}{}_{|y_{k=1}^{BASDAI}=\mu*} \sim N(\mu^{BASFI} + \frac{\sigma_{BASFI}}{\sigma_{BASDAI}}\rho_m(\mu^{BASDAI*} - \mu^{BASDAI}), (1-\rho_m{}^2)\sigma^2_{BASFI}). \quad (12)$$

8. Simulate the change from baseline for anti-TNF treatment:

$$\theta^{BASFI}{}_{|y^{BASDAI}=\mu*} \sim N(\mu^{BASFI} + d + \frac{\sigma_{BASFI}}{\sigma_{BASDAI}}\rho_m(x_{final}^{BASDAI*} - mean.x_{final}^{BASDAI*}), (1-\rho_m{}^2)\sigma^2_{BASFI}). \quad (13)$$

Note that d represents the mean of the predictive distribution from the synthesis model.

We used a simulation sample size of 10,000 patients. Given results depend on the baseline distributions of BASDAI and BASFI and on the change scores from baseline for placebo, we used the averages across trials (weighted by the number of patients in each trial) in AS. Baseline BASDAI scores were thus assumed normally distributed with mean 6.11 and SD of 1.56; change from baseline for placebo was simulated from a normal distribution with mean –0.61 and SD of 1.44. For BASFI, the baseline was assumed to have a mean of 5.27 and a SD of 1.79 and change from baseline for placebo a mean of –0.19 and a SD of 0.22. The correlation between baseline BASFI and BASDAI scores was valued at 0.7 (φ). Average scores from the RAPID-axSpA[64] trial for certolizumab were used for the nr-AxSpA analysis.

Results

Results of the prediction of conditional scores using the synthesis model in the AS population are presented in *Table 114* and for the nr-AxSpA population in *Table 115*.

TABLE 114 Conditional scores predicted for the AS population using the synthesis model

	BASDAI		BASFI	
	Control	Treat	Control	Treat
Scenario 1				
% responders to BASDAI 50	0.10	0.42	–	–
Change in score				
Responders	–2.70	–3.86	–1.41	–3.02
Non-responders	–0.45	–1.73	–0.17	–0.63
All	–0.66	–2.63	–0.29	–1.64
Baseline				
Responders	3.83	4.76	3.42	4.17
Non-responders	6.31	7.03	5.43	6.02
All	6.08	6.08	5.24	5.24

TABLE 115 Conditional scores predicted for the nr-AxSpA population using results and assumptions of the synthesis model

	BASDAI		BASFI	
	Control	Treat	Control	Treat
Scenario 1				
% responders to BASDAI 50	AiC information has been removed	AiC information has been removed	–	–
Change in score	*AiC information has been removed*[a]	*AiC information has been removed*[a]	–	–
Responders	AiC information has been removed	AiC information has been removed	AiC information has been removed	AiC information has been removed
Non-responders	AiC information has been removed	AiC information has been removed	AiC information has been removed	AiC information has been removed
All	AiC information has been removed	AiC information has been removed	AiC information has been removed	AiC information has been removed
Baseline	*AiC information has been removed*[a]	*AiC information has been removed*[a]	–	–
Responders	AiC information has been removed	AiC information has been removed	AiC information has been removed	AiC information has been removed
Non-responders	AiC information has been removed	AiC information has been removed	AiC information has been removed	AiC information has been removed
All	AiC information has been removed	AiC information has been removed	AiC information has been removed	AiC information has been removed
Scenario 2				
% responders to BASDAI 50	AiC information has been removed	AiC information has been removed	–	–
Change in score	*AiC information has been removed*[a]	*AiC information has been removed*[a]	–	–
Responders	AiC information has been removed	AiC information has been removed	AiC information has been removed	AiC information has been removed
Non-responders	AiC information has been removed	AiC information has been removed	AiC information has been removed	AiC information has been removed
All	AiC information has been removed	AiC information has been removed	AiC information has been removed	AiC information has been removed
Baseline	*AiC information has been removed*[a]	*AiC information has been removed*[a]	–	–
Responders	AiC information has been removed	AiC information has been removed	AiC information has been removed	AiC information has been removed
Non-responders	AiC information has been removed	AiC information has been removed	AiC information has been removed	AiC information has been removed
All	AiC information has been removed	AiC information has been removed	AiC information has been removed	AiC information has been removed

a Based on a BASDAI baseline score (AiC information has been removed) and a placebo change in BASDAI score of (AiC information has been removed) and a BASFI baseline score of (AiC information has been removed) and a placebo change in BASFI score of (AiC information has been removed) which represent the results seen in the certolizumab trial (RAPID-axSpA[64]).

Appendix 6 Bath Ankylosing Spondylitis Disease Activity Index and Bath Ankylosing Spondylitis Functional Index scores conditional on response data

TABLE 116 Baseline BASDAI score according to conditional on response at week 12 (or nearest time point)

Anti-TNF Population (trial)	Response criterion	Treatment	Response	*n*	Mean	SD
Adalimumab AS (ATLAS[61])	ASAS 20	Adalimumab 40 mg	Non-responder	83	6.23	1.929
			Responder	121	6.27	1.542
		Placebo	Non-responder	82	6.29	1.712
			Responder	22	6.64	1.468
	ASAS 40	Adalimumab 40 mg	Non-responder	119	6.21	1.802
			Responder	85	6.32	1.568
		Placebo	Non-responder	89	6.37	1.714
			Responder	15	6.34	1.362
	BASDAI 50	Adalimumab 40 mg	Non-responder	119	6.21	1.802
			Responder	85	6.32	1.568
		Placebo	Non-responder	89	6.37	1.714
			Responder	15	6.34	1.362
Golimumab AS (GO-RAISE[90])	ASAS 20	Golimumab 50 mg	Non-responder	56	6.51	1.687
			Responder	82	6.49	1.494
		Placebo	Non-responder	61	6.65	1.622
			Responder	17	6.46	1.120
	ASAS 40	Golimumab 50 mg	Non-responder	76	6.54	1.680
			Responder	62	6.45	1.433
		Placebo	Non-responder	66	6.65	1.579
			Responder	12	6.41	1.194
	BASDAI 50	Golimumab 50 mg	Non-responder	72	6.69	1.523
			Responder	61	6.25	1.638
		Placebo	Non-responder	66	6.63	1.581
			Responder	12	6.51	1.194

continued

TABLE 116 Baseline BASDAI score according to conditional on response at week 12 (or nearest time point) (*continued*)

Anti-TNF Population (trial)	Response criterion	Treatment	Response	*n*	Mean	SD
Etanercept AS (314-EU[167] study)	ASAS 20	Etanercept 25 mg twice weekly	Non-responder	43	CiC information has been removed	CiC information has been removed
			Responder	107	CiC information has been removed	CiC information has been removed
		Etanercept 50 mg once weekly	Non-responder	40	CiC information has been removed	CiC information has been removed
			Responder	115	CiC information has been removed	CiC information has been removed
		Placebo	Non-responder	32	CiC information has been removed	CiC information has been removed
			Responder	19	CiC information has been removed	CiC information has been removed
	ASAS 40	Etanercept 25 mg twice weekly	Non-responder	70	CiC information has been removed	CiC information has been removed
			Responder	80	CiC information has been removed	CiC information has been removed
		Etanercept 50 mg once weekly	Non-responder	65	CiC information has been removed	CiC information has been removed
			Responder	90	CiC information has been removed	CiC information has been removed
		Placebo	Non-responder	40	CiC information has been removed	CiC information has been removed
			Responder	11	CiC information has been removed	CiC information has been removed
	BASDAI 50	Etanercept 25 mg twice weekly	Non-responder	63	CiC information has been removed	CiC information has been removed
			Responder	87	CiC information has been removed	CiC information has been removed
		Etanercept 50 mg once weekly	Non-responder	62	CiC information has been removed	CiC information has been removed
			Responder	93	CiC information has been removed	CiC information has been removed
		Placebo	Non-responder	41	CiC information has been removed	CiC information has been removed
			Responder	10	CiC information has been removed	CiC information has been removed
Adalimumab ABILITY-1[58] (nr-AxSpA sub-population with a positive MRI and/or elevated CRP level)	ASAS 20	Adalimumab 40 mg	Non-responder	27	6.31	1.66
			Responder	41	6.46	1.49
		Placebo	Non-responder	46	6.49	1.37
			Responder	23	6.05	1.77
	ASAS 40	Adalimumab 40 mg	Non-responder	40	6.60	1.63
			Responder	28	6.13	1.41
		Placebo	Non-responder	59	6.41	1.55
			Responder	10	5.93	1.27
	BASDAI 50	Adalimumab 40 mg	Non-responder	41	6.53	1.69
			Responder	27	6.21	1.31
		Placebo	Non-responder	59	6.46	1.52
			Responder	10	5.64	1.34

TABLE 116 Baseline BASDAI score according to conditional on response at week 12 (or nearest time point) (*continued*)

Anti-TNF Population (trial)	Response criterion	Treatment	Response	*n*	Mean	SD
Etanercept (1031 study,[165] nr-AxSpA)	ASAS 20	Etanercept 50 mg	Non-responder	CiC information has been removed	CiC information has been removed	CiC information has been removed
			Responder	CiC information has been removed	CiC information has been removed	CiC information has been removed
		Placebo	Non-responder	CiC information has been removed	CiC information has been removed	CiC information has been removed
			Responder	CiC information has been removed	CiC information has been removed	CiC information has been removed
	ASAS 40	Etanercept 50 mg	Non-responder	CiC information has been removed	CiC information has been removed	CiC information has been removed
			Responder	CiC information has been removed	CiC information has been removed	CiC information has been removed
		Placebo	Non-responder	CiC information has been removed	CiC information has been removed	CiC information has been removed
			Responder	CiC information has been removed	CiC information has been removed	CiC information has been removed
	BASDAI 50	Etanercept 50 mg	Non-responder	CiC information has been removed	CiC information has been removed	CiC information has been removed
			Responder	CiC information has been removed	CiC information has been removed	CiC information has been removed
		Placebo	Non-responder	CiC information has been removed	CiC information has been removed	CiC information has been removed
			Responder	CiC information has been removed	CiC information has been removed	CiC information has been removed

TABLE 117 Baseline BASFI according to conditional on response at week 12 (or nearest time point)

Anti-TNF (trial)	Response criterion	Treatment	Response	*n*	Mean	SD
Adalimumab AS (ATLAS[61])	ASAS 20	Adalimumab 40 mg	Non-responder	83	53.03	23.881
			Responder	121	51.38	20.843
		Placebo	Non-responder	82	57.96	23.089
			Responder	22	52.27	16.661
	ASAS 40	Adalimumab 40 mg	Non-responder	119	53.05	22.864
			Responder	85	50.65	21.005
		Placebo	Non-responder	89	57.05	22.954
			Responder	15	54.98	14.996
	BASDAI 50	Adalimumab 40 mg	Non-responder	110	57.79	21.015
			Responder	94	45.34	21.514
		Placebo	Non-responder	87	59.06	21.989
			Responder	17	44.98	17.979
Golimumab AS (GO-RAISE[90])	ASAS 20	Golimumab 50 mg	Non-responder	56	5.35	2.530
			Responder	82	4.76	2.249
		Placebo	Non-responder	59	5.38	2.260
			Responder	17	4.13	1.985
	ASAS 40	Golimumab 50 mg	Non-responder	76	5.33	2.488
			Responder	62	4.60	2.184
		Placebo	Non-responder	64	5.33	2.247
			Responder	12	3.88	1.932
	BASDAI 50	Golimumab 50 mg	Non-responder	72	5.48	2.412
			Responder	61	4.45	2.288
		Placebo	Non-responder	64	5.39	2.179
			Responder	12	3.56	2.070
Etanercept AS (314-EU[167])	ASAS 20	Etanercept 25 mg twice weekly	Non-responder	43	CiC information has been removed	CiC information has been removed
			Responder	107	CiC information has been removed	CiC information has been removed
		Etanercept 50 mg once weekly	Non-responder	40	CiC information has been removed	CiC information has been removed
			Responder	115	CiC information has been removed	CiC information has been removed
		Placebo	Non-responder	32	CiC information has been removed	CiC information has been removed
			Responder	19	CiC information has been removed	CiC information has been removed

TABLE 117 Baseline BASFI according to conditional on response at week 12 (or nearest time point) (*continued*)

Anti-TNF (trial)	Response criterion	Treatment	Response	*n*	Mean	SD
	ASAS 40	Etanercept 25 mg twice weekly	Non-responder	70	CiC information has been removed	CiC information has been removed
			Responder	80	CiC information has been removed	CiC information has been removed
		Etanercept 50 mg once weekly	Non-responder	65	CiC information has been removed	CiC information has been removed
			Responder	90	CiC information has been removed	CiC information has been removed
		Placebo	Non-responder	40	CiC information has been removed	CiC information has been removed
			Responder	11	CiC information has been removed	CiC information has been removed
	BASDAI 50	Etanercept 25 mg twice weekly	Non-responder	63	CiC information has been removed	CiC information has been removed
			Responder	87	CiC information has been removed	CiC information has been removed
		Etanercept 50 mg once weekly	Non-responder	62	CiC information has been removed	CiC information has been removed
			Responder	93	CiC information has been removed	CiC information has been removed
		Placebo	Non-responder	41	CiC information has been removed	CiC information has been removed
			Responder	10	CiC information has been removed	CiC information has been removed
Adalimumab nr-AxSpA, ABILITY-1[58] (subpopulation with a positive MRI and/or elevated CRP level)	ASAS 20	Adalimumab 40 mg	Non-responder	27	45.17	22.07
			Responder	40	43.05	19.31
		Placebo	Non-responder	47	48.07	22.99
			Responder	23	47.91	23.75
	ASAS 40	Adalimumab 40 mg	Non-responder	40	47.61	22.60
			Responder	27	39.09	15.41
		Placebo	Non-responder	60	48.26	23.46
			Responder	10	46.54	21.67
	BASDAI 50	Adalimumab 40 mg	Non-responder	40	49.71	20.05
			Responder	27	35.97	18.12
		Placebo	Non-responder	59	49.06	23.25
			Responder	10	43.66	23.07

continued

TABLE 117 Baseline BASFI according to conditional on response at week 12 (or nearest time point) (*continued*)

Anti-TNF (trial)	Response criterion	Treatment	Response	*n*	Mean	SD
Etanercept (1031 study[166] nr-AxSpA)	ASAS 20	Etanercept 50 mg	Non-responder	CiC information has been removed	CiC information has been removed	CiC information has been removed
			Responder	CiC information has been removed	CiC information has been removed	CiC information has been removed
		Placebo	Non-responder	CiC information has been removed	CiC information has been removed	CiC information has been removed
			Responder	CiC information has been removed	CiC information has been removed	CiC information has been removed
	ASAS 40	Etanercept 50 mg	Non-responder	CiC information has been removed	CiC information has been removed	CiC information has been removed
			Responder	CiC information has been removed	CiC information has been removed	CiC information has been removed
		Placebo	Non-responder	CiC information has been removed	CiC information has been removed	CiC information has been removed
			Responder	CiC information has been removed	CiC information has been removed	CiC information has been removed
	BASDAI 50	Etanercept 50 mg	Non-responder	CiC information has been removed	CiC information has been removed	CiC information has been removed
			Responder	CiC information has been removed	CiC information has been removed	CiC information has been removed
		Placebo	Non-responder	CiC information has been removed	CiC information has been removed	CiC information has been removed
			Responder	CiC information has been removed	CiC information has been removed	CiC information has been removed

Summary

The mean baseline BASDAI and BASFI are presented by treatment response at week 12 (or 14 for golimumab) for three of the five anti-TNFs. This reveals that in patients with AS and patients with nr-AxSpA, on average baseline BASDAI score does not differ greatly between responders and non-responders either to placebo or to active anti-TNF therapy. In patients with AS or nr-AxSpA from the trials of adalimumab (ATLAS[61] and ABILITY-1[58]) and golimumab (GO-RAISE[90]) on average baseline BASFI score was higher in non-responders compared with responders. However, this was not seen in the etanercept trials.

Appendix 7 Relative effects of anti-tumour necrosis factors

Ankylosing spondylitis population

In the following tables, the intervention is stated in the top row and the comparator is in the left-hand column, which is reverse to normal.

TABLE 118 Relative effects relative risk BASDAI 50 AS

	ADA		CER		ETA		GOL		INF	
Anti-TNF	Mean difference	95% CrI	Mean difference	95% CrI	Mean difference	95% CrI	Mean difference	95% CrI	Mean difference	95% CrI
ADA	–	–	1.15	0.61 to 1.86	1.01	0.65 to 1.50	1.13	0.75 to 1.66	1.55	0.74 to 2.50
CER	0.87	0.54 to 1.63	–	–	0.88	0.50 to 1.70	0.99	0.58 to 1.89	1.34	0.61 to 2.74
ETA	0.99	0.67 to 1.53	1.14	0.59 to 1.98	–	–	1.12	0.71 to 1.78	1.53	0.72 to 2.66
GOL	0.88	0.60 to 1.33	1.01	0.53 to 1.74	0.89	0.56 to 1.40	–	–	1.37	0.65 to 2.30
INF	0.65	0.40 to 1.35	0.74	0.37 to 1.65	0.65	0.38 to 1.38	0.73	0.43 to 1.54	–	–

ADA, adalimumab; CER, certolizumab; ETA, etanercept; GOL, golimumab; INF, infliximab.

TABLE 119 Relative effects OR BASDAI 50 AS

	ADA		CER		ETA		GOL		INF	
Anti-TNF	Mean difference	95% CrI	Mean difference	95% CrI	Mean difference	95% CrI	Mean difference	95% CrI	Mean difference	95% CrI
ADA	–	–	1.28	0.47 to 3.48	1.01	0.51 to 2.02	1.25	0.62 to 2.48	2.58	0.62 to 10.60
CER	0.78	0.29 to 2.14	–	–	0.79	0.27 to 2.32	0.98	0.33 to 2.89	2.02	0.39 to 10.33
ETA	0.99	0.50 to 1.97	1.26	0.43 to 3.71	–	–	1.23	0.56 to 2.73	2.55	0.58 to 11.01
GOL	0.80	0.40 to 1.61	1.02	0.35 to 3.03	0.81	0.37 to 1.80	–	–	2.06	0.47 to 8.91
INF	0.39	0.09 to 1.62	0.50	0.10 to 2.56	0.39	0.09 to 1.72	0.49	0.11 to 2.12	–	–

ADA, adalimumab; CER, certolizumab; ETA, etanercept; GOL, golimumab; INF, infliximab.

APPENDIX 7

TABLE 120 Relative effects relative risk ASAS 20 AS

	ADA		CER		ETA		GOL		INF	
Anti-TNF	Mean difference	95% CrI	Mean difference	95% CrI	Mean difference	95% CrI	Mean difference	95% CrI	Mean difference	95% CrI
ADA	–	–	0.79	0.53 to 1.07	0.98	0.82 to 1.17	0.94	0.75 to 1.15	1.07	0.75 to 1.38
CER	1.27	0.93 to 1.88	–	–	1.24	0.91 to 1.83	1.19	0.85 to 1.77	1.35	0.88 to 2.09
ETA	1.03	0.86 to 1.22	0.81	0.55 to 1.10	–	–	0.96	0.76 to 1.18	1.10	0.77 to 1.41
GOL	1.07	0.87 to 1.34	0.84	0.56 to 1.18	1.04	0.85 to 1.31	–	–	1.14	0.79 to 1.53
INF	0.93	0.73 to 1.34	0.74	0.48 to 1.14	0.91	0.71 to 1.31	0.87	0.66 to 1.27	–	–

ADA, adalimumab; CER, certolizumab; ETA, etanercept; GOL, golimumab; INF, infliximab.

TABLE 121 Relative effects OR ASAS 20 AS

	ADA		CER		ETA		GOL		INF	
Anti-TNF	Mean difference	95% CrI	Mean difference	95% CrI	Mean difference	95% CrI	Mean difference	95% CrI	Mean difference	95% CrI
ADA	–	–	0.57	0.28 to 1.20	0.94	0.58 to 1.50	0.85	0.49 to 1.46	1.23	0.50 to 3.01
CER	1.74	0.84 to 3.57	–	–	1.62	0.78 to 3.35	1.47	0.67 to 3.16	2.13	0.74 to 6.13
ETA	1.07	0.67 to 1.71	0.62	0.30 to 1.28	–	–	0.90	0.52 to 1.55	1.31	0.54 to 3.20
GOL	1.18	0.69 to 2.05	0.68	0.32 to 1.49	1.11	0.65 to 1.91	–	–	1.46	0.57 to 3.70
INF	0.82	0.33 to 1.99	0.47	0.16 to 1.36	0.76	0.31 to 1.86	0.69	0.27 to 1.75	–	–

ADA, adalimumab; CER, certolizumab; ETA, etanercept; GOL, golimumab; INF, infliximab.

TABLE 122 Relative effects relative risk ASAS 40 AS

	ADA		CER		ETA		GOL	
Anti-TNF	Mean difference	95% CrI	Mean difference	95% CrI	Mean difference	95% CrI	Mean difference	95% CrI
ADA	–	–	0.74	0.41 to 1.22	0.80	0.51 to 1.20	0.91	0.61 to 1.32
CER	1.35	0.82 to 2.45	–	–	1.09	0.61 to 2.04	1.23	0.72 to 2.26
ETA	1.24	0.83 to 1.95	0.92	0.49 to 1.63	–	–	1.13	0.72 to 1.81
GOL	1.10	0.76 to 1.64	0.81	0.44 to 1.38	0.88	0.55 to 1.38	–	–

ADA, adalimumab; CER, certolizumab; ETA, etanercept; GOL, golimumab.

TABLE 123 Relative effects OR ASAS 40 AS

	ADA		CER		ETA		GOL	
Anti-TNF	Mean difference	95% CrI	Mean difference	95% CrI	Mean difference	95% CrI	Mean difference	95% CrI
ADA	–	–	0.59	0.25 to 1.45	0.68	0.33 to 1.40	0.84	0.42 to 1.67
CER	1.68	0.69 to 4.04	–	–	1.14	0.45 to 2.90	1.42	0.57 to 3.50
ETA	1.47	0.71 to 3.02	0.87	0.35 to 2.24	–	–	1.23	0.58 to 2.63
GOL	1.19	0.60 to 2.38	0.71	0.29 to 1.75	0.81	0.38 to 1.72	–	–

ADA, adalimumab; CER, certolizumab; ETA, etanercept; GOL, golimumab.

TABLE 124 Relative effects relative risk ASAS 50 AS

	ADA		ETA		INF	
Anti-TNF	Mean difference	95% CrI	Mean difference	95% CrI	Mean difference	95% CrI
ADA	–	–	1.24	0.60 to 3.31	2.00	0.73 to 5.87
ETA	0.81	0.30 to 1.66	–	–	1.63	0.68 to 2.95
INF	0.50	0.17 to 1.36	0.61	0.34 to 1.46	–	–

ADA, adalimumab; ETA, etanercept; INF, infliximab.

APPENDIX 7

TABLE 125 Relative effects OR ASAS 50 AS

	ADA		ETA		INF	
Anti-TNF	Mean difference	95% CrI	Mean difference	95% CrI	Mean difference	95% CrI
ADA	–	–	1.40	0.40 to 5.05	4.11	0.59 to 29.29
ETA	0.71	0.20 to 2.49	–	–	2.92	0.55 to 15.51
INF	0.24	0.03 to 1.71	0.34	0.06 to 1.81	–	–

ADA, adalimumab; ETA, etanercept; INF, infliximab.

TABLE 126 Relative effects mean difference BASDAI change from baseline AS

	ADA		CER		ETA		INF	
Anti-TNF	Mean difference	95% CrI	Mean difference	95% CrI	Mean difference	95% CrI	Mean difference	95% CrI
ADA	–	–	0.10	–0.68 to 0.88	–0.20	–0.71 to 0.30	–0.73	–1.69 to 0.24
CER	–0.10	–0.88 to 0.68	–	–	–0.30	–1.12 to 0.52	–0.82	–1.98 to 0.33
ETA	0.20	–0.30 to 0.71	0.30	–0.52 to 1.12	–	–	–0.53	–1.50 to 0.47
INF	0.73	–0.24 to 1.69	0.82	–0.33 to 1.98	0.53	–0.47 to 1.50	–	–

ADA, adalimumab; CER, certolizumab; ETA, etanercept; INF, infliximab.

TABLE 127 Relative effects mean difference BASFI change from baseline AS

	ADA		CER		ETA		GOL		INF	
Anti-TNF	Mean difference	95% CrI	Mean difference	95% CrI	Mean difference	95% CrI	Mean difference	95% CrI	Mean difference	95% CrI
ADA	–	–	0.15	–0.67 to 0.97	–0.18	–0.73 to 0.36	–0.20	–0.75 to 0.35	–0.91	–2.00 to 0.20
CER	–0.15	–0.97 to 0.67	–	–	–0.33	–1.16 to 0.49	–0.35	–1.17 to 0.47	–1.05	–2.31 to 0.22
ETA	0.18	–0.36 to 0.73	0.33	–0.49 to 1.16	–	–	–0.02	–0.57 to 0.55	–0.72	–1.83 to 0.39
GOL	0.20	–0.35 to 0.75	0.35	–0.47 to 1.17	0.02	–0.55 to 0.57	–	–	–0.71	–1.82 to 0.42
INF	0.91	–0.20 to 2.00	1.05	–0.22 to 2.31	0.72	–0.39 to 1.83	0.71	–0.42 to 1.82	–	–

ADA, adalimumab; CER, certolizumab; ETA, etanercept; GOL, golimumab; INF, infliximab.

TABLE 128 Relative effects mean difference BASMI change from baseline AS

	ADA		CER		ETA		GOL	
Anti-TNF	Mean difference	95% CrI	Mean difference	95% CrI	Mean difference	95% CrI	Mean difference	95% CrI
ADA	–	–	0.11	–0.21 to 0.42	0.00	–0.32 to 0.31	0.26	0.06 to 0.46
CER	–0.11	–0.42 to 0.21	–	–	–0.11	–0.51 to 0.30	0.15	–0.17 to 0.48
ETA	0.00	–0.31 to 0.32	0.11	–0.30 to 0.51	–	–	0.26	–0.06 to 0.58
GOL	–0.26	–0.46 to –0.06	–0.15	–0.48 to 0.17	–0.26	–0.58 to 0.06	–	–

ADA, adalimumab; CER, certolizumab; ETA, etanercept; GOL, golimumab.

TABLE 129 Relative effects mean difference SF-36 PCS change from baseline AS

	ADA		CER		GOL	
Anti-TNF	Mean difference	95% CrI	Mean difference	95% CrI	Mean difference	95% CrI
ADA	–	–	2.11	–0.20 to 4.44	1.52	–0.24 to 3.30
CER	–2.11	–4.44 to 0.20	–	–	–0.59	–2.99 to 1.85
GOL	–1.52	–3.30 to 0.24	0.59	–1.85 to 3.00	–	–

ADA, adalimumab; CER, certolizumab; GOL, golimumab.

TABLE 130 Relative effects mean difference MASES change from baseline AS

	GOL	
Anti-TNF	Mean difference	95% CrI
ADA	–0.20	–1.12 to 0.70

ADA, adalimumab; GOL, golimumab.

TABLE 131 Relative effects mean difference SF-36 MCS change from baseline AS

	ADA		CER		GOL	
Anti-TNF	Mean difference	95% CrI	Mean difference	95% CrI	Mean difference	95% CrI
ADA	–	–	–0.15	–3.83 to 3.54	1.33	–0.97 to 3.63
CER	0.15	–3.53 to 3.83	–	–	1.51	–2.24 to 5.21
GOL	–1.33	–3.63 to 0.98	–1.51	–5.20 to 2.24	–	–

ADA, adalimumab; CER, certolizumab; GOL, golimumab.

APPENDIX 7

Relative effects of anti-tumour necrosis factors: non-radiographic axial spondyloarthritis population

In the following tables, the intervention is stated in the top row and the comparator is in the left-hand column, which is reverse to normal.

TABLE 132 Relative effects relative risk BASDAI 50 nr-AxSpA

	ADA		CER		ETA	
Anti-TNF	Mean difference	95% CrI	Mean difference	95% CrI	Mean difference	95% CrI
ADA	–	–	1.11	0.62 to 1.96	0.76	0.44 to 1.30
CER	0.90	0.51 to 1.61	–	–	0.69	0.38 to 1.22
ETA	1.31	0.77 to 2.28	1.46	0.82 to 2.62	–	–

ADA, adalimumab; CER, certolizumab; ETA, etanercept.

TABLE 133 Relative effects OR BASDAI 50 nr-AxSpA

	ADA		CER		ETA	
Anti-TNF	Mean difference	95% CrI	Mean difference	95% CrI	Mean difference	95% CrI
ADA	–	–	1.24	0.42 to 3.75	0.62	0.25 to 1.55
CER	0.81	0.27 to 2.40	–	–	0.50	0.18 to 1.40
ETA	1.62	0.65 to 3.99	2.01	0.72 to 5.68	–	–

ADA, adalimumab; CER, certolizumab; ETA, etanercept.

TABLE 134 Relative effects relative risk ASAS 20 nr-AxSpA

	ADA		CER		ETA	
Anti-TNF	Mean difference	95% CrI	Mean difference	95% CrI	Mean difference	95% CrI
ADA	–	–	0.83	0.54 to 1.20	0.77	0.52 to 1.08
CER	1.20	0.84 to 1.87	–	–	0.92	0.60 to 1.44
ETA	1.31	0.93 to 1.94	1.09	0.70 to 1.67	–	–

ADA, adalimumab; CER, certolizumab; ETA, etanercept.

TABLE 135 Relative effects OR ASAS 20 nr-AxSpA

	ADA		CER		ETA	
Anti-TNF	Mean difference	95% CrI	Mean difference	95% CrI	Mean difference	95% CrI
ADA	–	–	0.62	0.25 to 1.59	0.52	0.23 to 1.19
CER	1.60	0.63 to 3.98	–	–	0.83	0.34 to 2.01
ETA	1.92	0.84 to 4.33	1.20	0.50 to 2.92	–	–

ADA, adalimumab; CER, certolizumab; ETA, etanercept.

TABLE 136 Relative effects relative risk ASAS 40 nr-AxSpA

	ADA		CER		ETA		INF	
Anti-TNF	Mean difference	95% CrI	Mean difference	95% CrI	Mean difference	95% CrI	Mean difference	95% CrI
ADA	–	–	0.97	0.51 to 1.78	0.66	0.35 to 1.21	1.16	0.42 to 2.29
CER	1.04	0.56 to 1.98	–	–	0.68	0.35 to 1.35	1.20	0.43 to 2.55
ETA	1.51	0.83 to 2.82	1.46	0.74 to 2.85	–	–	1.74	0.63 to 3.70
INF	0.86	0.44 to 2.37	0.84	0.39 to 2.33	0.57	0.27 to 1.58	–	–

ADA, adalimumab; CER, certolizumab; ETA, etanercept; INF, infliximab.

TABLE 137 Relative effects OR ASAS 40 nr-AxSpA

	ADA		CER		ETA		INF	
Anti-TNF	Mean difference	95% CrI	Mean difference	95% CrI	Mean difference	95% CrI	Mean difference	95% CrI
ADA	–	–	0.94	0.31 to 2.90	0.51	0.19 to 1.35	1.36	0.26 to 7.22
CER	1.07	0.34 to 3.25	–	–	0.54	0.18 to 1.58	1.45	0.25 to 8.10
ETA	1.98	0.74 to 5.23	1.86	0.63 to 5.51	–	–	2.68	0.52 to 13.91
INF	0.73	0.14 to 3.91	0.69	0.12 to 3.93	0.37	0.07 to 1.91	–	–

ADA, adalimumab; CER, certolizumab; ETA, etanercept; INF, infliximab.

TABLE 138 Relative effects mean difference BASDAI change from baseline nr-AxSpA

	ADA		CER		ETA		INF	
Anti-TNF	Mean difference	95% CrI	Mean difference	95% CrI	Mean difference	95% CrI	Mean difference	95% CrI
ADA	–	–	–0.63	–1.77 to 0.52	0.53	–0.51 to 1.56	–1.43	–3.08 to 0.22
CER	0.63	–0.52 to 1.77	–	–	1.15	–0.12 to 2.42	–0.81	–2.62 to 1.00
ETA	–0.53	–1.56 to 0.51	–1.15	–2.42 to 0.12	–	–	–1.97	–3.70 to –0.21
INF	1.43	–0.21 to 3.08	0.81	–1.00 to 2.62	1.97	0.21 to 3.70	–	–

ADA, adalimumab; CER, certolizumab; ETA, etanercept; INF, infliximab.

TABLE 139 Relative effects mean difference BASFI change from baseline nr-AxSpA

	ADA		CER		ETA		INF	
Anti-TNF	Mean difference	95% CrI	Mean difference	95% CrI	Mean difference	95% CrI	Mean difference	95% CrI
ADA	–	–	–1.00	–2.10 to 0.10	0.30	–0.48 to 1.08	–1.33	–2.86 to 0.19
CER	1.00	–0.10 to 2.10	–	–	1.30	0.19 to 2.41	–0.33	–2.05 to 1.38
ETA	–0.30	–1.08 to 0.48	–1.30	–2.41 to –0.19	–	–	–1.63	–3.15 to –0.09
INF	1.33	–0.19 to 2.86	0.33	–1.38 to 2.05	1.63	0.09 to 3.15	–	–

ADA, adalimumab; CER, certolizumab; ETA, etanercept; INF, infliximab.

APPENDIX 7

TABLE 140 Relative effects mean difference BASMI change from baseline nr-AxSpA

	ADA		CER		ETA	
Anti-TNF	Mean difference	95% CrI	Mean difference	95% CrI	Mean difference	95% CrI
ADA	–	–	–0.53	–0.93 to –0.12	0.02	–0.47 to 0.51
CER	0.53	0.12 to 0.93	–	–	0.55	–0.02 to 1.10
INF	–0.02	–0.51 to 0.47	–0.55	–1.10 to 0.02	–	–

ADA, adalimumab; CER, certolizumab; ETA, etanercept; INF, infliximab.

TABLE 141 Relative effects mean difference SF-36 PCS change from baseline nr-AxSpA

	ADA		CER		ETA	
Anti-TNF	Mean difference	95% CrI	Mean difference	95% CrI	Mean difference	95% CrI
ADA	–	–	2	–1.53 to 5.57	–2.88	–6.11 to 0.31
CER	–2.00	–5.57 to 1.53	–	–	–4.88	–8.52 to –1.29
INF	2.88	–0.31 to 6.11	4.88	1.29 to 8.52	–	–

ADA, adalimumab; CER, certolizumab; ETA, etanercept; INF, infliximab.

TABLE 142 Relative effects mean difference SF-36 MCS change from baseline nr-AxSpA

	ADA		CER	
Anti-TNF	Mean difference	95% CrI	Mean difference	95% CrI
ADA	–	–	2.87	–1.78 to 7.49
CER	–2.87	–7.49 to 1.78	–	–

ADA, adalimumab; CER, certolizumab.

Appendix 8 Long-term efficacy data

TABLE 143 Data from open label extensions of included RCTs

Study characteristics						Results			
						Number of responders (%)			Other outcomes
Trial cohort and references of open-label studies	Population	Treatment and dose	Time point	Number of patients	Imputation methods and withdrawal criteria	ASAS 20	ASAS 40	BASDAI 50	Other results ASAS 50, ASAS 70, BASDAI, BASFI, BASMI, mSASSS, MASES, SF-36 MCS, SF-36 PCS, EQ-5D
Haibel 2008[52–54,185–188]	Nr-axSpA with inflammation	Adalimumab, 40 mg every other week. Non-responders at the end of the double-blind trial (week 12) and after open-label therapy for at least 12 weeks were eligible for dose escalation to 40 mg/week	52 weeks	46	ITT. Patients who withdrew from the study were counted as non-responders for categorical data. LOCF was used for continuous variables	–	23/46 (50)	24/46 (52)	BASDAI change from baseline 2.8 (95% CI 2.1 to 3.6) BASFI change from baseline 2 (95% CI 1.4 to 2.6) BASMI change from baseline −0.4 (95% CI −0.7 to −0.04) EQ-5D change from baseline 0.22 (95% CI 0.13 to 0.31) SF-36 MCS change from baseline 4.9 (95% CI 1.6 to 8.1) SF-36 PCS change from baseline 10.3 (95% CI 6.9 to 13.8) ASQoL change from baseline 5.3 (95% CI 3.8 to 6.7) MASES change from baseline 0.9 (95% CI −0.02 to 1.9) In total, 26 patients with MRIs at baseline and 52 weeks showed no change in sclerosis or in erosions

Study characteristics						Results			
						Number of responders (%)			Other outcomes
Trial cohort and references of open-label studies	Population	Treatment and dose	Time point	Number of patients	Imputation methods and withdrawal criteria	ASAS 20	ASAS 40	BASDAI 50	Other results ASAS 50, ASAS 70, BASDAI, BASFI, BASMI, mSASSS, MASES, SF-36 MCS, SF-36 PCS, EQ-5D
ABILITY-1 2013[38,189–194]	Nr-axSpA with inflammation	Adalimumab, placebo/ adalimumab	52 weeks	61	–	–	–	–	SF-36 PCS change from baseline 10.0 (SD 9.91)
		Adalimumab, adalimumab/ adalimumab	52 weeks	55	–	–	–	–	SF-36 PCS change from baseline 11.0 (SD 9.93)
		Adalimumab, 40 mg every other week	68 weeks	111 (patients MRI positive or CRP positive)	Observed ($n = 142$ at week 12)		77/111 (69)	74/111 (67)	–
		Adalimumab, 40 mg every other week	104 weeks	102	–	–	–	–	–
		Adalimumab, 40 mg every other week	156 weeks	97 (patients MRI positive or CRP positive)	Observed ($n = 142$ at week 12)	83/97 (86)	67/97 (69)	70/97 (72)	ASAS 50 responders: $n = 58$ ASAS 70 responders: $n = 47$
ATLAS 2006[61,140,141,195–201]	AS	Adalimumab, 40 mg every other week	52 weeks	311 had at least one dose	Observed	193/276 (70)	138/276 (50)	167/276 (61)	BASDAI change from baseline –3.5 (SD 2.55), $n = 274$ BASFI change from baseline –2.6 (SD 2.04), $n = 274$ BASMI final value 3.2 (SD 2.2), $n = 273$ SF-36 MCS change from baseline 5.6 (SD 10.35), $n = 265$ SF-36 PCS change from baseline 10.19 (SD 9.5), $n = 265$ ASQoL change from baseline –4.8 (SD 4.41), $n = 274$ MASES final value 2.4 (SD 4.6), $n = 279$

continued

TABLE 143 Data from open label extensions of included RCTs (*continued*)

Study characteristics						Results			
						Number of responders (%)			Other outcomes
Trial cohort and references of open-label studies	Population	Treatment and dose	Time point	Number of patients	Imputation methods and withdrawal criteria	ASAS 20	ASAS 40	BASDAI 50	Other results ASAS 50, ASAS 70, BASDAI, BASFI, BASMI, mSASSS, MASES, SF-36 MCS, SF-36 PCS, EQ-5D
		Adalimumab, 40 mg every other week	76 weeks	–	Observed	–	–	–	BASDAI change from baseline −3.8 (SD 2.33), *n* = 270 BASFI change from baseline −2.8 (SD 2.1), *n* = 270 SF-36 MCS change from baseline 5.1 (SD 11.06), *n* = 263 SF-36 PCS change from baseline 10.8 (SD 9.88), *n* = 263 ASQoL change from baseline −5 (SD 4.32), *n* = 270
		Adalimumab, 40 mg every other week	104 weeks	173	Observed	135/173 (78)	109/173 (63)	122/173 (71)	BASDAI change from baseline −3.9 (SD 2.44), *n* = 262 BASFI change from baseline −2.9 (SD 2.14), *n* = 261 BASMI final value 3.1 (SD 2.2), *n* = 173 SF-36 MCS change from baseline 5.7 (SD 10.96), *n* = 255 SF-36 PCS change from baseline 11 (SD 9.88), *n* = 255 ASQoL change from baseline −5.4 (SD 4.28), *n* = 263 MASES change from baseline 2.2 (SD 4.4), *n* = 217

Study characteristics						Results			
						Number of responders (%)			Other outcomes
Trial cohort and references of open-label studies	Population	Treatment and dose	Time point	Number of patients	Imputation methods and withdrawal criteria	ASAS 20	ASAS 40	BASDAI 50	Other results ASAS 50, ASAS 70, BASDAI, BASFI, BASMI, mSASSS, MASES, SF-36 MCS, SF-36 PCS, EQ-5D
		Adalimumab, 40 mg every other week	128 weeks	–	Observed	–	–	–	BASDAI change from baseline –3.9 (SD 2.39), *n* = 242 BASFI change from baseline –2.9 (SD 2.17), *n* = 242 SF-36 MCS change from baseline 4.1 (SD 10.84), *n* = 229 SF-36 PCS change from baseline 11.3 (SD 9.68), *n* = 229 ASQoL change from baseline –5.3 (SD 4.35), *n* = 242
		Adalimumab, 40 mg every other week	156 weeks	–	Observed	–	–	–	BASDAI change from baseline –3.9 (SD 3.39), *n* = 236 BASFI change from baseline –3 (SD 2.1), *n* = 236 BASMI final value 3.7 (SD 1.8), *n* = 233 SF-36 MCS change from baseline 5.6 (SD 11.59), *n* = 227 SF-36 PCS change from baseline 11.6 (SD 9.65), *n* = 227 ASQoL change from baseline –5.4 (SD 4.36), *n* = 236

continued

TABLE 143 Data from open label extensions of included RCTs (*continued*)

Study characteristics						Results			
						Number of responders (%)			Other outcomes
Trial cohort and references of open-label studies	Population	Treatment and dose	Time point	Number of patients	Imputation methods and withdrawal criteria	ASAS 20	ASAS 40	BASDAI 50	Other results ASAS 50, ASAS 70, BASDAI, BASFI, BASMI, mSASSS, MASES, SF-36 MCS, SF-36 PCS, EQ-5D
		Adalimumab, 40 mg every other week	5 years	125 (patients randomised to adalimumab and completed 5 years)	Observed	111/125 (89)	88/125 (70)	96/124 (77)	BASDAI score final value 1.8 (SD 1.9), *n* = 124 BASFI score final value 2.1 (SD 2.1), *n* = 125 BASMI score final value 3.7 (SD 1.8), *n* = 124 SF-36 PCS score final value 44.4 (SD 10), *n* = 165 ASQoL score final value 4.8 (SD 4.8), *n* = 169
RAPID-axSpA 2014[64,202–204]	AS	Certolizumab pegol 200 mg every 2 weeks	48 weeks	65	NRI + LOCF	47/65 (72)	34/65 (52)	–	BASDAI score final value 3.3 BASFI score final value 3
		Certolizumab pegol 400 mg every 4 weeks	48 weeks	56		42/56 (75)	36/56 (64)	–	BASDAI score final value 3 BASFI score final value 3.2
		Certolizumab pegol, all	48 weeks	121		AiC information has been removed	AiC information has been removed	–	–

Study characteristics						Results			
						Number of responders (%)			Other outcomes
Trial cohort and references of open-label studies	Population	Treatment and dose	Time point	Number of patients	Imputation methods and withdrawal criteria	ASAS 20	ASAS 40	BASDAI 50	Other results ASAS 50, ASAS 70, BASDAI, BASFI, BASMI, mSASSS, MASES, SF-36 MCS, SF-36 PCS, EQ-5D
		Certolizumab pegol, 200 mg every 2 weeks	96 weeks	65	NRI	39	–	–	–
		Certolizumab pegol, 400 mg every 4 weeks	96 weeks	56	NRI	39	–	–	–
		Certolizumab pegol, all	96 weeks	121	NRI	78/121 (64)	61/121 (50)	–	–
		Certolizumab pegol, all	96 weeks	93	Observed case	78/93 (84)	61/93 (66)	–	–
	Nr-axSpA with inflammation	Certolizumab pegol, 200 mg every 2 weeks	48 weeks	46	NRI was used for categorical measures and LOCF for quantitative measures (48-week data)	32/46 (70)	25/46 (54)	–	BASDAI score final value 2.9
		Certolizumab pegol, 400 mg every 4 weeks	48 weeks	51		35/51 (69)	30/51 (59)	–	BASDAI score final value 3.3
		Certolizumab pegol, all	48 weeks	97		AiC information has been removed	AiC information has been removed	–	–
		Certolizumab pegol 200 mg every 2 weeks	96 weeks	46	NRI	30		–	–
		Certolizumab pegol 400 mg every 4 weeks	96 weeks	51	NRI	29		–	–
		Certolizumab pegol, all	96 weeks	97	NRI	59/97 (61)	49/97 (51)	–	–
		Certolizumab pegol, all	96 weeks	74	Observed case	59/74 (80)	49/74 (66)	–	–

continued

TABLE 143 Data from open label extensions of included RCTs (*continued*)

Study characteristics						Results			
						Number of responders (%)			Other outcomes
Trial cohort and references of open-label studies	Population	Treatment and dose	Time point	Number of patients	Imputation methods and withdrawal criteria	ASAS 20	ASAS 40	BASDAI 50	Other results ASAS 50, ASAS 70, BASDAI, BASFI, BASMI, mSASSS, MASES, SF-36 MCS, SF-36 PCS, EQ-5D
Davis 2003[72,119,144,145,205,206]	AS	Placebo then etanercept	72 weeks	105	Observed case	–	–	–	BASFI score final value 32.3 (SD 2.5)
		Etanercept then etanercept	96 weeks	95	Observed case	70/95 (74)	–	–	ASAS 70 responders 44 BASFI score final value 25.4 (SD 2.4)
		Combined groups	96 weeks	257	–	–	–	–	mSASSS change from baseline 0.91 (SD 2.45)
		Placebo then etanercept	168 weeks	127	LOCF	77/127 (61)	64/127 (50)	–	–
		Etanercept then etanercept	192 weeks	124	LOCF	83/124 (67)	61/124 (49)	–	–
Dougados 2014[76,207]	Nr-axSpA mixed	Etanercept then etanercept, 50 mg weekly	32 weeks	100	NRI	–	AiC information has been removed	–	–
		Placebo then etanercept, 50 mg weekly	32 weeks	105	NRI	–	AiC information has been removed	–	–
		Etanercept then etanercept, 50 mg weekly	40 weeks	100	NRI	–	AiC information has been removed	–	–
		Placebo then etanercept, 50 mg weekly	40 weeks	105	NRI	–	AiC information has been removed	–	–

Study characteristics						Results			
						Number of responders (%)			Other outcomes
Trial cohort and references of open-label studies	Population	Treatment and dose	Time point	Number of patients	Imputation methods and withdrawal criteria	ASAS 20	ASAS 40	BASDAI 50	Other results ASAS 50, ASAS 70, BASDAI, BASFI, BASMI, mSASSS, MASES, SF-36 MCS, SF-36 PCS, EQ-5D
		Etanercept plus placebo groups together, 50 mg weekly	48 weeks	AiC information has been removed (189 observed)	NRI + LOCF	147	108	128	AiC information has been removed
		Etanercept then etanercept, 50 mg weekly	48 weeks	100	–	–	AiC information has been removed	–	–
		Placebo then etanercept, 50 mg weekly	48 weeks	105	–	–	AiC information has been removed	–	–
Gorman 2002[79,206,208]		Etanercept then etanercept	28 weeks	19	NRI	AiC information has been removed	–	–	AiC information has been removed
		Placebo then etanercept	28 weeks	19	NRI	AiC information has been removed	–	–	AiC information has been removed
		Etanercept then etanercept	40 weeks	19	NRI	AiC information has been removed	–	–	AiC information has been removed
		Placebo then etanercept	40 weeks	19	NRI	AiC information has been removed	–	–	AiC information has been removed

continued

TABLE 143 Data from open label extensions of included RCTs (*continued*)

Study characteristics						Results			
						Number of responders (%)			Other outcomes
Trial cohort and references of open-label studies	Population	Treatment and dose	Time point	Number of patients	Imputation methods and withdrawal criteria	ASAS 20	ASAS 40	BASDAI 50	Other results ASAS 50, ASAS 70, BASDAI, BASFI, BASMI, mSASSS, MASES, SF-36 MCS, SF-36 PCS, EQ-5D
Calin 2004[83,146,147]	AS	Etanercept then etanercept	60 weeks	42	LOCF	–	–	–	BASDAI score final value 2.1 BASFI score final value 2.9 mSASSS change from baseline 0.36 (95% CI −0.1 to 0.8), $n = 33$
		Placebo then etanercept	60 weeks	39	LOCF	–	–	–	BASDAI final value 2.7 BASFI final value 3.4 mSASSS change from baseline −0.15 (95% CI −0.7 to 0.4), $n = 34$
		Combined group	108 weeks	81	LOCF	AiC information has been removed	44/81 (54)	AiC information has been removed	AiC information has been removed BASFI final value: 2.9
		Etanercept then etanercept	108 weeks	42	LOCF	–	–	–	BASDAI score final value 2.3 BASFI score final value 3
		Placebo then etanercept	108 weeks	39	LOCF	–	–	–	BASDAI score final value 2.9 BASFI score final value 3.5
		Combined group	264 weeks	59	LOCF		40/59 (68)	39/59 (66)	AiC information has been removed BASDAI score final value 2.7 BASFI score final value 3.2

Study characteristics						Results			
						Number of responders (%)			Other outcomes
Trial cohort and references of open-label studies	Population	Treatment and dose	Time point	Number of patients	Imputation methods and withdrawal criteria	ASAS 20	ASAS 40	BASDAI 50	Other results ASAS 50, ASAS 70, BASDAI, BASFI, BASMI, mSASSS, MASES, SF-36 MCS, SF-36 PCS, EQ-5D
Bao 2012[95,96,209,210]	AS	Golimumab, 50 mg	52 weeks	108	ITT	76/108 (70)	53/108 (49)	62/108 (57)	–
GO-RAISE 2008[90,120,142,143,211–222]	AS	Golimumab placebo, 50 mg	104 weeks	78	ITT	30/78 (38)	30/78 (38)	–	BASDAI score final value median 6 (IQR 1.36 to 7.79) BASFI score final value median 4.9 (IQR 0.98 to 7.07) mSASSS score change from baseline 1.6 (SD 4.6), $n = 66$
		Golimumab, 50 mg	104 weeks	138	ITT	83/138 (60)	77/138 (56)	–	BASDAI score final value median 2.7 (IQR 0.84 to 6.08) BASFI score final value median 2.2 (IQR 0.52 to 5.80) mSASSS score change from baseline 0.9 (SD 2.7), $n = 111$
		Golimumab, 100 mg	104 weeks	140	ITT	100/140 (71)	76/140 (54)	–	BASDAI score final value median 2.7 (IQR 1.08 to 5.34) BASFI score final value median 1.8 (IQR 0.49 to 4.79) mSASSS score change from baseline 0.9 (SD 3.9), $n = 122$
		All patients randomised (all golimumab from week 24)	104 weeks	356	NRI + LOCF	249/356 (70)	213/356 (60)	–	–

continued

TABLE 143 Data from open label extensions of included RCTs (*continued*)

Study characteristics						Results			
						Number of responders (%)			Other outcomes
Trial cohort and references of open-label studies	Population	Treatment and dose	Time point	Number of patients	Imputation methods and withdrawal criteria	ASAS 20	ASAS 40	BASDAI 50	Other results ASAS 50, ASAS 70, BASDAI, BASFI, BASMI, mSASSS, MASES, SF-36 MCS, SF-36 PCS, EQ-5D
		All patients randomised (all golimumab from week 24)	160 weeks	356	NRI + LOCF	246/356 (69)	208/356 (58)	–	–
		Golimumab placebo, 50 mg	208 weeks	78	–	–	–	–	mSASSS change from baseline 2.1 (SD 5.2), $n = 66$
		Golimumab, 50 mg	208 weeks	138	–	–	–	–	mSASSS change from baseline 1.3 (SD 4.1), $n = 111$
		Golimumab, 100 mg	208 weeks	140	–	–	–	–	mSASSS change from baseline 2 (SD 5.6), $n = 122$
		All patients randomised (all golimumab from week 24)	256 weeks	356	NRI + LOCF	235/356 (66)	203/356 (57)	199/356 (58)	–
Tam 2014[97]	AS	Golimumab, 50 mg monthly	54 weeks	19	UC	18	–	–	–
		Placebo/golimumab	54 weeks	17	UC	14	–	–	–
		Placebo/placebo	54 weeks	3	UC	1	–	–	–

Study characteristics						Results			
						Number of responders (%)			Other outcomes
Trial cohort and references of open-label studies	Population	Treatment and dose	Time point	Number of patients	Imputation methods and withdrawal criteria	ASAS 20	ASAS 40	BASDAI 50	Other results ASAS 50, ASAS 70, BASDAI, BASFI, BASMI, mSASSS, MASES, SF-36 MCS, SF-36 PCS, EQ-5D
PLANETAS 2013[110,223]	AS	CT-P13 (biosimilar to infliximab) 5 mg/kg	78 weeks	88	ITT	61/88 (69)	50/88 (57)	–	–
		CT-P13 then infliximab (switched at week 54) 5 mg/kg	78 weeks	86	ITT	64/86 (74)	43/86 (50)	–	–
		CT-P13 (biosimilar to infliximab) 5 mg/kg	102 weeks	88	ITT	67/88 (76)	53/88 (60)	–	–
		CT-P13 then infliximab (switched at week 54) 5 mg/kg	102 weeks	86	ITT	60/86 (70)	48/86 (56)	–	–
Braun 2002[86,148,164,224–230]	AS	Infliximab 5 mg/kg (infusion at 0, 2 and 6 weeks)	54 weeks	34	NRI for binary data. A completer analysis was conducted	–	–	47%	mSASSS reported for two groups: patients with worsening of BASFI score of > 1 and those with score of < 1
		Placebo/infliximab	54 weeks	35		–	–	51%	
		Aggregate	54 weeks	69		–	–	33/69 (48)	BASDAI score final value: 2.5 (SD 1.7), $n = 52$ BASFI score final value: 3.0 (SD 2.2), $n = 52$ BASMI score final value: 2.4 (SD 2.0), $n = 52$ SF-36 MCS score final value: 50.9 (SD 8.9), $n = 52$ SF-36 PCS score final value: 40.6 (SD 10.6), $n = 52$

continued

TABLE 143 Data from open label extensions of included RCTs (*continued*)

Study characteristics						Results			
						Number of responders (%)			Other outcomes
Trial cohort and references of open-label studies	Population	Treatment and dose	Time point	Number of patients	Imputation methods and withdrawal criteria	ASAS 20	ASAS 40	BASDAI 50	Other results ASAS 50, ASAS 70, BASDAI, BASFI, BASMI, mSASSS, MASES, SF-36 MCS, SF-36 PCS, EQ-5D
									BASDAI score final value: 2.4 (SD 1.6), *n* = 46 BASFI score final value: 3.1 (SD 2.2), *n* = 46 BASMI score final value: 2.4 (SD 1.9), *n* = 46 SF-36 MCS score final value: 51.5 (SD 8.6), *n* = 46 SF-36 PCS score final value: 40.2 (SD 10.8), *n* = 46
		Aggregate	102 weeks	69	NRI for binary data. A completer analysis was conducted	–	–	30/69 (43)	BASDAI score final value: 2.6 (SD 2), *n* = 52 BASFI score final value: 3.0 (SD 2.2), *n* = 52 BASMI score final value: 2.7 (SD 2.1), *n* = 52 SF-36 MCS score final value: 50.2 (SD 9.5), *n* = 52 SF-36 PCS score final value: 40.9 (SD 11.1), *n* = 52 BASDAI score final value: 2.6 (SD 2), *n* = 46 BASFI score final value: 3.1 (SD 2.3), *n* = 46 BASMI score final value: 2.7 (SD 2.1), *n* = 46

Study characteristics						Results			
						Number of responders (%)			Other outcomes
Trial cohort and references of open-label studies	Population	Treatment and dose	Time point	Number of patients	Imputation methods and withdrawal criteria	ASAS 20	ASAS 40	BASDAI 50	Other results ASAS 50, ASAS 70, BASDAI, BASFI, BASMI, mSASSS, MASES, SF-36 MCS, SF-36 PCS, EQ-5D
ASSERT 2005[102,231–235]	AS	Aggregate	156 weeks	46	Completer analysis. To calculate means, LOCF was used	–	–	28/46 (61)	SF-36 MCS score final value: 51.4 (SD 8.9), $n = 46$ SF-36 PCS score final value: 40.5 (SD 11.4), $n = 46$ BASDAI score final value: 2.7 (SD 2), $n = 46$? BASFI score final value: 3.1 (SD 2.5) BASMI score final value: 2.8 (SD 2.2) SF-36 MCS score final value: 48.8 (SD 10.4) SF-36 PCS score final value: 41.6 (SD 11.7)
		Infliximab (on placebo 0–24 weeks) 5 mg	Week 102	78 (remaining study patients may have taken high unlicensed dose of infliximab)	Completer analysis and ITT LOCF	–	Completer analysis 28/61 (46) ITT LOCF 33/78 (42)	–	BASMI score change from baseline: −1 (IQR −2.0 to 0.0) SF-36 MCS change from baseline: 2.3 (IQR −3.6 to 11.9) SF-36 PCS change from baseline: 8.3 (IQR 2.5 to 17.7)

IQR, interquartile range; NR, not reported; UC, unclear.
One trial which was extended evaluated only spinal and sacroiliac joint inflammation (Lambert *et al.*[57]); results not shown.

Appendix 9 Adverse events

TABLE 144 Adverse events in RCT placebo phases

Trial name	Population	Treatment arm	Dose	Time point (weeks)	Number of patients randomised	SAEs
Haibel 2008[52]	Nr-axSpA	Adalimumab	40 mg	12	22	0
	Nr-axSpA	Placebo	0	12	24	0
Hu 2012[55]	AS	Adalimumab	40 mg	12	26	
	AS	Placebo	0	12	20	
Huang 2014[56]	AS	Adalimumab	40 mg	12	229	1
	AS	Placebo	0	12	115	1
Lambert 2007[57]	AS	Adalimumab	40 mg	12	38	
	AS	Placebo	0	12	44	
ABILITY-1 2013 (licensed population)[58]	Nr-axSpA	Adalimumab	40 mg	12	95	3
	Nr-axSpA	Placebo	0	12	97	1
ATLAS 2006[61]	AS	Adalimumab	40 mg	12	208	
	AS	Placebo	0	12	107	
RAPID-axSpA 2014[64]	AS	Certolizumab pegol	200 mg	12	65	[a]
	AS	Certolizumab pegol	400 mg	12	56	
	AS	Placebo	0	12	57	
	Nr-axSpA	Certolizumab pegol	200 mg	12	46	
	Nr-axSpA	Certolizumab pegol	400 mg	12	51	
	Nr-axSpA	Placebo	0	12	50	
Barkham 2010[71]	AS	Etanercept	25 mg twice weekly	12	20	0
	AS	Placebo	0	12	20	0
Davis 2003[72]	AS	Etanercept	25 mg	12	138	
	AS	Placebo	0	12	139	
	AS	Etanercept	25 mg	24	138	
	AS	Placebo	0	24	139	
Dougados 2011[74]	AS	Etanercept	50 mg	12	39	
	AS	Placebo	0	12	43	
Dougados 2014[76]	Nr-axSpA mixed	Etanercept	50 mg	12	106	2
	Nr-axSpA mixed	Placebo	0	12	109	1
	Nr-axSpA	Etanercept	50 mg	12	94	
	Nr-axSpA	Placebo	0	12	95	
Gorman 2002[79]	AS	Etanercept	25 mg	16	20	0
	AS	Placebo	0	16	20	0

Serious infections	Tuberculosis (including tuberculosis reactivation)	Injection site reactions	Congestive heart failure	Malignancies	Non-melanoma skin cancer	Withdrawals because of AEs
0			0	0	0	0
0			0	0	0	0
1	0		0	0	0	4
0	0		0	0	0	0
0	0		0		1	
0	0		0		1	
[a]	[a]					
						0
						0
0	0	41				7
1	0	13				1
		3		1		1
		0		0		0
		5				0
		1				0

TABLE 144 Adverse events in RCT placebo phases (*continued*)

Trial name	Population	Treatment arm	Dose	Time point (weeks)	Number of patients randomised	SAEs
Calin 2004[83]	AS	Etanercept	25 mg	12	45	1
	AS	Placebo	0	12	39	0
van der Heijde 2006[86]	AS	Etanercept	25 mg	12	150	
	AS	Etanercept	50 mg	12	155	
	AS	Placebo	0	12	51	
Giardina 2010[88]	AS	Etanercept	50 mg	104	25	
	AS	Infliximab	5 mg/kg	104	25	
GO-RAISE 2008[90]	AS	Golimumab	50 mg	16	138	5
	AS	Golimumab	100 mg	16	140	7
	AS	Placebo	0	16	78	4
Bao 2014[95]	AS	Golimumab	50 mg	14	108	
	AS	Placebo	0	14	105	
Tam 2014[97]	AS	Golimumab	50 mg	24	20	
	AS	Placebo	0	24	21	
Barkham 2009[50]	Nr-axSpA	Infliximab	5 mg/kg	16	20	0
	Nr-axSpA	Placebo	0	16	20	
Braun 2002[98]	AS	Infliximab	5 mg/kg	12	34	3
	AS	Placebo	0	12	35	
Marzo-Ortega 2005[100]	AS	Infliximab + methotrexate	5 mg/kg	10	28	
	AS	Placebo + metotrexate	0	10	14	
	AS	Infliximab + methotrexate	5 mg/kg	30	28	0
	AS	Placebo + metotrexate	0	30	14	0
Van den Bosch 2002[101]	AS	Infliximab	5 mg/kg	12	9	Unclear
	AS	Placebo	0	12	12	Unclear
ASSERT 2005[102]	AS	Infliximab	5 mg/kg	24	202	7
	AS	Placebo	0	24	75	2
PLANETAS 2013[110]	AS	CT-P13	5 mg/kg	14	125	
	AS	Infliximab	5 mg/kg	14	125	
	AS	CT-P13	5 mg/kg	30	128	6
	AS	Infliximab	5 mg/kg	30	122	8
	AS	CT-P13	5 mg/kg	54	125	
	AS	Infliximab	5 mg/kg	54	125	

a Data only for whole group.
Blank fields indicate that data were not reported for that outcome.

Serious infections	Tuberculosis (including tuberculosis reactivation)	Injection site reactions	Congestive heart failure	Malignancies	Non-melanoma skin cancer	Withdrawals because of AEs
0		15				0
0		6				0
1	0	32		0		6
1	0	34		0		8
0	0	6		0		0
1	0	5	0	0		0
2	0	1	0	0		0
						0
						1
						1
	1	0				4
	0	0				0
		1				0
		0				0
		Unclear				
		Unclear				
2	0	22		0		2
0	0	7		0		1
	2	5				8
	1	6				5

Appendix 10 Quality assessment of studies included in the cost-effectiveness review

TABLE 145 Quality assessment of studies included in the cost-effectiveness review[a]

Question	Ara *et al.* 2007[161]	Botteman *et al.* 2007[162]	Kobelt *et al.* 2007[160]	McLeod *et al.* 2007[38]	Armstrong *et al.* 2013[163]
1. Was a well-defined question posed in answerable form?	Yes	Yes	Yes	Yes	Yes
2. Was a comprehensive description of the competing alternatives given (i.e. can you tell who did what to whom, where, and how often)?	Yes	Yes	No	Yes	Yes
3. Was the effectiveness of the programme or services established?	Yes (short–medium term)	Yes (short–medium term)	Yes (short term)	Yes (short–medium term)	Yes (short–medium term)
4. Were all the important and relevant costs and consequences for each alternative identified?	No	Yes	Yes	Yes	Yes (consequences); cannot tell (costs)
5. Were costs and consequences measured accurately in appropriate physical units (e.g. hours of nursing time, number of physician visits, lost work-days, gained life-years)?	Cannot tell	Cannot tell	Cannot tell	Yes	Cannot tell
6. Were the cost and consequences valued credibly?	Cannot tell	Cannot tell	Cannot tell	Yes	Cannot tell
7. Were costs and consequences adjusted for differential timing?	Yes	Yes	Yes	Yes	Cannot tell
8. Was an incremental analysis of costs and consequences of alternatives performed?	Yes	Yes	Yes	Yes	Yes
9. Was allowance made for uncertainty in the estimates of costs and consequences?	Yes	Yes	Yes	Yes	Yes
10. Did the presentation and discussion of study results include all issues of concern to users?	No	No	No	Yes	No

a Only stated publications were quality assessed and further materials (e.g. assessment group reports from the NICE website) were not consulted. The checklist used was Drummond *et al.*[237]

Appendix 11 Comparison of parameter inputs across manufacturer models

Tables 146 and *147* provide an overview of the main parameter inputs applied in each of the manufacturer models for the AS and nr-AxSpA populations.

TABLE 146 Summary of main model inputs in manufacturer models: AS population

Parameter	Merck Sharp & Dohme economic model[37] (infliximab, golimumab)	AbbVie economic model[34] (adalimumab)	UCB economic model[35] (certolizumab)	Pfizer economic model[36] (etanercept)
Time horizon	Lifetime	40 years	Lifetime	Lifetime
Discount rate (%)	3.5	3.5	3.5	3.5
Average age (years)	39	42	41	41
Percentage male	72	75	72	74
Average weight (kg)	70	81.1	81.7	76.4
Baseline BASDAI score	6.5	6.3	6.4	6.1
Baseline BASFI score	5.0	5.3	5.7	5.9
Source of baseline characteristics	GO-RAISE[90]	ATLAS[61]	RAPID-axSpA[64] trial	314-EU study[167]
Mortality (SMR)	Male 1.63; female 1.38	1.5	1.5	1.5
Response criterion	BASDAI 50 response at week 12	ASAS 20 response at week 12	ASAS 20 response at week 24	BASDAI 50 response at week 12
Percentage of responders	• Infliximab 79.3 • Golimumab 48.5 • Adalimumab 47.0 • Certolizumab 53.0 • Etanercept 48.2 • Placebo 14.5	• Infliximab 72.4 • Golimumab 59.3 • Adalimumab 63.2 • Certolizumab 46.2 • Etanercept 60.7 • Placebo 27.2	• Infliximab 65.7 • Golimumab 54.1 • Adalimumab 56.2 • Certolizumab 55.7 • Etanercept 56.4 • Placebo –	• Infliximab 68 • Golimumab 61 • Adalimumab 54 • Certolizumab 47 • Etanercept 54 • Placebo 22
Placebo response	Loss or maintenance of placebo response not clearly reported	BASDAI and BASFI score return to baseline at week 12	No placebo response	BASDAI and BASFI score return to baseline at 12 weeks
Annual long-term rate of anti-TNFs withdrawal	6.1% (GO-RAISE[90]), a common rate for all anti-TNFs	Time-dependent discontinuation; log-normal model fitted to adalimumab week 12 responder data (ATLAS[61]). Less than 15% projected to stay on treatment at year 40, a common rate for all anti-TNFs	7% (NICE TA143[17]), a common rate for all anti-TNFs	Exponential model fitted to etanercept data; model translates to 11% annual discontinuation for etanercept. Hazard ratios applied for other anti-TNFs (Glintborg 2010)[112]

continued

TABLE 146 Summary of main model inputs in manufacturer models: AS population (*continued*)

Parameter	Merck Sharp & Dohme economic model[37] (infliximab, golimumab)	AbbVie economic model[34] (adalimumab)	UCB economic model[35] (certolizumab)	Pfizer economic model[36] (etanercept)
Natural history: annual rate of BASFI progression	0.07 points (Kobelt *et al.* 2004[152])	0.056 points (ATLAS[61])	0.07 points (Kobelt *et al.* 2004[152])	0.07 points (Kobelt *et al.* 2004[152])
AEs included; annual probability/ rate	Serious AEs and ISRs included. Convent. care rates from GO-RAISE study. ORs from the NMA applied for each anti-TNF	Only infectious AEs included; excess proportion for adalimumab 29.7% annually (ATLAS[61] trial). Same rate applied to all anti-TNFs	No AEs included	Serious infections for etanercept: 3.8% annually. Relative effects from a published NMA (Singh 2011[137]) applied for other anti-TNFs
HRQoL algorithm (EQ-5D)	0.877121 − 0.03841 × BASDAI − 0.03225 × BASFI − 0.02789 × male + 0.00168 × age (NICE TA143[17])	0.899 − 0.031 × BASDAI-0.041 × BASFI (HUI-3, data from ATLAS[61])	2.126 − 0.132 × BASFI − 0.245 × BASDAI (RAPID-axSpA[64] study)	0.887 − 0.006030 × BASFI + 0.001030 × BASDAI + 0.000020 × BASFI2 − 0.0000064 × BASDAI2 (314-EU study[167])
Annual health-care resource use costs	1902.49 × exp (0.1832 × BASFI) (NICE TA143[17])	£1124.619 × exp (0.264 × BASDAI) (OASIS[118])	1909.33 × exp (0.1832 × BASFI) (NICE TA143[17])	• BASDAI score of < 4: annual cost: £151.96 • 4 ≤ BASDAI score < 6: annual cost: £311.08 • BASDAI score of ≥ 6: annual cost: £1039.16 (Rafia *et al.* 2012[168])

ISR, injection/infusion site reaction; NMA, network meta-analysis.

TABLE 147 Summary of main model inputs in manufacturer models: nr-AxSpA population

Parameter	AbbVie economic model[34] (adalimumab)	UCB economic model[35] (certolizumab)	Pfizer economic model[36] (etanercept)
Time horizon	40 years	Lifetime	Lifetime
Discount rate (%)	3.5	3.5	3.5
Average age (years)	38	37	32
Percentage male	45	48	60
Average weight (kg)	NR	82	74
Baseline BASDAI score	6.4	6.5	6.0
Baseline BASFI score	4.6	4.9	4.0
Source of baseline characteristics	ABILITY-1[58]	RAPID-axSpA[64] trial	1031 study[166]
Mortality (SMR)	1.0	1.5	1.0

TABLE 147 Summary of main model inputs in manufacturer models: nr-AxSpA population (*continued*)

Parameter	AbbVie economic model[34] (adalimumab)	UCB economic model[35] (certolizumab)	Pfizer economic model[36] (etanercept)
Response criterion	ASAS 40 response at week 12	ASAS 20 response at week 12	BASDAI 50 response at week 12
Percentage of responders	• Adalimumab 55.9 • Certolizumab 58.8 • Etanercept NR • Placebo 22.2	• Adalimumab 56.3 • Certolizumab 59.0 • Etanercept 47.1 • Placebo –	• Adalimumab 44 • Certolizumab 59 • Etanercept 38 • Placebo 27
Placebo response	BASDAI and BASFI score return to baseline at week 12	No placebo response	BASDAI and BASFI score return to baseline at 12 weeks
Annual long-term rate of anti-TNFs withdrawal	Time-dependent discontinuation; Log-normal model fitted to adalimumab week 12 responder data (ABILITY-1[58]). Less than 10% projected to stay on treatment at year 40, a common rate for all anti-TNFs	7% (NICE TA143), common rate for all anti-TNFs	Exponential model fitted to etanercept week 12 responder data; model translates to 5% annual discontinuation for etanercept. Hazard ratios applied for other anti-TNFs (Glintborg *et al.* 2010)[112]
Progression rate from nr-AxSpA to AS	–	3.84% per year	–
Natural history: annual rate of BASFI progression	0.084 points (ABILITY-1[58])	0.07 points (Kobelt *et al.* 2004[152])	0.07 points (Kobelt *et al.* 2004[152])
AEs included; annual probability/rate	Only tuberculosis AEs and non-tuberculosis SAEs included; excess rate for adalimumab 7.3% for non tuberculosis SAEs and 0.16% for tuberculosis AEs annually (ABILITY-1[58] trial). Same rate applied to all anti-TNFs	No AEs included	No AEs included
HRQoL algorithm	$0.922 - 0.039 \times$ BASDAI $- 0.041 \times$ BASFI (ABILITY-1[58])	$2.1262 - 0.1323 \times$ BASFI $- 0.2450 \times$ BASDAI (RAPID-axSpA[64] study)	$0.919 - 0.00431 \times$ BASFI $+ 0.000788 \times$ BASDAI $+ 0.0000511 \times$ BASFI2 $- 0.0000194 \times$ BASDAI2 $- 0.00102 \times$ Age $+ 0.0478 \times$ male $- 0.0000754 \times$ BASDAI $\times$ BASFI (1031 study[166])
Annual health-care resource use costs	£1124.62 × exp (0.264 × BASDAI) (OASIS[118])	1909.33 × exp (0.1832 × BASFI) (NICE TA143[17])	• BASDAI score of < 4: annual cost: £151.96 • 4 ≤ BASDAI score < 6: annual cost: £311.08 • BASDAI score of ≥ 6: annual cost: £1039.16 (Rafia *et al.* 2012[168])

Comparison of disease costs assumed for the ankylosing spondylitis and non-radiographic axial spondyloarthritis populations

A variety of alternative regressions were applied across the submissions to estimate the annual disease costs associated with BASDAI and BASFI scores. Merck Sharp & Dohme[37] and UCB[35] used the same exponential regression function estimated by LRiG, uprated to current prices, as part of NICE TA143[17] based on the OASIS study.[118]

Regression in NICE TA143[17] based on OASIS data and cost element uprated to current prices:
£1902.492 × exp(0.1832 × BASFI).

AbbVie undertook their own reanalysis of the OASIS data set based on current prices. In their base-case regression an exponential model based on BASDAI was assumed. However, results from separate linear and exponential models were also presented.

Base-case regression used by AbbVie:[34]

$$\text{Exp BASDAI: } £1124.619 \times \exp(0.264 \times \text{BASDAI}). \quad (14)$$

Alternative regressions presented by AbbVie:

$$\text{Linear BASFI: } £520.32102 + £804.64642 \times \text{BASFI} \quad (15)$$

$$\text{BASDAI: } £118.47088 + £943.21394 \times \text{BASDAI} \quad (16)$$

$$\text{Exp BASFI: } £1284.186 \times \exp(0.213 \times \text{BASFI}). \quad (17)$$

The submission by Pfizer[36] was based on a recent UK study by Rafia *et al.*[168] Rather than employing a regression approach, the manufacturer used results based on a categorical analysis of the annual costs for BASDAI: BASDAI score of $< 4 = £151.96$, $4 \leq$ BASDAI score $< 6 = £311.08$; and BASDAI score of $\geq 6 = £1039.16$.

However, the paper by Rafia *et al.*[170] also specified a separate two-part regression function which was not included within the Pfizer submission[36] but is used in the subsequent comparisons of regressions to provide a more comparable approach to assessing the alternative costs sources used across the manufacturer's submissions and the predictions across a range of different BASDAI and BASFI scores.

Two-part model in Rafia *et al.*[168]:

Logistic regression model to derive probability of incurring costs:

$$2.71795 + 0.16716 \times \text{BASFI} + 0.37053 \times \text{BASDAI} - 0.02468 \times \text{BASFI} \times \text{BASDAI} + 0.33778 \times \text{Male} - 0.04389 \times \text{Age} - 0.01373 \times \text{Disease Duration} \quad (18)$$

Generalised linear model to obtain 3-month costs:

$$6.79876 + 0.27548 \times \text{BASFI} + 0.13265 \times \text{BASDAI} - 0.01602 \times \text{BASFI} \times \text{BASDAI} + 0.46458 \times \text{Male} - 0.01656 \times \text{Age} + 0.00381 \times \text{Disease Duration} \quad (19)$$

Figures 22 and *23* provide a comparison of the predictions from the alternative cost regressions using the separate sources identified across the manufacturer models. The baseline characteristics (BASDAI, BASFI, age and disease duration) are derived from a weighted average of the baseline characteristics of the clinical trials for the AS population from the manufacturer's submissions.

In *Figure 22*, BASDAI scores are held constant at the mean value and the impact of varying BASFI across the range (0–10 scale) are reported. In *Figure 23*, BASFI scores are held constant at the mean value and the impact of varying BASDAI across the range (0–10 scale) are reported.

Figures 24 and *25* compare the alternative regression functions reported in the submission by AbbVie based on their reanalysis of the OASIS study.

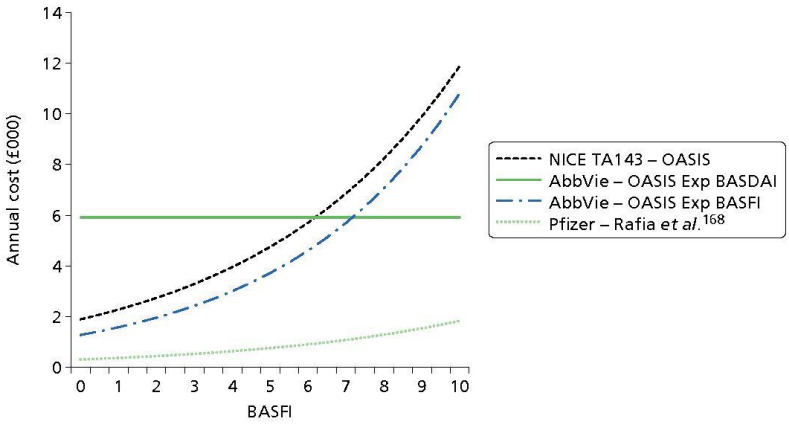


FIGURE 22 Comparison of main manufacturer cost regressions: assuming constant BASDAI score.

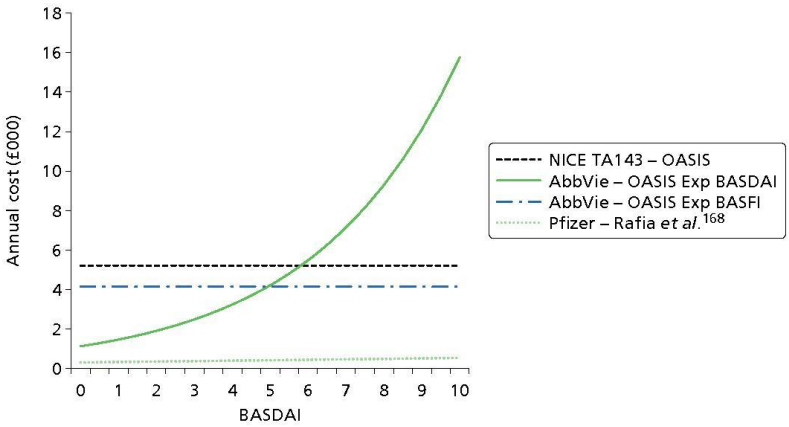


FIGURE 23 Comparison of main manufacturer cost regressions: assuming constant BASFI score.

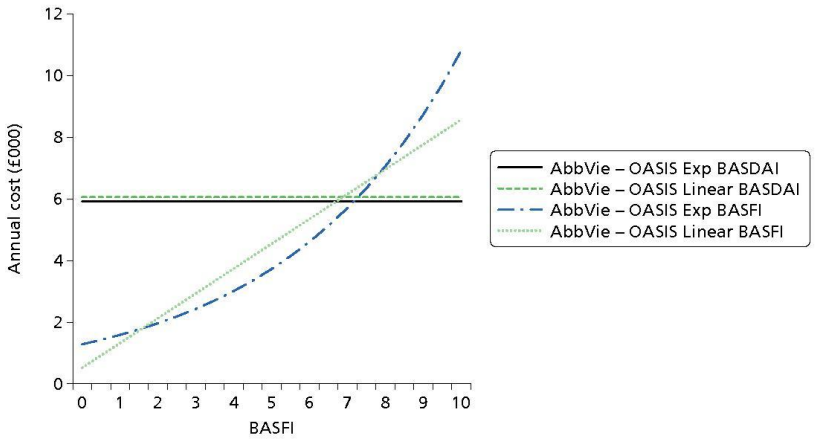


FIGURE 24 Comparison of AbbVie cost regressions: assuming constant BASDAI score.

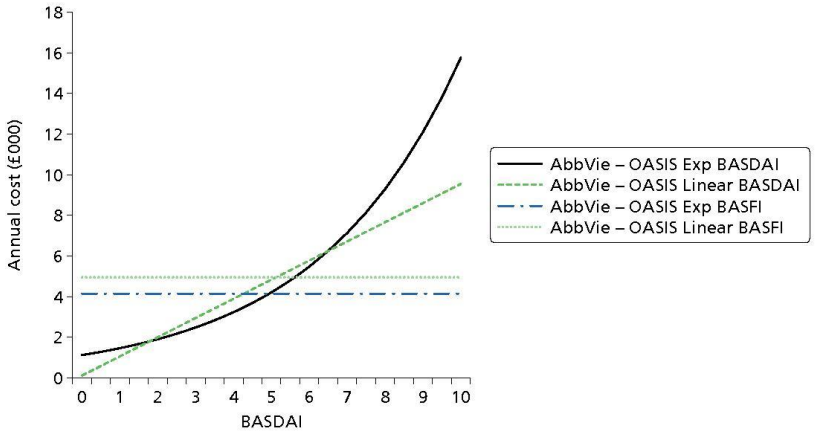


FIGURE 25 Comparison of AbbVie cost regressions: assuming constant BASFI score.

TABLE 148 Cost inputs in manufacturer's submissions (AS and nr-AxSpA population)

Parameter	Merck Sharp & Dohme economic model[37] (infliximab and golimumab)	AbbVie economic model[34] (adalimumab)	UCB economic model[35] (certolizumab)	Pfizer economic model[36] (etanercept)
Administration costs	Subcutaneous therapies: no administration cost	Subcutaneous therapies: no administration cost	Subcutaneous therapies: £49 cost of nurse training for self-administration (PSSRU[179])	Subcutaneous therapies: £49 cost of nurse training for self-administration (PSSRU[179])
	Intravenous therapies: cost of £109 per administration (no reference provided)	Intravenous therapies: cost of £99 per administration (no reference provided)	Intravenous therapies: cost of £398 per administration (PSSRU[179])	Intravenous therapies: cost of £302 per administration (NICE TA143[17])
Doses and unit costs	Costs estimated in line with licensed doses	Costs estimated in line with licensed doses	Costs estimated in line with licensed doses	Costs estimated in line with licensed doses
	PAS included for certolizumab and golimumab	PAS included for golimumab, not included for certolizumab	PAS included for certolizumab and golimumab	PAS included for certolizumab and golimumab
	Infliximab dosage: average weight of 70 kg assumed (four vials), subsequent administration every 7 weeks	Infliximab dosage: average weight of 81.1 kg assumed (five vials), subsequent administration every 6 weeks	Infliximab dosage: average weight of 81.7 kg assumed (4.88 vials), subsequent administration every 7 weeks	Infliximab dosage: average weight of 76.4 kg assumed (four vials), subsequent administration every 6 weeks
Monitoring costs	Short-term treatment costs applied in first cycle only for CC and anti-TNFs. Costs were informed by key opinion leader interviews. Anti-TNFs: £873.2. CC: £1459.5	Initiation and quarterly monitoring costs included. Common for all anti-TNFs comparators (York Model TA199[175]). Initiation: £470.09. Monitoring: £110.98 per cycle	No monitoring costs included	No monitoring costs included in the base case
Annual health-care resource use costs	1902.49 × exp (0.1832 × BASFI) (NICE TA143[17])	£1124.619 × EXP (0.264 × BASDAI) (OASIS[118])	1909.33 × exp (0.1832 × BASFI) (NICE TA143[17])	• BASDAI score of < 4: annual cost £151.96 • 4 ≤ BASDAI score < 6: annual cost £311.08 • BASDAI score of ≥ 6: annual cost £1039.16 (Rafia *et al.* 2012[168])

TABLE 149 Withdrawal inputs in manufacturer's submissions (AS and nr-AxSpA population)

Parameter	Merck Sharp & Dohme economic model[37] (infliximab, golimumab)	AbbVie economic model[34] (adalimumab)	UCB economic model[35] (certolizumab)	Pfizer economic model[36] (etanercept)
Annual long-term rate of anti-TNF withdrawal: AS population	6.1% (GO-RAISE[90] study, data of patients on treatment with golimumab from week 24 to week 256), common rate for all anti-TNFs	Time-dependent discontinuation rate; log-normal model fitted to adalimumab week-12 responder data up to week 260 (ATLAS[61]). Fewer than 15% of week-12 responders were projected to stay on treatment at year 40 for AS, a common rate for all anti-TNFs	7% (NICE TA143), common rate for all anti-TNFs	Exponential model fitted to etanercept data; model translates to 11% annual discontinuation for etanercept. Hazard ratios applied for other anti-TNFs (Glintborg 2010).[112] Annual discontinuation: • infliximab: 14.3% • golimumab: 12.3% • adalimumab: 12.3% • certolizumab: 12.3%
Annual long-term rate of anti-TNF withdrawal: nr-AxSpA population	Not applicable	Time-dependent discontinuation; log-normal model fitted to adalimumab week-12 responder data up to week 156 (ABILITY-1[58]). Fewer than 10% of week-12 responders were projected to stay on treatment at year 40, a common rate for all anti-TNFs	7% (NICE TA143), common rate for all anti-TNFs	Exponential model fitted to etanercept week-12 responder data; model translates to 5% annual discontinuation for etanercept. Hazard ratios applied for other anti-TNFs (Glintborg 2010).[112] Annual discontinuation: • infliximab: 6.5% • golimumab: 5.6% • adalimumab: 5.6% • certolizumab: 5.6%

TABLE 150 Adverse events inputs in manufacturer's submissions: AS population

Parameter	Merck Sharp & Dohme economic model[37] (infliximab and golimumab)	AbbVie economic model[34] (adalimumab)	UCB economic model[35] (certolizumab)	Pfizer economic model[36] (etanercept)
AEs included; annual probability	SAEs and ISRs included. CC rates from GO-RAISE[90] study at 24 weeks. OR of SAEs and ISRs from the NMA applied for each anti-TNF	Only infectious AEs included; excess proportion for adalimumab was 29.7% annually (ATLAS[61] trial)	No AEs included	Only serious infections included. Annual probability: 3.8% (312-EU[146])
	Annual probability (%) of SAEs: • placebo: 7.6 • infliximab: 21.4 • golimumab: 5.4 • adalimumab: 6.8 • certolizumab: 13.4 • etanercept: 20.5	Same rate applied to all anti-TNFs		Relative effects for other anti-TNF agents were applied in the model, obtained from a published NMA (Singh 2011)[137]
	Annual probability (%) of ISRs: • placebo: 19.7 • infliximab: 24.3 • golimumab: 51.0 • adalimumab: 38.5 • certolizumab: 38.5 • etanercept: 52.6			Annual probability (%): • infliximab: 4.1 • golimumab: 3.3 • adalimumab: 3.6 • certolizumab: 13.9 • etanercept: 3.8
Unit cost of AE	Cost per serious AE episode (weighted average): £214.26 anti-TNFs, £397.32 for CC (GO-RAISE[90]). Cost of injection site reaction £94.18 per episode	Cost per infectious AE episode: £45 (one GP visit assumed per infectious AE)	–	Cost per serious infection episode: £1457 (weighted average) (NHS Reference Costs 2012/13)[176]
Disutility of AE	Only disutility associated with SAEs applied; utility decrement of 0.01 applied for one cycle (NICE TA233[33])	No disutility applied	–	0.156 for 28 days

NMA, network meta-analysis.

TABLE 151 Adverse events inputs in manufacturer's submissions: nr-AxSpA population

Parameter	AbbVie economic model[34] (adalimumab)	UCB economic model[35] (certolizumab)	Pfizer economic model[36] (etanercept)
AEs included; annual probability	Only tuberculosis AEs and non-tuberculosis SAEs included; the excess percentage for adalimumab was 7.3% for non-tuberculosis SAEs and 0.16% for tuberculosis AEs annually (ABILITY-1[58] trial), same rate applied to all anti-TNFs	No AEs included	No AEs included
Unit cost of AE	Non-tuberculosis SAEs: £4216 per episode (NHS Reference Costs 2012/13)[176] Tuberculosis AEs: £6559.76 per episode (Botteman 2007)[162]	–	–
Disutility of AE	No disutility applied	–	–

Appendix 12 Extended synthesis models

In this appendix we describe in more detail the data and modelling approaches implemented in *Chapter 5*. Note that while this appendix aims to provide a methodological description of methods, a full description of findings and its interpretations are in *Chapter 5*.

General aspects of implementation and software

The synthesis was conducted from a Bayesian perspective, using WinBUGS (a MCMC simulation based software for Bayesian inference). For burn-in, we ran 100,000 simulations and another 100,000 were used in inferences. Convergence was assessed by running two chains and convergence was assumed if the Gelman–Rubin statistic was equal to 1. Goodness of fit was assessed using the DIC.[170] Models with smaller DIC are better supported by the data. In the presence of autocorrelation, the MCMC simulation for inference was increased to 200,000 and a thin of 20 was applied (yielding a sample for inference of 10,000 for each chain).

The main synthesis models will pool differences between treatment and control in change scores from baseline (BASDAI and BASFI). The treatment associated with the lowest (most negative) mean change score is expected to be best. However, it is important to quantify the uncertainty around the estimates and for this reason SDs will be reported alongside expected values. When ORs are presented, median values instead of means were used to summarise inferences.

When possible, meta-regression analyses were conducted to evaluate potential treatment effect modifiers. Meta-regression is a tool aimed at examining the impact of variables on effect size using regression-based techniques. In these explorations, the following baseline characteristics were considered: BASDAI score, BASFI score, age, sex, duration of symptoms (years) and CRP level.

Relative effectiveness estimates for models assuming exchangeability across treatments (model A5) are based on the predictive distribution, representing the distribution of the data averaged over all possible parameter values. This summary statistic best reflects the impact of uncertainty in the parameters of the model and is here judged as a more appropriate basis to be used in the decision model.[171]

Modelling approach A

Brief description of the data

Based on study populations and follow-up (i.e. around 12 weeks in duration), 16 of the RCTs are considered directly relevant to the decision problem for the AS population (studies 1 to 16 in *Table 152*). One of these studies did not report BASDAI or BASFI outcomes (study 3) and thus could not be included in analyses. The 15 remaining studies reported at least one outcome measure: BASDAI 50 score and/or change from baseline on BASDAI and BASFI scores.

This modelling approach directly evaluates relative treatment effects, that is log OR for BASDAI 50 response and the difference between treatment and placebo in change in BASDAI and BASFI scores from baseline. The data set analysed is shown in *Table 153*.

APPENDIX 12

TABLE 152 Evidence on BASDAI and BASFI score-related outcomes for the AS population

Study number	Trial name	Treatment	Number in treatment group	Number in placebo group	BASDAI 50 score	Change BASDAI score	Change BASFI score
1	Hu 2012[55]	1	26	20		✗	✗
2	Huang 2014[56]	1	229	115	✗	✗	✗
3	Lambert 2007[57]	1	38	44			
4	ATLAS 2006[61]	1	208	107	✗	✗	
5	RAPID-axSpA 2014[64]	2	121	57	✗	✗	✗
6	Barkham 2010[71]	3	20	20	✗	✗[a]	✗[a]
7	Davis 2003[72]	3	138	139		✗	✗
8	Dougados 2011[74]	3	39	43	✗	✗	✗
9	Gorman 2002[79]	3	20	20			✗
10	Calin 2004[83]	3	45	39		✗	✗
11	van der Heijde 2006[86]	3	305	51	✗		
12	GO-RAISE 2008[90]	4	138	78	✗		✗
13	Bao 2012[96]	4	108	105	✗		✗
14	Braun 2002[98]	5	34	35	✗	✗[a]	✗[a]
15	Marzo-Ortega 2005[100]	5	28	14		✗	✗[a]
16	Van den Bosch 2002[101]	5	9	12		✗[a]	✗[a]

a Do not report any measure of dispersion (such as SDs).
Treatment: 1, adalimumab; 2, certolizumab (certolizumab 200 mg and/or certolizumab 400 mg); 3, etanercept (etanercept 25 mg and/or etanercept 50 mg); 4, golimumab 50 mg; 5, infliximab.
Note that some studies only report one of the BASDAI measures. For example, the golimumab trials (studies 12 and 13) only report BASDAI 50 and not the absolute change in this score.

TABLE 153 Modelling approach A: data

Study, j	Treatment, t	Outcome, o	y	SE
1	1	1	–	–
2	1	1	1.61	0.28
3	1	1	–	–
4	1	1	1.47	0.30
5	2	1	1.79	0.42
6	3	1	2.30	1.13
7	3	1	–	–
8	3	1	1.04	0.48
9	3	1	–	–
10	3	1	–	–
11	3	1	1.78	0.37
12	4	1	1.47	0.36
13	4	1	2.34	0.50
14	5	1	2.45	0.69

TABLE 153 Modelling approach A: data (*continued*)

Study, j	Treatment, t	Outcome, o	y	SE
15	5	1	–	–
16	5	1	–	–
1	1	2	–1.60	0.67
2	1	2	–1.40	0.22
3	1	2	–	–
4	1	2	–1.80	0.28
5	2	2	–1.45	0.36
6	3	2	–1.87	0.90[a]
7	3	2	–1.91	0.26
8	3	2	–1.20	0.44
9	3	2	–	–
10	3	2	–1.87	0.49
11	3	2	–	–
12	4	2	–	–
13	4	2	–	–
14	5	2	–2.60	0.69[a]
15	5	2	–1.73	0.70
16	5	2	–2.97	1.26[a]
1	1	3	–0.90	0.68
2	1	3	–1.28	0.20
3	1	3	–	–
4	1	3	–	–
5	2	3	–1.10	0.37
6	3	3	–1.56	0.93[a]
7	3	3	–1.34	0.29
8	3	3	–1.20	0.40
9	3	3	–2.20	0.92
10	3	3	–1.73	0.45
11	3	3	–	–
12	4	3	–1.50	0.27
13	4	3	–1.37	0.32
14	5	3	–2.00	0.71[a]
15	5	3	–1.82	1.00[a]
16	5	3	–3.21	1.28[a]

a No SD was reported in the original studies, the highest SD from the other trials was used as a conservative estimate.
Outcome: 1, logOR for BASDAI 50; 2, difference between treatment and placebo on change in BASDAI from baseline; 3, difference between treatment and placebo on change in BASFI from baseline.
Treatment: 1, adalimumab; 2, certolizumab; 3, etanercept; 4, golimumab; 5, infliximab.
BASDAI, results from individual studies on difference between treatment and placebo in change from baseline in BASDAI scores; BASFI, results from individual studies on difference between treatment and placebo in change from baseline in BASFI scores; SE, SE associated with each outcome.

Description of synthesis methods for modelling approach A

Consider we have available information on J trials comparing an individual treatment, k (out of the total number of treatments T) to placebo. Trials report one or more outcomes, *o*. Information on outcome *o* for treatment k in a study j is represented by y_{jko} and is used alongside the SE for this measure, se^2_{jko}. In common with the approach implemented in *Chapter 3*, all outcomes are here assumed normally distributed, with mean θ_{jko}. We implemented alternative models that differ in the way treatment effects are considered; a summary of each is presented below. Note that at this stage each outcome was synthesised independently.

Model A1 (treatments; independent, studies; fixed effect): this model considers the *j* treatments to be independent, that is it assumes the effects to differ between treatments, d[k,o]. This is a fixed effect model in that multiple studies evaluating the same treatment are considered to measure the same treatment effect.

The model used was:

Likelihood:

$$y_{jko} \sim \text{dnorm}(\theta_{jko}, se^2_{jko}). \quad (20)$$

Model:

$$\theta_{jko} = d[k, o]. \quad (21)$$

Priors:

$$d[k, o] \sim N(0, 0.001). \quad (22)$$

Model A2 (treatments; independent, studies; random effects): this model differs from A1 in that a random effect is assumed to describe the findings of multiple studies evaluating the same treatment.

The model used was:

Likelihood:

$$y_{jko} \sim \text{dnorm}(\theta_{jko}, se^2_{jko}). \quad (23)$$

Model:

$$\theta_{jko} \sim N(d[k, o], \sigma^2_o). \quad (24)$$

Priors:

$$d[k, o] \sim N(0, 0.001);\ \sigma^2_o \sim dunif(0,10). \quad (25)$$

The random effect is defined using a variance parameter for each outcome but common across treatments, σ^2_o.

Model A3 (treatments; equal, studies; fixed effect): this model differs from model A1 in that treatments are not assumed to differ. The model thus evaluates a common relative effectiveness for all anti-TNFs, d[o], for each outcome.

The model used was:

Likelihood:

$$y_{jko} \sim \text{dnorm}(\theta_{jko}, se^2_{jko}). \quad (26)$$

Model:

$$\theta_{jko} = d[o]. \quad (27)$$

Priors:

$$d[k, o] \sim N(0, 0.001). \quad (28)$$

Model A4 (treatments; equal, studies; random effects): this model differs from model A3 in that a random effect is assumed to describe the findings of multiple studies evaluating the same treatment.

The model used was:

Likelihood:

$$y_{jko} \sim \text{dnorm}(\theta_{jko}, se^2_{jko}). \quad (29)$$

Model:

$$\theta_{jko} \sim N(d[o], \sigma^2_o). \quad (30)$$

Priors:

$$d[o] \sim N(0, 0.001); \ \sigma^2_o \sim dunif(0, 10). \quad (31)$$

Model A5 (treatments; exchangeable, studies; fixed effect): this model differs from model A1 in that a random effect is used to describe any differences between treatments (exchangeability is assumed). This model thus assumes the treatments to have a similar, but not equal, effectiveness; there are differences between the effectiveness of treatments that we may not be able to explain but that we should consider.

The model used was:

Likelihood:

$$y_{jko} \sim \text{dnorm}(\theta_{jko}, se^2_{jko}). \quad (32)$$

Model:

$$\theta_{jko} = d[k, o] \quad (33)$$

$$d[k, o] \sim N(D[o], \gamma_o^2). \quad (34)$$

Priors:

$$D[o] \sim N(0, 0.001);\ \gamma_1^2 \sim dunif(0, 2);\ \ \gamma_2^2, \gamma_3^2 \sim dunif(0, 10). \quad (35)$$

The parameter γ_o^2 is the variance parameter defining the random effect across treatment. The priors differ for outcome 1 because this is a log odds, while outcomes 2 and 3 are assumed continuous measures.

Within this modelling approach we explored potential heterogeneity in treatment effects using metaregression (i.e. potential treatment effect modifiers). We did so by extended the modelling approach in model A1 to include treatment effect interactions with baseline characteristics (centred on their means when relevant). We have explored the inclusion of alternative covariates by evaluating the DIC associated with alternative models.

Results of modelling approach A

The results of each modelling approach are shown in *Table 154*.

From model A5, drug-specific estimates can be retrieved (*Table 155*). Within this mode drug-specific inferences will borrow strength from the common class effect and estimates are thus shrunken towards the mean of this class effect (i.e. estimates are closer to the value reported for the class in *Table 153*).

Explorations of heterogeneity suggested only sex potentially modified the effect of anti-TNF treatment, specifically for change in BASDAI as outcome; however, when sex is used together with all covariates, such evidence on effect modification disappears (results not shown but available on request).

TABLE 154 Modelling approach A: results

	A1. Treatment: independent; studies: fixed effect	A2. Treatment: independent; studies: random effects	A3. Treatment: common; studies: fixed effect	A4. Treatment: common; studies: random effects	A5. Treatment: exchangeable; studies: fixed effect
Outcome 1: OR on BASDAI 50 score	***Median (SD)***	***Median (SD)***	***Median (SD)***	***Median (SD)***	***Median (SD)***
Adalimumab	4.71 (1.00)	4.69 (6.11)	5.21 (0.72)	5.30 (0.98)	5.34 (9.79)[a]
Certolizumab	6.02 (3.33)	6.04 (22.87)			
Etanercept	4.73 (1.43)	4.72 (3.32)			
Golimumab	5.86 (1.81)	6.10 (7.45)			
Infliximab	11.9 (11.94)	12.10 (44.00)			
σ_1	–	0.31 (0.30)	–	0.15 (0.14)	–
D_1	–	–	–	–	1.69 (0.23)
γ_1	–	–	–	–	0.27 (0.28)
Outcome 2: change in BASDAI score	***Mean (SD)***	***Mean (SD)***	***Mean (SD)***	***Mean (SD)***	***Mean (SD)***
Adalimumab	–1.56 (0.16)	–1.57 (0.27)	–1.66 (0.11)	–1.67 (0.15)	
Certolizumab	–1.45 (0.37)	–1.46 (0.51)			
Etanercept	–1.76 (0.20)	–1.73 (0.28)			–1.70 (0.87)[a]
Golimumab	N/A	N/A			
Infliximab	–2.28 (0.46)	–2.27 (–2.28)			
σ_2	–	0.25 (0.24)	–	0.25 (0.19)	–
D_2	–	–	–	–	–1.63 (0.57)
γ_2	–	–	–	–	0.43 (0.63)
Outcome 3: change in BASFI score	***Mean (SD)***	***Mean (SD)***	***Mean (SD)***	***Mean (SD)***	***Mean (SD)***
Adalimumab	–1.22 (0.18)	–1.18 (0.29)	–1.38 (0.11)	–1.39 (0.13)	
Certolizumab	–1.10 (0.37)	–1.11 (0.47)			
Etanercept	–1.48 (0.19)	–1.50 (0.24)			–1.41 (0.49)[a]
Golimumab	–1.45 (0.20)	–1.44 (0.29)			
Infliximab	–2.16 (0.53)	–2.17 (0.56)			
σ_3	–	0.22 (0.19)	–	0.14 (0.12)	–
D_3	–	–	–	–	–1.40 (0.22)
γ_3	–	–	–	–	0.28 (0.33)
DIC	52.4	57.0	39.1	44.3	43.6

N/A, not applicable.
a Predictive distribution.
Outcome: 1, logOR for BASDAI 50; 2, difference between treatment and placebo on change in BASDAI from baseline; 3, difference between treatment and placebo on change in BASFI from baseline; σ_2 is the variance parameter for outcome o of the random effect across studies; D_o is the mean of the random effect for outcome o; γ_3 is the variance parameter for outcome o of the random effect across treatments.

TABLE 155 Shrunken estimates of treatment effect from model A5

	Model A5		
	Outcome 1: OR on BASDAI 50 score, median (SD)	Outcome 2: change in BASDAI score, median (SD)	Outcome 3: change in BASFI score, median (SD)
Adalimumab	5.05 (0.87)	−1.60 (0.15)	−1.31 (0.16)
Certolizumab	5.42 (1.71)	−1.59 (0.26)	−1.31 (0.23)
Etanercept	5.13 (1.08)	−1.72 (0.17)	−1.43 (0.15)
Golimumab	5.47 (1.25)	−1.69 (0.84)	−1.42 (0.16)
Infliximab	5.70 (3.30)	−1.88 (0.34)	−1.55 (0.33)

Modelling approach B

In the previous section the two outcomes based on BASDAI scores were synthesised separately; however, BASDAI 50 is the probability of having a reduction in BASDAI score of 50%, and thus it should be possible to relate the proportion of BASDAI 50 responders to the change in absolute BASDAI scores from baseline observed in each study. Within this section, we use this structural relation within the synthesis, allowing change scores from baseline to be informed not only from direct data on this quantity but also from data on BASDAI 50.

Brief description of the data

The model implemented here pools the change in BASDAI score from baseline to evaluate the difference between treatment and placebo, using evidence reported in trials directly on the change scores for each arm and also data on BASDAI 50. The data modelled within this approach are shown in *Table 156*.

Description of synthesis methods

Consider we have available information on J trials comparing an individual treatment, k (out of the total number of treatments T) to placebo. Study j may report y_{jk}, the mean change in BASDAI from baseline, alongside the SE for this measure, se_{jk}. The likelihood for the data on change score was assumed normally distributed and was expressed as:

$$y_{jk} \sim N(\theta_{jk}, se_{jk}^2). \tag{36}$$

The mean of this distribution was the treatment effects, θ_{jk}, defined as the sum of the change score for the placebo arm plus the difference in change score for the treatments:

$$\theta_{jk} = \mu_j + \delta_{jk}. \tag{37}$$

Some studies also reported the number of responders to BASDAI 50 (a 50% reduction in BASDAI score), r_{jk}, out of the total number of individuals in the study, n_{jk}. The likelihood for the BASDAI 50 data was binomially distributed and thus expressed as:

$$r_{jk} \sim Bin(p_{jk}, n_{jk}). \tag{38}$$

TABLE 156 Data used in modelling approach B and C

s[]	t[]	n[]	r[]	b[]	sd[]	y[]	y.se[]	y.f[]	y.f.se[]
1	1	20	N/A	6.2	1.1	–2	0.560	–1	0.34
1	2	26	N/A	5.9	1.4	–3.6	0.377	–1.9	0.29
2	1	115	19	6.2	1.4	–1.4	0.177	–0.47	0.15
2	2	229	114	6	1.4	–2.8	0.126	–1.75	0.13
3	1	44	N/A	6.5	1.6	N/A	N/A	N/A	N/A
3	2	38	N/A	6.2	1.7	N/A	N/A	N/A	N/A
4	1	107	17	6.3	1.7	–0.8	0.2	N/A	N/A
4	2	208	94	6.3	1.7	–2.6	0.2	N/A	N/A
5	1	57	8	6.4	1.9	–1.0	0.3	–0.6	0.30
5	3	121	50	6.36	1.54	–2.45	0.206	–1.7	0.21
6	1	20	1	5.46	1.74	–0.1	0.632	0.21	0.71
6	4	20	7	6.05	1.71	–1.97	0.645	–1.35	0.56
7	1	139	N/A	5.96	1.65	–0.45	0.18	–0.33	0.21
7	4	138	N/A	5.81	1.76	–2.36	0.19	–1.67	0.20
8	1	43	10	5.8	1.5	–1.4	0.305	–1	0.27
8	4	39	18	6.4	1.2	–2.6	0.320	–2.2	0.29
9	1	20	N/A	N/A	N/A	N/A	N/A	–0.1	0.49
9	4	20	N/A	N/A	N/A	N/A	N/A	–2.3	0.36
10	1	39	N/A	5.86	2.05	–0.85	0.35	–0.33	0.31
10	4	45	N/A	6.1	1.87	–2.72	0.34	–2.06	0.33
11	1	51	10	6.11	1.37	N/A	N/A	N/A	N/A
11	4	305	180	6.09	1.69	N/A	N/A	N/A	N/A
12	1	78	12	6.6	1.49	N/A	N/A	0.1	0.19
12	5	138	61	6.6	1.49	N/A	N/A	–1.4	0.19
13	1	105	5	6.5	1.54	N/A	N/A	0.11	0.20
13	5	108	37	6.6	1.31	N/A	N/A	–1.26	0.25
14	1	35	3	6.3	1.4	–0.6	0.478	–0.1	0.55
14	6	34	18	6.5	1.2	–3.2	0.495	–2.1	0.44
15	1	14	N/A	6.57	2.05	–1.38	0.564	0.1	0.88
15	6	28	N/A	6.45	1.87	–3.11	0.42	–1.72	0.49
16	1	12	N/A	5.27	2.05	–0.26	0.816	1.3	0.95
16	6	9	N/A	5.89	1.87	–3.23	0.961	–1.91	0.86

N/A, not applicable.
Outcome: 1, logOR for BASDAI 50; 2, difference between treatment and placebo on change in BASDAI from baseline; 3, difference between treatment and placebo on change in BASFI from baseline
s[] = study, t[] = treatment; 1 = placebo, 2 = adalimumab, 3 = certolizumab pegol, 4 = etanercept, 5 = golimumab, 6 = infliximab; n[] = total number of patients, r[] = number of patients showing a BASDAI 50 response, y[] = vector of results from studies on change from baseline on BASDAI score; y.se[] = standard error associated with each y; y.f[] = vector of results from studies on change from baseline on BASFI score; y.f.se[] = standard error associated with each y.f.

APPENDIX 12

Consider the BASDAI score at baseline for study j and treatment k, X_{jk}, as normally distributed, with a mean score at baseline of ν_{jk} and variability on BASDAI score at baseline represented by σ^2_{jk}:

$$X_{jk} \sim N(\nu_{jk}, \sigma^2_{jk}). \tag{39}$$

The probability parameter of the binomial distribution can be expressed as a function of the baseline and final BASDAI scores:

$$p_{jk} = P\left[\frac{Y_{jk}}{X_{jk}} < -0.5\right] = P[Y_{jk} + X_{jk}/2 < 0]. \tag{40}$$

This can help us establish an algebraic relation between p_{jk} and the change score Y_{jk}, for a given baseline value, X_{jk}. This requires some assumptions over the distribution of scores, which are described next.

Across individuals, the BASDAI scores at baseline and the change score are assumed correlated and are described using a bivariate normal distribution:

$$\begin{pmatrix} X_{1jk} \\ Y_{jk} \end{pmatrix} \sim N\left(\begin{pmatrix} \nu_{1jk} \\ \theta_{jk} \end{pmatrix}, \begin{pmatrix} \sigma^2_{jk} & \rho\sigma^2_{jk} \\ \rho\sigma^2_{jk} & \sigma^2_{jk} \end{pmatrix}\right). \tag{41}$$

For simplicity, the variability on BASDAI score at baseline, σ^2_{jk}, was assumed equal to that of the change score. The correlation parameter is represented by ρ.

We would like to further explore the following relationship:

$$P_{jk} = P[Y_{jk} + X_{jk}/2 < 0]. \tag{42}$$

To do so, first consider expressing Y by conditioning on the baseline value, $X_{jk} = x$ (for simplicity we will drop the *jk* subscript in the next few formulas):

$$Y|X \sim N(\theta + \rho(x - \nu), (1 - \rho^2)\sigma^2). \tag{43}$$

So, we can standardise and relate this probability to a standard Normal distribution

$$p_{jX_1=x} = P(Y_{jk} + x/2 < 0_{X_1=x}) = \Phi\left(\frac{-\left(\frac{x}{2} + \theta + \rho(x - \nu)\right)}{\sigma\sqrt{(1 - \rho^2)}}\right). \tag{44}$$

To obtain the joint distribution, one needs to average over $X_{jk} \sim N(\nu_{jk}, \sigma^2_{jk})$, which means integrating over this distribution with respect to x:

$$P_{jk} = \int_{-\infty}^{+\infty} \Phi\left(\frac{-\left(\frac{x}{2} + \theta + \rho(x - \nu)\right)}{\sigma\sqrt{(1 - \rho^2)}}\right) f_x(x) dx. \tag{45}$$

Note that one can express the expectation over the cdf of a normal distribution as:

$$E[\Phi(aX+b)] = \Phi\left(\frac{b+a\nu}{\sqrt{1-a^2\sigma^2}}\right) \text{ when } x \sim N(\upsilon, \sigma^2). \quad (46)$$

Here, $a = \frac{-(1/2+\rho)}{\sigma\sqrt{(1-\rho^2)}}$ and $b = \frac{-\theta+\rho\nu}{\sigma\sqrt{(1-\rho^2)}}$. Therefore:

$$p_{jk} = \Phi\left(\frac{-\theta+\rho\nu-(1/2+\rho)\nu}{\sigma\sqrt{(1-\rho^2)}\sqrt{1-\frac{(1/2-\rho)^2}{(1-\rho^2)}}}\right) = \Phi\left(-\frac{\theta+\nu/2}{\sigma\sqrt{5/4+\rho}}\right). \quad (47)$$

The relations established above thus allow the probability parameter from BASDAI 50 data to be expressed algebraically as a function of the change score:

$$\text{probit}(p_{jk}) = \frac{-\theta_{jk}-\nu_{jk}/2}{\sigma_{jk}\sqrt{5/4+\rho}}. \quad (48)$$

In computations, we used the mean score at baseline, v_{jk}, and the associated SD, σ_{jk}, as reported in the data (these were thus assumed known). The correlation between baseline and change score was estimated within the model by assuming this quantity to be independent of study but assumed to differ between placebo and anti-TNF treatments.

In what concerns the treatment effects, all trials in our evidence base compare against CC: $\delta_{jk} = d_k$. Our preferred approach to model these was to assume a common class effect (i.e. exchangeable effects across treatments, analogous assumption to model A5 above). This means:

$$d_k \begin{cases} =0 & \text{if } k=1 \\ \sim N(D, \sigma_{re}^2) & \text{if } k \neq 1 \end{cases}, \quad (49)$$

where k = 1 is standard care.

The priors used to implement this model were:

$$D \sim N(0, 0.001), \quad \mu_j \sim N(0, 0.001), \quad \rho_{pla} \sim U(-1,1), \rho_{anti-TNF} \sim U(-1,1). \quad (50)$$

WinBUGS code for modelling approach B

```
model{
        for (i in 1:10) {
                y[i] ~ dnorm(theta[i], y.prec[i])          #change in score
                theta[i] <- mu[s[i]] + d[t[i]]
                }
        for (i in 11:18) {
                r[i] ~ dbin(p[i], n[i])
                aux[i] <- equals(t[i],1)+1
                probit(p[i]) <- -(b[i]*0.5 + theta[i])/(pow(prec[i],-
0.5)*pow(5/4+rho[aux[i]],0.5))
                theta[i] <- mu[s[i]] + d[t[i]]
                }
        for (i in 19:28) {
                r[i] ~ dbin(p[i], n[i])
                y[i] ~ dnorm(theta[i], prec[i])            #change in score
                aux[i] <- equals(t[i],1)+1
                probit(p[i]) <- -(b[i]*0.5 + theta[i])/(pow(prec[i],-
0.5)*pow(5/4+rho[aux[i]],0.5))
                theta[i] <- mu[s[i]] + d[t[i]]
        }
        for (j in 1:14) {
                mu[j] ~ dnorm(0,0.001)
                }
        d[1] <- 0
        for (k in 2:6) {
                d[k] ~ dnorm(re,intau)
        }
        re ~ dnorm(0, 0.01)
        intau <- 1/tau
        tau <- pow(sd,2)
        sd ~ dunif(0,2)
        re.pred ~ dnorm(re,intau)
        rho[1] ~ dunif(-1,1)
        rho[2] ~ dunif(-1,1)
}
```

FIGURE 26 WinBUGS code for modelling approach B.

Results of modelling approach B

The summary results regarding relative treatment effects from this modelling approach are reported in *Table 157* for model B.

Drug-specific (shrunken) estimates from model B are shown in *Table 158*.

TABLE 157 Modelling approach B: results

	Estimated	Assumed[a]	Predicted	
	Difference in change score from baseline, mean (SD)	Probability of having a BASDAI 50 response, placebo, mean (SD)	Probability of having a BASDAI 50 response, anti-TNF, mean (SD)	OR for BASDAI 50 response, anti-TNF vs. placebo, median (SD)
Anti-TNFs	−1.91 (0.48)[b]	0.10 (–)	0.40 (0.08)	5.94 (4.06)
Other model summaries				
D	−1.91 (0.28)	–	–	–
γ	0.30 (0.28)	–	–	–
$\rho_{placebo}$	0.26 (0.33)	–	–	–
$\rho_{anti\text{-}TNF}$	0.69 (0.26)	–	–	–
DIC	146.3	–	–	–

a This figure is based on a BASDAI baseline score of 6.11 (SD 1.56) and a placebo change score of −0.61 (SD 1.44), which represent the average across trials (weighted by number of patients).
b Predictive distribution.

TABLE 158 Shrunken estimates of treatment effect from model B

Treatment	Change in BASDAI score, mean (SD)
Adalimumab	−1.77 (0.25)
Certolizumab	−2.01 (0.37)
Etanercept	−1.88 (0.18)
Golimumab	−1.92 (0.30)
Infliximab	−2.02 (0.32)

Modelling approach C

The models implemented here extend those in the previous section by adding the syntheses of changes in BASFI score. The data used are presented in *Table 156*.

Description of synthesis methods

Data on mean change in BASFI score reported in some of the studies available have been described as normally distributed (the likelihood):

$$y_{jk}^{BASFI} \sim N\left(\theta_{jk}^{BASFI}, \left(se_{jk}^{BASFI}\right)^2\right). \tag{51}$$

The treatment effects over BASFI θ_{jk}^{BASFI} were then defined as:

$$\theta_{jk}^{BASFI} = \mu_j^{BASFI} + \delta_{jk}^{BASFI}. \tag{52}$$

Treatment effects on BASFI were assumed correlated to those on BASDAI across trials:

$$\begin{pmatrix} \delta_{jk}^{BASDAI} \\ \delta_{jk}^{BASFI} \end{pmatrix} \sim N\left(\begin{pmatrix} d_k^{BASDAI} \\ d_k^{BASFI} \end{pmatrix}, \begin{pmatrix} \tau^2_{BASDAI} & \rho_m \tau^2_{BASDAI} \tau^2_{BASFI} \\ \rho_m \tau^2_{BASDAI} \tau^2_{BASFI} & \tau^2_{BASFI} \end{pmatrix}\right) \tag{53}$$

$$d_k^o \begin{cases} =0 & if\ k = 1 \\ \sim N(D_o, \sigma^2_{re,o}) & if\ k \neq 1 \end{cases}, \tag{54}$$

with o = {BASDAI, BASFI} and $k = 1$ is placebo.

The additional priors used to implement this model were:

$$D_o \sim N(0, 0.001), \quad \sigma^2_{re} \sim U(0,2) \quad \rho_m \sim U(-1,1). \tag{55}$$

The variation in treatment effects for both BASDAI and BASFI and the correlation parameter between these were estimated from the data. As in model B, we assumed exchangeability across the effects of the different treatments.

WinBUGS code for modelling approach C

```
model{
        for (i in 1:10) {
                y[i] ~ dnorm(theta[i,1], y.prec[i])              #change in score
                y.f[i] ~ dnorm(theta[i,2], y.prec.f[i])          #change in score BASFI
                }
        for (i in 11:14) {
                r[i] ~ dbin(p[i], n[i])
                aux[i] <- equals(t[i],1)+1
                probit(p[i]) <- -(b[i]*0.5 + theta[i,1])/(pow(prec[i],-
0.5)*pow(5/4+rho[aux[i]],0.5))
                y.f[i] ~ dnorm(theta[i,2], y.prec.f[i])          #change in score BASFI
                }
        for (i in 15:16) {
                r[i] ~ dbin(p[i], n[i])
                aux[i] <- equals(t[i],1)+1
                probit(p[i]) <- -(b[i]*0.5 + theta[i,1])/(pow(prec[i],-
0.5)*pow(5/4+rho[aux[i]],0.5))
                }
        for (i in 17:26) {
                r[i] ~ dbin(p[i], n[i])
                y[i] ~ dnorm(theta[i,1], prec[i])          #change in score
                aux[i] <- equals(t[i],1)+1
                probit(p[i]) <- -(b[i]*0.5 + theta[i,1])/(pow(prec[i],-
0.5)*pow(5/4+rho[aux[i]],0.5))
                y.f[i] ~ dnorm(theta[i,2], y.prec.f[i])          #change in score BASFI
        }
        for (i in 27:28) {
                y.f[i] ~ dnorm(theta[i,2], y.prec.f[i])          #change in score BASFI
        }
        for (i in 29:30) {
                r[i] ~ dbin(p[i], n[i])
                y[i] ~ dnorm(theta[i,1], prec[i])          #change in score
                aux[i] <- equals(t[i],1)+1
                probit(p[i]) <- -(b[i]*0.5 + theta[i,1])/(pow(prec[i],-
0.5)*pow(5/4+rho[aux[i]],0.5))
        }
        for (i in 1:30) {
                theta[i,1:2] ~ dmnorm(delta[i,1:2],B[1:2,1:2])
                delta[i,1] <- mu1[s[i]] + d1[t[i]]
                delta[i,2] <- mu2[s[i]] + d2[t[i]]
        }
        d1[1] <- 0
        d2[1] <- 0
        for (k in 2:6) {
                d1[k] ~ dnorm(re1,intau)
                d2[k] ~ dnorm(re2,intau)
        }
        B[1,1]<- 1/(pow(sd[1],2)*(1-pow(cor,2)))
        B[2,2]<- 1/(pow(sd[2],2)*(1-pow(cor,2)))
        B[1,2]<- -cor/(sd[1]*sd[2]*(1-pow(cor,2)))
        B[2,1]<- B[1,2]
        sd[1] ~ dunif(0,5)
```

FIGURE 27 WinBUGS code for modelling approach C. (*continued*)

```
                sd[2] ~ dunif(0,5)
                cor~dunif(0,1)
                for (j in 1:15) {
                        mu1[j] ~ dnorm(0,0.01)I(-5,5)
                        mu2[j] ~ dnorm(0,0.01)I(-5,5)
                        }
                re1 ~ dnorm(0, 0.01)I(-10,10)
                re.pred1 ~ dnorm(re1,intau)
                re2 ~ dnorm(0, 0.01)I(-10,10)
                re.pred2 ~ dnorm(re2,intau)
                intau <- 1/tau
                tau <- pow(sd.re,2)
                sd.re ~ dunif(0,2)
                rho[1] ~ dunif(0,1)
                rho[2] ~ dunif(0,1)
                for (k in 2:6) {
                        d1.pred[k] ~ dnorm(re1,intau)
                }
}
```

FIGURE 27 WinBUGS code for modelling approach C.

Results of modelling approach C

The results on differences between treatment and placebo on change score form baseline are reported in *Table 159*, both for BASDAI and BASFI scores.

TABLE 159 Modelling approach C: results

	Estimated	Assumed[a]	Predicted	
	Difference in change score from baseline, mean (SD)	Probability of having a BASDAI 50 response, placebo, mean (SD)	Probability of having a BASDAI 50 response, anti-TNF, mean (SD)	OR for BASDAI 50 response, anti-TNF vs. placebo, mean (SD)
Effect of anti-TNFs on BASDAI	−1.95 (0.30)	0.10 (–)	0.41 (0.05)	6.30 (1.56)
Effect of anti-TNFs on BASFI	−1.40 (0.28)	–	–	–
Other model summaries				
D_{BASDAI}	−1.99 (0.20)	–	–	–
D_{BASFI}	−1.40 (0.16)	–	–	–
γ_{BASDAI}	0.13 (0.10)	–	–	–
γ_{BASFI}	0.11 (0.09)	–	–	–
$\rho_{placebo}$	0.42 (0.26)	–	–	–
$\rho_{anti\text{-}TNF}$	0.71 (0.23)	–	–	–
ρ_m	0.51 (0.29)	–	–	–
σ^2_{re}	0.16 (0.14)	–	–	–
DIC	181.9	–	–	–

a Based on a BASDAI baseline score of 6.11 (SD = 1.56) and a placebo change score of −0.61 (SD = 1.44), which represent the average across trials (weighted by number of patients).

Drug-specific (shrunken) estimates from model C are shown in *Table 160*.

TABLE 160 Shrunken estimates of treatment effect from model C

Treatment	Change in BASDAI score, mean (SD)	Change in BASFI score, mean (SD)
Adalimumab	−1.89 (0.22)	−1.34 (0.17)
Certolizumab	−2.02 (0.28)	−1.36 (0.21)
Etanercept	−1.94 (0.18)	−1.43 (0.16)
Golimumab	−1.98 (0.25)	−1.42 (0.17)
Infliximab	−2.03 (0.27)	−1.49 (0.25)

Appendix 13 Synthesis of evidence on the non-radiographic axial spondyloarthritis population

This section analyses the evidence on the effectiveness of anti-TNFs on the nr-AxSpA population.

Brief description of the data

On the nr-AxSpA population, five RCTs were considered directly relevant to the decision problem (studies 17–21 in *Table 161*). All studies reported BASFI outcomes and one study did not report BASDAI 50 (study 21).

The data on these five studies are shown in *Table 162*.

TABLE 161 Evidence on BASDAI and BASFI-related outcomes for the nr-AxSpA population

Study number	Trial name	Treatment	Number in treatment group	Number in placebo group	BASDAI 50 score	Change BASDAI score	Change BASFI score
17	Haibel 2008[52]	Adalimumab	22	24	✗	✗	✗
18	ABILITY-1 2013[58]	Adalimumab	69	73	✗	✗	✗
19	RAPID-axSpA 2014[64]	Certolizumab pegol	46 + 51	50	✗	✗	✗
20	Dougados 2014[76]	ETA50	106	109	✗	✗	✗
21	Barkham 2009[50]	Infliximab	20	20		✗	✗

TABLE 162 Data on the nr-AxSpA population

s[]	t[]	n[]	r[]	b[]	sd[]	y[]	y.se[]	y.f[]	y.f.se[]
1	1	24	5	6.20	0.59	–1.20	7.79	–0.80	6.87
1	2	22	11	6.50	0.69	–2.70	6.30	–2.40	7.24
2	1	73	10	6.38	0.44	–1.10	19.00	–0.63	22.78
2	2	69	27	6.43	0.42	–2.20	11.04	–1.28	16.91
3	1	50	8	6.40	0.44	–1.50	6.25	–0.40	6.25
3	3	97	47	6.55	0.43	–3.35	11.64	–2.30	12.21
4	1	109	26	6.00	0.28	–1.30	11.11	–0.80	25.00
4	4	106	46	6.00	0.31	–2.00	11.11	–1.40	25.00
5	1	20	N/A	5.76	0.28	–0.75	3.42	–0.47	3.95
5	5	20	N/A	5.85	0.31	–3.41	3.12	–2.70	3.59

N/A, not applicable.
s[] = study, t[] = treatment: 1 = placebo, 2 = adalimumab, 3 = certolizumab pegol, 4 = etanercept, 5 = infliximab; n[] = total number of patients, r[] = number of patients showing a BASDAI 50 response, y[] = vector of results from studies on change from baseline on BASDAI score; y.se[] = standard error associated with each y; y.f[] = vector of results from studies on change from baseline on BASFI score; y.f.se[] = standard error associated with each y.f.

Description of approaches to the synthesis

To synthesise these data we used the same implementation and software specifications as described in *Appendix 12*. Analyses explored two different scenarios to consider these data:

- scenario 1: data from nr-AxSpA trials were considered in isolation
- scenario 2: data from AS population were also used, no difference between the populations was assumed.

All models implemented here jointly synthesise BASDAI and BASFI outcomes (our preferred modelling approach, C).

Results of the synthesis

Results of the analysis are in *Table 163*.

TABLE 163 Non-radiographic axial spondyloarthritis population: results

	Estimated	Assumed[a]	Predicted	
	Difference in change score from baseline, mean (SD)	Probability of having a BASDAI 50 response, placebo, mean (SD)	Probability of having a BASDAI 50 response, anti-TNF, mean (SD)	OR for BASDAI 50 response, anti-TNF vs. placebo, median (SD)
Scenario 1: data from nr-AxSpA trials				
Effect of anti-TNFs on BASDAI	−1.86 (0.79)	0.20 (–)	0.53 (0.13)	4.39 (6.59)
Effect of anti-TNFs on BASFI	−1.30 (0.84)	–	–	–
Other model summaries				
D_{BASDAI}	−1.86 (0.53)	–	–	–
D_{BASFI}	−1.30 (0.65)	–	–	–
γ_{BASDAI}	0.41 (0.43)	–	–	–
γ_{BASFI}	0.68 (0.53)	–	–	–
$\rho_{placebo}$	0.60 (0.27)	–	–	–
$\rho_{anti\text{-}TNF}$	0.57 (0.28)	–	–	–
ρ_m	0.51 (0.29)	–	–	–
σ^2_{re}	0.55 (0.29)	–	–	–
DIC	88.6	–	–	–

TABLE 163 Non-radiographic axial spondyloarthritis population: results (*continued*)

	Estimated	Assumed[a]	Predicted	
	Difference in change score from baseline, mean (SD)	Probability of having a BASDAI 50 response, placebo, mean (SD)	Probability of having a BASDAI 50 response, anti-TNF, mean (SD)	OR for BASDAI 50 response, anti-TNF vs. placebo, median (SD)
Scenario 2: data from AS and nr-AxSpA trials, no difference between the populations				
Effect of anti-TNFs on BASDAI	–1.97 (0.32)	0.20 (–)	0.55 (0.06)	4.94 (1.48)
Effect of anti-TNFs on BASFI	–1.37 (0.3)	–	–	–
Other model summaries				
D_{BASDAI}	–1.97 (0.20)	–	–	–
D_{BASFI}	–1.37 (0.18)	–	–	–
γ_{BASDAI}	0.12 (0.09)	–	–	–
γ_{BASFI}	0.18 (0.11)	–	–	–
$\rho_{placebo}$	0.50 (0.26)	–	–	–
$\rho_{anti\text{-}TNF}$	0.74 (0.22)	–	–	–
ρ_{m}	0.54 (0.29)	–	–	–
σ^2_{re}	0.19 (0.16)	–	–	–
DIC	269.0	–	–	–

a Based on a BASDAI baseline score of (AiC information has been removed) and a placebo change score of (AiC information has been removed), which represent the results seen in the certolizumab trial (RAPID-axSpA[64]).

Appendix 14 Utility review

In accordance with the NICE reference case,[238] utility values should be based on the EQ-5D instrument. Therefore, a systematic review of utility studies was carried out to identify relevant studies which (1) directly estimate EQ-5D utility values; and (2) establish the relationship between generic measures of utility (in particular the EQ-5D) and measures of disease progression (including mapping studies). The review of utility studies focuses on anti-TNFs for AS and AxSpA without radiographic evidence of AS (nr-AxSpA).

Methods

Searches were undertaken in EMBASE and MEDLINE/MEDLINE In-Process & Other Non-Indexed Citations (Ovid). A combination of disease terms and terms associated with the EQ-5D were used. Upon initial review, it was evident that the results of the search did not identify the studies found in the cost-effectiveness review that also reported on the quality of life of AS patients, for example Ara *et al.*[161] Therefore, a separate search in NHS EED, MEDLINE and EMBASE for published modelling studies was also subsequently undertaken. No language and date limits were applied. Full details of the search strategy used are presented in *Appendix 1*.

Studies that reported utility values consistent with the NICE reference case were included in the review, that is studies reporting utilities for AS or nr-AxSpa patients generated using:

- the EQ-5D
- HRQoL or changes in HRQoL measured directly by patients
- changes in HRQoL should be valued using public preferences from a representative sample of the UK population using a choice-based method (or this could be reasonably assumed from the publication).

When a mapping algorithm was reported, eligibility of studies was restricted to those that mapped from BASDAI score and/or BASFI score to EQ-5D.

Results

Identified studies

The combined search retrieved 210 citations. After screening titles and abstracts, 28 citations were retrieved for full review. The abstract by Pumford *et al.*[239] was excluded, as the full publication by Wade *et al.*[240] reported on the same study. The abstract by Lee *et al.*[241] was excluded as a more recent full publication of the study (Lee *et al.*[242]) reported that a non-UK valuation set was used. Joore *et al.*[243] was also excluded, as primary data were reported in Van Tubergen *et al.*[244] A further three studies were excluded because the manuscripts were in a language other than English.

Kobelt *et al.* have reported costs/quality of life/cost-effectiveness of AS patients in multiple references (e.g. Kobelt *et al.* 2004,[152] Kobelt *et al.* 2006,[245] Kobelt *et al.* 2007[160] and Kobelt *et al.* 2008[246]). Kobelt *et al.*[152,160] are relevant to a UK population and are preferred to the other Kobelt publications that are relevant to non-UK populations. Of these, Kobelt *et al.*[152] reports utility data collected and used in the analysis and is, therefore, included in this review.

In total, 12 studies were deemed to meet the NICE reference case[238] and are summarised in *Table 164*.

TABLE 164 Summary of utility studies that meet the NICE reference case

Study	Population characteristics	Interventions	Utility assessment methods	Utilities reported, mean (SD) [95% CI]	Reviewer comments
Ara *et al.* (2007)[161] The cost-effectiveness of etanercept in patients with severe ankylosing spondylitis in the UK (mapping algorithm to EQ-5D values also reported)	• AS, diagnosed using mNY criteria defined by a VAS for mean morning stiffness ≥ 30 units, and by at least two of the following: VAS score for patient global assessment of disease activity ≥ 30, average VAS score for nocturnal and total pain ≥ 30 or BASFI score of ≥ 30, patients from two etanercept RCTs • European RCT: 356 patients randomised to receive placebo ($n = 51$), etanercept 25 mg twice weekly ($n = 150$) and etanercept 50 mg once weekly ($n = 155$) for 12 weeks. Data from the etanercept arms were combined as no significant differences in outcomes were found • Mainly US RCT: 277 patients randomised to receive placebo ($n = 139$) or etanercept 25 mg twice weekly ($n = 138$) for 24 weeks plus a 3-year open-label extension • Age: 41 years (European RCT), 42 years (US RCT) • Disease duration: 9.3 years (European RCT), 10.3 years (US RCT) • BASDAI score: 6.1 (European RCT), 5.9 (US RCT) • BASFI score: 5.9 (European RCT), 5.4 (US RCT)	• Placebo • Etanercept 25 mg twice weekly • Etanercept 50 mg once weekly	EQ-5D • Completed by patients in 11 European countries (including the UK) • UK population valuation set is assumed to have been used as this is a UK study • European RCT data were used to derive an algorithm between BASDAI/BASFI and EQ-5D. Methods were not reported	European RCT week 12 (observed) for patient with a BASDAI score of ≥ 4: • Anti-TNF responder: 0.79 (NR) • Anti-TNF non-responder: 0.48 (NR) • Placebo responder: 0.74 (NR) • Placebo non-responder: 0.46 (NR) US RCT week 24 (predicted using algorithm): • Anti-TNF responder: 0.80 (NR) • Anti-TNF non-responder: 0.46 (NR) • Placebo responder: 0.79 (NR) • Placebo non-responder: 0.42 (NR) Algorithm % (BASDAI/BASFI scores are on the 0–100 scale): • Utility = 0.923 (0.0170) − 0.004 (0.0007) × BASFI − 0.004 (0.0008) × BASDAI • $R^2 = 0.52$	• Observed values may be generalisable to an AS population that has been treated with etanercept. However, it is not clear how generalisable the outputs are to a UK population • Responders were categorised using BSR guidelines, that is BASDAI 50 • Baseline values were not reported • The generalisability of the algorithm is unclear as the methods have not been reported • A UK population valuation set is assumed to have been used

Study	Population characteristics	Interventions	Utility assessment methods	Utilities reported, mean (SD) [95% CI]	Reviewer comments
Boonen *et al.* (2002)[247] and (2003)[248] 2002: Work status and productivity costs due to ankylosing spondylitis: comparison of three European countries 2003: Costs of ankylosing spondylitis in three European countries: the patient's perspective	AS patients diagnosed using mNY criteria There were 130 patients from the Netherlands. Patients were sampled from the Dutch standard diagnosis register of rheumatic diseases: • Age: 46 years • Disease duration since diagnosis: 12 years • BASDAI score: 3.7 • BASFI score: 3.9 There were 53 patients from France. Consecutive in- and outpatients at a hospital rheumatology department • Age: 38 years • Disease duration since diagnosis: 9 years • BASDAI score: 2.8 • BASFI score: 2.5 There were 26 patients from Belgium. Consecutive outpatients at a hospital rheumatology department • Age: 42 years • Disease duration since diagnosis: 11 years • BASDAI score: 3.1 BASFI score: 2.6	N/A	EQ-5D • Completed by patients in Europe (not including the UK) It is assumed that the UK population valuation set was used as the authors' reference. Dolan *et al.*[249] and Boonen *et al.*[248] say the 'York weighting' was used	Baseline: • Netherlands 0.69 (0.16) • France 0.63 (0.29) • Belgium 0.67 (0.14) Time averaged across 2-year follow-up period: • Netherlands 0.68 (0.16) • France 0.63 (0.23) • Belgium 0.67 (0.14) • All patients 0.67 (0.19)	• Results may be generalisable to an AS population; however, generalisability to a UK population is unknown • A high proportion of missing data (84% 'were missing at least one bimonthly questionnaire) • UK population valuation set is assumed to have been used

continued

TABLE 164 Summary of utility studies that meet the NICE reference case (*continued*)

Study	Population characteristics	Interventions	Utility assessment methods	Utilities reported, mean (SD) [95% CI]	Reviewer comments
Boonen *et al.* (2007)[250] How do the EQ-5D, SF-6D and the well-being rating scale compare in patients with ankylosing spondylitis?	AS patients diagnosed using mNY criteria There were 134 patients from the prevalence-based OASIS cohort (Boonen *et al.* 2002/3)[247,248] There were 120 patients from a RCT comparing spa treatment ($n = 80$) with usual care ($n = 40$) (Van Tuburgen *et al.* 2002[244]) Both data sets were merged as authors found that QoL instruments provided similar results in the two populations: • Age: 48 years • Disease duration since diagnosis: 13 years • BASDAI score: 4.2 • BASFI score: 4.2	OASIS: N/A (prevalence cohort) RCT: spa treatment (3 weeks) and usual care	EQ-5D • Completed by patients in Europe (not including the UK) • UK population valuation set used • Outputs from the EQ-5D rating scale and SF-6D are also reported in this study but are not summarised here	• Combined data sets ($n = 254$): 0.64 (0.23) • BASDAI score of < 4 ($n = 125$): 0.73 (0.16) • BASDAI score of ≥ 4 ($n = 137$): 0.55 (0.26) • BASFI score of < 4 ($n = 121$): 0.74 (0.16) • BASFI score of ≥ 4 ($n = 143$): 0.55 (0.25)	• Results may be generalisable to an AS population; however, generalisability to a UK population is unknown • It is not clear if the utilities reported are baseline values (baseline and post intervention at 4 weeks EQ-5D results were included in the RCT) • EQ-5D discriminates more between lower and higher BASDAI patients (and lower and higher BASFI patients) than the SF-6D. The authors suggest there is a ceiling effect between EQ-5D values 0.6–0.8 (these patients showed a wide range of values on the SF-6D and rating scale)

Study	Population characteristics	Interventions	Utility assessment methods	Utilities reported, mean (SD) [95% CI]	Reviewer comments
Boonen *et al.* (2008)[205] Rapid and sustained improvement in health-related quality of life and utility for 72 weeks in patients with ankylosing spondylitis receiving etanercept	In total 257 AS patients were diagnosed using mNY criteria who had completed 24 weeks of treatment in a previous RCT (277 patients enrolled) comparing etanercept with placebo. Patients were treated with etanercept in the open-label extension study • Age: 41 years • Disease duration: 10.8 years • BASDAI: not reported • BASFI: not reported	• Etanercept 25 mg twice weekly	EQ-5D • Completed by patients in 28 centres across Europe and North America • UK population valuation set was used	Baseline ($n = 232$): • Previously treated with etanercept in the RCT ($n = 128$): 0.69 (0.2) • Previously treated with placebo in the RCT ($n = 129$): 0.49 (0.3) Figure 3(a) in Boonen *et al.*[205] shows that patients who were previously on etanercept maintained their baseline utility up to week 72 (105 patients completed 72 weeks of treatment). Patients who were previously on placebo achieved a similar utility to those patients previously on etanercept by week 12 and maintained this to week 72 (115 patients completed 72 weeks of treatment)	• Results may be generalisable to an AS population, however, generalisability to a UK population is unknown • Negative utility values were imputed as 0 • Patients eligible for the open-label study were those who completed the initial RCT, patients who discontinued because of lack of efficacy but completed follow-up evaluations and patients who discontinued because of AEs which subsequently resolved • Figure 3(a) in Boonen *et al.*[205] refers to 'combined' EQ-5D scores but it is not clear what 'combine' denotes

continued

TABLE 164 Summary of utility studies that meet the NICE reference case (*continued*)

Study	Population characteristics	Interventions	Utility assessment methods	Utilities reported, mean (SD) [95% CI]	Reviewer comments
Braun *et al.* (2007)[85] Improvement in patient-reported outcomes for patients with ankylosing spondylitis treated with etanercept 50 mg once-weekly and 25 mg twice-weekly	In total 356 active AS patients were diagnosed using mNY criteria defined by a VAS score for mean morning stiffness ≥ 30, and by at least two of the following: VAS score for patient global assessment of disease activity ≥ 30, average VAS score for nocturnal and total pain ≥ 30 or BASFI score of ≥ 30 • Age: 40 years • Disease duration: 9 years • BASDAI score: 6.1 • BASFI score: 6.0	• Placebo • Etanercept 25 mg twice weekly (12 weeks) • Etanercept 50 mg once weekly (12 weeks)	EQ-5D • Completed by patients in 11 European countries (including the UK) • It is assumed that the UK population valuation set was used as the authors' reference, Dolan *et al.* (1997)[249]	Mean increase between 0 to 12 weeks reported in figure 2 in Braun *et al.*:[205] • placebo, patient's utility increase at 12 weeks: 0.13 • etanercept 25 mg, patient's utility increase at 12 weeks: 0.25 • etanercept 50 mg, patient's utility increase at 12 weeks: 0.3	• Results may be generalisable to an AS population that has been treated with etanercept. However, it is not clear how generalisable the outputs are to a UK population • Baseline values were not reported • A rapid improvement in utilities was seen within 2 weeks • In total, 90% of patients completed 12 weeks of treatment • A UK population valuation set is assumed to have been used
Gordeev *et al.* (2010)[251] Role of contextual factors in health-related quality of life in ankylosing spondylitis	In total 764 patients with AS were diagnosed using mNY criteria, in Canada and Australia were sent a questionnaire in the post. Overall, 522 (68%) responded and were included in the analysis • Age: 43 years (Australian), 53 years (Canadian) • Diagnosis duration: 13 years (Australian), 19 years (Canadian) • BASDAI score: 3.5 (Australian), 4.1 (Canadian) • BASFI score: 3.3 (Australian), 3.9 (Canadian)	• N/A	EQ-5D • Completed by patients in Canada and Australia • UK population valuation set is used	• Australian cohort ($n = 105$): 0.68 (0.27) • Canadian cohort ($n = 417$): 0.62 (0.29)	• This study may be generalisable to patients with AS. However, generalisability to a UK population is unknown • Contextual factors explained 37% of the variance in EQ-5D • Helplessness (measured using the Rheumatoid Attitudes Index Helplessness Subscale), employment and education were the most important contextual factors. Their role was independent of the strong effect of BASDAI and BASFI

Study	Population characteristics	Interventions	Utility assessment methods	Utilities reported, mean (SD) [95% CI]	Reviewer comments
Haywood *et al.* (2002)[252] Generic measures of health-related quality of life in ankylosing spondylitis: reliability, validity and responsiveness	A random sample of 451 patients with AS, diagnosed using mNY criteria, were sent a postal questionnaire • *n* = 349 (77%) patients returned the questionnaire at baseline • *n* = 349 patients returned the questionnaire at baseline • *n* = 303 patients returned the questionnaire at 2 weeks • *n* = 289 patients returned the questionnaire at 6 months • Age: 46 years • Symptom duration: 20 years • BASDAI: NR • BASFI: NR	NA	EQ-5D • Completed by patients in the UK • It is assumed that the UK population valuation set was used as the authors reference Kind *et al.* (1998)[253] • Outputs from the EQ-5D VAS and SF-12 are also reported in this study but are not summarised here	Reliability analysis using data from patients whose health remained the same at 2 weeks (*n* = 321): 0.53 (0.35) Longitudinal construct validity analysis at 6 months AS: • Patients whose AS health was better (*n* = 57): improved by 0.30 (1.2) • Patients whose AS health stayed the same (*n* = 120): –0.25 (1.5) • Patients whose AS health was worse (*n* = 77): improved by –0.09 (1.6) General health: • Patients whose general health was better (*n* = 49): improved by 0.35 (1.3) • Patients whose AS health stayed the same (*n* = 132): –0.21 (1.4) • Patients whose AS health was worse (*n* = 67): –0.15 (1.7)	• This study may be generalisable to UK patients with AS • UK population valuation set is assumed to have been used • BASDAI/BASFI values for this cohort are not reported

continued

TABLE 164 Summary of utility studies that meet the NICE reference case (*continued*)

Study	Population characteristics	Interventions	Utility assessment methods	Utilities reported, mean (SD) [95% CI]	Reviewer comments
Healey *et al.* (2013)[14] Patients with well-established ankylosing spondylitis show limited deterioration in a ten-year prospective cohort study	In total 269 patients with AS, diagnosed using mNY criteria, were invited to participate at a rheumatology centre 159 patients participated at baseline 69 patients participated at the 10 year assessment • Age: 49 years • Disease duration: 16 years • BASDAI score: 4.1 • BASFI score: not reported	N/A	EQ-5D • Completed by patients in the UK • It is assumed that the UK population valuation set was used as this is a UK study Outputs from the SF-12 are also reported in this study but are not summarised here	Baseline assessment in 1998 ($n = 159$): 0.64 (0.28) 10-year follow-up assessment: 0.61 (0.30)	• This study may be generalisable to UK patients with AS • Only 69 patients participated in both assessments • A UK population valuation set is assumed to have been used
Kobelt *et al.* (2004)[152] The burden of ankylosing spondylitis and the cost-effectiveness of treatment with infliximab (Remicade®)	Clinical trial, hospital cohort and survey data for AS patients were utilised in this study Utilities were estimated from a survey of 3000 patients. 1413 (57%) patients responded and were included in the analysis	N/A	EQ-5D • Completed by patients in the UK • It is assumed that the UK population valuation set was used as the study references Dolan *et al.* (1995)[254] and was conducted in the UK	Survey mean: 0.67 (0.21) BASDAI subgroups: • Patients with a BASDAI score of < 3 (mean BASFI score = 2.4): 0.8 • Patients with a BASDAI 3–3.99 (mean BASFI score = 3.7): 0.7	• This study may be generalisable to UK patients with AS • Patients from across the spectrum of possible BASDAI/BASFI values (0–10) responded to the survey • Measures of uncertainty not reported

Study	Population characteristics	Interventions	Utility assessment methods	Utilities reported, mean (SD) [95% CI]	Reviewer comments
	Survey responders had the following characteristics: • Age: 57 years • Disease duration: 30 years • BASDAI score: 4.2 • BASFI score: 4.5			• Patients with a BASDAI 4–4.99 (mean BASFI score = 4.5): 0.64 • Patients with a BASDAI 5–5.99 (mean BASFI score = 5.4): 0.60 • Patients with a BASDAI 6–6.99 (mean BASFI score = 6.4): 0.51 • Patients with a BASDAI score of > 7 (mean BASFI score = 7.8): 0.39 BASFI subgroups: • Patients with a BASFI score of < 3 (mean BASDAI score = 2.5): 0.8 • Patients with a BASFI 3–3.99 (mean BASDAI score = 3.8): 0.71 • Patients with a BASFI 4–4.99 (mean BASDAI score = 4.2): 0.67 • Patients with a BASFI 5–5.99 (mean BASDAI score = 4.7): 0.57 • Patients with a BASFI 6–6.99 (mean BASDAI score = 5.5): 0.53 • Patients with a BASFI score of > 7 (mean BASDAI score = 8.4): 0.47	• UK population valuation set is assumed to have been used

continued

TABLE 164 Summary of utility studies that meet the NICE reference case (*continued*)

Study	Population characteristics	Interventions	Utility assessment methods	Utilities reported, mean (SD) [95% CI]	Reviewer comments
McLeod *et al.* (2007)[38] Adalimumab, etanercept and infliximab for the treatment of ankylosing spondylitis: a systematic review and economic evaluation (mapping algorithm to EQ-5D values reported)	Utilities were estimated from a reanalysis of the Kobelt *et al.*[152] survey data by the manufacturer of infliximab ($n = 1144$) • Age: NR • Disease duration: NR • BASDAI: not reported • BASFI: NR	N/A	EQ-5D • Completed by patients in the UK • It is assumed that the UK population valuation set was used as the study references Dolan *et al.*[254] and was conducted in the UK Methods for mapping algorithm used by the assessment group NR	Algorithm used in the assessment group (LRiG) model: • Utility = 0.8772129 − 0.0384087 × BASDAI − 0.0322519 × BASFI − 0.0278913 × Male + 0.0016809 × Age The algorithms used in the manufacturer's submissions are also reported but not reproduced here	• Generalisability of the algorithm is unclear as methods have not been reported • Report states that the manufacturer analysis is based on 1144 patients from Kobelt *et al.*[152] Utility values in Kobelt *et al.*[152] were calculated using data from 1413 patients • UK AS patients from across the spectrum of possible BASDAI/BASFI values (0–10) are likely to have been included in the analysis

Study	Population characteristics	Interventions	Utility assessment methods	Utilities reported, mean (SD) [95% CI]	Reviewer comments
Van Tubergen *et al.* (2002)[244] Cost effectiveness of combined spa-exercise therapy in ankylosing spondylitis: a randomized controlled trial	120 AS patients, diagnosed using mNY criteria 111 included in the analysis • Age: 48 years • Disease duration: 11 years • BASDAI score: NR • BASFI score: 4.4	• Spa treatment (3 weeks) • Usual care	EQ-5D • Completed by patients in Europe (not including the UK) • UK population valuation set is assumed to have been used as the study references Dolan *et al.* (1996)[255] Outputs from the SF-6D are also reported in this study but are not summarised here	Spa treatment in Austria ($n = 36$): • Baseline (2 weeks before treatment): 0.650 (0.22) • Change at 4 weeks: 0.02 (0.2) • Change at 16 weeks: 0.04 (0.21) • Change at 28 weeks: –0.03 (0.23) • Change at 40 weeks: –0.01 (0.27) Spa treatment in the Netherlands ($n = 38$): • Baseline (2 weeks before treatment): 0.64 (0.22) • Change at 4 weeks: 0.1 (0.24) • Change at 16 weeks: 0.12 (0.24) • Change at 28 weeks: 0.1 (0.21) • Change at 40 weeks: 0.03 (0.23) Usual care ($n = 37$): • Baseline (2 weeks before treatment): 0.72 (0.1) • Change at 4 weeks: –0.06 (0.18) • Change at 16 weeks: –0.04 (0.19) • Change at 28 weeks: –0.08 (0.28) • Change at 40 weeks: –0.03 (0.19)	• Results may be generalisable to an AS population, however, generalisability to a UK population is unknown • Patients were allowed to continue taking their usual medication throughout the study period. Medication could be changed if needed. This may bias the results • A UK population valuation set is assumed to have been used

continued

TABLE 164 Summary of utility studies that meet the NICE reference case (*continued*)

Study	Population characteristics	Interventions	Utility assessment methods	Utilities reported, mean (SD) [95% CI]	Reviewer comments
Wade *et al.* (2011)[240] Baseline characteristics and patient reported outcome data of patients prescribed etanercept: web-based and telephone evaluation	There were 43 patients prescribed etanercept for AS (diagnostic criteria not reported) RA, PsA and psoriasis patients were also included in the study • Age: 49 years • Disease duration: NR • BASDAI score: NR • BASFI score: NR	• Etanercept	EQ-5D • Completed by patients in the UK • A UK population valuation set is assumed to have been used as this is a UK study	Baseline: 0.37 (0.37)[a]	• This study may be generalisable to UK patients with AS • Overall, 23% of AS patients were previously treated with a TNF-α inhibitor • A UK population valuation set is assumed to have been used • Differences in characteristics between telephone and web-based responders were observed for the entire sample (for all conditions)

QoL, quality of life; mNY, modified New York criteria; N/A, not applicable; NR, not reported; PsA, psoriatic arthritis; RA, rheumatoid arthritis; SF-6D, Short Form questionnaire-6 Dimensions.
a Only baseline reported in this study.

The main reasons for excluding studies at the title/abstract and at full review stage were (1) utilities were not reported (e.g. Haywood *et al.*[256]), (2) valuation set not reported or a non-UK valuation was used (e.g. Kvamme *et al.*[257]), (3) utilities were reported for a mixed population with different inflammatory arthropies or in a population not relevant to the decision problem (e.g. Osnes-Ringen *et al.*[258]).

Studies meeting the National Institute for Health and Care Excellence reference case

The 12 studies meeting the reference case[238] have been summarised in *Table 164*. The table includes a primary study to Boonen *et al.*,[250] reported in Boonen *et al.*[247,248] The study by Boonen *et al.*[250] has been retained as it reports utility values for patients with a BASDAI score of ≥ 4.

Ankylosing spondylitis population

All studies included in *Table 164* are of AS patients. Five studies reported utility values (or mapping algorithms) generated from data specifically collected from the UK population (Haywood *et al.*,[252] Healey *et al.*,[14] Kobelt *et al.*,[152] McLeod *et al.*[38] and Wade *et al.*[240]). Four studies included interventions specific to this appraisal, all of these studies were of etanercept (Ara *et al.*,[161] Boonen *et al.*,[205] Braun *et al.*[87] and Wade *et al.*[240]). Utility values reported ranged from values at baseline to at 10 years' follow-up.

Non-radiographic axial spondyloarthritis population

Two citations were identified in the review that reported utilities for nr-AxSpA patients (Dougados *et al.*[78] and Lindstrom *et al.*[259]). However, these studies did not explicitly report which population valuation sets were used and, therefore, were excluded from the review.

Mapping algorithms

Of the 12 studies in *Table 164*, two report mapping algorithms between disease-specific measures and the EQ-5D (Ara *et al.*[161] and McLeod *et al.*[38]). Both have been reported as part of a cost-effectiveness analysis and provide limited information on methodology employed (e.g. covariates tested, correlation considerations and goodness of fit). McLeod *et al.*[38] reports on an algorithm generated using data from UK AS patients.

Appendix 15 Additional cost-effectiveness results

Following the consultation process to the NICE appraisal, additional analyses were undertaken to address comments received on *Chapters 5* and 6. These focused on the conditional baseline BASDAI scores used in *Chapter 5* and on the existence of a biosimilar product for infliximab with a lower list price.

A. Truncated baseline Bath Ankylosing Spondylitis Disease Activity Index scores

One of the consultees identified that the simulation procedure undertaken to evaluate conditional scores in *Chapter 5* was using a non-truncated distribution for the baseline BASDAI score. This meant that it was possible for simulated individuals to have a baseline BASDAI score of < 4, which is inconsistent with clinical practice for which treatment with anti-TNFs is only provided to patients with a baseline BASDAI score of > 4. This was a result of assuming a normal distribution for baseline BASDAI scores. Together with the SD applied, this resulted in the simulation model sampling population characteristics which we acknowledge would subsequently fall outside of the decision problem. We have rerun our base-case results assuming a truncated distribution (i.e. excluding the possibility of sampling patients with a baseline BASDAI score of < 4). The results are presented in *Table 165* and confirm only a minimal impact on the conditional scores.

Tables 166 and *167* report the revised base-case cost-effectiveness results for the AS population using a truncated distribution for baseline BASDAI (the original base-case cost-effectiveness results are reported in *Tables 95* and *96* for the AS population).

These results demonstrate that the ICERs show only small variation using a truncated baseline BASDAI distribution. The ICERs appear marginally less favourable than the original base-case results. Hence, any potential bias in the original analysis appears to work in favour of the TNF-inhibitors. However, the magnitude is small and could equally be a result of simulation error as opposed to any bias.

TABLE 165 Conditional scores predicted for the AS population using the synthesis model (truncated baseline BASDAI)

	BASDAI		BASFI	
Patient population	Control	Treatment	Control	Treatment
BASDAI baseline score truncated to a minimum value of 4				
% responders to BASDAI 50	0.08	0.45	–	–
Change in score				
Responders	–2.93	–3.77	–1.49	–3.01
Non-responders	–0.39	–1.56	–0.06	–0.39
All	–0.60	–2.56	–0.18	–1.58
Baseline				
Responders	4.52	4.86	3.92	4.24
Non-responders	6.30	7.23	5.40	6.15
All	6.15	6.15	5.28	5.28

TABLE 166 Revised base-case cost-effectiveness results (truncated normal): AS (rebound equal to gain)

Strategy	Total QALYs	Incremental QALYs	Total costs (£)	Incremental costs (£)	ICER (£)	Probability of CE £20,000	Probability of CE £30,000
Conventional therapy	7.240	–	111,532	–	–	–	–
Certolizumab pegol PAS	8.249	1.009	131,909	20,377	20,195	0.482	0.877
Golimumab	8.249	1.009	133,543	22,011	21,814	0.381	0.830
Adalimumab	8.249	1.009	133,637	22,105	21,907	0.376	0.827
Etanercept	8.249	1.009	134,054	22,522	22,321	0.345	0.811
Certolizumab pegol	8.249	1.009	135,483	23,951	23,737	0.258	0.758
Infliximab	8.249	1.009	153,255	41,723	41,350	0.000	0.081

CE, cost-effectiveness.
The probability of CE £20,000/30,000 is the probability that the TNF-α inhibitor, compared with CC, is a cost-effective option at the stated threshold.

TABLE 167 Revised base-case cost-effectiveness results (truncated normal): AS (rebound to CC)

Strategy	Total QALYs	Incremental QALYs	Total costs (£)	Incremental costs (£)	ICER (£)	Probability of CE £20,000	Probability of CE £30,000
Conventional therapy	7.251	–	111,666	–	–	–	–
Certolizumab pegol PAS	7.898	0.647	134,955	23,289	35,982	0.013	0.333
Golimumab	7.898	0.647	136,585	24,919	38,500	0.009	0.245
Adalimumab	7.898	0.647	136,679	25,013	38,646	0.008	0.239
Etanercept	7.898	0.647	137,100	25,434	39,296	0.006	0.222
Certolizumab pegol	7.898	0.647	138,528	26,862	41,503	0.004	0.172
Infliximab	7.898	0.647	156,420	44,754	69,146	0.000	0.000

CE, cost-effectiveness.
The probability of CE £20,000/30,000 is the probability that the TNF-α inhibitor, compared with CC, is a cost-effective option at the stated threshold.

Non-radiographic axial spondyloarthritis population

We have rerun our base-case results assuming a truncated distribution (i.e. excluding the possibility of sampling patients with a baseline BASDAI score of < 4) for the nr-AxSpA population. The results are presented in *Table 168* and confirm only a minimal impact on the conditional scores.

Tables 169 and *170* report the revised base-case cost-effectiveness results for the nr-AxSpA population using a truncated distribution for baseline BASDAI (the original base-case cost-effectiveness results are reported in *Tables 97* and *98*).

TABLE 168 Conditional scores predicted for the nr-AxSpA population using the synthesis model (truncated baseline BASDAI)

	BASDAI		BASFI	
Patient population	Control	Treatment	Control	Treatment
BASDAI baseline score truncated to a minimum value of 4				
% responders to BASDAI 50	0.21	0.55	–	–
Change in score				
Responders	–3.23	–4.29	–1.77	–3.23
Non-responders	–1.05	–2.25	0.02	0.17
All	–1.50	–3.37	–0.35	–1.70
Baseline				
Responders	4.80	5.40	3.24	3.86
Non-responders	6.83	7.64	5.34	6.20
All	6.41	6.41	4.91	4.91

TABLE 169 Revised base-case cost-effectiveness results (truncated normal): nr-AxSpA (rebound equal to gain)

Strategy	Total QALYs	Incremental QALYs	Total costs (£)	Incremental costs (£)	ICER (£)	Probability of CE £20,000	Probability of CE £30,000
Conventional therapy	9.906	–	90,850	–	–	–	–
Certolizumab pegol PAS	11.291	1.386	130,974	40,124	28,958	0.102	0.576
Adalimumab	11.291	1.386	132,373	41,523	29,968	0.069	0.506
Etanercept	11.291	1.386	133,119	42,269	30,506	0.059	0.484
Certolizumab pegol	11.291	1.386	134,547	43,696	31,536	0.045	0.454

CE, cost-effectiveness.
The probability of CE £20,000/30,000 is the probability that the TNF-α inhibitor, compared with CC, is a cost-effective option at the stated threshold.

TABLE 170 Revised base-case cost-effectiveness results (truncated normal): nr-AxSpA (rebound to CC)

Strategy	Total QALYs	Incremental QALYs	Total costs (£)	Incremental costs (£)	ICER (£)	Probability of CE £20,000	Probability of CE £30,000
Conventional therapy	9.963	–	90,219	–	–	–	–
Certolizumab pegol PAS	11.200	1.237	131,714	41,495	33,555	0.057	0.396
Adalimumab	11.200	1.237	133,109	42,890	34,684	0.038	0.343
Etanercept	11.200	1.237	133,859	43,640	35,290	0.035	0.318
Certolizumab pegol	11.200	1.237	135,286	45,067	36,444	0.029	0.284

CE, cost-effectiveness.
The probability of CE £20,000/30,000 is the probability that the TNF-α inhibitor, compared with CC, is a cost-effective option at the stated threshold.

As *Tables 169* and *170* demonstrate, the ICERs show only small variation employing a truncated distribution for baseline BASDAI. The ICERs appear marginally less favourable than the original base-case results. Hence, any potential bias in the original analysis appears to work in favour of the TNF-inhibitors. However, the magnitude is small and could equally be a result of simulation error as opposed to any bias.

B. Biosimilar

One of the consultees provided the list price for Remsima™, a biosimilar for infliximab marketed by Celltrion Healthcare. In view of this, we updated the base-case analysis for the AS population using the list price for Remsima. The results, of rerunning the analysis excluding infliximab and replacing it with Remsima using the list price provided, are reported in *Tables 171* and *172*. Minor differences in the ICERs for the other TNF inhibitors compared with the original base case are a result of sapling variation (i.e. all ICERs are derived from the probabilistic analysis and rerunning the simulation results in minor differences each time).

TABLE 171 Revised base-case results using list price for Remsima: AS population (rebound equal to gain)

Strategy	Total QALYs	Incremental QALYs	Total costs (£)	Incremental costs (£)	ICER (£)	Probability of CE £20,000	Probability of CE £30,000
Conventional therapy	7.266	–	111,696	–	–	–	–
Certolizumab pegol PAS	8.179	0.913	129,281	17,586	19,257	0.551	0.899
Golimumab	8.179	0.913	130,969	19,274	21,106	0.428	0.848
Adalimumab	8.179	0.913	131,053	19,357	21,197	0.421	0.846
Etanercept	8.179	0.913	131,426	19,731	21,606	0.390	0.835
Certolizumab pegol	8.179	0.913	132,855	21,159	23,171	0.295	0.762
Remsima™	8.179	0.913	145,256	33,561	36,751	0.004	0.204

CE, cost-effectiveness.
The probability of CE £20,000/30,000 is the probability that the TNF-α inhibitor, compared with CC, is a cost-effective option at the stated threshold.

TABLE 172 Revised base-case results using list price for Remsima: AS population (rebound to CC)

Strategy	Total QALYs	Incremental QALYs	Total costs (£)	Incremental costs (£)	ICER (£)	Probability of CE £20,000	Probability of CE £30,000
Conventional therapy	7.250	–	111,647	–	–	–	–
Certolizumab pegol PAS	7.854	0.604	131,922	20,275	33,578	0.040	0.397
Golimumab	7.854	0.604	133,605	21,958	36,366	0.016	0.307
Adalimumab	7.854	0.604	133,690	22,043	36,506	0.016	0.305
Etanercept	7.854	0.604	134,067	22,420	37,131	0.012	0.287
Certolizumab pegol	7.854	0.604	135,496	23,849	39,497	0.007	0.215
Remsima™	7.854	0.604	148,010	36,363	60,222	0.000	0.003

CE, cost-effectiveness.
The probability of CE £20,000/30,000 is the probability that the TNF-α inhibitor, compared with CC, is a cost-effective option at the stated threshold.

Appendix 16 Full incremental cost-effectiveness ratio tables for scenarios

Ankylosing spondylitis population

TABLE 173 Summary of cost-effectiveness scenarios: AS population

Strategy number	Parameter/structural	Approach in scenario	Approach in base case
1	CC ('placebo') response	No response to CC assumed at 12 weeks	Response to CC included at 12 weeks
2	Different baselines assumed for responders and non-responders and change in BASDAI/BASFI scores	Separate baselines based on pooled estimates provided by manufacturers. Changes in BASDAI/BASFI score conditioned on response also based on pooled estimates provided by manufacturers	Separate baselines and changes in BASDAI/BASFI conditioned on responses estimated via extended synthesis model
3	BASFI progression	No effect of anti-TNFs on BASFI progression	Treatment effect applied from year 4 onwards
4	BASFI progression	Treatment effect of anti-TNFs applied from start of model	Treatment effect applied from year 4 onwards
5	Utilities	Linear BASDAI/BASFI model (based on Kobelt *et al.*[152])	Non-linear BASDAI/BASFI model (Pfizer submission[36])
6	Baseline BASDAI score truncated at 4	Baseline BASDAI scores used in the NMA are limited to be between 4 and 10	Baseline BASDAI scores used in the NMA are limited to be between 0 and 10
7	Price of biosimilar infliximab	The price of the biosimilar for infliximab provided by the manufacturer was used in the model	BNF price of infliximab

BNF, *British National Formulary*; NMA, network meta-analysis.

Ankylosing spondylitis scenario results: rebound equal to gain

TABLE 174 Ankylosing spondylitis: scenario 1 (rebound equal to gain)

Strategy	Total QALYs	Incremental QALYs	Total costs (£)	Incremental costs (£)	ICER (£)	Probability of CE £20,000	Probability of CE £30,000
Conventional therapy	7.262	–	111,702	–	–	–	–
Certolizumab pegol PAS	7.952	0.691	125,734	14,033	20,319	0.462	0.861
Golimumab	7.952	0.691	127,531	15,829	22,920	0.313	0.764
Adalimumab	7.952	0.691	127,594	15,893	23,013	0.308	0.761
Etanercept	7.952	0.691	127,879	16,178	23,425	0.292	0.741
Certolizumab pegol	7.952	0.691	129,308	17,607	25,495	0.188	0.651
Infliximab	7.952	0.691	141,750	30,048	43,510	0.000	0.063

CE, cost-effectiveness.
The probability of CE £20,000/30,000 is the probability that the TNF-α inhibitor, compared with CC, is a cost-effective option at the stated threshold.

TABLE 175 Ankylosing spondylitis: scenario 2 (rebound equal to gain)

Strategy	Total QALYs	Incremental QALYs	Total costs (£)	Incremental costs (£)	ICER (£)	Probability of CE £20,000	Probability of CE £30,000
Conventional therapy	7.100	–	112,768	–	–	–	–
Certolizumab pegol PAS	8.120	1.019	127,856	15,088	14,803	0.840	0.988
Golimumab	8.120	1.019	129,536	16,768	16,451	0.727	0.979
Adalimumab	8.120	1.019	129,621	16,853	16,535	0.720	0.977
Etanercept	8.120	1.019	130,001	17,233	16,907	0.696	0.975
Certolizumab pegol	8.120	1.019	131,430	18,662	18,309	0.561	0.955
Infliximab	8.120	1.019	147,674	34,906	34,246	0.015	0.232

CE, cost-effectiveness.
The probability of CE £20,000/30,000 is the probability that the TNF-α inhibitor, compared with CC, is a cost-effective option at the stated threshold.

TABLE 176 Ankylosing spondylitis: scenario 3 (rebound equal to gain)

Strategy	Total QALYs	Incremental QALYs	Total costs (£)	Incremental costs (£)	ICER (£)	Probability of CE £20,000	Probability of CE £30,000
Conventional therapy	7.253	–	109,379	–	–	–	–
Certolizumab pegol PAS	8.128	0.875	127,455	18,075	20,655	0.462	0.843
Golimumab	8.128	0.875	129,140	19,760	22,581	0.348	0.775
Adalimumab	8.128	0.875	129,224	19,845	22,677	0.341	0.771
Etanercept	8.128	0.875	129,600	20,220	23,106	0.319	0.760
Certolizumab pegol	8.128	0.875	131,028	21,649	24,739	0.234	0.698
Infliximab	8.128	0.875	147,118	37,739	43,125	0.001	0.063

CE, cost-effectiveness.
The probability of CE £20,000/30,000 is the probability that the TNF-α inhibitor, compared with CC, is a cost-effective option at the stated threshold.

TABLE 177 Ankylosing spondylitis: scenario 4 (rebound equal to gain)

Strategy	Total QALYs	Incremental QALYs	Total costs (£)	Incremental costs (£)	ICER (£)	Probability of CE £20,000	Probability of CE £30,000
Conventional therapy	7.239	–	111,036	–	–	–	–
Certolizumab pegol PAS	8.201	0.962	128,804	17,767	18,466	0.589	0.929
Golimumab	8.201	0.962	130,485	19,448	20,213	0.462	0.878
Adalimumab	8.201	0.962	130,570	19,533	20,301	0.453	0.875
Etanercept	8.201	0.962	130,949	19,912	20,695	0.429	0.862
Certolizumab pegol	8.201	0.962	132,377	21,341	22,180	0.345	0.808
Infliximab	8.201	0.962	148,597	37,560	39,037	0.005	0.124

CE, cost-effectiveness.
The probability of CE £20,000/30,000 is the probability that the TNF-α inhibitor, compared with CC, is a cost-effective option at the stated threshold.

TABLE 178 Ankylosing spondylitis: scenario 5 (rebound equal to gain)

Strategy	Total QALYs	Incremental QALYs	Total costs (£)	Incremental costs (£)	ICER (£)	Probability of CE £20,000	Probability of CE £30,000
Conventional therapy	10.272	–	111,187	–	–	–	–
Certolizumab pegol PAS	11.043	0.771	129,139	17,953	23,290	0.217	0.891
Golimumab	11.043	0.771	130,819	19,632	25,469	0.099	0.755
Adalimumab	11.043	0.771	130,904	19,717	25,579	0.094	0.750
Etanercept	11.043	0.771	131,285	20,098	26,073	0.074	0.724
Certolizumab pegol	11.043	0.771	132,713	21,526	27,926	0.048	0.593
Infliximab	11.043	0.771	148,974	37,787	49,021	0.000	0.003

CE, cost-effectiveness.
The probability of CE £20,000/30,000 is the probability that the TNF-α inhibitor, compared with CC, is a cost-effective option at the stated threshold.

TABLE 179 Ankylosing spondylitis: scenario 6 (rebound equal to gain)

Strategy	Total QALYs	Incremental QALYs	Total costs (£)	Incremental costs (£)	ICER (£)	Probability of CE £20,000	Probability of CE £30,000
Conventional therapy	7.240	–	111,532	–	–	–	–
Certolizumab pegol PAS	8.249	1.009	131,909	20,377	20,195	0.482	0.877
Golimumab	8.249	1.009	133,543	22,011	21,814	0.381	0.830
Adalimumab	8.249	1.009	133,637	22,105	21,907	0.376	0.827
Etanercept	8.249	1.009	134,054	22,522	22,321	0.345	0.811
Certolizumab pegol	8.249	1.009	135,483	23,951	23,737	0.258	0.758
Infliximab	8.249	1.009	153,255	41,723	41,350	0.000	0.081

CE, cost-effectiveness.
The probability of CE £20,000/30,000 is the probability that the TNF-α inhibitor, compared with CC, is a cost-effective option at the stated threshold.

TABLE 180 Ankylosing spondylitis: scenario 7 (rebound equal to gain)

Strategy	Total QALYs	Incremental QALYs	Total costs (£)	Incremental costs (£)	ICER (£)	Probability of CE £20,000	Probability of CE £30,000
Conventional therapy	7.266	–	111,696	–	–	–	–
Certolizumab pegol PAS	8.179	0.913	129,281	17,586	19,257	0.551	0.899
Golimumab	8.179	0.913	130,969	19,274	21,106	0.428	0.848
Adalimumab	8.179	0.913	131,053	19,357	21,197	0.421	0.846
Etanercept	8.179	0.913	131,426	19,731	21,606	0.390	0.835
Certolizumab pegol	8.179	0.913	132,855	21,159	23,171	0.295	0.762
Remsima	8.179	0.913	145,256	33,561	36,751	0.004	0.204

CE, cost-effectiveness.
The probability of CE £20,000/30,000 is the probability that the TNF-α inhibitor, compared with CC, is a cost-effective option at the stated threshold.

Ankylosing spondylitis scenario results: rebound to conventional care

TABLE 181 Ankylosing spondylitis: scenario 1 (rebound to CC)

Strategy	Total QALYs	Incremental QALYs	Total costs (£)	Incremental costs (£)	ICER (£)	Probability of CE £20,000	Probability of CE £30,000
Conventional therapy	7.295	–	112,675	–	–	–	–
Certolizumab pegol PAS	7.762	0.467	128,654	15,979	34,229	0.038	0.385
Golimumab	7.762	0.467	130,446	17,771	38,068	0.014	0.257
Adalimumab	7.762	0.467	130,511	17,836	38,207	0.013	0.256
Etanercept	7.762	0.467	130,799	18,124	38,824	0.010	0.245
Certolizumab pegol	7.762	0.467	132,228	19,553	41,885	0.004	0.161
Infliximab	7.762	0.467	144,800	32,125	68,815	0.000	0.000

CE, cost-effectiveness.
The probability of CE £20,000/30,000 is the probability that the TNF-α inhibitor, compared with CC, is a cost-effective option at the stated threshold.

TABLE 182 Ankylosing spondylitis: scenario 2 (rebound to CC)

Strategy	Total QALYs	Incremental QALYs	Total costs (£)	Incremental costs (£)	ICER (£)	Probability of CE £20,000	Probability of CE £30,000
Conventional therapy	7.092	–	114,772	–	–	–	–
Certolizumab pegol PAS	7.756	0.664	132,257	17,485	26,348	0.136	0.666
Golimumab	7.756	0.664	133,945	19,174	28,892	0.076	0.531
Adalimumab	7.756	0.664	134,029	19,257	29,018	0.075	0.521
Etanercept	7.756	0.664	134,402	19,631	29,580	0.068	0.493
Certolizumab pegol	7.756	0.664	135,831	21,059	31,733	0.042	0.380
Infliximab	7.756	0.664	151,831	37,059	55,842	0.000	0.008

CE, cost-effectiveness.
The probability of CE £20,000/30,000 is the probability that the TNF-α inhibitor, compared with CC, is a cost-effective option at the stated threshold.

TABLE 183 Ankylosing spondylitis: scenario 3 (rebound to CC)

Strategy	Total QALYs	Incremental QALYs	Total costs (£)	Incremental costs (£)	ICER (£)	Probability of CE £20,000	Probability of CE £30,000
Conventional therapy	7.252	–	110,930	–	–	–	–
Certolizumab pegol PAS	7.818	0.566	131,610	20,679	36,518	0.021	0.339
Golimumab	7.818	0.566	133,289	22,359	39,483	0.009	0.249
Adalimumab	7.818	0.566	133,374	22,444	39,634	0.008	0.245
Etanercept	7.818	0.566	133,755	22,824	40,306	0.006	0.230
Certolizumab pegol	7.818	0.566	135,183	24,253	42,828	0.003	0.166
Infliximab	7.818	0.566	151,457	40,526	71,565	0.000	0.000

CE, cost-effectiveness.
The probability of CE £20,000/30,000 is the probability that the TNF-α inhibitor, compared with CC, is a cost-effective option at the stated threshold.

TABLE 184 Ankylosing spondylitis: scenario 4 (rebound to CC)

Strategy	Total QALYs	Incremental QALYs	Total costs (£)	Incremental costs (£)	ICER (£)	Probability of CE £20,000	Probability of CE £30,000
Conventional therapy	7.268	–	108,817	–	–	–	–
Certolizumab pegol PAS	7.894	0.626	128,999	20,182	32,222	0.047	0.429
Golimumab	7.894	0.626	130,683	21,866	34,910	0.022	0.341
Adalimumab	7.894	0.626	130,767	21,951	35,045	0.020	0.339
Etanercept	7.894	0.626	131,144	22,327	35,647	0.016	0.310
Certolizumab pegol	7.894	0.626	132,573	23,756	37,928	0.008	0.234
Infliximab	7.894	0.626	148,706	39,889	63,684	0.000	0.000

CE, cost-effectiveness.
The probability of CE £20,000/30,000 is the probability that the TNF-α inhibitor, compared with CC, is a cost-effective option at the stated threshold.

TABLE 185 Ankylosing spondylitis: scenario 5 (rebound to CC)

Strategy	Total QALYs	Incremental QALYs	Total costs (£)	Incremental costs (£)	ICER (£)	Probability of CE £20,000	Probability of CE £30,000
Conventional therapy	10.272	–	112,648	–	–	–	–
Certolizumab pegol PAS	10.967	0.695	133,103	20,455	29,414	0.012	0.511
Golimumab	10.967	0.695	134,781	22,133	31,827	0.005	0.340
Adalimumab	10.967	0.695	134,866	22,218	31,950	0.005	0.333
Etanercept	10.967	0.695	135,248	22,600	32,499	0.004	0.300
Certolizumab pegol	10.967	0.695	136,677	24,028	34,554	0.002	0.165
Infliximab	10.967	0.695	152,997	40,349	58,022	0.000	0.000

CE, cost-effectiveness.
The probability of CE £20,000/30,000 is the probability that the TNF-α inhibitor, compared with CC, is a cost-effective option at the stated threshold.

TABLE 186 Ankylosing spondylitis: scenario 6 (rebound to CC)

Strategy	Total QALYs	Incremental QALYs	Total costs (£)	Incremental costs (£)	ICER (£)	Probability of CE £20,000	Probability of CE £30,000
Conventional therapy	7.251	–	111,666	–	–	–	–
Certolizumab pegol PAS	7.898	0.647	134,955	23,289	35,982	0.013	0.333
Golimumab	7.898	0.647	136,585	24,919	38,500	0.009	0.245
Adalimumab	7.898	0.647	136,679	25,013	38,646	0.008	0.239
Etanercept	7.898	0.647	137,100	25,434	39,296	0.006	0.222
Certolizumab pegol	7.898	0.647	138,528	26,862	41,503	0.004	0.172
Infliximab	7.898	0.647	156,420	44,754	69,146	0.000	0.000

CE, cost-effectiveness.
The probability of CE £20,000/30,000 is the probability that the TNF-α inhibitor, compared with CC, is a cost-effective option at the stated threshold.

TABLE 187 Ankylosing spondylitis: scenario 7 (rebound to CC)

Strategy	Total QALYs	Incremental QALYs	Total costs (£)	Incremental costs (£)	ICER (£)	Probability of CE £20,000	Probability of CE £30,000
Conventional therapy	7.250	–	111,647	–	–	–	–
Certolizumab pegol PAS	7.854	0.604	131,922	20,275	33,578	0.040	0.397
Golimumab	7.854	0.604	133,605	21,958	36,366	0.016	0.307
Adalimumab	7.854	0.604	133,690	22,043	36,506	0.016	0.305
Etanercept	7.854	0.604	134,067	22,420	37,131	0.012	0.287
Certolizumab pegol	7.854	0.604	135,496	23,849	39,497	0.007	0.215
Remsima	7.854	0.604	148,010	36,363	60,222	0.000	0.003

CE, cost-effectiveness.
The probability of CE £20,000/30,000 is the probability that the TNF-α inhibitor, compared with CC, is a cost-effective option at the stated threshold.

Non-radiographic axial spondyloarthritis population

TABLE 188 Summary of cost-effectiveness scenarios: nr-AxSpA population

Strategy number	Parameter/structural	Approach in scenario	Approach in base case
1	CC ('placebo') response	No response to CC assumed at 12 weeks	Response to CC included at 12 weeks
2	Different baselines assumed for responders and non-responders and change in BASDAI/BASFI scores	Separate baselines based on pooled estimates provided by manufacturers. Changes in BASDAI/BASFI score conditioned on response also based on pooled estimates provided by manufacturers	Separate baselines and changes in BASDAI/BASFI conditioned on responses estimated via extended synthesis model
3	BASFI progression	No effect of anti-TNFs on BASFI progression	Treatment effect applied from year 4 onwards
4	BASFI progression	Treatment effect of anti-TNFs applied from start of model	Treatment effect applied from year 4 onwards
5	Utilities	Linear BASDAI/BASFI model (based on Kobelt *et al.*[152])	Non-linear BASDAI/BASFI model (Pfizer submission[36])
6	Treatment effect of anti-TNFs	Trials in nr-AxSpA and AS populations combined	Only trials in nr-AxSpA included
7	Baseline BASDAI score truncated at 4	Baseline BASDAI scores used in the NMA are limited to be between 4 and 10	Baseline BASDAI scores used in the NMA are limited to be between 0 and 10

NMA, network meta-analysis.

Non-radiographic axial spondyloarthritis scenario results: rebound equal to gain

TABLE 189 Non-radiographic axial spondyloarthritis: scenario 1 (rebound equal to gain)

Strategy	Total QALYs	Incremental QALYs	Total costs (£)	Incremental costs (£)	ICER (£)	Probability of CE £20,000	Probability of CE £30,000
Conventional therapy	10.012	–	87,215	–	–	–	–
Certolizumab pegol PAS	11.864	1.852	122,480	35,265	19,040	0.550	0.861
Adalimumab	11.864	1.852	123,883	36,668	19,797	0.500	0.844
Etanercept	11.864	1.852	124,625	37,410	20,198	0.481	0.838
Certolizumab pegol	11.864	1.852	126,052	38,837	20,968	0.421	0.815

CE, cost-effectiveness.
The probability of CE £20,000/30,000 is the probability that the TNF-α inhibitor, compared with CC, is a cost-effective option at the stated threshold.

TABLE 190 Non-radiographic axial spondyloarthritis: scenario 2 (rebound equal to gain)

Strategy	Total QALYs	Incremental QALYs	Total costs (£)	Incremental costs (£)	ICER (£)	Probability of CE £20,000	Probability of CE £30,000
Conventional therapy	9.988	–	87,947	–	–	–	–
Certolizumab pegol PAS	11.666	1.678	124,455	36,508	21,757	0.377	0.764
Adalimumab	11.666	1.678	125,858	37,910	22,593	0.337	0.739
Etanercept	11.666	1.678	126,600	38,653	23,036	0.317	0.722
Certolizumab pegol	11.666	1.678	128,027	40,080	23,886	0.276	0.683

CE, cost-effectiveness.
Probability of CE £20,000/30,000 is the probability that the TNF-α inhibitor, when compared with CC, is a cost-effective option at the stated threshold.

TABLE 191 Non-radiographic axial spondyloarthritis: scenario 3 (rebound equal to gain)

Strategy	Total QALYs	Incremental QALYs	Total costs (£)	Incremental costs (£)	ICER (£)	Probability of CE £20,000	Probability of CE £30,000
Conventional therapy	9.891	–	91,479	–	–	–	–
Certolizumab pegol PAS	11.262	1.370	130,734	39,254	28,643	0.138	0.576
Adalimumab	11.262	1.370	132,141	40,662	29,670	0.102	0.528
Etanercept	11.262	1.370	132,879	41,399	30,208	0.093	0.505
Certolizumab pegol	11.262	1.370	134,306	42,827	31,250	0.076	0.460

CE, cost-effectiveness.
The probability of CE £20,000/30,000 is the probability that the TNF-α inhibitor, compared with CC, is a cost-effective option at the stated threshold.

TABLE 192 Non-radiographic axial spondyloarthritis: scenario 4 (rebound equal to gain)

Strategy	Total QALYs	Incremental QALYs	Total costs (£)	Incremental costs (£)	ICER (£)	Probability of CE £20,000	Probability of CE £30,000
Conventional therapy	9.923	–	90,625	–	–	–	–
Certolizumab pegol PAS	11.338	1.415	129,492	38,867	27,471	0.154	0.627
Adalimumab	11.338	1.415	130,899	40,274	28,466	0.127	0.574
Etanercept	11.338	1.415	131,637	41,012	28,988	0.116	0.549
Certolizumab pegol	11.338	1.415	133,064	42,440	29,996	0.087	0.501

CE, cost-effectiveness.
The probability of CE £20,000/30,000 is the probability that the TNF-α inhibitor, compared with CC, is a cost-effective option at the stated threshold.

TABLE 193 Non-radiographic axial spondyloarthritis: scenario 5 (rebound equal to gain)

Strategy	Total QALYs	Incremental QALYs	Total costs (£)	Incremental costs (£)	ICER (£)	Probability of CE £20,000	Probability of CE £30,000
Conventional therapy	10.973	–	89,400	–	–	–	–
Certolizumab pegol PAS	12.527	1.554	128,760	39,361	25,324	0.120	0.781
Adalimumab	12.527	1.554	130,165	40,765	26,227	0.086	0.725
Etanercept	12.527	1.554	130,905	41,506	26,704	0.071	0.692
Certolizumab pegol	12.527	1.554	132,333	42,933	27,622	0.053	0.629

CE, cost-effectiveness.
The probability of CE £20,000/30,000 is the probability that the TNF-α inhibitor, compared with CC, is a cost-effective option at the stated threshold.

TABLE 194 Non-radiographic axial spondyloarthritis: scenario 6 (rebound equal to gain)

Strategy	Total QALYs	Incremental QALYs	Total costs (£)	Incremental costs (£)	ICER (£)	Probability of CE £20,000	Probability of CE £30,000
Conventional therapy	9.944	–	88,563	–	–	–	–
Certolizumab pegol PAS	11.382	1.437	129,592	41,030	28,282	0.068	0.612
Adalimumab	11.382	1.437	130,978	42,415	29,228	0.040	0.570
Etanercept	11.382	1.437	131,737	43,175	29,753	0.032	0.546
Certolizumab pegol	11.382	1.437	133,165	44,602	30,732	0.020	0.483

CE, cost-effectiveness.
The probability of CE £20,000/30,000 is the probability that the TNF-α inhibitor, compared with CC, is a cost-effective option at the stated threshold.

TABLE 195 Non-radiographic axial spondyloarthritis: scenario 7 (rebound equal to gain)

Strategy	Total QALYs	Incremental QALYs	Total costs (£)	Incremental costs (£)	ICER (£)	Probability of CE £20,000	Probability of CE £30,000
Conventional therapy	9.906	–	90,850	–	–	–	–
Certolizumab pegol PAS	11.291	1.386	130,974	40,124	28,958	0.102	0.576
Adalimumab	11.291	1.386	132,373	41,523	29,968	0.069	0.506
Etanercept	11.291	1.386	133,119	42,269	30,506	0.059	0.484
Certolizumab pegol	11.291	1.386	134,547	43,696	31,536	0.045	0.454

CE, cost-effectiveness.
The probability of CE £20,000/30,000 is the probability that the TNF-α inhibitor, compared with CC, is a cost-effective option at the stated threshold.

Non-radiographic axial spondyloarthritis scenario results: rebound to conventional care

TABLE 196 Non-radiographic axial spondyloarthritis: scenario 1 (rebound to CC)

Strategy	Total QALYs	Incremental QALYs	Total costs (£)	Incremental costs (£)	ICER (£)	Probability of CE £20,000	Probability of CE £30,000
Conventional therapy	9.998	–	88,600	–	–	–	–
Certolizumab pegol PAS	11.646	1.648	124,101	35,501	21,537	0.386	0.787
Adalimumab	11.646	1.648	125,519	36,919	22,397	0.342	0.746
Etanercept	11.646	1.648	126,246	37,646	22,839	0.318	0.733
Certolizumab pegol	11.646	1.648	127,674	39,074	23,705	0.267	0.712

CE, cost-effectiveness.
The probability of CE £20,000/30,000 is the probability that the TNF-α inhibitor, compared with CC, is a cost-effective option at the stated threshold.

TABLE 197 Non-radiographic axial spondyloarthritis: scenario 2 (rebound to CC)

Strategy	Total QALYs	Incremental QALYs	Total costs (£)	Incremental costs (£)	ICER (£)	Probability of CE £20,000	Probability of CE £30,000
Conventional therapy	9.989	–	87,391	–	–	–	–
Certolizumab pegol PAS	11.460	1.472	124,660	37,268	25,326	0.230	0.618
Adalimumab	11.460	1.472	126,073	38,682	26,287	0.193	0.590
Etanercept	11.460	1.472	126,805	39,414	26,784	0.182	0.573
Certolizumab pegol	11.460	1.472	128,232	40,841	27,754	0.163	0.542

CE, cost-effectiveness.
The probability of CE £20,000/30,000 is the probability that the TNF-α inhibitor, compared with CC, is a cost-effective option at the stated threshold.

TABLE 198 Non-radiographic axial spondyloarthritis: scenario 3 (rebound to CC)

Strategy	Total QALYs	Incremental QALYs	Total costs (£)	Incremental costs (£)	ICER (£)	Probability of CE £20,000	Probability of CE £30,000
Conventional therapy	9.891	–	91,602	–	–	–	–
Certolizumab pegol PAS	11.066	1.175	132,047	40,445	34,416	0.052	0.396
Adalimumab	11.066	1.175	133,456	41,854	35,615	0.036	0.348
Etanercept	11.066	1.175	134,192	42,590	36,241	0.031	0.330
Certolizumab pegol	11.066	1.175	135,620	44,017	37,456	0.026	0.290

CE, cost-effectiveness.
The probability of CE £20,000/30,000 is the probability that the TNF-α inhibitor, compared with CC, is a cost-effective option at the stated threshold.

TABLE 199 Non-radiographic axial spondyloarthritis: scenario 4 (rebound to CC)

Strategy	Total QALYs	Incremental QALYs	Total costs (£)	Incremental costs (£)	ICER (£)	Probability of CE £20,000	Probability of CE £30,000
Conventional therapy	9.948	–	90,402	–	–	–	–
Certolizumab pegol PAS	11.223	1.275	131,015	40,613	31,841	0.063	0.456
Adalimumab	11.223	1.275	132,416	42,014	32,940	0.047	0.415
Etanercept	11.223	1.275	133,160	42,758	33,523	0.040	0.395
Certolizumab pegol	11.223	1.275	134,587	44,185	34,642	0.027	0.337

CE, cost-effectiveness.
The probability of CE £20,000/30,000 is the probability that the TNF-α inhibitor, compared with CC, is a cost-effective option at the stated threshold.

TABLE 200 Non-radiographic axial spondyloarthritis: scenario 5 (rebound to CC)

Strategy	Total QALYs	Incremental QALYs	Total costs (£)	Incremental costs (£)	ICER (£)	Probability of CE £20,000	Probability of CE £30,000
Conventional therapy	10.975	–	90,413	–	–	–	–
Certolizumab pegol PAS	12.462	1.487	130,404	39,991	26,900	0.069	0.678
Adalimumab	12.462	1.487	131,817	41,404	27,850	0.050	0.599
Etanercept	12.462	1.487	132,549	42,136	28,343	0.042	0.572
Certolizumab pegol	12.462	1.487	133,976	43,563	29,303	0.028	0.498

CE, cost-effectiveness.
The probability of CE £20,000/30,000 is the probability that the TNF-α inhibitor, compared with CC, is a cost-effective option at the stated threshold.

TABLE 201 Non-radiographic axial spondyloarthritis: scenario 6 (rebound to CC)

Strategy	Total QALYs	Incremental QALYs	Total costs (£)	Incremental costs (£)	ICER (£)	Probability of CE £20,000	Probability of CE £30,000
Conventional therapy	9.953	–	89,196	–	–	–	–
Certolizumab pegol PAS	11.228	1.275	131,515	42,319	33,184	0.013	0.398
Adalimumab	11.228	1.275	132,901	43,704	34,270	0.007	0.353
Etanercept	11.228	1.275	133,661	44,464	34,866	0.002	0.332
Certolizumab pegol	11.228	1.275	135,088	45,891	35,985	0.001	0.284

CE, cost-effectiveness.
The probability of CE £20,000/30,000 is the probability that the TNF-α inhibitor, compared with CC, is a cost-effective option at the stated threshold.

TABLE 202 Non-radiographic axial spondyloarthritis: scenario 7 (rebound to CC)

Strategy	Total QALYs	Incremental QALYs	Total costs (£)	Incremental costs (£)	ICER (£)	Probability of CE £20,000	Probability of CE £30,000
Conventional therapy	9.963	–	90,219	–	–	–	–
Certolizumab pegol PAS	11.200	1.237	131,714	41,495	33,555	0.057	0.396
Adalimumab	11.200	1.237	133,109	42,890	34,684	0.038	0.343
Etanercept	11.200	1.237	133,859	43,640	35,290	0.035	0.318
Certolizumab pegol	11.200	1.237	135,286	45,067	36,444	0.029	0.284

CE, cost-effectiveness.
The probability of CE £20,000/30,000 is the probability that the TNF-α inhibitor, compared with CC, is a cost-effective option at the stated threshold.

Additional validation scenarios assuming same baselines for responders and non-responders

TABLE 203 Ankylosing spondylitis: rebound equal to gain

Strategy	Total QALYs	Incremental QALYs	Total costs (£)	Incremental costs (£)	ICER (£)	Probability of CE £20,000	Probability of CE £30,000
Conventional therapy	7.262	–	111,636	–	–	–	–
Certolizumab pegol PAS	8.317	1.054	126,238	14,601	13,851	0.803	0.975
Golimumab	8.317	1.054	127,917	16,281	15,444	0.732	0.958
Adalimumab	8.317	1.054	128,002	16,366	15,525	0.730	0.958
Etanercept	8.317	1.054	128,383	16,746	15,886	0.708	0.952
Certolizumab pegol	8.317	1.054	129,811	18,175	17,241	0.645	0.931
Infliximab	8.317	1.054	146,079	34,443	32,673	0.044	0.376

CE, cost-effectiveness.
The probability of CE £20,000/30,000 is the probability that the TNF-α inhibitor, compared with CC, is a cost-effective option at the stated threshold.

TABLE 204 Ankylosing spondylitis: rebound to CC

Strategy	Total QALYs	Incremental QALYs	Total costs (£)	Incremental costs (£)	ICER (£)	Probability of CE £20,000	Probability of CE £30,000
Conventional therapy	7.274	–	109,511	–	–	–	–
Certolizumab pegol PAS	7.958	0.684	127,164	17,654	25,809	0.277	0.632
Golimumab	7.958	0.684	128,850	19,339	28,273	0.183	0.554
Adalimumab	7.958	0.684	128,934	19,423	28,396	0.178	0.550
Etanercept	7.958	0.684	129,309	19,799	28,945	0.165	0.534
Certolizumab pegol	7.958	0.684	130,738	21,227	31,034	0.107	0.473
Infliximab	7.958	0.684	146,808	37,298	54,528	0.000	0.010

CE, cost-effectiveness.
The probability of CE £20,000/30,000 is the probability that the TNF-α inhibitor, compared with CC, is a cost-effective option at the stated threshold.

TABLE 205 Non-radiographic axial spondyloarthritis: rebound equal to gain

Strategy	Total QALYs	Incremental QALYs	Total costs (£)	Incremental costs (£)	ICER (£)	Probability of CE £20,000	Probability of CE £30,000
Conventional therapy	9.977	–	88,692	–	–	–	–
Certolizumab pegol PAS	11.551	1.574	125,205	36,513	23,199	0.390	0.759
Adalimumab	11.551	1.574	126,606	37,914	24,089	0.341	0.733
Etanercept	11.551	1.574	127,350	38,658	24,562	0.319	0.720
Certolizumab pegol	11.551	1.574	128,777	40,085	25,469	0.272	0.702

CE, cost-effectiveness.
The probability of CE £20,000/30,000 is the probability that the TNF-α inhibitor, compared with CC, is a cost-effective option at the stated threshold.

TABLE 206 Non-radiographic axial spondyloarthritis: rebound to CC

Strategy	Total QALYs	Incremental QALYs	Total costs (£)	Incremental costs (£)	ICER (£)	Probability of CE £20,000	Probability of CE £30,000
Conventional therapy	10.030	–	88,389	–	–	–	–
Certolizumab pegol PAS	11.391	1.361	126,116	37,727	27,721	0.218	0.617
Adalimumab	11.391	1.361	127,525	39,136	28,756	0.176	0.586
Etanercept	11.391	1.361	128,261	39,872	29,297	0.160	0.574
Certolizumab pegol	11.391	1.361	129,689	41,299	30,345	0.133	0.537

CE, cost-effectiveness.
The probability of CE £20,000/30,000 is the probability that the TNF-α inhibitor, compared with CC, is a cost-effective option at the stated threshold.

EME
HS&DR
HTA
PGfAR
PHR

Part of the NIHR Journals Library
www.journalslibrary.nihr.ac.uk

This report presents independent research funded by the National Institute for Health Research (NIHR). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health

Published by the NIHR Journals Library

Paper 5.

Soares MO, Bojke L, Dumville J, Iglesias C, Cullum N, Claxton K. Methods to elicit experts' beliefs over uncertain quantities: application to a cost effectiveness transition model of negative pressure wound therapy for severe pressure ulceration. Statistics in Medicine 2011; 30(19): 2363-*2380.*

Research Article

Statistics in Medicine

Received 19 October 2010, Accepted 11 April 2011 Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.4288

Methods to elicit experts' beliefs over uncertain quantities: application to a cost effectiveness transition model of negative pressure wound therapy for severe pressure ulceration

Marta O. Soares,[a*†] Laura Bojke,[a] Jo Dumville,[b] Cynthia Iglesias,[b] Nicky Cullum[b] and Karl Claxton[a]

We can use decision models to estimate cost effectiveness, quantify uncertainty regarding the adoption decision and provide estimates of the value of further research. In many cases, the existence of only limited data with which to populate a decision model can mean that a cost-effectiveness analysis either does not proceed or may misrepresent the degree of uncertainty associated with model inputs. An example is the case of negative pressure wound therapy (NPWT) used to treat severe pressure ulceration, for which the evidence base is limited and sparse. There is, however, substantial practical experience of using this treatment and its comparators. We can capture this knowledge quantitatively to inform a cost-effectiveness model by eliciting beliefs from experts.

This paper describes the design and conduct of an elicitation exercise to generate estimates of multiple uncertain model inputs and validate analytical assumptions for a decision model on the use of NPWT. In designing the exercise, the primary focus was the use of elicitation to inform decision models (multistate models), where representations of uncertain beliefs need to be probabilistically coherent. This paper demonstrates that it is feasible to collect formally elicited evidence to inform decision models. Copyright © 2011 John Wiley & Sons, Ltd.

Keywords: cost-effectiveness; expert; elicitation; decision analytic models; negative pressure wound therapy

1. Introduction

Decision makers often require a quantitative evaluation of both the costs and benefits of alternative health technologies in order to choose treatments that offer the best value for money. We can estimate costs and benefits by synthesising data from multiple sources using a decision model; however, the input information is almost always associated with a degree of uncertainty. We must identify and characterise this uncertainty not only to produce accurate estimates of expected cost and benefits but also to evaluate whether existing evidence is sufficient to make a decision and/or to assess the possible consequences of an uncertain decision [1]. In many cases, the existence of only limited inputs (i.e. research data) with which to populate a decision model means that a cost-effectiveness analysis, which is often crucial in determining how scarce health care resources are allocated, either does not proceed or is at risk of misrepresenting the degree of uncertainty associated with model inputs.

An alternative option is to obtain quantitative representations of uncertainty from expert opinion. The process by which experts formulate a quantitative judgement based on their own beliefs (independently of the quality of such knowledge) for an uncertain quantity is called 'elicitation' [2]. Elicitation is an

[a] *Centre for Health Economics, University of York, York, UK*
[b] *Department of Health Sciences, University of York, York, UK*
**Correspondence to: Marta O. Soares, Alcuin A Block, Centre for Health Economics, University of York, YO10 5DD, York.*
[†] *E-mail: marta.soares@york.ac.uk*

Contract/grant sponsor: Medical Research Council; contract/grant number: G0501814

appropriate method by which to characterise uncertainties either by augmenting existing knowledge or by formulating a probabilistic judgement over a quantity when no data exist [3]. Although elicited evidence have huge potential, their use in health technology assessment (HTA) has been minimal [4, 5]. Much of the literature on elicitation focuses on deriving appropriate prior distributions for the parameters of a statistical model. Nonetheless, in HTA, we may use a decision model [6], rather than a statistical model to structure evidence and describe disease progression. Decision models can be complex, in turn requiring a complex elicitation process that may be in conflict with the desire for a user-friendly elicitation process [4].

This paper describes, in detail, the design and conduct of an elicitation exercise to generate estimates of multiple uncertain parameters and validate analytical assumptions for a decision model on the use of negative pressure wound therapy (NPWT) for severe pressure ulcers. To our knowledge, this is one of the most comprehensive elicitation exercises undertaken for use in HTA to date. Here, the primary focus was the use of elicitation to inform decision models, where representations of uncertain beliefs need to be probabilistically coherent.

2. Description of the case study: NPWT for severe pressure ulcers

Negative pressure wound therapy, also known as topical negative pressure, is a medical device used to treat full thickness wounds such as severe pressure ulcers. It has been claimed that NPWT speeds healing and reduces infection rates and costs as well as assists in the practicalities of wound management; however, there is very little actual evidence for its clinical or cost effectiveness [7–22]. NPWT is also a relatively expensive treatment used widely in the developed world; thus, it likely incurs a significant burden on health care resources. Therefore, there is a need to evaluate the cost effectiveness of NPWT and alternative treatments for its various indications, including severe pressure ulcers. Additionally, given the expected uncertainty surrounding the choice of treatment, it is important to explore whether investing in further research regarding the use of NPWT is worthwhile and, if so, what type of future research is most likely to offer the most value for money. We will describe the comprehensive report of this work separately (Soares MO, unpublished data), and hence we next give only a brief overview.

We designed a Markov decision model evaluated in discrete time as part of the above study. Clinical experts identified the relevant alternatives to NWPT for the patient population of people with severe ulcers to be spun hydrocolloid (HC), alginate and foam dressings. Clinical experts also helped design the model by identifying relevant events (independently of data availability); this information is crucial in determining the costs and effects of alternative treatments for severe pressure ulceration. We show the model structure in Figure 1, and it essentially tracks patients through three health states (unhealed, healed and dead) and determines their associated costs and quality-adjusted life years. Although we considered ulcer healing to be the main outcome, we also deemed other events to be relevant for inclusion in the model. We split transitions to healing by the method by which patients achieved healing [surgery to close the ulcer ('closure surgery') or secondary healing; see Figure 1]. We assumed that complications and treatment discontinuation occur in a proportion of unhealed patients (varying over time). We did not consider pressure ulcer recurrence in the economic model.

Table I further summarises the way in which the relevant events were represented in the model and the available information for each. For some inputs of the model, for example, occurrence of ulcer healing, limited data existed in the literature; however, for others there was no relevant evidence, for example, rates and outcomes of closure surgery and complications (from a UK perspective).

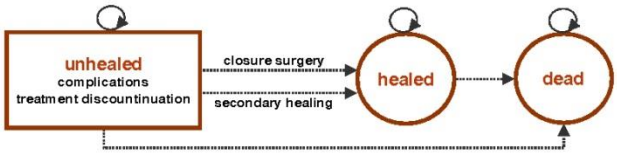


Figure 1. State transition diagram for the theoretical model designed to represent treatment of grades 3 and 4 pressure ulcers with and without NPWT. Death is an absorbent state.

Table I. Description of information needs and general requirements to incorporate in the model.

Brief description	How were these quantities modelled?	Impacts on	Does evidence exist?	Elicited?
Events of interest				
Occurrence of healing	Associated with transitions from state 'unhealed' to 'healed'	Costs and benefits	Yes, sparse evidence exists on multiple studies focussing on alternative treatments. The evidence on the hazard of healing for alternative treatments (hazard and hazard ratios) was formally synthesised	Yes
Impact of treatments on healing	Associated with transitions from state 'unhealed' to 'healed'	Costs and benefits		Yes
Discontinuation of treatments, frequency of application of treatments	Associated with the stay in state 'unhealed'	Costs	No robust evidence exists besides manufacturer recommendations	Yes
Closure surgery	Associated with transitions from state 'unhealed' to 'healed'	Costs	No evidence relevant for the UK exists	Yes
Complications	Associated with the stay in state 'unhealed'	Costs	No robust evidence exists	Yes
Death	Associated with transitions to state 'dead'	Costs and benefits	Yes, general population estimates were used because there is no robust evidence that pressure ulceration affects mortality	No
Other quantities				
Resource use and/or costs	Associated with the permanence in each of the health states or transition between states (e.g. surgery)	Costs	An informal elicitation of expected total costs exists, although uncertainty was not considered.	No
Utilities	Associated with the permanence in each of the health states or transition between states	Benefits	Primary data exist from an observational study	No

3. Methods

We captured experts' knowledge through formal elicitation of their beliefs, translating these beliefs into an explicit, probabilistic language for inclusion in the decision analytic model. We may define designing the elicitation exercise by a number of sequential, yet interrelated, tasks (see Figure 2). This diagram is specific to the case study presented here; however, there may be other tasks relevant to other elicitation exercises [2]. The following sections describe the issues and the practical solutions relevant to the design and conduct of the elicitation exercise used to inform the NPWT case study.

3.1. Defining the aims of the elicitation exercise

Correctly identifying the parameters to elicit is crucial to ensure that the results are fit for purpose [2]. For this case study, data characterising transitions and relevant events of the multistate model were either sparse or non-existent. We thus elicited all transitions and related events (except death) relevant to the model, including beliefs about the impact of the alternative treatments on the occurrence of events, that is, relative effectiveness. We also elicited uncertainty over the quantities of interest. We did not elicit resource use or cost parameters. Although the methods would have been appropriate, we deemed that eliciting these parameters would have a significant impact on the burden of the exercise.

Eliciting multiple quantities from an expert, however, raises the issue of possible dependence between responses. Accounting for this would require eliciting beliefs not only on the quantities themselves but also on how the experts' judgements would change if values for other parameters were known. Methods to elicit this are complex, and it has been suggested that experts may be unable to supply the required level of detail [23]. Here, we assumed independence and elicited only marginal distributions [3].

3.2. Selecting the experts

There is no ideal method for identifying experts for elicitation. O'Hagan [2] describes an expert simply as someone possessing knowledge about the uncertain quantities of interest. Furthermore, there is no clear guidance on how to recruit experts or how many experts to recruit, only that the process must be transparent [24].

Substantive experts [24] in HTA are likely to be health care professionals. Using such experts raises a number of issues that need to be considered further. Firstly, because we need a variety of quantities to inform a decision model, there may be subsets where different experts are appropriate. A second, but possibly related, issue is that there may be some heterogeneity between experts, for example, if their clinical case mix experience differs (e.g. between specialists and generalists or geographical areas). To represent this heterogeneity in the results, Clemen and Winkler [25] argue for the use of multiple and heterogeneous experts. Thirdly, it is unlikely that the proposed experts have knowledge of elicitation or relevant mathematical skills. Health care professionals, such as nurses, have in general limited training in quantitative subjects [26]. Because the elicitation task requires experts to express their beliefs (and associated uncertainty, i.e. the strength of belief) in a quantitative manner, it is essential that we fully consider participants' training needs at the planning stage.

In the UK, it is mainly nurses that treat and manage pressure ulcers. Their experience in observing and managing the progression of pressure ulceration makes them substantive experts in the topic. We undertook this exercise on a sample of nurses with experience and knowledge of tissue viability and wound

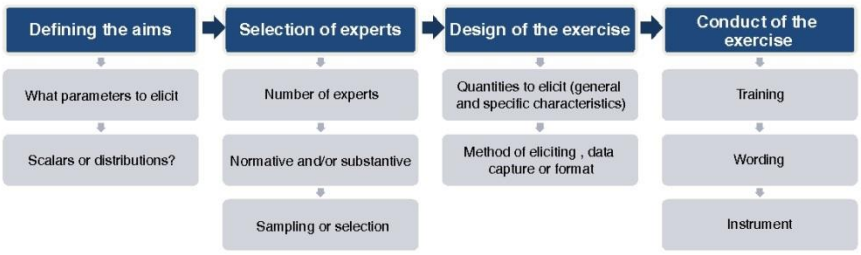


Figure 2. Schematic representation of general steps to design an elicitation exercise to inform one or more parameters of a decision model.

management. In inviting nurses to participate in the elicitation exercise, we ensured that we included nurses from all relevant settings (both hospital and community) and both specialists and generalists. We offered a small honorarium (£100) to the participating nurses.

3.3. Designing the elicitation exercise

In designing the elicitation exercise, we considered a number of specific issues regarding the characteristics of the quantities to be elicited and the data capture method.

3.3.1. General characteristics of the quantities to be elicited. Once we have identified the parameters of interest, we must make a decision about the specific quantities that we will elicit in order to inform such parameters. There are often several alternatives that we can use. For example, when eliciting a transition probability regarding ulcer healing, we could ask experts to indicate their beliefs regarding the transition probability itself, the time required for $x\%$ patients to heal or the proportion of patients healed after y amount of time. Conditional on certain assumptions, evidence on each of these quantities can inform the same parameter. The choice regarding the quantities to be elicited should consider the following factors:

Format of existing evidence and fitness for purpose. If evidence obtained through experimentation exists, the way in which these data were previously collected, analysed, synthesised or reported can help define the quantities to elicit. In the case study, we also considered the way in which these quantities were to be included in the model, for example, as transitions describing a multistate model.

Directly observable. It may be difficult to derive experts' beliefs on some of the quantities of interest because these may not be directly observable [27]. Previous research suggests that an expert is unlikely to be able to specify non-observable quantities reliably [2], and hence we have only elicited observable quantities [28,29]. Because we only elicit observable quantities, we likely need further parameterisation (i.e. transformation) for inclusion in the decision model.

Homogeneity of quantities elicited. There is a risk of excessive complexity where we require multiple quantities using 'distinct' summaries (e.g. asking for beliefs over proportions and over means). In these instances, we may need more intensive training to ensure that experts comprehend the elicitation task appropriately. To alleviate this additional strain, in the case study we expressed all questions in a consistent way, that is, we elicited proportions throughout the exercise to obtain estimates of parameters of binomial variables.

3.3.2. Method of data capture. Key to the success of any elicitation exercise is the format in which we pose questions to the experts (or the method of data capture). The method of data capture must be clear, concise and comprehensive, must capture the quantity of interest and must be used in a user-friendly interface.

A comprehensive review of alternative methods of data capture is available in the literature [2] and is beyond the scope of the current paper. However, when choosing a format of data capture for elicitation in HTA, it is important to consider that the elicitation of distributions will normally be necessary, and methods for these will differ from those used to elicit unique values, that is, point estimates.

The characteristics of the quantities of interest are also relevant. Within the case study, all elicited parameters were binomial variables. Various methods of eliciting these quantities are available in the literature [2,24,30,31]. Because of the non-mathematical background of our experts, we chose a graphical format, the histogram technique (sometimes referred to as probability grid) to undertake this elicitation [32].

The histogram method is well described in the literature [33] and has been previously used in elicitation exercises developed for HTA [34,35]. It is a fixed interval method based on the probability density function [2]. We partitioned into intervals the range of values the quantity may take, and for each, we collected information on the probability of observing values in such interval. For ease, we used a discrete version where we asked experts to indicate their strength of belief over exact numbers (i.e. 0, 5, 10, ..., 100) rather than intervals (i.e. 0 – 5, 5 – 10, 10 – 15, ..., 95 – 100).

All quantities elicited were probabilities; thus, we used a common scale (from zero to 100). We asked individual experts to place 21 crosses on a grid defined to have 21 × 21 cells (Figure 3). By placing the 21 crosses in the grid, the expert is, in reality, attributing a probability mass to each of the possible

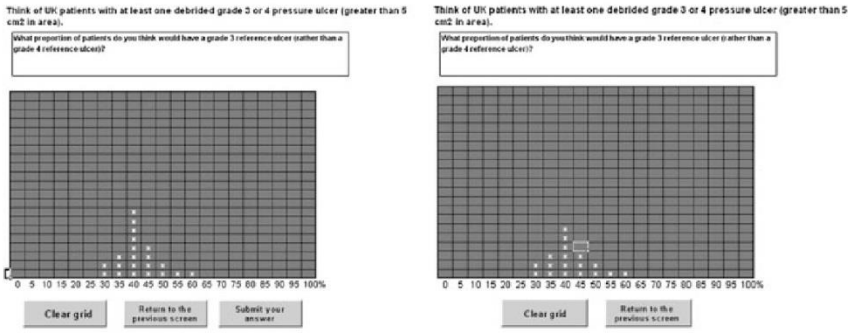


Figure 3. Graphic set-up for the data capture histogram.

values, where each cross represents 4.76% probability. The expert can then express certainty by stacking all of the crosses in the same value (vertical column) or express full certainty that a value is not possible by not attributing any crosses to it. By attributing one cross to each possible value, the expert expresses that any value could be possible, that is, full uncertainty.

3.3.3. Judgements defining a multistate process. It is important to achieve coherence in the probability judgements when eliciting quantities to inform transitions of a multistate model. We made the following considerations when further defining the quantities being elicited.

Informing multiple transitions occurring through time. The probabilities of individuals being in each one of the mutually exclusive states of the multistate process, at any given time point, should sum to one. A Dirichlet distribution can describe this scenario. However, for the case study, we regarded methods to elicit Dirichlet distributions directly as too complex for experts [2, 24]. As an alternative, we elicited a set of $N-1$ conditionally independent binomial variables. We depicted an example of the questions posed in the final exercise for quantities 1 and 2 in Table II. Here, we defined only forward transitions; further thought is warranted in exercises where backward transitions also need to be elicited [36].

Time dependency. Time dependency occurs when the rate at which transitions occur depends on time elapsed, a common feature of decision models. There are known challenges in eliciting the parameters of the Weibull distribution commonly used to represent time dependency [37–41]. We thus adopted a non-parametric approach by eliciting the proportion of patients healed for two separate and non-overlapping time periods (questions 2 and 4 in Table II). If time dependency is not important, we expect the rate of healing to be very similar in both periods; however, the experts may still be more uncertain about outcomes attained over the longer term. By assuming independence between the elicited quantities, we can integrate time dependency in the decision model without loss of probabilistic coherence.

Eliciting other judgements related to transition probabilities. There are a number of alternative parameterisation strategies that we can use to avoid complexity when specifying decision models. In our case study, although the type of healing was used to split transitions to ulcer healing (Figure 1), existing healing data did not distinguish between these healing types (both being included under the umbrella term of healing with no further information). Since no relevant data for the UK were found on the rates of closure surgery and their associated outcomes, elicitation was the only source of evidence. In cases such as this, it is important that the elicited beliefs do not affect judgements over the overall transitions to healing. For this reason, the rate of occurrence of closure surgery was elicited by conditioning it on healing outcomes. If S represents the closure surgery event and H healing, we elicited the following:

- $\mathrm{P}[S \mid H]$, the probability of patients having had surgery knowing that they healed (question 9 in Table 2), and
- $\mathrm{P}[H \mid S]$, the probability of healing in patients that received surgery (question 10 in Table 2).

By knowing the unconditional probability of healing (elicited through question 2 in Table II), $P[H]$, and applying the law of total probability, we calculated the probability of receiving closure surgery as

Table II. Examples of the questions used to elicit experts' beliefs. Distributions obtained by pooling (linear pooling) the individual experts' replies and fitted distributions.

ID	Elicitation question	Box plots of individual replies (pooled distribution presented in red when appropriate +)	Transformation	Pooled and fitted distribution
1	*p_g3* What proportion of patients do you think would have a grade 3 reference ulcer (rather than a grade 4 reference ulcer)?	*p_g3*	No	*p_g3* Beta, mean = 0.50, CI = [0.13 to 0.87]
2	*p_mu* (informs the hazard of healing) Think of UK patients with at least one debrided grade 3 or 4 pressure ulcer (greater than 5 cm^2 in area). Patients start treatment with a non-silver spun hydrocolloid/hydrofibre dressing as the primary contact layer. If patients have multiple grade 3 or 4 ulcers, assume that you are treating the deepest ulcer (we will refer to this as the reference ulcer). Once patients have started this treatment, they continue to receive standard pressure ulcer care (i.e. support surfaces). Over time, treatment with spun hydrocolloid/hydrofibre may or may not stop depending on clinical decisions made. Six months after starting treatment with HC, what proportion of patients who are alive do you think would have a healed reference ulcer? This is regardless of whether patients are still receiving treatment with HC at this 6- month point.	*p_mu*	Yes *	*mu*, Log hazard of healing for HC (0 to 6 months) Normal, mean = -3.74, CI = [-5.96 to -1.52]
3	*p_comp* Now think of those patients who still have a grade 3 or 4 pressure ulcer after 6 months of treatment (the unhealed patients). At this 6-month point, what proportion of these unhealed patients do you think would have osteomyelitis (OM) or systemic infection (SI)?	*p_comp*	No	*p_comp* Beta, mean = 0.43, CI = [0.05 to 0.88]

Table II. Continued

ID	Elicitation question	Box plots of individual replies (pooled distribution presented in red when appropriate +)	Transformation	Pooled and fitted distribution
4	*p_mu6to12* (informs the hazard of healing) Think only of those patients who still have a grade 3 or 4 pressure ulcer 6 months after starting treatment with HC. What proportion of them do you think would heal their reference ulcer between 6 and 12 months?	*p_mu6to12* *Red triangles indicate reference values*	Yes *	*mu6to12* Log hazard of healing for HC (6 to 12 months) Normal, mean =-4.28, CI = [-10.24 to 1.67]
5	** *p_F* (informs the relative effectiveness on healing for foam dressings, *F*) *[Statement 1] - Your strongest belief was that <<ref>>% of patients had a healed ulcer 6 months after starting HC. Assume that this value is true.* Six months after starting treatment with foam dressings, what proportion of patients who are alive do you think would have a healed reference ulcer?	*p_F* *Red triangles indicate reference values*	Yes *	*d[F]* Log HR of healing for *F* vs HC Normal, mean =-0.96, CI = [-6.32 to 4.40]
6	** *p_ALG* (informs the relative effectiveness on healing for alginate dressings, ALG) *[Statement 1]* Six months after starting treatment with alginate dressings, what proportion of patients who are alive do you think would have a healed reference ulcer?	*p_ALG* *Red triangles indicate reference values*	Yes*	*d[ALG]* Normal, mean =0.003, CI = [-0.63 to 0.64]

Table II. Continued

7	** *p_NPWT* (informs the relative effectiveness on healing for topical negative pressure therapy, NPWT) *[Statement 1]* Six months after starting treatment with topical negative pressure therapy, what proportion of patients who are alive do you think would have a healed reference ulcer?	*p_NPWT* *Red triangles indicate reference values*	Yes*	*d[NPWT]* Normal, mean =0.45, CI = [-0.66 to 1.56]	
8	*maint* Think of the unhealed patients who do not have OM or SI at this 6-month point. What proportion do you think will still be being treated with HC?	*maint*	No	*maint* Beta, mean = 0.35, CI = [0.01 to 0.89]	
9	*p_SH* Imagine that these patients had their deepest grade 3 or 4 pressure ulcer (reference ulcer) treated and it healed completely at some point. What proportion of patients with a healed reference ulcer achieved this healing as a direct result of closure surgery?	*p_SH*	No	*p_SH* Beta, mean = 0.15, CI = [0.00 to 0.69]	

Table II. Continued

ID	Elicitation question	Box plots of individual replies (pooled distribution presented in red when appropriate +)	Transformation	Pooled and fitted distribution
10	*p_HS* Consider patients with a grade 3 or 4 reference ulcer on which closure surgery has been performed. What proportion of patients who received this surgery do you think will still have intact skin on the closure surgery site 1 month post-surgery?	*p_HS*	No	*p_HS* Beta, mean = 0.61, CI = [0.08 to 0.99]

+For questions 5, 6 and 7, it is not appropriate to pool directly the individual judgements as the individual replies were conditioned to reference values.

*We calculated Log hazard rates assuming a constant transition rate. As an example, the transition rate for mu was derived as follows: *mu* = log[1-exp(*p_mu* *6)].
The log hazard ratios (HR) were calculated by dividing the hazard elicited for the treatment of interest by the specific value of the hazard for the comparator (HC) in which the question was conditioned. Exemplifying, d[*NPWT*] = log[1-exp(*p_NPWT* *6)/(1-exp(*ref*6))].

**These histograms were preceded by filter questions: "Think of UK patients with at least one debrided grade 3 or 4 pressure ulcer (greater than 5 cm^2 in area). Assume that the deepest ulcer was treated with spun hydrocolloid/hydrofibre as the primary contact layer and a certain proportion healed. Do you think the proportion healed would be different if, instead of a HC, patients were treated with F/ALG/NPWT". Only experts suggesting treatments to be different proceeded to elicit relative effectiveness and its uncertainty. In this way, we avoided eliciting through the grid when experts expressed that they were fully certain that the treatments were similar or they were fully uncertain about this. In this case, we assumed the density to be fully allocated to the reference value or to be equally distributed by the range of values available, respectively.

follows: $P[S] = \frac{P[S|H] \cdot P[H]}{P[H|S]}$. By eliciting occurrence and outcomes of surgery in this way, we guaranteed consistency between these and the existing data regarding healing.

3.3.4. Eliciting relative effectiveness. To elicit the effectiveness of the three comparators and active treatment (NPWT), we can specify either absolute or relative measures of effectiveness. Absolute effectiveness measures may be constrained through correlation. To remedy this, O'Hagan [2] suggests eliciting both baseline and relative measures of effectiveness and assuming independence between these quantities. However, relative effectiveness measures (e.g. hazard ratio) are non-observables and are therefore difficult to elicit.

Additionally, the elicitation format chosen (the histogram) requires quantities to be bounded; that is, an idea of the magnitude of the uncertainty is required to pick a sensible range of values for probabilities. We therefore used a different approach to elicit relative effectiveness based on conditional independence. As before, we first asked experts to record the probability (and uncertainty) of a patient being healed when they received treatment with HC dressing (question 2, Table II). Following this, we asked the experts if they believed another treatment would return different outcomes (filter question, example in footnote of Table II). If individuals replied that the effectiveness of a specific treatment was not different from HC, we elicited no further information and attributed all of the density to the conditioning value. Alternatively, when individuals stated that they did not know whether the treatments were different, again we did not use the histogram and assumed full uncertainty. Otherwise, the experts expressed their beliefs through the histogram. In doing this, we asked the expert to assume that the value they believe best represented their knowledge about the effectiveness of the comparator treatment, HC, was true (reference value). The reference value was the mode (or one of multiple modes, selected at random). We then elicited the absolute effectiveness of the treatment of interest and its uncertainty, conditional to the value assumed known for the comparator treatment. See questions 5, 6 and 7 (Table II) for examples of how for examples of how this was implemented.

Because the experts conditioned their judgement about the effectiveness of a treatment on a plausible value for the comparator, when analysing this evidence we need to assume that this relationship sustains for any other possible values of the absolute effectiveness of the comparator. This is an assumption commonly undertaken when analysing effectiveness data [42].

3.4. Implementing the elicitation exercise

We designed a computer-based instrument, built in Microsoft Office Excel (Company: Microsoft Redmond, Washington, USA) using Visual Basic for Applications, to conduct the elicitation. Its use had several advantages, including the ability to avoid inconsistencies in completing the histogram, providing graphical representations of the elicited distributions to experts and allowing them to easily review and amend their answers. Also, we easily implemented a programme to condition beliefs on previous responses (by the same expert). When replying to conditional statements on competing events defining the multistate model, the questions appeared sequentially in the same sheet, unravelling as respondents answered them. It was possible to implement filter questions before conditional statements on relative effectiveness (see section 3.3.4). We used several strategies to ease the task of eliciting, such as grouping related questions together and having auto click buttons for completing the histograms. Multiple team members carefully considered the instrument and wording of questions, and then one nurse piloted these prior to the elicitation task.

3.5. Combining judgements from multiple experts and distribution fitting

3.5.1. Approach used to combine judgments. When we obtain judgements from multiple experts, it is often appropriate to obtain a unique distribution that reflects the combined judgements of all experts. Methods to achieve this fall into two categories, behavioural and mathematical approaches. In behavioural aggregation methods, we encourage experts to interact in order for them to achieve a level of agreement for a particular parameter [43]. Mathematical approaches to elicitation instead focus on generating individual valuations which we then combine to generate a single estimate or distribution using various mathematical techniques. These differ substantially and quite often produce dissimilar results. They also have different requirements in terms of subsequent synthesis [25]; in particular, behavioural methods avoid issues of synthesis that are inevitable with the mathematical method, namely the choice of synthesis method and whether or not to calibrate experts [2].

Prior to the case study elicitation, we conducted a pilot exercise to evaluate the use of the two approaches to elicitation. The elicitation again focussed on the development, treatment and healing of pressure ulcers for UK patients, and we elicited beliefs from nurses. We designed a questionnaire to elicit distributions for six quantities using the histogram. We elicited beliefs using both methods, mathematical and behavioural. There were mixed messages from experts regarding the ease of the two approaches, and the results showed that there was some variation in elicited distributions between elicitation approaches. The behavioural method produced responses with narrower confidence/credible intervals but also generated incoherent probability statements between two related quantities in suggesting that the median healing time was greater than the time taken for 70% of patients to heal. These results are in line with the general view that consensus promotes the use of heuristics and may produce overconfident judgements [25]. As a result of this pilot, we chose the mathematical approach for the case study elicitation exercise.

3.5.2. Synthesis. In the mathematical approach, we must undertake the aggregation of judgements from individual experts explicitly via a mathematical procedure. There are two main ways to do this: opinion pooling and Bayesian methods [2]. Bayesian methods rely on the idea that the aggregate distribution represents the belief of the decision maker and are thus difficult to implement. In pooling methods, the aggregate distribution is defined as a function of the individual assessments, where judgements from different experts may be weighted. It is not clear from the literature which pooling method performs best [2]; however, linear opinion pooling is the method most commonly applied in the synthesis of elicited evidence [44], and we have used this method in the case study. In linear pooling, we aggregate experts' assessments using simple linear combinations. If $p(\boldsymbol{\theta})$ is the probability distribution for unknown parameter θ, in linear pooling, experts' assessments are aggregated as follows: $p(\theta) = \sum_i w_i \cdot p_i(\theta)$ where w_i is the weight attributed to the ith expert.

In weighting the experts unequally, 'better' experts contribute the most to the aggregate distribution. We can derive weights by eliciting distributions for seed variables, variables whose distribution is known to the facilitator but not to the expert [44]. We assessed the use of a calibration procedure within the pilot exercise. We used four known seed questions. Results showed that alternative seed questions generated disparate weights for the experts, leading to differences in the aggregate distribution when we used each of the seed questions and when we used the average weight. Although it is generally accepted that we should use as many known parameters as possible to generate weights [44], the appropriateness of using particular seed questions is difficult to assess, especially in the context of a complex exercise, where unknown questions of varied nature are elicited simultaneously by nurses with differing expertise.

Because further research is needed in defining the process by which weights are generated and used and because, for the case study eliciting from, numerous types of seed variables (relating to treatment effectiveness, healing rates etc) would have been required, in the full elicitation exercise we instead equally weighted the experts.

Where needed, we transformed each expert's judgements over probabilities attained in the full elicitation to reflect an appropriate measure for use in the decision model (i.e. hazard or hazard ratios). Before transforming beliefs over relative effectiveness, we translated replies to the filter questions into judgements for the relevant distribution (see Section 3.3.4). We then aggregated beliefs expressed on the transformed scale using linear pooling.

3.5.3. Distribution fit. After obtaining a pooled, discretised distribution for each parameter, we conducted fitting of a parametric continuous density. Alternatively, we could have used the empiric distribution; however, such a non-parametric approach can complicate, if not impede, further analyses, for example, by not allowing the use of conjugacy in Bayesian updating. This was a relevant issue in the case study where an analysis of the value of further research was required.

The fitting procedure may itself introduce uncertainty [45]. To reduce this uncertainty, some exercises in the literature conduct the elicitation by doing face-to-face interviews where the expert validates and tunes the fitting of distributions [46,47]. In the current study, we elicited multiple quantities from multiple experts (we elicited 18 uncertain quantities from 23 experts, returning an overall of 414 distributions); therefore, the added task of fitting a distribution to each quantity in the presence of the expert was considered significantly burdensome. We have also assumed that distributions fitted according to statistical criteria represent the beliefs of the expert appropriately and that no imprecision is added in the process of fitting. Specifically, we calculated summary measures (mean and variance) for each of the pooled distributions, and we fitted parametric, continuous distributions using the method of moments. These

were either beta (for untransformed quantities) or normal distributions (for log hazards or log hazard ratios).

3.6. Conduct of the exercise

An experienced facilitator led the elicitation process. We conducted an extensive training session prior to the elicitation. The training took the form of a talk on the concepts and questions covered in the exercise followed by an opportunity to complete example exercises. Specifically, the training session covered the following:

- Overview of the method of expert elicitation and its use in the project
- Overview of sections included in the exercise
- Explanation of uncertainty and how to express this uncertainty in their beliefs
- Overconfidence and other potential biases (e.g. motivational bias caused by conflicts of interest)
- Example exercises
- How to use the computer-based instrument.

Following the training session, each of the experts completed the exercise independently, and after they replied to a first round, we asked the experts to review their replies. During the exercise, the facilitator and trained tutors were available at all times to resolve any concerns. We urged experts to make as much use of the tutors and facilitator as possible.

At the end of each section of the exercise, we asked experts whether they thought the answers they had given reflected their views. This allowed an assessment of face validity. Additionally, we asked experts to comment on the clarity of the questions and state if it was difficult or challenging to reply to the questions. To understand the challenges posed by the exercise, we also recorded additional comments at the end of the exercise.

4. Results of the elicitation

The elicitation was conducted in April 2009. Here we briefly provide an overview of the conduct of the exercise: feasibility, adequacy and face validity and comparability with existing data. We provide examples of results in Table II.

Twenty-three expert nurses participated in the exercise. Fourteen participants were specialist nurses (tissue viability or wound care) and the remaining district, staff or ward nurses; eight were acute care nurses, 14 were community based, and one worked across both acute and community settings. All participants had experience of treating people with pressure ulcers (median = 13 years; min–max = 2–30 years) and had treated 2–250 patients with pressure ulcers (grade 2 or more) in the last 6 months. Although all had experience of using NPWT, this varied: six nurses reported not having used this treatment for pressure ulcers in the last 6 months (median = 1 wound; min–max = 0–20 wounds), whereas another six reported not having used NPWT for other wound types over the last 6 months (median = 4 wounds; min–max = 0–50 wounds).

4.1. Feasibility

The exercise ran smoothly, and the conduct of the exercise did not pose any major challenges. The elicitation session lasted for approximately 4 h and the preparatory training for around 2 h, after which the experts declared themselves happy about both using the computer-based tool and expressing their judgements using the histogram technique. In total, each expert answered more than 30 questions, 18 of which were uncertain parameters elicited through the histogram. We depicted the wording and elicited descriptions of an example set of questions in Table II. There were very few missing or invalid data—only two experts gave some missing values. The responses from one of the experts regarding relative effectiveness were invalid as this individual did not assign any density to the reference value and instead placed the crosses around this value. The remaining experts expressed their judgements coherently.

4.2. Adequacy and face validity

The quality of elicited evidence is impossible to measure as we cannot know the expert's true beliefs. However, we can make a critical assessment regarding the extent to which the exercise was adequate

and whether it was fit for purpose. The results for the case study show that there was variation in elicited distributions between individuals and between questions (see Table II for examples). In elicitation, we request individual beliefs, and we thus expect variation between experts. For some questions, experts expressed discordant views; for example, the mode varied between 10% and 75% for question 1 in Table II. This variation is also desirable as, after synthesising these judgements, all views will be represented and thus used further in the decision making process. Due to the nature of this specific exercise, we also expected between-questions variation as different experts had different levels of expertise. Experts used a variety of distributional shapes to characterise their strength of belief, showing that they felt comfortable in using the histogram method.

We also show the linearly pooled discretised distributions and the corresponding fitted distributions in Table II for some of the elicited quantities. The observed variation between experts seems to be mirrored in the wide uncertainty over the pooled distributions. Experts (as a group) expressed different judgements over the relative effectiveness of the treatments: the pooled evidence suggested that we expect foam dressings to be slightly less effective than the comparator, alginate dressings to have the same effectiveness and NPWT to be slightly beneficial. This was a reflection of their individual replies.

The vast majority of experts found the questions clear and stated that the task was challenging (instead of easy or extremely difficult). In terms of face validity, throughout the sections, the majority of experts stated that their answers reflected their views. In general, experts found the exercise to be well organised, the training session to be useful and very clear and the computer-based approach to have been helpful. Some experts stated that for those questions where they had extensive clinical experience, providing the responses was somewhat easier. For most, the task was challenging not because of any issues with the technicalities of the elicitation task but because of the difficulties in making the judgements. Two nurses expressed that sometimes it was difficult to formalise their judgements because, in practice, they do not care for the patient through to healing. The difficulty arises because intuitively they have to account for censoring (the greater the level of censoring, the greater difficulty experts are expected to have in formalising judgements).

4.3. Comparison with existing data

In a separate stream of work, evidence from the literature was sought to inform the input parameters of the decision model (Soares MO, unpublished data) Although sparse, we can compare the available evidence with the elicited judgements. Literature reviews focused on specific questions, reflecting the data required [48]. We could only identify existing data for the effects of NPWT and its comparators, quality of life and costs. As we did not elicit beliefs on costs and quality of life, we will solely do the comparison of elicited judgements with existing data for effectiveness.

We found one randomised controlled trial (RCT) investigating NPWT and 11 investigating dressings. We linked data from these trials within an evidence network and synthesised the data using Bayesian indirect and mixed treatment comparisons (IMTC) [49] (methods and results are not presented here in detail). Because most links in the network were informed by a single study and the number of healing events in some trials was small or zero, it was not possible to obtain inferences on foam using unconstrained treatment effects. Therefore, we initially assumed a common but random effect of any dressing treatment [50]. We presented results as log hazard ratios (Table III, second column), with HC used as a reference treatment.

Both existing (Table III, second column) and elicited evidence (Table III, third column) showed wide credible intervals reflecting current uncertainties. The only marked difference between data sources was on the relative effectiveness of foam, where elicited expert opinion judged foam to be less effective than the scenario where existing evidence only was considered.

Assuming the mechanism through which the experts formulate their judgements does not rely on information available from the literature, we can collate the two sources of evidence. We used Bayesian updating for this purpose. We generated a combined, posterior distribution (Table III, fourth column) incorporating both the prior distribution (elicited beliefs) and observations from the existing evidence.[‡] In this scenario, the use of elicited data allowed the implausible assumption of exchangeability to be dropped, and thus we estimated treatment effects for each dressing individually. The inclusion of elicited evidence allows formal clinical beliefs to inform uncertainties in estimating relative cost effectiveness

[‡]*We used elicited judgements as prior distributions to the relevant unconstrained treatment effects and updated existing data through the MTC.*

Table III. Summary of existing evidence on the relative effectiveness of treatments derived from the literature (second column, 'Existing evidence'), elicited evidence (third column) and posterior distribution generated by updating existing with elicited evidence (fourth column, 'Existing and elicited evidence collated'). *

	Existing evidence Mean [95% CrI]	Elicited evidence Mean [95% CrI]	Existing and elicited evidence collated Mean [95% CrI]
Log hazard of healing for HC	−3.95 [−4.50 to −3.46]	−3.74 [−5.96 to −1.52]	−3.97 [−4.59 to −3.46]
Relative effectiveness in relation to HC			
Log hazard ratio of healing for F	0.03 [−1.97 to 1.86]	−0.96 [−6.32 to 4.40]	−0.91 [−2.14 to 0.21]
Log hazard ratio of healing for ALG	−0.19 [−1.76 to 1.13]	0.003 [−0.63 to 0.64]	−0.27 [−2.12 to 1.57]
Log hazard ratio of healing for NPWT	0.18 [−2.17 to 2.63]	0.45 [−0.66 to 1.56]	0.47 [−1.18 to 2.10]

CrI – Credible interval.

* Posterior means are not necessarily in between the means of the distributions used to generate it. This is because we did not directly combine the distributions describing existing and elicited evidence. Instead, we combined evidence from the literature either with uninformative or vague priors (to generate the distribution described as 'Existing evidence') or with elicited evidence (to generate the distribution described as 'Existing and elicited evidence collated').

and the value associated with further research; otherwise, an analytic assumption would have influenced results. As identified previously, the collation of the two sources of evidence assumes that experts have not considered existing information from the literature. To justify the use of such an assumption, we argue that within the current case study, existing evidence is very sparse (with few studies including a small number of patients) and dispersed in the literature. We did not incentivise nurses to review the existing literature before the elicitation exercise, and we did not divulge which treatments we were aiming to evaluate before the exercise. Moreover, the results regarding the synthesis of existing evidence (the IMTC) used to generate existing evidence were not available to the public. Although there is no way to know how much existing evidence the nurses had considered when formulating their judgements, we do believe their own clinical experience was the main influence.

5. Discussion

The evidence base for NPWT is limited and sparse; however, there is substantial practical experience of using this treatment and its comparators. Although the use of formally elicited evidence to inform decision models has not been widespread, this paper demonstrates that it can be regarded as a key source of evidence and that excluding relevant clinical experience would have misrepresented current knowledge over the effectiveness of alternative treatments for severe pressure ulcers. Alternatively, the evaluation could be postponed until further evidence is collected, leaving clinicians uninformed about which treatments might be most appropriate for their patients.

By using clinical experience to inform adoption decisions, we draw conclusions that are dependent on the experts' judgements.Applications of elicitation should show how elicited data compare with other available data and the impact of using these data in the results, where possible. In this case study, elicited data allowed analytical assumptions needed under sparseness of data to be relaxed. We showed this clinical experience to be valuable in describing parameters for which little information was available (e.g. relative effectiveness of the treatment foam).

Although elicitation is often used informally for decision making in health care, there are few published examples of formal elicitation exercises. The examples to date relate to the elicitation of prior beliefs in clinical trials [23, 51] and survival analysis [52]. Elicited data has also been seldom used in both stochastic cost effectiveness analysis [4, 53, 54] (where the data on net costs and net effects of alternative treatments generated within a trial is updated with prior information on these) and in decision modelling [5, 34, 55, 56]. Although there is increasing awareness of the advantages of using elicitation [57], the design and conduct of elicitation needs to be further explored to assess its potential feasibility and to facilitate the transference of guidance from the existing elicitation literature (mainly relevant for eliciting parameters of statistical models rather than decision models). This paper reports the design of

probably one of the largest elicitation exercises to have been undertaken in ' the field of HTA. Also, to our knowledge, no elicitation exercise had previously attempted to elicit parameters of a multistate model, ensuring both probabilistic coherence and ease of completion.

In designing the elicitation exercise, we had to make a number of choices regarding methodology, including whether to elicit observable or unobservable quantities, how to elicit distributions and how to ensure that we obtained coherent probability judgements. In this paper, we describe the rationale for our decisions. One important decision was to elicit beliefs for only observable quantities. Given that it is difficult to directly elicit transition probabilities, related quantities such as the proportions of patients healed were considered. These related quantities warrant further transformation for inclusion in the decision model, which may impact on the precision of the estimates. However, available research suggests that an expert is unlikely to be able to specify non-observables reliably [2], especially if they have little experience in elicitation and are not particularly mathematical.

Eliciting distributions that reflect uncertainty in experts' beliefs is a requirement of an explicit HTA process [1]. This imposes additional complexity to the elicitation exercise and restricts the format of the data capture instrument. Here we used the histogram technique, a very intuitive and flexible method that has been previously used in this context [34, 35]. This method allowed the exercise itself to be quick and relatively straightforward for the experts to complete. The assumption of conditional independence to elicit coherent probability judgements for structurally related quantities, specifically for the set of transition probabilities, for relative treatment effects and to evaluate time dependency also made the elicitation more feasible.

To provide a unique description of each parameter to use in the decision model, it was necessary to use methods to appropriately pool the distributions elicited by individual experts. We then fitted parametric densities to the pooled distributions. We acknowledge that in some cases, the quality of the fitting of pre-specified parametric distributions was not ideal. When further Bayesian analyses are required, it is common to fit a probability distribution using standard parametric families of distributions [3]. There are methods that allow fitting of non-parametric distributions [58]; however, these are very complex methods not commonly applied in the elicitation literature. The use of more flexible distributions could have compromised further analyses (cost effectiveness and value of information).

In this case study, elicited evidence was used alongside published evidence under the assumption that the experts did not consider existing evidence when formulating their judgements. This assumption may not be sustained in other cases, where the aggregation of both sources could lead to an incorrect specification of uncertainty (through double counting). Further thought and research is needed on how to deal with such a situation. This could involve trying to understand how experts use both existing evidence and their own experience in formulating judgements, to then be able to formulate strategies to either account for possible double counting in the aggregation or to minimise the experts' use of evidence from the literature when formulating judgements.

Compared with many other forms of evidence, elicitation constitutes a reasonably low cost source of evidence; however, the potential biases in elicited evidence cannot be ignored. Elicitation is, by definition, highly subjective and entirely dependent on the sample of experts chosen for the exercise. Whereas some of the issues of representativeness can be reduced or even avoided by ensuring a generalisable sample of reasonable size, we can never avoid its influence with this type of evidence. What elicited evidence does not do is replace good quality experimental evidence, for example, from RCTs. What it can do is provide preliminary estimates of the extent of uncertainty for particular model parameters or assumptions, which can help to inform the decision to acquire further evidence.

Acknowledgements

We thank Una Adderley, Elizabeth McGinnis and Nikki Stubbs for their invaluable advice in designing the elicitation exercises. We also thank the nurses and tutors who took part in the elicitation exercises.

References

1. Claxton K. Exploring uncertainty in cost-effectiveness analysis. *Pharmacoeconomics* 2008; **9**:781–798.
2. O'Hagan A. *Uncertain Judgements : Eliciting Experts' Probabilities*. John Wiley & Sons, 2006.
3. Garthwaite PH, Kadane JB, O'Hagan A. Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association* 2005; **100**:680–701.
4. O'Hagan A, Stevens JW. Bayesian methods for design and analysis of cost-effectiveness trials in the evaluation of health care technologies. *Statistical Methods in Medical Research* 2002; **11**(6):469–490.

5. Stevenson MD, Oakley JE, Chick SE, Chalkidou K. The cost-effectiveness of surgical instrument management policies to reduce the risk of vCJD transmission to humans. *Journal of the Operational Research Society* 2008; **60**(4):506–518.
6. Claxton K, Sculpher M, Drummond M. A rational framework for decision making by the National Institute For Clinical Excellence (NICE). *Lancet* 2002; **360**(9334):711–715. DOI: S0140-6736(02)09832-X [pii] 10.1016/S0140-6736(02)09832-X.
7. Philbeck TE, Jr, Whittington KT, Millsap MH, Briones RB, Wight DG, Schroeder WJ. The clinical and cost effectiveness of externally applied negative pressure wound therapy in the treatment of wounds in home healthcare Medicare patients. *Ostomy Wound Management* 1999; **45**(11):41–44, 46–50.
8. Witkowski W, Jawien A, Witkiewicz W, Zon B. Initial multi-centre observations upon the effect of a new topical negative pressure device upon patient and clinician experience and the treatment of wounds. *International Wound Journal* 2009; **6**(2):167–174. DOI: IWJ586 [pii] 10.1111/j.1742-481X.2009.00586.x.
9. Seternes A, Myhre HO, Dahl T. Early results after treatment of open abdomen after aortic surgery with mesh traction and vacuum-assisted wound closure. *European Journal of Vascular and Endovascular Surgery* 2010; **40**(1):60–64. DOI: S1078-5884(10)00134-6 [pii] 10.1016/j.ejvs.2010.02.018.
10. Dosluoglu HH, Loghmanee C, Lall P, Cherr GS, Harris LM, Dryjski ML. Management of early (<30 day) vascular groin infections using vacuum-assisted closure alone without muscle flap coverage in a consecutive patient series. *Journal of Vascular Surgery* 2010; **51**(5):1160–1166. DOI: S0741-5214(09)02341-6 [pii] 10.1016/j.jvs.2009.11.053.
11. Apelqvist J, Armstrong DG, Lavery LA, Boulton AJ. Resource utilization and economic costs of care based on a randomized trial of vacuum-assisted closure therapy in the treatment of diabetic foot wounds. *American Journal of Surgery* 2008; **195**(6):782–788. DOI: S0002-9610(08)00006-8 [pii] 10.1016/j.amjsurg.2007.06.023.
12. Gabriel A, Heinrich C, Shores J, Cho D, Baqai W, Moores D, Miles D, Gupta S. Outcomes of vacuum-assisted closure for the treatment of wounds in a paediatric population: case series of 58 patients. *Journal of Plastic, Reconstructive & Aesthetic Surgery* 2009; **62**(11):1428–1436. DOI: S1748-6815(08)00770-5 [pii] 10.1016/j.bjps.2008.06.033.
13. Kelm J, Schmitt E, Anagnostakos K. Vacuum-assisted closure in the treatment of early hip joint infections. *International Journal of Medical Sciences* 2009; **6**(5):241–246.
14. Pham C, Middleton P, Maddern G. Vacuum-assisted closure for the management of wounds: an accelerated systematic review. in *Surgical 2003*, A.S.a.E.R.o.N.I. Procedures.
15. Fisher A, Brady B. Vacuum assisted wound closure therapy. in *Issues in Emerging Health Technologies* C.C.C.f.H.T. Assessment, Editor. 2003.
16. Centre for Evidence Based Purchasing, *Evidence Review: Vacuum Assisted Closure Therapy*, N.P.a.S. Agency, Editor, 2008.
17. Ubbink DT, Westerbos SJ, Evans D, Land L, Vermeulen H. Topical negative pressure for treating chronic wounds. *Cochrane Database of Systematic Reviews* 2008; **(3)**:CD001898. DOI: 10.1002/14651858.CD001898.pub2.
18. Gregor S, Maegele M, Sauerland S, Krahn JF, Peinemann F, Lange S. Negative pressure wound therapy: a vacuum of evidence? *Archives of Surgery* 2008; **143**(2):189–196. DOI: 143/2/189 [pii] 10.1001/archsurg.2007.54.
19. Vikatmaa P, Juutilainen V, Kuukasjarvi P, Malmivaara A. Negative pressure wound therapy: a systematic review on effectiveness and safety. *European Journal of Vascular and Endovascular Surgery* 2008; **36**(4):438–448. DOI: S1078-5884(08)00333-X [pii] 10.1016/j.ejvs.2008.06.010.
20. Costa V, Brophy JMM. Vacuum-assisted wound closure therapy (V.A.C (R)), 2005. Technology Assessment Unit of the McGill University Health Centre (MUHC).
21. Ontario Health Technology Advisory Committee. Negative Pressure Wound Therapy: Update, 2006. Health Technology Policy Assessment.
22. Topical negative pressure for chronic wounds? *Drug and Therapeutics Bulletin* 2007; **45**(8):57–61.
23. Chaloner K, Rhame FS. Quantifying and documenting prior beliefs in clinical trials. *Statistics in Medicine* 2001; **20**(4):581–600.
24. Jenkinson D. The Elicitation of Probabilities—A Review of the Statistical Literature, 2005. BEEP working paper. Available at http://www.sheffield.ac.uk/beep/publications.html [consulted in 23-02-20 11].
25. Clemen RT, Winkler RL. Combining probability distributions from experts in risk analysis. *Risk Analysis* 1999; **19**(2):187–203.
26. Sabin M. *Competence in Practice Based Calculation: Issues for Nurse Education*. Napier University: Edinburgh, 2001.
27. Zuashkiani A, Banjevic D, Jardine AKS. Estimating parameters of proportional hazards model based on expert knowledge and statistical data. *Journal of the Operational Research Society* 2008; **60**(12):1621–1636.
28. Kadane JB, Wolfson LJ. Experiences in elicitation. *Journal of the Royal Statistical Society. Series D (The Statistician)* 1998; **47**(1):3–19.
29. Choy SL, O'Leary R, Mengersen K. Elicitation by design in ecology: using expert opinion to inform priors for Bayesian statistical models. *Ecology* 2009; **90**(1):265–277.
30. Gavasakar U. A comparison of two elicitation methods for a prior distribution for a binomial parameter. *Management Science* 1988; **34**(6):784–790.
31. Chaloner KM, Duncan GT. Assessment of a beta prior distribution: PM elicitation. *Journal of the Royal Statistical Society. Series D (The Statistician)* 1983; **32**(1/2):174–180.
32. van Noortwijk J, Dekker A, Cooke R, Mazzuchi T. Expert judgment in maintenance optimization. *IEEE Transactions on Reliability* 1992; **41**:427–432.
33. Roosen J, Hennessy DA. Capturing experts' uncertainty in welfare analysis: an application to organophosphate use regulation in U.S. apple production. *American Journal of Agricultural Economics* 2001; **83**(1):166–182.
34. McKenna C, McDaid C, Suekarran S, Hawkins N, Claxton K, Light K, Chester M, Cleland J, Woolacott N, Sculpher M. Enhanced external counterpulsation for the treatment of stable angina and heart failure: a systematic review and economic analysis. *Health Technology Assessment* 2009; **13**(24).

35. Speight PM, Palmer S, Moles DR, Downer MC, Smith DH, Henriksson M, Augustovski F. The cost-effectiveness of screening for oral cancer in primary care. *Health Technology Assessment* 2006; **10**(14):1–144, iii–iv. DOI: 99-46-02 [pii].
36. Welton NJ, Ades AE. Estimation of Markov chain transition probabilities and rates from fully and partially observed data: uncertainty propagation, evidence synthesis, and model calibration. *Medical Decision Making* 2005; **25**(6):633–645. DOI: 10.1177/0272989x05282637.
37. Singpurwalla ND, Song MS. Analysis of Weibull lifetime data incorporating expert opinion. In *Probability and Bayesian Statistics*, Viertl R (ed.). Plenum: New York, 1987; 431–442.
38. Singpurwalla ND. An interactive PC-based procedure for reliability assessment incorporating expert opinion and survival data. *Journal of the American Statistical Association* 1988; **83**(43–51).
39. Kaminsky MP, Krivtsov VV. An Approach to Evaluating the Joint Prior Distribution of Weibull Parameters. *3rd International Conference on Mathematical Methods in Reliability*, Santa Fe, New Mexico, 2004.
40. Jager P, Bertsche B. A new approach to gathering failure behavior information about mechanical components based on expert knowledge. In *Reliability and Maintainability, 2004 Annual Symposium - RAMS*, 2004.
41. Kadane JB. Predictive and structural methods for eliciting prior distributions. In *Bayesian Analysis in Econometrics and Statistics*, Zellner A (ed.). North-Holland: Amsterdam, 1980; 89–93.
42. McAlister FA. Commentary: relative treatment effects are consistent across the spectrum of underlying risks ... usually. *International Journal of Epidemiology* 2002; **31**(1):76–77. DOI: 10.1093/ije/31.1.76.
43. Ouchi F. A literature review on the use of expert opinion in probabilistic risk analysis. In *World Bank Research Working Paper 3201*, 2004.
44. Cooke R. *Experts in Uncertainty: Opinion and Subjective Probability in Science*. Oxford University Press: Oxford, 1991.
45. Gosling J, Oakley J, O'Hagan A. Nonparametric elicitation for heavy-tailed prior distributions. *Bayesian Analysis* 2007; **2**(693-718).
46. Leal J, Wordsworth S, Legood R, Blair E. Eliciting expert opinion for economic models: an applied example. *Value in Health* 2007; **10**(3):195–203.
47. Gosling J. On the elicitation of continuous, symmetric, unimodal distributions, 2008. Available from: http://www.jpgosling.co.uk/Pub/Tail.pdf.
48. Philips Z, Ginnelly L, Sculpher M, Claxton K, Golder S, Riemsma R, Woolacoot N, Glanville J. Review of guidelines for good practice in decision-analytic modelling in health technology assessment. *Health Technology Assessment* 2004; **8**(36).
49. Caldwell DM, Ades AE, Higgins JPT. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ* 2005; **331**(7521):897–900. DOI: 10.1136/bmj.331.7521.897.
50. Nixon RM, Bansback N, Brennan A. Using mixed treatment comparisons and meta-regression to perform indirect comparisons to estimate the efficacy of biologic treatments in rheumatoid arthritis. *Statistics in Medicine* 2007; **26**(6):1237–1254.
51. White IR, Pocock SJ, Wang D. Eliciting and using expert opinions about influence of patient characteristics on treatment effects: a Bayesian analysis of the CHARM trials. *Statistics in Medicine* 2005; **24**(24):3805–3821.
52. Winkler RL, Poses RM. Evaluating and combining physicians' probabilities of survival in an intensive care unit. *Management Science* 1993; **39**(12):1526–1543. DOI: 10.1287/mnsc.39.12.1526.
53. Stevens JW, O'Hagan A, Miller P. Case study in the Bayesian analysis of a cost-effectiveness trial in the evaluation of health care technologies: depression. *Pharmaceutical Statistics* 2003; **2**(1):51–68. DOI: 10.1002/Pst.043.
54. Stevens JW, O'Hagan A. Incorporation of genuine prior information in cost-effectiveness analysis of clinical trial data. *International Journal of Technology Assessment in Health Care* 2002; **18**(4):782–790.
55. Stevenson MD, Oakley JE, Lloyd Jones M, Brennan A, Compston JE, McCloskey EV, Selby PL. The cost-effectiveness of an RCT to establish whether 5 or 10 years of bisphosphonate treatment is the better duration for women with a prior fracture. *Medical Decision Making* 2009; **29**(6):678–689. DOI: 0272989X09336077 [pii] 10.1177/0272989X09336077.
56. Bojke L, Claxton K, Bravo-Vergel Y, Sculpher M, Palmer S, Abrams K. Eliciting distributions to populate decision analytic models. *Value in Health* 2010; **13**(5):557–564. DOI: 10.1111/j.1524-4733.2010.00709.x.
57. National Institute for Health and Clinical Excellence. *Guide to the methods of technology appraisal*. NICE: London, 2008.
58. Oakley JE, O'Hagan A. Uncertainty in prior elicitations: a nonparametric approach. *Biometrika* 2007; **94**(2):427–441. DOI: 10.1093/biomet/asm031.

Paper 6.

Soares MO; Dumville J; Ashby R; Iglesias C; Bojke L; Adderley U; McGinnis E; Stubbs N; Torgerson D; Claxton K; Cullum N. Methods to assess cost effectiveness and value of further research when data are sparse: negative pressure wound therapy for severe pressure ulcers. Medical Decision Making. 33(3), 2013, 415-36.

Methods to Assess Cost-Effectiveness and Value of Further Research When Data Are Sparse: Negative-Pressure Wound Therapy for Severe Pressure Ulcers

Marta O. Soares, MSc, Jo C. Dumville, PhD, Rebecca L. Ashby, PhD, Cynthia P. Iglesias, PhD, Laura Bojke, PhD, Una Adderley, MSc, Elizabeth McGinnis, PhD, Nikki Stubbs, MSc, David J. Torgerson, PhD, Karl Claxton, PhD, Nicky Cullum, PhD

Health care resources are scarce, and decisions have to be made about how to allocate funds. Often, these decisions are based on sparse or imperfect evidence. One such example is negative-pressure wound therapy (NPWT), which is a widely used treatment for severe pressure ulcers; however, there is currently no robust evidence that it is effective or cost-effective. This work considers the decision to adopt NPWT given a range of alternative treatments, using a decision analytic modeling approach. Literature searches were conducted to identify existing evidence on model parameters. Given the limited evidence base, a second source of evidence, beliefs elicited from experts, was used. Judgments from experts on relevant (uncertain) quantities were obtained through a formal elicitation exercise. Additionally, data derived from a pilot trial were also used to inform the model. The 3 sources of evidence were collated, and the impact of each on cost-effectiveness was evaluated. An analysis of the value of further information indicated that a randomized controlled trial may be worthwhile in reducing decision uncertainty, where from a set of alternative designs, a 3-arm trial with longer follow-up was estimated to be the most efficient. The analyses presented demonstrate how allocation decisions about medical technologies can be explicitly informed when data are sparse and how this kind of analyses can be used to guide future research prioritization, not only indicating whether further research is worthwhile but what type of research is needed and how it should be designed. ***Key words:*** *Markov model; elicited evidence; pilot trial; negative pressure wound therapy; sparse; evidence synthesis; expected value of information; research design; cost-effectiveness analysis.* ***(Med Decis Making 2013;33:415–436)***

Received 10 February 2011 from the Centre for Health Economics, The University of York, York, UK (MOS, LB, KC); Department of Health Sciences, The University of York, York, UK (JCD, RLA, CI, DT); School of Health and Social Care, Teesside University, Middlesbrough, UK (UA); Leeds Teaching Hospitals National Health Service (NHS) Trust, Leeds General Infirmary, Leeds, UK (EM); NHS Leeds Community Healthcare, St Mary's Hospital, Leeds, UK (NS); and School of Nursing, Midwifery and Social Work, University of Manchester, Manchester, UK (NC). This work was supported by the Medical Research Council (grant number G0501814). Kinetic Concepts Inc. supplied the NPWT machines used in the pilot trial at no cost. The manufacturer had no role in the design of the pilot trial or in the collection, analysis, and interpretation of the data. Revision accepted for publication 24 March 2012.

Address correspondence to Marta O. Soares, The University of York, Centre for Health Economics, Alcuin A Block, Heslington, York, YO10 5DD, UK; e-mail: marta.soares@york.ac.uk.

DOI: 10.1177/0272989X12451058

Health professionals and decision-making bodies should use existing research evidence (regarding clinical and cost-effectiveness) to inform their decisions regarding the adoption of health care interventions.[1] However, there are many situations where only sparse data are available to inform these decisions. This common problem places decision makers in a difficult position. Limited data may delay a formal evaluation, leaving people uninformed, with ad hoc use of a technology that may offer less value than alternatives.

In this article, we demonstrate how analyses to inform adoption and research decisions can be conducted when the evidence base is sparse. This process requires characterization of all sources of uncertainty and the assessment of decision uncertainty and its

consequences; therefore, all available evidence must be utilized. In the absence of robust research data but the presence of substantial practical experience of using the treatment of interest and its comparators, evidence may include expert opinion. Expert knowledge can be captured quantitatively and hence used to inform a cost-effectiveness model by applying formal methods to elicit experts' beliefs.[2]

MOTIVATING EXAMPLE

Negative-pressure wound therapy (NPWT) is a medical device used to treat full-thickness wounds, such as severe (grades 3 and 4) pressure ulcers. It has been claimed that NPWT speeds healing and reduces infection rates and costs as well as facilitating wound management.[3,4] However, there is very little research evidence for the clinical or cost-effectiveness of NPWT,[3,5] even though it is a relatively expensive treatment, used widely in the developed world. Therefore, there is a need to evaluate the cost-effectiveness of NPWT and alternative treatments for its various indications, including severe pressure ulcers.

In this case study, the population of interest was UK patients with severe pressure ulcers. The alternatives to NPWT for this patient population are dressings and topical treatments, with several different treatment options and limited evidence to guide treatment choice. Thus, we conducted an electronic survey of nurses involved in wound care (in the UK, most wound care is nurse led). Respondents (n = 28) were asked to identify possible alternatives to NPWT and to rank these in order of preference. Subsequently, we established a panel of senior expert nurses (n = 7) and used the nominal group technique[6] to establish the level of consensus regarding the alternative treatment options identified in the survey. This panel, in common with the survey results, identified spun hydrocolloid (hydrofiber), alginate, and foam dressings as relevant comparators. Alongside the cost-effectiveness study, we also conducted a pilot randomized controlled trial (RCT) to assess the feasibility of conducting a further study on NPWT. Dressing use in the standard care arm of this pilot trial was limited to spun hydrocolloid (hydrofiber), alginate, and foam.

ANALYTIC FRAMEWORK

Cost-Effectiveness and Decision Uncertainty

Decisions based on expected cost-effectiveness should not only consider the expected health gain associated with a specific technology but also the potential health gains forgone elsewhere due to additional costs associated with the technology displacing other National Health Service (NHS) activities. This can be expressed using the net health benefit (NHB) statistic for a defined willingness to pay value.[7,8] The NHB of a treatment t informed by the set of input parameters, $\boldsymbol{\theta}$, can be defined as $\mathrm{NHB}_t^{\boldsymbol{\theta}} = \mathrm{effects}_t^{\boldsymbol{\theta}} - \mathrm{costs}_t^{\boldsymbol{\theta}}/\lambda$, where λ is the cost-effectiveness threshold. Unless otherwise stated, a cost-effectiveness threshold of £20,000 per quality-adjusted life year (QALY) was assumed.[7]

The process of explicitly structuring and combining all available evidence in these evaluations often relies on the use of decision analytic models (DAMs). A DAM aims to represent disease progression and often model transitions between key health states; the costs and consequences associated with these states (in relation to the relevant health technologies) can then be assessed. If we knew the true values of costs and effects, the optimal decision D would be made by choosing the treatment with the highest NHB. Yet, evidence used to inform the DAM is often uncertain; an adopt or reject decision rooted in available evidence should be based on expected values: in this case, the optimal treatment is that which offers the highest expected NHB; that is, $\mathrm{D} : \max_{t \in T} \left(\mathrm{E}_{\boldsymbol{\theta}} \left[\mathrm{NHB}_t^{\boldsymbol{\theta}} \right] \right)$, where t represents the optimal treatment out of a set of treatments T.

To evaluate the cost-effectiveness of NPWT for the treatment of severe pressure ulcers, a DAM was designed and literature searches conducted to identify evidence to inform model parameters. Evidence from the literature was found to be few and sparse, failing to adequately characterize the events that clinical experts identified as relevant for inclusion in the DAM (see the next section). Limited evidence meant that we would either be unable to proceed with the evaluation or that further analytic assumptions would be required. We were aware, however, that the treatments of interest were being used in clinical practice, and thus, there was some expertise regarding the characteristics and outcomes of relevant patients. To collect this evidence, we used formal methods for the elicitation of expert opinion. A full report of this process is reported separately[8]; however, a brief summary is provided in a later section. The evidence generated by the pilot trial was also used as an additional evidence source for the DAM.

Because 3 sources of evidence (literature, elicitation, and pilot trial data) were potentially available to inform a single model parameter, alternative sources were

combined so that the impact of each source of evidence on cost-effectiveness could be evaluated.

The perspective of the analysis was that of the UK NHS, with costs presented in pounds sterling (£) at 2008-2009 prices. Costs and health benefits were discounted at 3.5%.[7] We implemented an incremental analysis. The DAM was probabilistic, where uncertainty was propagated using a Monte Carlo simulation.[9] Decision uncertainty was presented as the probability that each treatment has the highest expected NHB (i.e., is expected to be cost-effective).

Value of Further Research

Given the uncertainty around treatment decisions, it is important to explore whether investing in further research to reduce such decision uncertainty is worthwhile and, if so, what type of future research is most likely to offer the most value for the money.

To explore these issues, value of further research analyses were implemented.[10–13] The maximum potential value of additional evidence to inform this decision problem can be represented by the expected value of perfect information (EVPI). The EVPI is defined as the difference between the NHB, which might be achieved when the uncertainties are resolved (and different treatment choices made), and the expected NHB based on current evidence; that is, $\text{EVPI} = \underset{\boldsymbol{\theta}}{\text{E}}\left[\max_t(\text{NHB}_t^{\boldsymbol{\theta}})\right] - \max_t\left(\underset{\boldsymbol{\theta}}{\text{E}}\left[\text{NHB}_t^{\boldsymbol{\theta}}\right]\right)$.

The value of resolving the uncertainty associated with one, or a subset of parameters, (parameter EVPI) can also be estimated by $\text{EVPI}(\upsilon) = \underset{\upsilon}{\text{E}}\left[\max_t\left(\underset{\upsilon^c|\upsilon}{\text{E}}\left[\text{NHB}_t^{\boldsymbol{\theta}}\right]\right)\right] - \max_t\left(\underset{\boldsymbol{\theta}}{\text{E}}\left[\text{NHB}_t^{\boldsymbol{\theta}}\right]\right)$, where υ is the subset of parameters, from $\boldsymbol{\theta}$ being evaluated and υ^c the remaining parameters on $\boldsymbol{\theta}$. Investigating the parameter EVPI can highlight the type of evidence that would be most valuable, indicating the research designs that might be required.

Estimates of the EVPI place an upper bound on the value of research, assuming that all uncertainty could be resolved. However, research will generate data from a sample of the population of interest that can only be expected to reduce, rather than eliminate, uncertainty. By predicting the range of possible sample results for a particular research design and sample size, and calculating the NHB of subsequent treatment choices for each possible sample result, the expected value of sample information (EVSI) can also be established.[12] The EVSI represents the expected benefits of particular research designs.[10] The EVSI can be estimated by $\text{EVSI}(\upsilon) = \underset{data}{\text{E}}\left[\max_t\left(\underset{\upsilon^c,(\upsilon|data)}{\text{E}}\left[\text{NHB}_t^{\boldsymbol{\theta}}\right]\right)\right] - \max_t\left(\underset{\boldsymbol{\theta}}{\text{E}}\left[\text{NHB}_t^{\boldsymbol{\theta}}\right]\right)$. The first term in this equation represents the NHB associated with the range of possible, unknown results of research (data). How data were generated is detailed ahead in this article.

To choose between alternative designs, the costs as well as the benefits of research need to be considered. The difference between the EVSI and the estimated costs of that chosen particular research design/study is the expected net benefit of sampling (ENBS). Further research is expected to be worthwhile if the ENBS is positive, indicating that more efficient research designs (including choice of sample size) will generate a higher ENBS.

Calculations of the value of further research can be computationally burdensome, especially when nested expectations need to be evaluated. Specificities of the implementation of these analyses (simplifying assumptions and estimation strategy used) are detailed in the relevant sections below.

STRUCTURING THE DECISION PROBLEM

A DAM was developed with clinical input to ensure that it adequately represented the clinical trajectory of patients with severe pressure ulceration, including all events relevant to cost-effectiveness. The model was based on 3 transition states: unhealed, healed, and dead (Figure 1).

Surgery to close the wound ("closure surgery"), wound-related complications (osteomyelitis or systemic infection), and treatment changes were included as relevant events associated with the transition states. Patients remaining in state 1, unhealed, were assumed to incur costs related to the treatment of their ulcer and the possible presence of wound-related complications. Longer stays in this state would result in higher incurred costs. Transition to state 2, healed, could be a consequence of either healing by secondary intention or closure surgery, with the occurrence, outcomes, and costs of each form of closure being considered separately. Closure surgery was not considered as an alternative to NPWT in treating severe pressure ulcers because the clinicians advising on the development of the DAM noted that a decision to undertake surgery would occur downstream of treatment with NPWT and/or dressings. Transitions to death could occur for healed and unhealed patients. Details of all

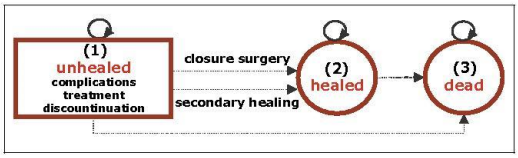


Figure 1 Decision analytic model for the treatment of severe (grades 3 and 4) pressure ulcers. In secondary healing, the wound closes spontaneously by contraction and re-epithelialization.

model transitions, relevant events, and associated costs and outcomes are in Tables 1, 2, and 3. A time horizon of 2 years and a cycle length of 4 weeks were used. Outcomes were measured as health care costs per QALY gained.

SOURCES OF EVIDENCE

To inform the DAM, 3 sources of evidence were considered: existing data (from literature searches), elicited beliefs, and data derived from a pilot trial. This section describes the available sources of evidence and the methods used to collect and/or synthesize the evidence.

Existing Evidence

Existing evidence was identified using literature reviews. The reviews focused on specific questions, reflecting the data required.[14] Existing data were sparse and only identified for the following parameters.

Relative effectiveness. Systematic searches were conducted to identify all studies that contained data on the relative effectiveness of NPWT and dressings to treat severe (grade 3 and/or grade 4) pressure ulcers. The outcome of interest was an objective measure of wound healing. One RCT evaluating NPWT[15] and 11 evaluating dressings[16–26] were found. Overall, these studies were small (mean = 42 patients per study). Data from these trials were linked within an evidence network and synthesized using Bayesian indirect and mixed treatment comparisons (IMTCs).[27] The analysis assumed a binomial likelihood. Relative, rather than absolute, measures of effectiveness were aggregated due to likely heterogeneity between trial populations. The binomial parameter of probability was described as a function of the absolute healing hazard, assuming a constant healing rate over time. Because most links in the network were informed by a single study and the number of healing events in some trials was small or zero, inferences on foam dressings were not obtained. Therefore, a common but random effect of any dressing treatment was initially assumed.[28] A more detailed description of the synthesis of relative effectiveness evidence is reported separately (see Appendix online). Results are presented in Table 1 (column 1) as log hazard ratios with hydrocolloid as a reference treatment.

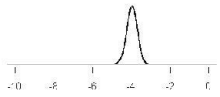
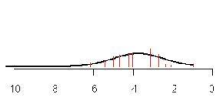
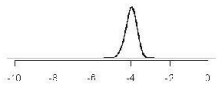
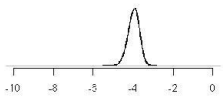
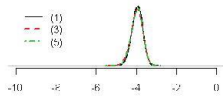
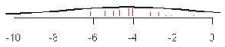
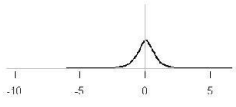
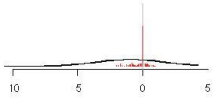
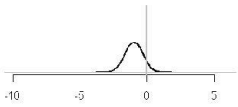
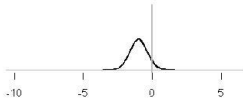
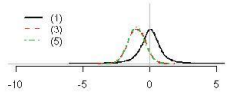
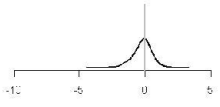
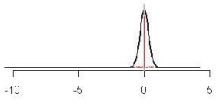
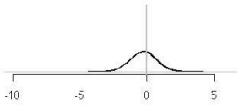
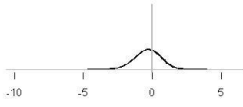
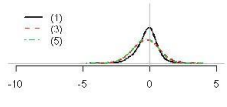
Costs and utilities. A broad literature search on health-related quality of life data yielded one publication[29] reporting Short Form-36 data from 218 people with and 2289 people without pressure ulceration (any grade) in the UK. Patient-level data from this study were used to calculate patient utility values (via the SF-6D[30]). To estimate mean utility in patients with and without pressure ulceration, a Bayesian linear regression was applied (adjusting for sex, age, and presence of comorbidities[29]). Results are presented in Table 3 (column 1). The literature search of cost data yielded one relevant study.[31] The study did not collect primary data regarding the cost of treating severe pressure ulcers but rather developed treatment protocols that reflected good practice. Data from this study were used to derive weekly costs incurred by patients with an unhealed, severe pressure ulcer with and without complications (Table 3, column 1).

Despite the literature reviews being conducted by defining broad search criteria, we did not find evidence on the impact of complications, closure surgery, and recurrence in UK patients with pressure ulcers or on the impact of severe pressure ulceration on death. Standard UK life table data were used to describe mortality.[32,33]

Elicited Evidence

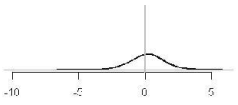
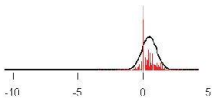
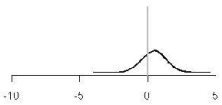
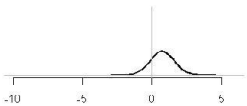
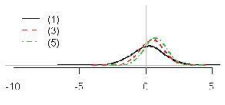
The existing evidence base on the treatments assessed was sparse; however, NPWT and comparators are used extensively within the NHS. Excluding such experience could misrepresent the current level of knowledge regarding these treatments. Thus, a formal exercise was designed to systematically capture experts' knowledge and uncertainty around the treatment and progression of severe pressure ulcers; evidence was collected in the form of probabilistic judgments around various quantities.[34] A detailed report of the methods used in this elicitation exercise is published elsewhere.[8]

Table 1 Parameters Regarding Transitions: Density Functions and Summary Descriptions of the Evidence Derived from Distinct Sources

Parameter	Existing Evidence (1) Distribution, Mean, and Credible Interval	Elicited Evidence (2) Distribution, Mean, and Credible Interval	Existing + Elicited Evidence (3) Distribution, Mean, and Credible Interval	Pilot Trial[a] (4) *n* of *N*[b]	Existing + Elicited + Pilot Trial Evidence (5) Distribution, Mean, and Credible Interval	(1), (3), and (5)
Log hazard of healing when treated with HC (up to 6 months)	 MCMC, −3.95 (−4.50 to −3.46)	 Normal, −3.74 (−5.96 to −1.52)	 MCMC, −3.97 (−4.59 to −3.46)	0 of 4	 MCMC, −3.96 (−4.55 to −3.47)	 (1) (3) (5)
Log hazard of healing after 6 months, assumed independent of treatment[c]	NA	 Normal, −4.28 (−10.24 to 1.67)	Same as (2)	NA	Same as (2)	Same as (2)
Log hazard ratio of healing for F relative to HC[d]	 MCMC, 0.03 (−1.97 to 1.86)	 Normal, −0.96 (−6.32 to 4.40)	 MCMC, −0.91 (−2.14 to 0.21)	0 of 1	 MCMC, −0.98 (−2.12 to 0.18)	 (1) (3) (5)
Log hazard ratio of healing for ALG relative to HC	 MCMC, −0.19 (−1.76 to 1.13)	 Normal, 0.003 (−0.63 to 0.64)	 MCMC, −0.27 (−2.12 to 1.57)	0 of 1	 MCMC, −0.34 (−2.21 to 1.47)	 (1) (3) (5)

(*continued*)

Table 1 (continued)

Parameter	Existing Evidence (1) Distribution, Mean, and Credible Interval	Elicited Evidence (2) Distribution, Mean, and Credible Interval	Existing + Elicited Evidence (3) Distribution, Mean, and Credible Interval	Pilot Trial[a] (4) *n* of *N*[b]	Existing + Elicited + Pilot Trial Evidence (5) Distribution, Mean, and Credible Interval	(1), (3), and (5)
Log hazard ratio of healing for NPWT relative to HC	 MCMC, 0.18 (–2.17 to 2.63)	 Normal, 0.45 (–0.66 to 1.56)	 MCMC, 0.47 (–1.18 to 2.10)	1 of 6	 MCMC, 0.75 (–0.88 to 2.34)	

Note: Columns (1), (2), and (4) relate to information derived from existing, elicited, and pilot trial evidence, respectively, while densities in (3) and (5) relate to the collation of these individual sources of evidence using Bayesian updating. NPWT = negative-pressure wound therapy; HC = spun hydrocolloid; ALG = alginate; F = foam; OM = osteomyelitis; SI = systemic infection; MCMC = Markov chain Monte Carlo; NA = not available.

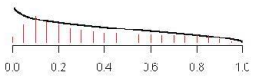
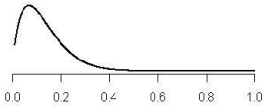
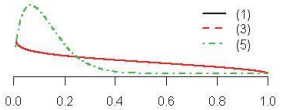
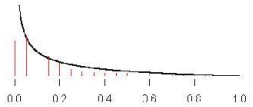
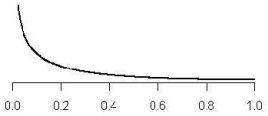
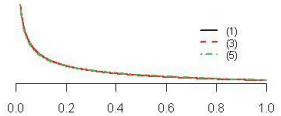
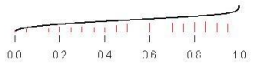
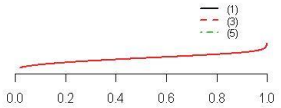
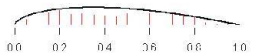
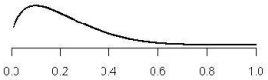
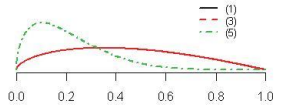
a. Within the pilot trial, participants in the standard care arm could be treated with spun hydrocolloid, alginate, and foam; for the purpose of this analysis, these were assumed to be allocated at random.

b. Number of patients for whom the event of interest (healing) occurred, *n*, out of the total patients of interest (patients randomized to the treatment of interest), *N*.

c. This parameter was elicited using the following question: "Think only of those patients who still have a grade 3 or 4 pressure ulcer 6 months after starting treatment with HC. What proportion of them do you think would heal their reference ulcer between 6 and 12 months?"[2] The upper bound of 12 months was used to ease the task of experts. In the model, however, the log hazard derived was applied to the end of the time horizon.

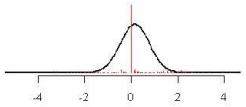
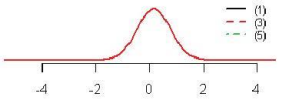
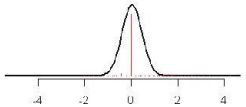
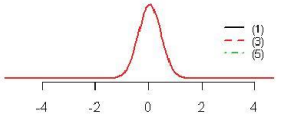
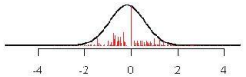
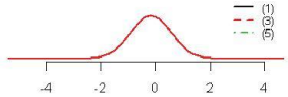
d. The existing evidence of foam, F, consisted of a trial where zero events were reported in both arms. These data are not informative, and a relative treatment effect was not estimable. To allow obtaining an estimate of the treatment effect of foam under this scenario, a common (but random) effect of dressing treatments was assumed.

Table 2 Parameters Regarding Relevant Events: Density Functions and Summary Descriptions of the Evidence Derived from Distinct Sources

	Existing Evidence (1)	Elicited Evidence (2)	Existing + Elicited Evidence (3)	Pilot Trial (4)	Existing + Elicited + Pilot Trial Evidence (5)	
Parameter	**Distribution, Mean, and Credible Interval**	**Distribution and Specified Parameter Values**		***n* of *N*[a]**	**Distribution, Mean, and Credible Interval**	**(1), (3), and (5)**
Maintained treatment, % of patients being treated with the same treatment 6 months after starting treatment	NA	 Beta(0.85, 1.58)	Same as (2)	1 of 12	 Beta(1.85, 12.58)	
Probability of having had surgery conditional on having healed (proportion of healed patients who achieved healing through surgery, $P[S \mid H]$)	NA	 Beta(0.39, 2.16)	Same as (2)	0 of 1	 Beta(0.39, 3.16)	
Probability of healing conditional on having undertaken closure surgery (proportion of patients healed after surgery, $P[H \mid S]$)	NA	 Beta(1.42, 0.92)	Same as (2)	NA	Same as (2)	
Unhealed patients presenting with complications after starting treatment with HC	NA	 Beta(1.57, 2.06)	Same as (2)	0 of 4	 Beta(1.57, 6.06)	

(*continued*)

Table 2 (continued)

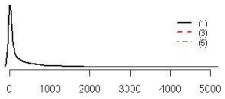
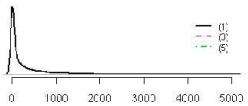
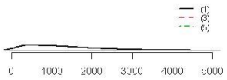
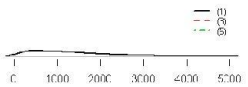
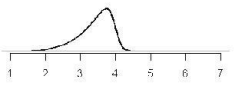
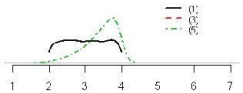
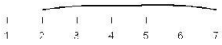
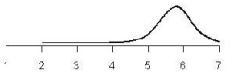
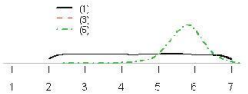
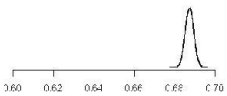
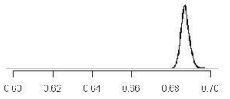
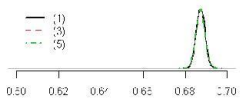
Parameter	Existing Evidence (1)	Elicited Evidence (2)	Existing + Elicited Evidence (3)	Pilot Trial (4)	Existing + Elicited + Pilot Trial Evidence (5)	
	Distribution, Mean, and Credible Interval	Distribution and Specified Parameter Values		*n* of *N*[a]	Distribution, Mean, and Credible Interval	(1), (3), and (5)
Hazard ratio of complication occurrence in unhealed patients treated with F in relation to HC[b]	NA	 Normal(0.16, 0.62)	Same as (2)	0 of 1	Same as (2)	
Hazard ratio of complication occurrence in unhealed patients treated with ALG relative to HC[b]	NA	 Normal(0.04, 0.43)	Same as (2)	0 of 1	Same as (2)	
Log hazard ratio of complication occurrence in unhealed patients treated with NPWT relative to HC[b]	NA	 Normal(−0.16, 0.74)	Same as (2)	0 of 6	Same as (2)	

Note: Columns (1), (2), and (4) relate to information derived from existing, elicited, and pilot trial evidence, respectively, while densities in (3) and (5) relate to the collation of these individual sources of evidence using Bayesian updating. NPWT = negative-pressure wound therapy; HC = spun hydrocolloid; ALG = alginate; F = foam; OM = osteomyelitis; SI = systemic infection; MCMC = Markov chain Monte Carlo; NA = not available.

a. Number of patients for whom the event (e.g., complications) occurred, *n*, out of the total patients of interest, *N*.

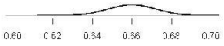
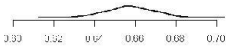
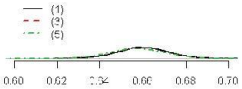
b. Pilot trial data were not included due to tractability issues.

Table 3 Evidence from Distinct Sources Used to Describe Stochastic (Uncertain) Parameters of the Decision Model Regarding Costs and Utilities

	Existing Evidence (1)	Elicited Evidence (2)	Existing + Elicited Evidence (3)	Pilot Trial (4)	Existing + Elicited + Pilot Trial Evidence (5)	
Parameter	**Distribution and Specified Parameter Values, or, for MCMC, Mean and Credible Interval**	**Distribution and Specified Parameter Values, or, for MCMC, Mean and Credible Interval**	**Distribution and Specified Parameter Values, or, for MCMC, Mean and Credible Interval**	**Mean, SE, and Distribution**	**Distribution and Specified Parameter Values, or, for MCMC, Mean and Credible Interval**	**(1), (3), and (5)**
Costs per week (£) incurred by unhealed patients, no complications	 Gamma(0.275–1000)	NA	Same as (1)	NA	Same as (1)	
Additional weekly costs (£) incurred by unhealed patients with complications, in relation to patients without	 Gamma(1.212–1000)	NA	Same as (1)	NA	Same as (1)	
Number of dressing changes per week for patients on NPWT[a]	 Normal(3–100), I(2,4)	NA	Same as (1)	Mean = 4.08; SE = 0.84; normal	 MCMC, 3.44 (2.37–3.98)	
Number of changes of dressings (HC, ALG, F) per week[a]	 Normal(4–100), I(2,7)	NA	Same as (1)	Mean = 6.15; SE = 0.47; normal	 MCMC, 5.74 (4.70–6.67)	
Utility associated with being pressure ulcer free, healed	 MCMC: 0.686 (0.682–0.690)	NA	Same as (1)	NA	 MCMC, 0.687 (0.684–0.692)	

(*continued*)

Table 3 (continued)

Parameter	Existing Evidence (1)	Elicited Evidence (2)	Existing + Elicited Evidence (3)	Pilot Trial (4)	Existing + Elicited + Pilot Trial Evidence (5)	
	Distribution and Specified Parameter Values, or, for MCMC, Mean and Credible Interval	Distribution and Specified Parameter Values, or, for MCMC, Mean and Credible Interval	Distribution and Specified Parameter Values, or, for MCMC, Mean and Credible Interval	Mean, SE, and Distribution	Distribution and Specified Parameter Values, or, for MCMC, Mean and Credible Interval	(1), (3), and (5)
Utility associated with having a pressure ulcer, unhealed	 MCMC: 0.660 (0.645–0.675)	NA	Same as (1)	Mean = 0.110; SE = 0.097; normal	 MCMC, 0.657 (0.634–0.680)	

Note: Columns (1), (2), and (4) relate to information derived from existing, elicited, and pilot trial evidence, respectively, while densities in (3) and (5) relate to the collation of these individual sources of evidence using Bayesian updating. SE = standard error; NPWT = negative-pressure wound therapy; HC = spun hydrocolloid; ALG = alginate; F = foam; OM = osteomyelitis; SI = systemic infection; MCMC = Markov chain Monte Carlo; NA = not available.

a. While dressing changes for comparator treatments were assumed to take place every 1 to 3 days (standard practice, informed by clinical opinion), for NPWT, dressings were assumed to be changed every 2 to 3 days following manufacturers' recommendations.[50]

As literature searches identified only limited existing data with which to characterize model transitions regarding healing rates and other relevant events, these quantities along with their uncertainty were elicited. While the elicitation methods would also have been appropriate for the collection of utilities, resource use, and costs data, the collection of these in addition to transition and related events was anticipated to be too arduous for experts given the available time and was not undertaken.

Because experts are unlikely to be able to specify nonobservable quantities reliably[2] (e.g., hazard of healing), we only elicited observable quantities[35,36] (e.g., proportion of patients healed): the exercise elicited all responses as proportions. It was thus anticipated that further parameterization (i.e., transformation) would be required for inclusion in the DAM. The data capture method used was the histogram method, which has been previously used in health technology assessment.[37,38]

A sample of nurses with experience and knowledge of tissue viability and wound management were invited to participate in the exercise; they were also asked to invite colleagues with experience in treating pressure ulcers. In total, 23 nurses participated. The exercise was undertaken using a specially designed Microsoft Office Excel 2007 program (Redmond, WA). It was led by an experienced facilitator and included an extensive training session. A mathematical, rather than a consensus, approach was adopted,[2] requiring separate elicitation from each expert, with results from experts combined (linear pooling was used) after elicitation. The choice of the mathematical approach was based on a prior pilot exercise where beliefs were elicited using both mathematical and consensus methods. In this pilot, the consensus method (where experts were encouraged to interact to achieve a level of agreement for a particular parameter[39]) produced more accurate responses but also generated incoherent probability statements between 2 related quantities. These results are in line with the general view that consensus promotes the use of heuristics and may produce overconfident judgments.[40] To the pooled empiric distribution obtained after eliciting, parametric distributions were fitted using the method of moments. Results of the elicitation exercise for transition parameters and other relevant events (pooled discretized distribution, fitted distribution, and summary measures) are shown in Tables 1 and 2 (column 2), respectively.

Pilot Trial–Derived Evidence

The third source of evidence was a pilot RCT. This pilot was conducted to generate further data and evaluate the feasibility of conducting large-scale primary research. It took place in one UK community health care trust and in one hospital. Patients were screened for eligibility during a 12-month recruitment period starting in September 2008. Consenting participants with grade 3 or 4 pressure ulcers were randomized to receive either NPWT (Vacuum Assisted Closure, Kinetic Concepts Inc., San Antonio, TX) or standard care (spun hydrocolloid, alginate, or foam dressings). Patients were followed up for 6 months. Twelve patients were recruited (6 per arm). The design of the trial was also tailored to collect data that would contribute to the DAM: data were collected on recruitment rates; occurrence of important events relevant to the model including complications, healing, and death; treatment changes and number of dressing changes; and EQ-5D. Data are presented descriptively in Tables 1, 2, and 3 (column 4).

INCORPORATING ALL AVAILABLE EVIDENCE

This section details how the 3 alternative sources of evidence were combined, followed by a description of the evidence and how this was used in the decision model (parameterization from the evidence). Evidence is presented for 3 sets of model parameters: parameters regarding transitions (speed at which healing or death occurs), other relevant events (complications and closure surgery), and costs and utilities.

Collating the Sources of Evidence

The impact of each source of evidence on cost-effectiveness was evaluated by considering 3 scenarios: 1) existing evidence used alone, 2) existing evidence combined with elicited evidence, and 3) the most comprehensive evidence base that combined data from all 3 sources. The collation of evidence sources, necessary to generate the second and third scenarios, required Bayesian updating.[41] This was a 2-step procedure (illustrated in Figure 2 using hypothetical distributions):

Step 1: Elicited evidence was used as prior distribution (EL) and combined with data from existing evidence (EX) to generate the posterior density EX + EL; it is

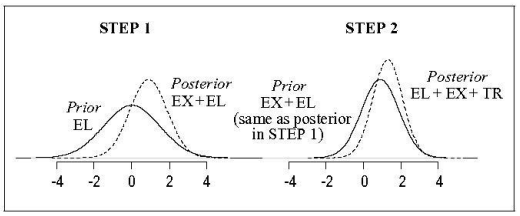


Figure 2 Illustration of the 2-step procedure based on Bayesian updating used to generate alternative evidence scenarios.

this posterior density that is further used in scenario 2 above.

Step 2: The posterior density from step 1 (EX + EL) was here considered prior distribution and was combined with trial-derived data (TR) to generate the posterior density EX + EL + TR. This posterior density collates all 3 sources of evidence and was used in scenario 3 above.

For existing and elicited evidence, the availability of data from these 2 sources conditioned how the data were collated. For all parameters except relative treatment effects, where only one evidence source was available, this alone was used to describe the parameter of interest. In estimating relative treatment effects, if no elicited evidence existed, Bayesian updating was used with vague or uninformative distributions as prior distribution. If both sources were available, expert opinion was considered subjective prior distribution and was combined with existing data to obtain inferences. When elicited prior data were included in the IMTC model, treatment effects could be estimated for each dressing individually.

The scenario collating all 3 sources of evidence (existing, elicited, and pilot trial derived) was also generated using Bayesian updating: posterior distributions derived from the collation of existing and elicited evidence were considered as priors distribution and updated with data from the pilot trial, where available.

Evidence on Transitions

All 3 sources of evidence provided data on healing. Table 1 summarizes the available evidence using probability density functions for existing and elicited data (columns 1 and 2, respectively) and the data from the pilot trial (column 3). Existing and elicited evidence on relative treatment effects showed wide credible intervals (the Bayesian equivalent of confidence intervals), reflecting current uncertainties. The only marked difference between data sources was for foam dressings, which elicited expert opinion judged less effective than existing evidence indicated. The synthesis of existing RCT evidence required an assumed common but random effect of dressings. However, the elicited data allowed treatment effects to be estimated for each dressing individually, reflecting experts' opinions. Overall, the inclusion of data from the pilot trial was not influential due to its small sample size and the fact that only one participant healed. Death data were obtained from UK life tables and updated with pilot trial, but not elicited, data (results not shown).

Parameterizing the evidence. Consider that transition rates for patients treated with spun hydrocolloid dressing are denoted by γ_{ij}, where the indexes i and j indicate, respectively, the departure and arrival state (1 = unhealed; 2 = healed; 3 = dead) (Figure 1).

Inferences over transition parameters were obtained by assuming a continuous time evolvement, and hazard rates were estimated from the data (Table 1). Despite parameters being estimated in continuous time, a discretized version of the Markov model was used to establish cost-effectiveness. Transition probabilities for a prespecified cycle length, $\pi_{ij}(t)$, needed to be evaluated from the corresponding transition rates. For the 3 state models, the solution to the Kolmogorov equations[42] regarding the unknown probability transition parameters is

$$\begin{cases} \pi_{11}(t) = e^{-(\gamma_{12}+\gamma_{13})\cdot t} \\ \pi_{12}(t) = \gamma_{12} e^{-\gamma_{23}\cdot t} \cdot \left(1 - e^{-(\gamma_{12}+\gamma_{13}-\gamma_{23})\cdot t}\right) / (\gamma_{12}+\gamma_{13}-\gamma_{23}) \cdot \\ \pi_{22}(t) = e^{-\gamma_{23}\cdot t} \end{cases}$$

The rates of death for healed and unhealed patients were assumed equal, $\gamma_{13}=\gamma_{23}$, and therefore, the probability of healing in a certain period of time was simplified to $\pi_{12}(t)= \exp(-\gamma_{13}\cdot t)(1-\exp(-\gamma_{12}\cdot t))$. Available trial evidence was used to estimate the rates; however, most trials disregarded individuals who died when reporting healing outcomes. Consequently, these trials reported the probability that a patient heals is conditional on not having died. This quantity equates to $\pi_{12}(t)=1-\exp(-\gamma_{12}\cdot t)$ when death rates are assumed to be equal for healed and unhealed patients. This functional relation was used in the synthesis of available evidence to aggregate the outcomes of the distinct trials (see Appendix online).

It was assumed that the treatments being evaluated would impact healing; their effectiveness was parameterized as a treatment effect relative to spun hydrocolloid dressings (hazard ratios). After 6 months unhealed, patients were assumed to have a different rate of healing, informed by expert opinion only (Table 1). Given there is no evidence regarding the effectiveness of treatments after 6 months, treatments were assumed not to impact the rate of healing.

Evidence on Relevant Events

Relevant events were the occurrence and resolution of complications related to severe pressure ulceration (osteomyelitis or systemic infection), occurrence and success of closure surgery in the UK, and discontinuation of treatments. As no UK data were identified from the literature regarding these events, they were elicited. In general, experts expressed substantial uncertainty regarding these parameters, possibly reflecting their lack of direct experience (Table 2, distribution (3)). The pilot trial evidence had a significant impact on those estimates where elicited data were very uncertain (Table 2, distribution (5)); for example, trial data suggested that the duration of treatment was much shorter than assumed under existing evidence. No complications occurred during pilot trial follow-up.

Parameterizing the evidence. Complications and discontinuation of treatment occur only in unhealed patients. Thus, a measure of the incidence of these events was applied to the pool of unhealed patients at each time interval of the Markov model. Complications were allowed to differ by treatment type.

Classifying a closure surgery as "successful" meant that an ulcer was considered healed. This required further consideration because uncertain judgments obtained from expert opinion regarding the occurrence and success of closure surgery had the potential to generate probabilistically incoherent results regarding healing. To prevent this, a different parameterization was adopted, and 2 alternative quantities (quantities 1 and 2 as defined in Table 2) were considered for inclusion in the model: $P[S|H]$ and $P[H|S]$, where S is the occurrence of closure surgery and H is the occurrence of healing. These quantities were used, alongside the overall probability of healing, to determine the occurrence of closure surgery by using the relation $P[S] = \frac{P[S|H] \cdot P[H]}{P[H|S]}$.

Evidence on Costs and Utilities

Costs and resource use data were derived from the literature,[31] with diffuse distributions assigned to reported point estimates (Table 2). No cost or utility data were elicited. The data derived from the pilot trial were influential in describing these uncertain quantities, that is, for the number of dressing changes for NPWT and comparator dressing treatments. Utilities for unhealed patients were estimated from the available data set and updated with pilot trial–derived EQ-5D scores (Table 3). Utilities for unhealed patients in the trial were substantially lower (mean = 0.110) than values derived from the literature (mean = 0.660). However, because of its small sample size, trial-derived utility data did not have a significant impact on the posterior distribution. These parameters were applied directly to the model; no further parameterization was conducted.

Nonstochastic Parameters

Other model parameters, mainly unit costs, were assumed to be nonstochastic (Table 4). Unit cost parameters were derived from national figures. The incidence and prevalence of severe pressure ulcers were used in calculating the value of conducting further research.

COST-EFFECTIVENESS AND DECISION UNCERTAINTY

The model was run probabilistically using 1000 Monte Carlo simulations. The number of simulations was defined using assessments of convergence.

Cost-Effectiveness Based on Existing Evidence Alone (Scenario 1)

In Figure 3A, where cost-effectiveness calculations were informed by data from existing literature, foam appeared to be the most effective dressing treatment. This finding was, however, attributed to the assumption of common (random) effects employed for the synthesis of existing trial data. In this scenario, NPWT does not appear cost-effective: it had the lowest expected NHB (Figure 3A). If treatment choice was based on existing data (combined with the necessary assumptions), then foam dressings might be recommended. However, such a decision would be highly uncertain (P = 0.32 of being cost-effective),

Table 4 Nonstochastic Parameters

Description	Value	Source
Treatments, cost (£)		
NPWT		
Cost of VAC machine and canister, £/week	£299	
Cost of VAC dressings, £/dressing change	£25.49	
Foam, £/dressing change	£3.79	
Alginate, £/dressing change	£2.10	
Spun hydrocolloid, £/ dressing change	£2.48	
Other unit costs		
Costs associated with treatment change excluding treatments (additional dressing costs), £/dressing change	£21.19	Bennett and others[31]
Costs of closure surgery, £/surgery	£2795.72	Department of Health[51]
Other parameters		
Resident population of the UK (mid-2008)	61,383,000	Office for National Statistics[33]
Prevalence of pressure ulceration, per 1000 individuals	0.74	Vowden and Vowden[52]
Proportion of severe ulcers out of total ulcers	33%	Vowden and Vowden[52]
Incidence of severe ulcers	4997 cases/year[a]	

Note: NPWT = negative-pressure wound therapy; VAC = vacuum assisted closure.
a. The estimate of 4997 incident cases per year was based on the total population of the UK and on the incidence of severe pressure ulceration. The latter derived from the prevalence of pressure ulceration, the proportion of grade 3 and 4 pressure ulcers, and the expected time to healing of a pressure ulcer (61,383,000 × 0.74/1000 × 0.33 / 3 years = 14,990 / 3 years = 4997, discounted at 3.5% as 1/(1 + 0.035)^(time in years − 1).

and it would be very likely that any of the other treatments might offer similar or better value ($P = 0.16$ for spun hydrocolloid dressings, 0.30 for alginate dressings, and 0.22 for NPWT).

Cost-Effectiveness Based on Existing and Elicited Evidence (Scenario 2)

Analysis based on combined existing and elicited data suggested that spun hydrocolloid dressing had the highest expected NHB (Figure 3B). The inclusion of elicited evidence, which suggested that foam dressings were the least effective treatment, meant that this treatment was very unlikely to be cost-effective ($P = 0.02$). Again, NPWT was estimated to be more effective than the dressing comparators. Uncertainty in choosing the optimal treatment was also substantial in this scenario ($P = 0.37$ for spun hydrocolloid dressings, 0.32 for alginate dressings, and 0.29 for NPWT).

Cost-Effectiveness Based on Existing, Elicited, and Pilot Trial–Derived Evidence (Scenario 3)

When all evidence sources were combined, NPWT was expected to be less costly and more effective than other treatments (Figure 3C). This suggests that this treatment could be adopted for the treatment of severe pressure ulcers (Table 5). The improved cost-effectiveness of NPWT reflected the shorter duration of treatment in the pilot trial than that reported by manufacturers and with fewer dressing changes. However, even using all available evidence, a decision is still very uncertain (Table 5); there is a significant chance (approximately 0.55) that NPWT will in fact not offer the best value.

VALUE OF FURTHER RESEARCH

All available evidence was used to conduct a value of further research analysis. Before presenting the results of these analyses, specificities to their implementation are detailed.

Implementation of EVPI Analysis

Estimates of the EVPI for subsets of parameters require that these parameters are not correlated with the remaining model parameters. Given that relative treatment effects were estimated jointly, the existence and impact of correlation were assessed prior to implementing the EVPI analyses. For such, correlation between the estimated log hazard ratios and between transformations of these parameters (as used in the decision model) was assessed using correlation coefficients and graphical displays (scatterplots): there was no indication of the presence of correlation between the input parameters. We also

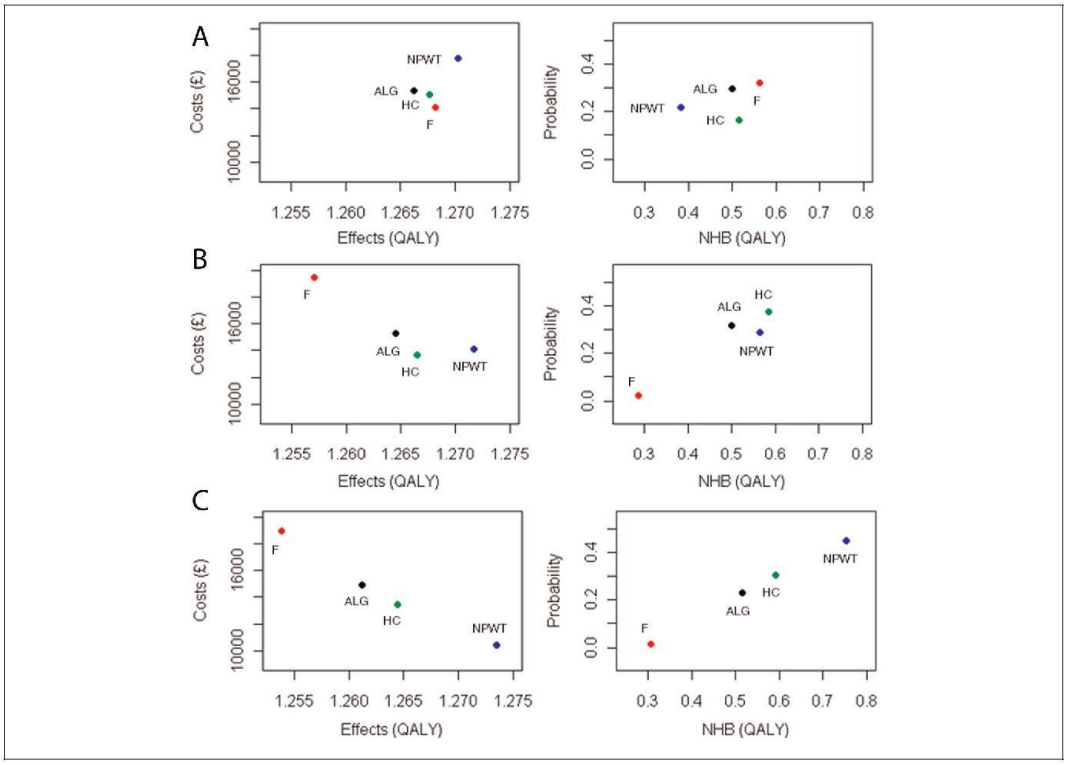


Figure 3 Cost-effectiveness and decision uncertainty considering existing evidence (A), existing and elicited evidence (B), and existing, elicited, and trial-derived evidence (C). For each scenario, cost-effectiveness planes are presented on the left and expected net benefit plotted against the probability of each of the treatments being cost-effective at a willingness to pay of £20,000/QALY on the right. NPWT = negative-pressure wound therapy; HC = spun hydrocolloid; ALG = alginate; F = foam; NHB = net health benefit; QALY = quality-adjusted life year.

Table 5 Cost-Effectiveness Estimates When Existing, Elicited, and Pilot Trial Evidence Are Considered

					Probability of a Treatment Being Cost-Effective	
Treatment	**Costs (£)**	**Effectiveness (QALY)**	**NHB (QALY)**	**Next Best ICER (£/QALY)**	**£20,000**	**£30,000**
NPWT	10,399	1.273	0.754	—	0.451	0.460
HC	13,461	1.264	0.591	Dominated	0.304	0.296
ALG	14,898	1.261	0.516	Dominated	0.230	0.231
F	18,969	1.254	0.305	Dominated	0.015	0.013

Note: NPWT = negative-pressure wound therapy; HC = spun hydrocolloid; ALG = alginate; F = foam; NHB = net health benefit; ICER = incremental cost-effectiveness ratio; QALY = quality-adjusted life year.

evaluated the potential impact of correlation in the outcomes: 1) in the NHBs associated with the treatments, 2) in the decision regarding which treatment to adopt, and 3) in the maximum NHBs attained. We used only graphical methods to evaluate item 1: distinct values of a conditioning parameter (1 of the

Table 6 Estimates of the Value of Further Research

	Individual Value: NHB, QALY (NMB, £)	Population Values[a]: NHB, QALY (NMB, £)
EVPI	0.114 (£2273)	4888 (£97.8 million)
EVPI for transition parameters		
Absolute healing parameters and death rates	0 (£0)	0 (£0)
Relative treatment effects	0.101 (£2010)	4327 (£87 million)
F v. HC	0.001 (£13)	29 (£0.6 million)
ALG v. HC	0.041 (£817)	1757 (£35.1 million)
NPWT v. HC	0.056 (£1114)	2395 (£47.9 million)
EVPI for other relevant events and payoffs		
Related events (surgery and complications)	0 (£0)	0 (£0)
Costs, discontinuation, and number of dressing changes	0.018 (£363)	780 (£15.6 million)
Utilities	0 (£0)	0 (£0)

Note: NPWT = negative-pressure wound therapy; HC = spun hydrocolloid; ALG = alginate; F = foam; NHB = net health benefit; NMB = net monetary benefit; EVPI = expected value of perfect information; QALY = quality-adjusted life year.
a. Benefits from research are assumed to sustain for 10 years.

3 treatment effect parameters) were plotted against the NHBs associated with the remaining treatments. This was applied in turn to all 3 treatment effects. To evaluate the impact over the adoption decision (item 2 above) for distinct quartiles of the conditioning parameter, the probabilities that the remaining treatments were chosen as the optimum treatment were evaluated. These were not expected to differ significantly between the quartiles. To evaluate the impact over maximum net benefits (item 3 above) for each quartile of the variable of interest, we summarized the maximum net benefits. These were again not expected to differ significantly. The results did not show correlation to have a significant impact, and independence was therefore assumed, allowing parameter EVPI to be estimated separately for each treatment effect parameter.

We also note that in calculating parameter EVPI, model linearity was not assumed, and 2 nested Monte Carlo simulation procedures were conducted to evaluate the nested expectations. Convergence was assessed, and 1000 simulations were used in both simulation procedures.

Results of EVPI Analysis

The EVPI for the choice between NPWT and the dressing treatments was estimated to be 0.114 QALYs per patient or £2273, which represents the monetary value of resolving the uncertainty surrounding treatment choice for each patient (Table 6). The incident UK population of grade 3 or 4 pressure ulcer patients was estimated at 4997 per year (Table 4), and assuming benefits from research would last 10 years, the EVPI for this population of patients was approximately £97.8 million. This represents the maximum value of further research that might resolve existing uncertainties. Because this value exceeds the likely costs of further investigation, additional research in this area is potentially worthwhile.

In the case study, the EVPI associated with the relative effectiveness parameters, specifically for NPWT and alginate dressing compared with spun hydrocolloid dressing, was estimated to be £83 million, suggesting that investment in an RCT to provide unbiased estimates of these parameters might be worthwhile. However, additional evidence about cost parameters might also be valuable because they were associated with an EVPI of £15.6 million. The analyses suggested that there was little value associated with resolving uncertainties in the other parameters, such as utilities, and baseline progression.

DESIGN OF FURTHER RESEARCH

In the analyses of parameter EVPI, uncertainty over the relative effectiveness parameters was identified to have the highest consequences; we have thus taken this forward by evaluating the EVSI, focusing on the design of an RCT.

Research Designs Evaluated

The EVSI associated with 3 alternative trial designs was established: 1) a 2-arm trial evaluating NPWT versus spun hydrocolloid dressings, 2) a 2-arm trial evaluating NPWT versus alginate dressings,

Table 7 Breakdown of Assumed Annual Costs for a Future Trial

	Annual Cost	Phase of Study Where Annual Cost Is Incurred
Site staff		
Principal investigator	£2389.8 per site	Recruitment, start-up, and follow-up
Research nurse	£33,727 per site	Recruitment, start-up, and follow-up
Treating nurse	£516.18 per patient	Recruitment and follow-up
Site administrator	£698.04 per site	Recruitment, start-up, and follow-up
University staff		
Chief investigator	£4926.45	Recruitment, start-up, and follow-up
Senior statistician	£4926.45	Recruitment, start-up, and follow-up
Junior statistician	£14,370.8	Recruitment, start-up, and follow-up
Health economist	£14,370.8	Recruitment, start-up, and follow-up
Methodologist	£4926.45	Recruitment, start-up, and follow-up
Trial coordinator	£35,927	Recruitment, start-up, and follow-up
Administrative staff	£5085	Recruitment, start-up, and follow-up
Data management	£12,944.2	Recruitment, start-up, and follow-up
Excess treatment costs	£3232 per patient	
Other consumables	£12,000	Recruitment, start-up, and follow-up
Overheads		70% of costs with research staff at university

and 3) a 3-arm trial evaluating the 3 treatments. In the case study, primary research results were assumed to inform the relative treatment effect but not the baseline hazard. For each trial design, the sample size of each arm was assumed equal.

Within the framework of analyses implemented, the outcomes measured within a future trial were numbers of patients achieving healing (the data mentioned in the previous section on the analytic framework). This quantity was modeled using a Poisson variable where a given sample size and follow-up period need to be assumed, an approach that is proposed by Ades and others[12] in calculating the EVSI for hazard ratios using conjugacy. In the case study, a range of sample sizes and the impact of alternative follow-up times (i.e., duration of data collection for each participant after randomization) of 0.5, 1, and 2 years were also evaluated. Given a Poisson distribution is used, the rate of occurrence of these events is assumed constant with exposure, where exposure combines sample size with follow-up time. A study with, for example, 200 patients followed up for 0.5 years thus provides the same information as a study with 100 patients but with a follow-up time of 1 year.

Predictions of the costs of possible research designs were based on those of previously funded wound care projects (Table 7). Potential recruitment rates (1 patient per month/site) and staffing levels required were based on data from the pilot study, where recruitment occurred at a low rate.

Implementation of EVSI and ENBS Analysis

Calculations of the EVSI assume that prior information is updated using the data made available to generate the posterior distribution $\upsilon|data$. This requires Bayesian updating; conjugacy was used (the prior distribution and data are conjugate distributions), and thus, the posterior distribution was known in closed form.

Despite this, the computational burden of estimating the EVSI for all research designs without assuming linearity would have been significant (7.5 days to evaluate 20 alternative sample sizes). Linearity can be assumed when the model outputs (here, the NHBs) are a linear (or multilinear under certain additional conditions[12]) function of the input parameters ($\boldsymbol{\theta}$). Assuming linearity means expected NHBs can be computed in a single calculation with the parameters set to their expected values (because $E[NHB_t^{\boldsymbol{\theta}}]$ is equal to $NHB_t^{E[\boldsymbol{\theta}]}$). In practice, 1 of the 2 nested simulation procedures needed for computations of parameter EVPI and EVSI can be eliminated, reducing significantly the computational burden.

Within this work, the extent and impact of nonlinearity were evaluated by comparing the results of nonlinear computations (for parameter EVPI) with those based on an assumption of linearity. Results were similar between approaches, and thus, linearity was used to obtain a first estimate of optimal sample size (N*) for each trial design. To confirm the linear estimate of the sample size leading to the maximum

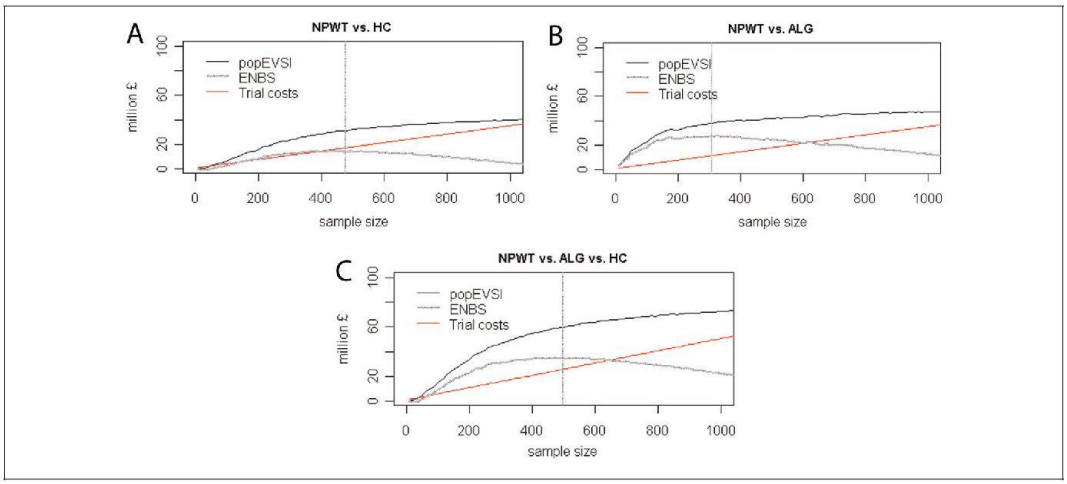


Figure 4 Population EVSI, costs of a trial, and ENBS for alternative designs of a randomized controlled trial of 1 year of follow-up: (A) two arm trial of NPWT v. HC, (B) two arm trial of NPWT v. ALG, (A) three arm trial of NPWT v. HC v. ALG. Vertical gray line represents the sample size at which the maximum ENBS is obtained. NPWT = negative-pressure wound therapy; HC = spun hydrocolloid; ALG = alginate; F = foam; EVSI = expected value of sample information; ENBS = expected net benefit of sampling.

ENBS, nonlinear ENBS estimates were obtained for N* and for sample sizes representing ±20% of N*. For all alternative designs evaluated, the value of the ENBS calculated nonlinearly was similar to the one calculated linearly and confirmed the estimate of N*.

A second consideration in calculating the EVSI is correlation, and following the results of investigations undertaken for parameter EVPI in the case study, independence was assumed throughout. Convergence was evaluated for all simulation procedures and was used to define the number of samples with which to conduct the EVSI (1000 for each simulation procedure).

Results of EVSI and ENBS Analysis

The results of EVSI analysis indicate that, across distinct sample sizes, a 3-arm trial would produce more value than a 2-arm trial (EVSI in Figure 4). Within 2-arm trials, comparing NPWT with alginate dressings was of more value than comparing with spun hydrocolloid dressings (due to the higher uncertainty regarding the use of alginate dressings).

Trials following up patients for 0.5 years were shown to be less efficient research designs or not worthwhile (Table 8). Longer follow-up times provided the best value (higher ENBS); however, a trial considering a 2-year follow-up is likely to be difficult to implement in this patient population. Table 8 shows that the maximum ENBS for a 3-arm trial with 1-year follow-up is reached at a sample size of 497 patients (in each arm), with an estimated net value of £34.7 million, while the maximum ENBS for a 2-arm trial (NPWT v. alginate dressings) is returned with a sample size of 306 patients, with a net value of £27.2 million.

DISCUSSION

This research demonstrates the use of explicit methods to estimate the cost-effectiveness of a range of treatments while characterizing the considerable uncertainty regarding treatment choice. The use of NPWT in severe pressure ulceration was utilized as a case study; this is a medical device widely used in the UK to treat these ulcers, despite a dearth of evidence regarding its relative effectiveness.

A key feature of this work is that multiple sources of available evidence were included in the evaluation in a transparent way: data from the literature (using IMTCs to estimate relative treatment effects in the absence of relevant head-to-head trials), elicited data, and pilot trial data. The presentation of cost-effectiveness estimates after the addition of data

Table 8 Optimal Sample Size and ENBS for Alternative Designs of Further Research

	NPWT v. Spun Hydrocolloid		NPWT v. Alginate		NPWT v. Spun Hydrocolloid v. Alginate	
Follow-up Time	Maximum ENBS	Optimal Sample Size, N*	Maximum ENBS	Optimal Sample Size, N*	Maximum ENBS	Optimal Sample Size, N*
0.5 years	—	—	£12.3 million	272	£154,028	403
1 year	£14.0 million	476	£27.2 million	306	£34.7 million	497
2 years	£27.1 million	389	£35.2 million	234	£54.6 million	411

Note: The maximum ENBS was calculated from smoothed ENBS functions using a polynomial function of degree 5. Smoothing did not provide a good fit in one scenario (3-arm trial following up patients for 0.5 years), in which case the observed maximum ENBS and correspondent sample size are presented.
NPWT = negative-pressure wound therapy; ENBS = expected net benefit of sampling.

from each source into the model emphasizes how different conclusions would have been drawn if analysis were limited to only one source of data. This highlights the importance of including all relevant data in any DAM.[43]

To collate the evidence from each of the sources, we used Bayesian updating. This methodology has the potential of being used sequentially, for example, in updating the results of DAMs every time new evidence becomes available. This can be extremely useful when 1) adoption and research decisions are linked, as the adoption decision needs to be revised after research reports, or 2) when research is conducted sequentially. An example of the latter could have been implemented in our case study because existing evidence alone could have been firstly used to establish the need for elicited evidence. This was not done (and all sources were collected without looking at the results of the DAM) because many parameters of the DAM had no evidence available in the literature and vague or uninformative distributions were thus used to describe these. In specifying these distributions, a judgment on the value of the mean is required, which can, as a consequence, affect cost-effectiveness and the valuation of further research needs.

One feature of the DAM implemented was that while it considered closure surgery and the occurrence of complications on healing, it did not explicitly evaluate the impact of these on healing rates. There was, however, an absence of data that suggests such a relationship. Additionally, studies in the literature evaluating the effectiveness of distinct treatments report the proportion of patients who have healed over time, but patients may have had complications or closure surgery before healing. Analytically, these data are identified as partially observed data.[42] Methods to adequately analyze these data are complex, and the computational burden increases with model complexity. A more complex model could thus have been specified; however, given the absence of evidence and the complexity of implementation, this was deemed unreasonable.

Another feature of the current analyses was that recurrence was not considered. We found no existing evidence to indicate that the alternative treatments impacted ulcer recurrence, and there is also no clinical rationale for an association. Even when assuming no impact of treatment on recurrence, if recurrence rates were high, cost-effectiveness may have been overestimated. This is because the time spent in the state healed would be inferior due to recurrence. Because there were only small differences in utility between healed and unhealed states, we would not expect this to have a significant impact on the results. Despite not having incorporated recurrence in the decision model (due to an expected increase in model complexity), we did elicit recurrence rates: the experts expected 31% of healed patients to have a recurrence after 6 months (SD = 24%), and 24% of patients were expected to have a recurrence in the following 6 months (SD = 22%). This was elicited using 2 questions: 1) Think of UK patients with at least one debrided grade 3 or 4 pressure ulcer (greater than 5 cm^2 in area). Imagine that these patients had their deepest grade 3 or 4 pressure ulcer (reference ulcer) treated and that it healed completely at some point. Six months after healing, what proportion of patients do you think would have a recurrent pressure ulcer (grade 2 to 4) in the same site as their previous grade 3 or 4 reference ulcer? 2) Now think of only those patients who still do not have a recurrent pressure ulcer 6 months after healing. What proportion of them do you think would have a recurrent pressure ulcer (grade 2 to 4) in the same site between 6 and 12 months after healing of the reference ulcer?

Previous assessments of cost-effectiveness in wound care evaluated decision models using both patient-level simulation[44] and cohort models.[45] In the case study, we used a discrete time cohort model

because the use of a patient-level simulation model could have either precluded the analyses of the value of further research or it could have meant that a metamodel or other approximation needed to be used.[46,47]

As well as facilitating immediate decision making, the results of this study suggest that the implications of continuing to use NPWT in the UK in the absence of robust evidence are likely to be even more costly than conducting further research to reduce the uncertainty. The value of further research analysis was possible when it was most important, that is, when the existing evidence base was clearly insufficient, and further research will need to be prioritized. In guiding further research, alternative study designs were evaluated, including not only the sample size of an RCT but also possible comparator arms and alternative follow-up times. By conducting a pilot trial, we were able to assess both the feasibility of future research and its likely costs (through, for example, prediction of recruitment rates). This information is invaluable to research commissioners often faced with the aftermath of unrealistic, anticipated recruitment rates and the high cost of study extensions.[48]

A key source of data was the elicitation of clinical experience, and therefore, the conclusions drawn are to some extent dependent on the experts' judgments. Ignoring experts' beliefs as a source of evidence could be regarded as wasteful (which often translates into money).[49] Despite the elicitation used here, representing one of the largest exercises of its type to have been undertaken in the field of health technology assessment, performing such an extensive quantitative elicitation exercise required a number of assumptions, including independence between the quantities elicited by each expert. However, such assumptions are common in statistical or mathematical modeling.

When combining experimental and elicited evidence, it is important to consider the potential for double counting. Although there is no way to know how much the nurse experts have considered existing evidence when formulating their judgments, it is likely that their own experience was the main source. This is because published research evidence was very sparse, and results of IMTCs were not made available to them. Moreover, the treatments of interest were not divulged before the elicitation exercise, and nurses were not incentivized to review the existing literature.

The analyses presented here demonstrate how decisions about medical technologies can be explicitly informed when data are sparse and, most importantly, guiding future research prioritization, not only indicating whether further research is worthwhile but also what "kind" of research is needed and how it should be designed.

ACKNOWLEDGMENTS

We thank all the nurses and tutors who took part in the pilot and full elicitation exercises and all the staff and patients involved in the pilot trial, especially research nurse Sally Blundell. We also thank Holly Essex and Mike Clark for supplying quality of life data and John Posnett for supplying data from a previous study he conducted. We also thank Tony Ades and Nicky Welton for analytical help and advice. The literature searches were carried out by Ruth Foxlee.

REFERENCES

1. Claxton K, Sculpher M, Drummond M. A rational framework for decision making by the National Institute For Clinical Excellence (NICE). Lancet. 2002;360(9334):711–5.

2. O'Hagan A. Uncertain Judgements: Eliciting Experts' Probabilities. Hoboken, NJ: John Wiley & Sons; 2006.

3. Philbeck TE Jr., Whittington KT, Millsap MH, Briones RB, Wight DG, Schroeder WJ. The clinical and cost effectiveness of externally applied negative pressure wound therapy in the treatment of wounds in home healthcare Medicare patients. Ostomy Wound Manage. 1999;45(11):41–50.

4. Kinetic Concepts Inc. Science Behind the Therapy: Understanding the Science of V.A.C. Therapy. San Antonio, TX: Kinetic Concepts Inc. 2010.

5. Witkowski W, Jawien A, Witkiewicz W, Zon B. Initial multicentre observations upon the effect of a new topical negative pressure device upon patient and clinician experience and the treatment of wounds. Int Wound J. 2009;6(2):167–74.

6. Vella K, Goldfrad C, Rowan K, Bion J, Black N. Use of consensus development to establish national research priorities in critical care. BMJ. 2000;320(7240):976–80.

7. National Institute for Health and Clinical Excellence. Guide to the Methods of Technology Appraisal. London: NICE; 2008.

8. Soares MO, Bojke L, Dumville JC, Iglesias CP, Cullum NA, Claxton K. Methods to elicit experts' beliefs over uncertain quantities: application to a cost effectiveness transition model of negative pressure wound therapy for severe pressure ulceration. Stat Med. 2011. ;30(19):2363–2380.

9. Briggs AH, Claxton K, Sculpher MJ. Decision Modelling for Health Economic Evaluation. Oxford: Oxford University Press; 2006.

10. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. J Health Econ. 1999;18(3):341–64.

11. Sculpher M, Claxton K. Establishing the cost-effectiveness of new pharmaceuticals under conditions of uncertainty: when is there sufficient evidence? Value Health. 2005;8(4):433–46.

12. Ades AE, Lu G, Claxton K. Expected value of sample information calculations in medical decision modelling. Med Decis Making. 2004;24(2):207–27.

13. Stevenson MD, Jones ML. The cost-effectiveness of a randomized controlled trial to establish the relative efficacy of vitamin K1 compared with alendronate. Med Decis Making. 2011;31(1): 43–52.

14. Philips Z, Ginnelly L, Sculpher M, et al. Review of guidelines for good practice in decision-analytic modelling in health technology assessment. Health Technol Assess. 2004;8(36):1–158.

15. Ford CN, Reinhard ER, Yeh D, et al. Interim analysis of a prospective, randomized trial of vacuum-assisted closure versus the healthpoint system in the management of pressure ulcers. Ann Plast Surg. 2002;49(1):55–61.

16. Graumlich JF, Blough LS, McLaughlin RG, et al. Healing pressure ulcers with collagen or hydrocolloid: a randomized, controlled trial. J Am Geriatr Soc. 2003;51(2):147–54.

17. Belmin J, Meaume S, Rabus M-T, Bohbot S. Sequential treatment with calcium alginate dressings and hydrocolloid dressings accelerates pressure ulcer healing in older subjects: a multicenter randomized trial of sequential versus nonsequential treatment with hydrocolloid dressings alone. J Am Geriatr Soc. 2002;50(2): 269–74.

18. Kloth LC, Berman JE, Nett M, Papanek PE, Dumit-Minkel S. A randomized controlled clinical trial to evaluate the effects of noncontact normothermic wound therapy on chronic full-thickness pressure ulcers. Adv Skin Wound Care. 2002;15(6):270–6.

19. Sopata M, Luczak J, Ciupinska M. Effect of bacteriological status on pressure ulcer healing in patients with advanced cancer. J Wound Care. 2002;11(3):107–10.

20. Honde C, Derks C, Tudor D. Local treatment of pressure sores in the elderly: amino acid copolymer membrane versus hydrocolloid dressing. J Am Geriatr Soc. 1994;42(11):1180–3.

21. Matzen S, Peschardt A, Alsbjørn B. A new amorphous hydrocolloid for the treatment of pressure sores: a randomised controlled study. Scand J Plast Reconstr Surg Hand Surg. 1999;33(1): 13–5.

22. Burgos A, Gimenez J, Moreno E, et al. Cost, efficacy, efficiency and tolerability of collagenese ointment versus hydrocolloid occlusive dressing in the treatment of pressure ulcers. Clin Drug Invest. 2000;19:357–65.

23. Whitney JD, Salvadalena G, Higa L, Mich M. Treatment of pressure ulcers with noncontact normothermic wound therapy: healing and warming effects. J Wound Ostomy Continence Nurs. 2001;28(5):244–52.

24. Muller E, van Leen MW, Bergemann R. Economic evaluation of collagenase-containing ointment and hydrocolloid dressing in the treatment of pressure ulcers. Pharmacoeconomics. 2001;19(12): 1209–16.

25. Price P, Bale S, Crook H, Harding K. The effect of a radiant heat dressing on pressure ulcers. J Wound Care. 2000;9(4):201–5.

26. Thomas DR, Diebold MR, Eggemeyer LM. A controlled, randomized, comparative study of a radiant heat bandage on the healing of stage 3-4 pressure ulcers: a pilot study. J Am Med Dir Assoc. 2005;6(1):46–9.

27. Caldwell DM, Ades AE, Higgins JPT. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. BMJ. 2005;331(7521):897–900.

28. Nixon RM, Bansback N, Brennan A. Using mixed treatment comparisons and meta-regression to perform indirect comparisons to estimate the efficacy of biologic treatments in rheumatoid arthritis. Stat Med. 2007;26(6):1237–54.

29. Essex HN, Clark M, Sims J, Warriner A, Cullum N. Health-related quality of life in hospital inpatients with pressure ulceration: assessment using generic health-related quality of life measures. Wound Repair Regen. 2009;17(6):797–805.

30. Brazier J, Usherwood T, Harper R, Thomas K. Deriving a preference-based single index from the UK SF-36 Health Survey. J Clin Epidemiol. 1998;51(11):1115–28.

31. Bennett G, Dealey C, Posnett J. The cost of pressure ulcers in the UK. Age Ageing. 2004;33:230–5.

32. Office for National Statistics. Interim life tables. 2007. http://www.statistics.gov.uk.

33. Office for National Statistics. Population estimates. August 2009. http://www.statistics.gov.uk.

34. Garthwaite PH, Kadane JB, O'Hagan A. Statistical methods for eliciting probability distributions. J Am Stat Assoc. 2005;100:680–701.

35. Kadane JB, Wolfson LJ. Experiences in elicitation. Statistician. 1998;47(1):3–19.

36. Choy SL, O'Leary R, Mengersen K. Elicitation by design in ecology: using expert opinion to inform priors for Bayesian statistical models. Ecology. 2009;90(1):265–77.

37. McKenna C, McDaid C, Suekarran S, et al. Enhanced external counterpulsation for the treatment of stable angina and heart failure: a systematic review and economic analysis. Health Technol Assess. 2009;13(24):1–90.

38. Speight PM, Palmer S, Moles DR, et al. The cost-effectiveness of screening for oral cancer in primary care. Health Technol Assess. 2006;10(14):1–144.

39. Ouchi F. A Literature Review on the Use of Expert Opinion in Probabilistic Risk Analysis. World Bank Policy Research Working Paper 3201. Washington, DC: World Bank; 2004.

40. Clemen RT, Winkler RL. Combining probability distributions from experts in risk analysis. Risk Anal. 1999;19(2):187–203.

41. Spiegelhalter DJ, Myles JP, Jones DR, Abrams KR. Bayesian methods in health technology assessment: a review. Health Technol Assess. 2000;4(38):1–130.

42. Welton NJ, Ades AE. Estimation of Markov chain transition probabilities and rates from fully and partially observed data: uncertainty propagation, evidence synthesis, and model calibration. Med Decis Making. 2005;25(6):633–45.

43. Claxton K, Cohen JT, Neumann PJ. When is evidence sufficient? Health Aff. 2005;24(1):93–101.

44. Michaels JA, Campbell WB, King BM, et al. A prospective randomised controlled trial and economic modelling of antimicrobial silver dressings versus non-adherent control dressings for venous leg ulcers: the VULCAN trial. Health Technol Assess. 2009;13(56):1–114.

45. Iglesias CP, Claxton K. Comprehensive decision-analytic model and Bayesian value-of-information analysis: pentoxifylline in the treatment of chronic venous leg ulcers. Pharmacoeconomics. 2006;24(5):465–78.

46. Brennan A, Kharroubi SA. Efficient computation of partial expected value of sample information using Bayesian approximation. J Health Econ. 2007;26(1):122–48.

47. Tappenden P, Chilcott JB, Eggington S, Oakley J, McCabe C. Methods for expected value of information analysis in complex health economic models: developments on the health economics of interferon-beta and glatiramer acetate for multiple sclerosis. Health Technol Assess. 2004;8(27):1–78.

48. McDonald AM, Knight RC, Campbell MK, et al. What influences recruitment to randomised controlled trials? A review of trials funded by two UK funding agencies. Trials. 2006;7:9.

49. Bolstad WM. Introduction to Bayesian Statistics. Hoboken, NJ: Wiley; 2004.

50. Kinetic Concepts Inc. V.A.C. Therapy Clinical Guidelines: A Reference Source for Clinicians. San Antonio, TX: Kinetic Concepts Inc.; 2007.

51. Department of Health. Reference Costs: 2007-2008. London: Department of Health; 2008.

52. Vowden KR, Vowden P. The prevalence, management and outcome for acute wounds identified in a wound care survey within one English health care district. J Tissue Viability. 2009;18(1):7–12.

Bibliographical references

1. Palmer S, Raftery J. Economic Notes: opportunity cost. BMJ. 1999;318(7197):1551-1552.

2. Claxton K, Martin S, Soares M, et al. Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold. Health Technol Assess. 2015;19(14):1-503, v-vi.

3. Stinnett AA, Mullahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. Med Decis Making. 1998;18(2 Suppl):S68-80.

4. McCabe C, Claxton K, Culyer AJ. The NICE cost-effectiveness threshold: what it is and what that means. Pharmacoeconomics. 2008;26(9):733-744.

5. National Institute for Health and Care Excellence. https://www.nice.org.uk/. Accessed 24.03, 2017.

6. Canadian Agency for Drugs and Technologies in Health. CADTH. https://www.cadth.ca/, 2017.

7. Pharmaceutical Benefits Advisory Committee. https://pbac.pbs.gov.au/home.html. Accessed 24.03, 2017.

8. Infarmed -- Autoridade Nacional do Medicamento e Produtos de Saúde I.P. Infarmed. http://www.infarmed.pt/. Accessed 24.03, 2017.

9. Sculpher MJ, Claxton K, Drummond M, McCabe C. Whither trial-based economic evaluation for health care decision making? Health Econ. 2006;15(7):677-687.

10. Banta D. The development of health technology assessment. Health Policy. 2003;63(2):121-132.

11. Drummond M. Methods for the economic evaluation of health care programmes. Fourth edition. ed. Oxford, United Kingdom ; New York, NY, USA: Oxford University Press; 2015.

12. Sutton AJ. Methods for meta-analysis in medical research. Chichester ; New York: J. Wiley; 2000.

13. Caldwell DM, Welton NJ, Ades AE. Mixed treatment comparison analysis provides internally coherent treatment effect estimates based on overviews of reviews and can reveal inconsistency. J Clin Epidemiol. 2010;63(8):875-882.

14. Sutton AJ, Higgins JP. Recent developments in meta-analysis. Stat Med. 2008;27(5):625-650.

15. Dias S, Welton NJ, Sutton AJ, Ades AE. Evidence synthesis for decision making 1: introduction. Med Decis Making. 2013;33(5):597-606.

16. Saramago P, Manca A, Sutton AJ. Deriving input parameters for cost-effectiveness modeling: taxonomy of data types and approaches to their statistical synthesis. Value Health. 2012;15(5):639-649.

17. Briggs AH, Claxton K, Sculpher MJ. Decision modelling for health economic evaluation. Oxford: Oxford University Press; 2006.

18. Weinstein MC, O'Brien B, Hornberger J, et al. Principles of good practice for decision analytic modeling in health-care evaluation: Report of the ISPOR task force on good research practices-modeling studies. Value in Health. 2003;6(1):9-17.

19. Briggs A, Sculpher M. An introduction to Markov modelling for economic evaluation. Pharmacoeconomics. 1998;13(4):397-409.

20. Briggs A. Probabilistic analysis of cost-effectiveness models: statistical representation of parameter uncertainty. Value Health. 2005;8(1):1-2.

21. Doubilet P, Begg CB, Weinstein MC, Braun P, McNeil BJ. Probabilistic sensitivity analysis using Monte Carlo simulation. A practical approach. Med Decis Making. 1985;5(2):157-177.

22. Cooper NJ, Spiegelhalter D, Bujkiewicz S, Dequen P, Sutton AJ. Use of implicit and explicit bayesian methods in health technology assessment. Int J Technol Assess Health Care. 2013;29(3):336-342.

23. Bojke L, Claxton K, Sculpher M, Palmer S. Characterizing structural uncertainty in decision analytic models: a review and application of methods. Value Health. 2009;12(5):739-749.

24. Mahon R, Manca A, Palmer S. Where Did the Time Go?: Temporal Uncertainty in Cost-Effectiveness Decision Models. Value in Health. 2012;15(7):A373-A373.

25. Griffin S, Claxton K, Hawkins N, Sculpher M. Probabilistic analysis and computationally expensive models: Necessary and required? Value Health. 2006;9(4):244-252.

26. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. J Health Econ. 1999;18(3):341-364.

27. Claxton K, Cohen JT, Neumann PJ. When is evidence sufficient? Health Affair. 2005;24(1):93-101.

28. Claxton K, Ginnelly L, Sculpher M, Philips Z, Palmer S. A pilot study on the use of decision theory and value of information analysis as part of the NHS Health Technology Assessment programme. Health Technol Assess. 2004;8(31):1-103, iii.

29. Ades AE, Lu G, Claxton K. Expected value of sample information calculations in medical decision modeling. Medical Decision Making. 2004;24(2):207-227.

30. Meltzer DO, Hoomans T, Chung JW, Basu A. Minimal modeling approaches to value of information analysis for health research. Med Decis Making. 2011;31(6):E1-E22.

31. Strong M, Oakley JE, Brennan A. Estimating multiparameter partial expected value of perfect information from a probabilistic sensitivity analysis sample: a nonparametric regression approach. Med Decis Making. 2014;34(3):311-326.

32. Chalkidou K, Hoy A, Littlejohns P. Making a decision to wait for more evidence: when the National Institute for Health and Clinical Excellence recommends a technology only in the context of research. J R Soc Med. 2007;100(10):453-460.

33. Tunis SR, Chalkidou K. Coverage with evidence devlopment: A very good beginning, but much to be done. Commentary to Hutton et al. Int J Technol Assess. 2007;23(4):432-435.

34. Agency EM. Adaptive pathways. http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000601.jsp. Accessed 24.03, 2017.

35. Government U. Accelerated access review. https://www.gov.uk/government/organisations/accelerated-access-review. Accessed 24.03, 2017.

36. Eichler HG, Pignatti F, Flamion B, Leufkens H, Breckenridge A. Balancing early market access to new drugs with the need for benefit/risk data: a mounting dilemma. Nat Rev Drug Discov. 2008;7(10):818-826.

37. Department of Health Australian Government. Co-dependent and Hybrid Technologies. http://www.health.gov.au/internet/hta/publishing.nsf/Content/co-1. Accessed 24.03, 2017.

38. National Institute for Health and Care Excellence. Diagnostics Assessement Programme Manual. 2011.

39. Meckley LM, Neumann PJ. Personalized medicine: factors influencing reimbursement. Health Policy. 2010;94(2):91-100.

40. NHS England. Cancer Drugs Fund. https://www.england.nhs.uk/cancer/cdf/.

41. Systemic Anti-Cancer Therapy Dataset. http://www.chemodataset.nhs.uk/home. Accessed 24.03, 2017.

42. National Institute for Health and Care Excellence. Summary of decisions 2017; https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-technology-appraisal-guidance/summary-of-decisions. Accessed 10 February, 2017.

43. Gregson N, Sparrowhawk K, Mauskopf J, Paul J. Pricing medicines: theory and practice, challenges and opportunities. Nat Rev Drug Discov. 2005;4(2):121-130.

44. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal. London: NICE;2013.

45. Fenwick E, Claxton K, Sculpher M. Representing uncertainty: the role of cost-effectiveness acceptability curves. Health Econ. 2001;10(8):779-787.

46. Andronis L, Barton P, Bryan S. Sensitivity analysis in economic evaluation: an audit of NICE current practice and a review of its use and value in decision-making. Health Technol Asses. 2009;13(29):1-+.

47. Hutton J, Trueman P, Henshall C. Coverage with evidence development: An examination of conceptual and policy issues. Int J Technol Assess. 2007;23(4):425-432.

48. Raftery J. Multiple sclerosis risk sharing scheme: a costly failure. BMJ. 2010;340:c1672.

49. Griffin SC, Claxton KP, Palmer SJ, Sculpher MJ. Dangerous omissions: the consequences of ignoring decision uncertainty. Health Econ. 2011;20(2):212-224.

50. Palmer S, Smith PC. Incorporating option values into the economic evaluation of health care technologies. J Health Econ. 2000;19(5):755-766.

51. Eckermann S, Willan AR. Expected value of information and decision making in HTA. Health Econ. 2007;16(2):195-209.

52. Stafinski T, McCabe CJ, Menon D. Funding the unfundable: mechanisms for managing uncertainty in decisions on the introduction of new and innovative technologies into healthcare systems (vol 28, pg 113, 2010). Pharmacoeconomics. 2010;28(7):608-608.

53. Rogowski W, Burch J, Palmer S, Craigs C, Golder S, Woolacott N. The effect of different treatment durations of clopidogrel in patients with non-ST-segment elevation acute coronary syndromes: a systematic review and value of information analysis. Health Technol Asses. 2009;13(31):1-+.

54. Philips Z, Claxton K, Palmer S. The half-life of truth: what are appropriate time horizons for research decisions? Med Decis Making. 2008;28(3):287-299.

55. Higgins JPT, Green S, Cochrane Collaboration. Cochrane handbook for systematic reviews of interventions. Chichester, England ; Hoboken, NJ: Wiley-Blackwell; 2008.

56. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. Res Synth Methods. 2010;1(2):97-111.

57. DerSimonian R, Kacker R. Random-effects model for meta-analysis of clinical trials: an update. Contemp Clin Trials. 2007;28(2):105-114.

58. Higgins JP, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. J R Stat Soc Ser A Stat Soc. 2009;172(1):137-159.

59. Welton N, Ades AE. Research decisions in the face of heterogeneity: what can a new study tell us? Health Econ. 2012;21(10):1196-1200.

60. Ades AE, Lu G, Higgins JP. The interpretation of random-effects meta-analysis in decision models. Med Decis Making. 2005;25(6):646-654.

61. Soares MO, Welton NJ, Harrison DA, et al. An evaluation of the feasibility, cost and value of information of a multicentre randomised controlled trial of intravenous immunoglobulin for sepsis (severe sepsis and septic shock): incorporating a systematic review, meta-analysis and value of information analysis. Health Technol Assess. 2012;16(7):1-186.

62. Welton NJ, Soares MO, Palmer S, et al. Accounting for Heterogeneity in Relative Treatment Effects for Use in Cost-Effectiveness Models and Value-of-Information Analyses. Med Decis Making. 2015;35(5):608-621.

63. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med. 2002;21(11):1539-1558.

64. Shuster JJ, Jones LS, Salmon DA. Fixed vs random effects meta-analysis in rare event studies: the rosiglitazone link with myocardial infarction and cardiac death. Stat Med. 2007;26(24):4375-4385.

65. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Stat Med. 2004;23(20):3105-3124.

66. Nam IS, Mengersen K, Garthwaite P. Multivariate meta-analysis. Stat Med. 2003;22(14):2309-2333.

67. Jackson D, Riley R, White IR. Multivariate meta-analysis: potential and promise. Stat Med. 2011;30(20):2481-2498.

68. Davis S, Tappenden P, Cantrell A. A review of studies examining the relationship between progression-free survival and overall survival in advanced or metastatic cancer. Sheffield;2012.

69. Owen RK, Tincello DG, Keith RA. Network meta-analysis: development of a three-level hierarchical modeling approach incorporating dose-related constraints. Value Health. 2015;18(1):116-126.

70. Dakin HA, Welton NJ, Ades AE, Collins S, Orme M, Kelly S. Mixed treatment comparison of repeated measurements of a continuous endpoint: an example using topical treatments for primary open-angle glaucoma and ocular hypertension. Stat Med. 2011;30(20):2511-2535.

71. Warren FC, Abrams KR, Sutton AJ. Hierarchical network meta-analysis models to address sparsity of events and differing treatment classifications with regard to adverse outcomes. Stat Med. 2014;33(14):2449-2466.

72. Ades AE, Sculpher M, Sutton A, et al. Bayesian methods for evidence synthesis in cost-effectiveness analysis. Pharmacoeconomics. 2006;24(1):1-19.

73. Ades AE, Cliffe S. Markov chain Monte Carlo estimation of a multiparameter decision model: consistency of evidence and the accurate assessment of uncertainty. Med Decis Making. 2002;22(4):359-371.

74. Ades AE, Welton NJ, Caldwell D, Price M, Goubar A, Lu G. Multiparameter evidence synthesis in epidemiology and medical decision-making. J Health Serv Res Policy. 2008;13 Suppl 3:12-22.

75. Welton NJ. Evidence synthesis for decision making in healthcare. Chichester, West Sussex: John Wiley & Sons; 2012.

76. Jackson D, White IR, Price M, Copas J, Riley RD. Borrowing of strength and study weights in multivariate and network meta-analysis. Stat Methods Med Res. 2015.

77. Achana FA, Cooper NJ, Bujkiewicz S, et al. Network meta-analysis of multiple outcome measures accounting for borrowing of information across outcomes. BMC Med Res Methodol. 2014;14:92.

78. Garthwaite PH, Kadane JB, O'Hagan A. Statistical methods for eliciting probability distributions. J Am Stat Assoc. 2005;100(470):680-700.

79. Montibeller G, von Winterfeldt D. Cognitive and Motivational Biases in Decision and Risk Analysis. Risk Anal. 2015;35(7):1230-1251.

80. O'Hagan A. Uncertain judgements : eliciting experts' probabilities. London ; Hoboken, NJ: John Wiley & Sons; 2006.

81. Grigore B, Peters J, Hyde C, Stein K. Methods to elicit probability distributions from experts: a systematic review of reported practice in health technology assessment. Pharmacoeconomics. 2013;31(11):991-1003.

82. Babuscia A, Cheung KM. An approach to perform expert elicitation for engineering design risk analysis: methodology and experimental results. J R Stat Soc a Stat. 2014;177(2):475-497.

83. Bojke L, Claxton K, Bravo-Vergel Y, Sculpher M, Palmer S, Abrams K. Eliciting distributions to populate decision analytic models. Value Health. 2010;13(5):557-564.

84. Garthwaite PH, Chilcott JB, Jenkinson DJ, Tappenden P. Use of expert knowledge in evaluating costs and benefits of alternative service provisions: a case study. Int J Technol Assess Health Care. 2008;24(3):350-357.

85. Vannoortwijk JM, Dekker R, Cooke RM, Mazzuchi TA. Expert Judgment in Maintenance Optimization. Ieee T Reliab. 1992;41(3):427-432.

86. Chaloner K, Rhame FS. Quantifying and documenting prior beliefs in clinical trials. Stat Med. 2001;20(4):581-600.

87. French S. Aggregating expert judgement. Racsam Rev R Acad A. 2011;105(1):181-206.

88. Cooke RM. Experts in uncertainty : opinion and subjective probability in science. New York: Oxford University Press; 1991.

89. Bolger F, Rowe G. The aggregation of expert judgment: do good things come to those who weight? Risk Anal. 2015;35(1):5-11.

90. Clemen RT, Winkler RL. Combining probability distributions from experts in risk analysis. Risk Analysis. 1999;19(2):187-203.

91. Stone M. Opinion Pool. Ann Math Stat. 1961;32:1339-&.

92. Annemans L, Geneste B, Jolain B. Early modelling for assessing health and economic outcomes of drug therapy. Value Health. 2000;3(6):427-434.

93. Garrison LP, Jr., Towse A, Briggs A, et al. Performance-based risk-sharing arrangements-good practices for design, implementation, and evaluation: report of the ISPOR good practices for performance-based risk-sharing arrangements task force. Value Health. 2013;16(5):703-719.

94. Welton NJ, Thom HH. Value of Information: We've Got Speed, What More Do We Need? Med Decis Making. 2015;35(5):564-566.

95. Claxton K, Griffin S, Koffijberg H, McKenna C. How to estimate the health benefits of additional research and changing clinical practice. BMJ. 2015;351:h5987.

96. Hettle R, Corbett M, Hinde S, et al. The assessment and appraisal of regenerative medicines and cell therapy products: an exploration of methods for review, economic evaluation and appraisal. Health Technol Assess. 2017;21(7):1-204.