Overview

Adalimumab, etanercept and infliximab for ankylosing spondylitis

The overview is written by members of the Institute’s team of technical analysts. It forms part of the information received by the Appraisal Committee members before the first committee meeting. The overview summarises the evidence and views that have been submitted by consultees and evaluated by the Assessment Group, and highlights key issues and uncertainties. To allow sufficient time for the overview to be circulated to Appraisal Committee members before the first Appraisal Committee meeting, it is prepared before the Institute receives consultees’ comments on the assessment report. These comments are therefore not addressed in the overview.

A list of the sources of evidence used in the preparation of this document is given in appendix A.

1 Background

1.1 The condition

Ankylosing spondylitis (AS) is an inflammatory disease of unknown cause. It is one of a group of conditions known as seronegative spondyloarthropathies. The principal feature of AS is inflammation of the sacroiliac joint at the base of the spine (sacroiliitis) followed by inflammation rising along the spine. The result is back pain and stiffness. Inflammation at entheses (the sites where ligaments and tendons attach to bone) can lead to new bone development and joint fixation (ankylosis). The large peripheral joints (hips, shoulders and knees) may also be involved, and the eye and cardiovascular system can also be affected. Systemic involvement in addition to eye and cardiovascular disease may be significant. Disease damage is progressive and irreversible. Osteoporosis can occur early in the disease, contributing to the increased risk of spinal fracture later in life.
Although symptoms can occur at any stage of life, AS typically affects people in their late teenage years and twenties. AS is nearly three times as common in men as it is in women and men are also more likely to develop severe spinal disease.

The course of AS is highly variable and is characterised by mild or moderate flares of active disease, alternating with periods of near or total quiescence. The disease can be severe, with spinal fusion, pronounced incapacity and significant deformities. There is a need for joint replacement surgery in some patients. About a third of people with AS may be unable to work altogether, with a further 15% reporting some changes to their working lives. In addition, AS is associated with an increased risk of death: the assessment report states that patients have a standardised mortality ratio of 1.5 or greater. According to British Society for Rheumatology (BSR) guidelines, the excess mortality is mainly accounted for by cardiac valvular disease, amyloidosis and fractures.

The prevalence of AS is unknown, although it has been estimated to range from 0.05% to 0.23% (based on data from the UK and Hungary). Evidence from the USA suggests that the annual incidence is 7.3 per 100,000 person–years. The Assessment Group considered that data from a Finnish study provide more accurate estimates of the prevalence and incidence of AS for a UK population (see assessment report section 8.1). In that study, the prevalence of ‘clinically significant AS’ was estimated to be 0.15%. The annual incidence was calculated to be 6.9 per 100,000. Using mid-2004 population figures for England and Wales, this suggests that there are approximately 2300 new cases each year.

There are three key elements to AS that are assessed in clinical trials: disease activity, physical function and structural damage. A number of assessment tools have been developed to measure these. For example, the Bath ankylosing spondylitis disease activity index (BASDAI) is the most commonly used instrument to measure the inflammatory activity of AS. The BASDAI is a validated, composite index that records patients’ responses to six questions relating to the five major symptoms of AS: fatigue, axial and peripheral pain, stiffness and enthescopathy. Responses are recorded on 10-cm visual analogue scales (VAS).

Physical function is widely assessed through the use of the Bath ankylosing spondylitis functional index (BASFI). The BASFI is a patient-assessed, validated, composite index made
up of 10 questions that address function and the patient’s ability to cope with AS. As with the BASDAI, responses are recorded on a 10-cm VAS.

Structural damage and disease progression are primarily evaluated using radiography. Two instruments used to assess structural damage are the Bath ankylosing spondylitis radiology index (BASRI) and the modified Stoke ankylosing spondylitis spinal score (mSASSS).

The Assessment in Ankylosing Spondylitis (ASAS) working group has developed a set of response criteria commonly used in AS clinical trials. The ASAS improvement criteria relate to improvement across a set of four domains:

- patient global assessment (measured on a 10-mm VAS)
- spinal pain (measured on a 100-mm VAS)
- physical function (measured using the BASFI)
- inflammation (mean of the last two questions from the BASDAI, concerning the intensity and duration of morning stiffness).

Scores in each domain range from 0 (best) to 100 (worst). An ASAS 20 response (a common primary efficacy outcome in clinical trials) is defined as an improvement of greater than 20% and an absolute change of 10 VAS or more points on the 0–100 scale in at least three of the four domains. On the fourth domain there must be no worsening by a similar amount. Other definitions of ASAS response (ASAS 50 and ASAS 70) based on improvements of 50% and 70%, respectively, are also used to measure outcomes in clinical studies.

1.2 Current management

Therapies aim to provide symptom relief and improve spinal mobility. Conventional therapy for AS includes agents such as non-steroidal anti-inflammatory drugs (NSAIDs), drugs that modify the disease process in rheumatoid arthritis (for example, sulfasalazine and methotrexate), and non-drug interventions (for example, physiotherapy). These treatments may be used concomitantly with tumour necrosis factor (TNF) alpha inhibitors.

Physiotherapy, exercise and NSAIDs are often first-line therapies. From the available evidence discussed in the assessment report, the benefits of sulfasalazine and methotrexate in AS are somewhat limited (see section 2.1.10 of the assessment report for further details).
Furthermore, consultees have indicated that disease-modifying anti-rheumatic drugs (DMARDs) may help with peripheral joint involvement, but not with spinal symptoms.

According to guidelines produced by the BSR, there is no clear evidence that NSAIDs alter the structural progression of the disease. This, and the side-effect profile of these drugs, have led clinicians to use NSAIDs for clinical relapses rather than as continuous therapy in the majority of patients.

2 The technologies

<table>
<thead>
<tr>
<th>Table 1  Summary description of technologies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generic name</td>
</tr>
<tr>
<td>Proprietary name</td>
</tr>
<tr>
<td>Manufacturer</td>
</tr>
<tr>
<td>Dose</td>
</tr>
<tr>
<td>Acquisition cost excluding VAT (BNF edition 51)</td>
</tr>
</tbody>
</table>

Adalimumab, etanercept and infliximab inhibit the activity of TNF alpha. TNF alpha is a naturally occurring cytokine involved in normal immune responses. Excess TNF alpha activity can cause severe inflammation and tissue damage.

Adalimumab and etanercept are both licensed for the treatment of adults with severe active AS which has responded inadequately to conventional therapy. In addition, both treatments are administered via subcutaneous injection, with patients able to self-inject. Infliximab is
licensed for treatment of AS in patients who have severe axial symptoms, elevated serological markers of inflammatory activity and whose AS has responded inadequately to conventional therapy. It is administered as an intravenous infusion.

Both etanercept and infliximab, and to a lesser extent adalimumab, have been used for some time in the treatment of rheumatoid arthritis (RA), and therefore their adverse-event profiles are relatively well known. Experience of their use in AS is more limited. All three drugs have the potential to cause serious adverse events and to cause the development of autoimmune antibodies. All three drugs can also cause adverse effects associated with the mode of administration: injection-site reactions with adalimumab and etanercept, and acute infusion-related reactions with infliximab. As these drugs attenuate the immune response, the SPCs carry warnings about the risk of infections and neurological events.

Guidelines from the BSR¹ state that at least two NSAIDs each taken sequentially at the maximum tolerated/recommended dosage for 4 weeks should have been tried before anti-TNF therapy is initiated. The Scottish Medicines Consortium (SMC) has issued guidance on the use of etanercept and infliximab for the treatment of AS. Both technologies have been recommended for use in Scotland as long as prescribing decisions are in accordance with BSR guidelines.

3 The evidence

3.1 Clinical effectiveness

The Assessment Group found nine randomised controlled trials (RCTs) that met the inclusion criteria for their review. All these studies were placebo controlled and all participants continued with some form of standard treatment such as NSAIDs with or without DMARDs.

3.1.1 Adalimumab studies

There were two RCTs comparing adalimumab with placebo. In the larger of the two studies (known as ATLAS ['Adalimumab trial evaluating long-term efficacy and safety in ankylosing spondylitis'] or M03-607) 315 patients were randomised to adalimumab or

---

placebo in a 2:1 ratio. ATLAS was a multicentre trial conducted at 43 centres in the USA and Europe. In the smaller study (known as M03-606 and referred to in the assessment report as the ‘Canadian AS’ study) 82 patients were randomised to adalimumab or placebo in a 1:1 ratio. The entry criteria for both studies specified that participants must have active AS which had responded inadequately to one or more NSAIDs. People in whom one or more DMARDs had failed were also included.

Both studies were double blind in design and in both adalimumab was given at a dose of 40 mg by subcutaneous injection every other week. Treatment was for 24 weeks after which both active and placebo groups were switched to open-label treatment with adalimumab. The primary endpoint was the proportion of ASAS 20 responders at week 12. Those who had not reached an ASAS 20 response on assessment at week 12, 16, or 20 were offered the option of ‘early escape’ open-label treatment with adalimumab. Participants who chose this option were counted only as ‘non-responders’ in the analysis of primary and secondary endpoints. The manufacturers note that this has implications when comparing the results of these studies with previous studies of anti-TNF drugs. Earlier studies did not include this ‘early escape’ option and therefore would include those who responded later than 12 weeks in their analyses of outcomes. For this reason the Assessment Group did not include 24-week data for adalimumab in their meta-analysis. Only 12-week data are presented below. In general, the 24-week outcome data suggest that differences versus placebo are maintained.

3.1.1.1 Composite binary outcomes (ASAS response rates)

In the ATLAS study the primary endpoint (ASAS 20 at week 12) was reached by 58.2% of the adalimumab group and 20.6% of the placebo group (p < 0.001). In the Canadian AS study, the respective corresponding ASAS 20 response rates were 47.2% and 21.5%. The difference did not reach conventional levels of statistical significance in the smaller study (p = 0.06). In the meta-analysis performed by the Assessment Group, the pooled relative risk of an ASAS 20 response was 2.43 (95% confidence interval [CI] 1.76 to 3.35), see figure 4-1 in the assessment report. The corresponding relative risks for ASAS 50 and ASAS 70 responses were 3.22 (95% CI 1.98 to 5.23) and 5.47 (95% CI 2.43 to 12.31), respectively.
3.1.1.2  Disease activity

The mean reduction in BASDAI at 12 weeks in the ATLAS study was 2.6 (based on 0–10 cm VAS) in the adalimumab group compared with 0.8 in the placebo group (p < 0.001). In the Canadian AS study the corresponding reductions were *** and **. In the Assessment Group meta-analysis the additional reduction in BASDAI achieved with adalimumab versus placebo (weighted mean difference) was **** (95% CI **** to ****), see figure 4-2 in assessment report.

3.1.1.3  Function

Adjusted BASFI score at 12 weeks in the ATLAS study was reduced by a mean of 1.72 (scale of 0-10) in the adalimumab group compared with 0.55 (8.0% reduction from baseline) in the placebo group (p < 0.001). In the Canadian study the corresponding reductions in BASFI scores were *** and ***, respectively. In the Assessment Group meta-analysis the additional reduction in BASFI achieved with adalimumab versus placebo (weighted mean difference) was **** (95% CI **** to ****), see figure 4-3 in assessment report.

3.1.1.4  Structural damage/radiological progression

In both adalimumab studies, the manufacturer’s submission reports that potential biomarkers of synovitis (matrix metalloproteinase-3) and type II collagen degradation (urinary type II collagen C-telopeptide) were significantly suppressed by adalimumab in the Canadian AS study. However, further research is required to demonstrate the relationship of these markers with structural damage.

3.1.2  Etanercept studies

The assessment group included five studies of etanercept versus placebo in their review. There were three phase III trials (Davis et al. 2003 [n = 277], Calin et al. 2004 [n = 84], and a third, unpublished study [n = 356] referred to as the ‘Wyeth study’ in the assessment report) and two smaller studies (Brandt et al. 2003 [n = 33], and Gorman et al. 2002 [n = 40]). The studies compared etanercept at a dose of 25 mg twice weekly with placebo. The unpublished study also had a third arm in which etanercept was given at a dose of 50 mg once a week. In
this study the participants appear to have been randomised to etanercept 50 mg weekly, 25 mg twice weekly and placebo in a 3:3:1 ratio. In the other studies, randomisation between the two arms was in a 1:1 ratio.

All the studies were double blind, and ASAS 20 response at 12 weeks was the stated primary endpoint in all but one of them (in Brandt et al., the primary end point was 50% improvement in disease activity using BASDAI). The duration of the studies varied from 6 weeks (Brandt et al.) to 24 weeks (Davis et al.).

3.1.2.1 Composite binary outcomes (ASAS response rates)

The ASAS response rates from the included studies are shown in Table 2 below. In the meta-analysis performed by the Assessment Group, the pooled relative risk of an ASAS 20 response was 2.13 at 12 weeks (95% CI 1.73 to 2.63) and 2.53 at 24 weeks (95% CI 1.80 to 3.57), see figure 4-1 in the assessment report. The corresponding relative risks for ASAS 50 and ASAS 70 response at 12 weeks were 3.53 (95% CI 2.50 to 4.98) and 3.38 (95% CI 2.10 to 5.45), respectively, and at 24 weeks the relative risks for ASAS 50 and ASAS 70 response were 3.96 (95% CI 2.37 to 6.63) and 4.59 (95% CI 2.32 to 9.07), respectively.

### Table 2  ASAS response rates in etanercept studies

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Placebo</td>
<td>ETN 25 mg twice weekly</td>
<td>ETN 50 mg once weekly</td>
<td>Placebo</td>
<td>ETN</td>
</tr>
<tr>
<td></td>
<td>n = 51</td>
<td>n = 150</td>
<td>n = 155</td>
<td>n = 139</td>
<td>n = 138</td>
</tr>
<tr>
<td>ASAS 20</td>
<td>6 weeks</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>25%</td>
</tr>
<tr>
<td></td>
<td>12 weeks</td>
<td>37%</td>
<td>71%</td>
<td>74%</td>
<td>27%</td>
</tr>
<tr>
<td></td>
<td>24 weeks</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>22%</td>
</tr>
<tr>
<td>ASAS 50</td>
<td>6 weeks</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>12 weeks</td>
<td>16%</td>
<td>45%</td>
<td>51%</td>
<td>13%</td>
</tr>
<tr>
<td>ASAS 70</td>
<td>12 weeks</td>
<td>8%</td>
<td>23%</td>
<td>32%</td>
<td>7%</td>
</tr>
</tbody>
</table>

*p value compared with placebo reported as 0.05 or less.
ETN = etanercept

3.1.2.2 Disease activity

Four of the five etanercept studies reported BASDAI scores (see Table 3). According to the Assessment Group meta-analysis based on the three studies that reported BASDAI at 12 weeks, the additional reduction in BASDAI score (on a scale of 0–10) achieved with
etanercept versus placebo (weighted mean difference) was 1.67 (95% CI 1.24 to 2.10), see figure 4-2 in the assessment report. At 24 weeks the additional reduction was 2.0 (95% CI 1.39 to 2.61), based on the study by Davis et al. only.

### Table 3  Mean BASDAI scores in etanercept studies

<table>
<thead>
<tr>
<th></th>
<th>Wyeth</th>
<th>Davis et al.</th>
<th>Calin et al.</th>
<th>Brandt et al.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Placebo ETN</td>
<td>Placebo ETN</td>
<td>Placebo ETN</td>
<td>Placebo ETN</td>
</tr>
<tr>
<td>Baseline</td>
<td>n = 51</td>
<td>n = 150</td>
<td>n = 155</td>
<td>n = 139</td>
</tr>
<tr>
<td>6 weeks</td>
<td>61.1</td>
<td>59.4</td>
<td>62.4</td>
<td>59.6</td>
</tr>
<tr>
<td>12 weeks</td>
<td>46</td>
<td>29.2</td>
<td>27.5</td>
<td>52.3</td>
</tr>
<tr>
<td>24 weeks</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>55.1</td>
</tr>
</tbody>
</table>

BASDAI was reported on a scale of 0–100 except for Brandt et al. which was reported on a scale of 0–10

\*p value for difference in change from baseline compared with placebo reported as 0.003 or less

ETN = etanercept

### 3.1.2.3  Function

BASFI data from the etanercept studies are shown in Table 4. The Assessment Group meta-analysis found an additional lowering in score of 1.48 (95% CI 1.13 to 1.83) with etanercept compared with placebo at 23 weeks and 1.42 (95% CI 0.95 to 1.89) at 24 weeks (on a scale of 0–10).

### Table 4  Mean BASFI scores in etanercept studies

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Placebo ETN</td>
<td>Placebo ETN</td>
<td>Placebo ETN</td>
<td>Placebo ETN</td>
<td>Placebo ETN</td>
</tr>
<tr>
<td>Baseline</td>
<td>n = 51</td>
<td>n = 150</td>
<td>n = 155</td>
<td>n = 139</td>
<td>n = 16</td>
</tr>
<tr>
<td>6 weeks</td>
<td>59.7</td>
<td>57.7</td>
<td>60.6</td>
<td>56.3</td>
<td>5.3</td>
</tr>
<tr>
<td>12 weeks</td>
<td>49</td>
<td>32.9</td>
<td>31.8</td>
<td>53.0</td>
<td>-</td>
</tr>
<tr>
<td>16 weeks</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>24 weeks</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>54.7</td>
<td>3.1</td>
</tr>
</tbody>
</table>

BASFI is reported on a scale of 0–100 except for Brandt et al. and Gorman et al which report it on a scale of 0–10

\*p value for difference in change from baseline compared with placebo reported as 0.008 or less

ETN = etanercept
3.1.2.4 **Radiographic progression**

The manufacturer notes that long-term placebo-controlled trials investigating the effects of anti-TNF therapies on radiographic progression are problematic for ethical and practical reasons. Radiographic outcome data are only available from observational follow-up. Patients in the study published by Calin et al. were eligible to enter a 2-year open-label follow-up study to provide longer-term data including radiographic outcomes. X-rays taken at baseline in the double-blind study were compared with films at week 48 of the extension or at early termination. There appeared to be no change in mSASSS over this period of time. In another extension to a study, 18 etanercept-treated patients had X-rays at baseline and after 1 year. The mean change in mSASSS was 1.4 points. However, given the relatively short observation periods and the lack of a control group it is difficult to draw conclusions about the effect of etanercept on radiological progression.

3.1.3 **Infliximab studies**

The Assessment Group included two randomised, placebo-controlled studies of infliximab in their review. In the larger of the two studies, ‘Ankylosing spondylitis study for the evaluation of recombinant infliximab therapy’ (ASSERT), participants were randomised in a 3:8 ratio to placebo or infliximab infusion at a dose of 5 mg/kg at weeks 0, 2, 6 and every 6 weeks thereafter. The primary endpoint was the proportion of ASAS 20 responders at week 24. The smaller study (Braun et al.) included 70 patients randomised to placebo or infliximab in a 1:1 ratio. The dose of infliximab was 5 mg/kg at weeks 0, 2 and 6, and the primary endpoint was the proportion of patients achieving a 50% reduction in BASDAI.

In both studies the inclusion criteria specified that patients had to have a BASDAI score of 4 or greater (scale of 0–10) and spinal pain score of 4 or greater measured on a 10-cm VAS.

3.1.3.1 **Composite binary outcomes (ASAS response rates)**

In the ASSERT study the primary endpoint (ASAS 20 at week 24) was reached by 61.2% of the infliximab group and 19.2% of the placebo group (p < 0.001). In the study by Braun et al., ASAS 20 response rates were reported only in graphic form. For the purposes of their meta-analysis, ASAS 20 results from the Braun et al. study were estimated from these plots. In the Assessment Group meta-analysis, the pooled relative risk of an ASAS 20 response with infliximab versus placebo was 4.11 at 12 weeks (95% CI 2.62 to 6.44), see figure 4-1 in the...
assessment report. At 24 weeks the relative risk of an ASAS 20 response was 3.18 (95% CI 1.99 to 5.08), based on ASSERT only.

3.1.3.2 Disease activity

The mean BASDAI score (0–10) at 12 weeks in the ASSERT study was in the infliximab group versus 0.4 in the placebo group. The corresponding figures from the Braun study were and 0.6. In the Assessment Group meta-analysis, the additional reduction in BASDAI score achieved with infliximab versus placebo (weighted mean difference) was 2.46 (see figure 4-2 in the assessment report). At 24 weeks the additional reduction was 2.3 (based on the study by ASSERT only).

3.1.3.3 Function

The mean in BASFI score (0–10) at 12 weeks in the ASSERT study was in the infliximab group versus in the placebo group. The corresponding figures from the Braun study were and . In the Assessment Group meta-analysis, the additional reduction in BASFI achieved with infliximab versus placebo (weighted mean difference) was (see figure 4-2 in the assessment report). At 24 weeks the additional reduction in BASFI score versus placebo was (based on the study by ASSERT only).

3.1.3.4 Radiographic progression

The manufacturer’s submission mentions one study that showed less radiographic progression in people treated with infliximab for 2 years than in people treated with conventional treatment. This was based on a comparison of 41 patients who had received infliximab as part of a clinical trial with 41 patients from a cohort study. The change in mSASSS was change was 0.4 in the infliximab-treated group and 0.7 in the patients from the cohort study. This difference was not statistically significant.

The main evidence of potential for an effect on structural progression comes from MRI imaging including some recent evidence from the ASSERT study. This study suggested a 73% reduction in spinal inflammation score (using a specially developed scoring system) at 2 years in the infliximab group versus no change in the placebo group.
3.1.4 Pooled results

The Assessment Group also assessed the anti-TNF agents as a class versus placebo. The results of this pooled analysis are shown in Table 5.

Table 5 Results of pooled analysis of anti-TNF agents as a class compared with placebo

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Pooled relative risk (95% CI)</th>
<th>Weighted mean difference (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>12 weeks</td>
<td>24 weeks</td>
</tr>
<tr>
<td>ASAS 20</td>
<td>2.52 (2.14 to 2.98)</td>
<td>2.80 (2.11 to 3.71)</td>
</tr>
<tr>
<td>ASAS 50</td>
<td>3.58 (2.72 to 4.71)</td>
<td>3.96 (2.37 to 6.63)</td>
</tr>
<tr>
<td>ASAS 70*</td>
<td>3.94 (2.61 to 5.95)</td>
<td></td>
</tr>
<tr>
<td>Continuous</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BASDAI</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>BASFI</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

* Infliximab studies did not report ASAS 70 so did not contribute to this estimate

The Assessment Group also undertook a longitudinal meta-analysis based on correlation data between two time points. This produced near-identical results to the standard meta-analysis (see section 4.4.5).

3.1.5 Indirect comparison

The Assessment Group attempted indirect comparisons of the agents (see section 4.4.4 and appendix 3 of the assessment report). No statistically significant differences were found in ASAS response rates. For BASDAI, the difference between infliximab and adalimumab was statistically significant at 12 weeks, as was the difference between infliximab and etanercept at the same time point. In both cases the comparison favoured infliximab, but the differences were described as marginal by the Assessment Group. For the comparison of infliximab and etanercept, a 24-week comparison was also possible, and this found that the difference in BASDAI was no longer statistically significant. For BASFI there was a significant difference between infliximab and adalimumab at 12 weeks and between infliximab and etanercept at 24 weeks. Again, both differences favoured infliximab, but were judged to be marginal.

3.1.6 Adverse events

There were few serious adverse events reported in the clinical studies (see section 4.4.6 of the assessment report). For the drugs given subcutaneously, injection-site reactions were the most commonly reported adverse events.
3.2 Cost effectiveness

3.2.1 Published literature

In their literature search, the Assessment Group identified six full economic evaluations (two peer-reviewed published papers and four abstracts) that met the inclusion criteria.

Three of the studies carried out cost–utility analyses, one study was a cost-minimisation analysis and the remainder were cost-effectiveness analyses. Only one study compared anti-TNF alpha drugs with each other (etanercept with infliximab). None of the economic evaluations considered adalimumab. Only one of the studies was UK based and compared infliximab with placebo; the remaining studies originated from Holland, Canada and the USA. Three of the studies attempted to extend beyond 1 year using observational data.

The cost-effectiveness results were difficult to compare because of the varying approaches adopted and limited information available from the abstracts. (See section 5.3 of the assessment report for further details.)

The Assessment Group also reviewed studies that examined the cost impact of AS. The Assessment Group found that the majority of the total costs of AS were indirect (that is, costs due to lost work productivity). Full details can be found in section 5.2 of the assessment report.

3.2.2 Manufacturer submissions

All three manufacturers provided economic evaluations. All evaluations adopted an NHS perspective, although Schering-Plough also considered a societal perspective in sensitivity analysis. Costs and benefits in all cases were discounted at 3.5%. None of the models attempted to make any direct or indirect comparisons between the TNF inhibitors.

Unlike the two other manufacturer models, which were constructed using Microsoft Excel, the Schering-Plough model was implemented within the TreeAge software package. In order to assess the Schering-Plough model on a comparable basis with the other two manufacturers, the Assessment Group attempted to replicate the model using Microsoft Excel.
3.2.2.1 Abbott model – adalimumab

The Abbott economic evaluation – structured as a patient-based transition-state model – compares the use of adalimumab plus NSAIDs versus treatment with NSAIDs alone. This model incorporates patient-level data from the Canadian AS and ATLAS phase III trials, and aims to simulate treatment decisions based on the BSR guidelines. The trial populations included patients who would not have met BSR eligibility criteria; for example, study patients included those who were intolerant to or whose AS had responded inadequately to fewer than two NSAIDs.

The model consists of two components. The first uses short-term trial data (first 48 weeks) as noted above. Patients are categorised as responders or non-responders at weeks 8, 12, 20, 24, 30, 36, 40 and 48. Regardless of response at week 8, all patients remained on therapy until week 12 in accordance with BSR guidelines. The second component simulates long-term outcomes for responders for up to 30 years.

In the short-term component of the model, response to treatment was defined as a reduction of BASDAI to 50% of the pre-treatment value or a fall of 2 units or more, plus a reduction of the spinal pain VAS score (over the last week) by 2 cm or more. From week 48 onwards, the BASDAI and spinal pain VAS scores of each patient (including those on standard therapy) were assumed to remain stable. In contrast, BASFI scores of patients on standard therapy were assumed to increase by 0.05 units/year. For adalimumab-treated patients, BASDAI and BASFI scores remained stable as long as the patient remained on adalimumab therapy. A separate scenario analysis was conducted to assess the impact of assuming no effect of adalimumab on BASFI progression.

After 48 weeks it was assumed that patients would discontinue adalimumab treatment at a rate of 10% per year. When patients on adalimumab discontinued therapy, it was assumed that their BASDAI and BASFI scores would return to the average values of those patients in the model managed by conventional therapy. In the economic evaluation the rebound occurs rapidly (that is, within the next measurement period). The Assessment Group was concerned that the values of BASDAI and BASFI beyond the observed data were calculated differently for the two groups, particularly in relation to adalimumab patients who discontinue treatment. (For further details, see page 91 of the assessment report.)
Disease-specific costs were based on ordinary least squares (OLS) regression of BASDAI and BASFI data from OASIS (‘Outcome in ankylosing spondylitis international study’)². Only BASDAI measurements were used to predict costs in the base-case analysis. In a secondary analysis, BASFI scores only were used to estimate costs. Unlike the two other manufacturer models, the costs of managing adverse events were included in the model.

In the base-case analysis, BASDAI and BASFI scores were used jointly to estimate health-related quality of life. A mapping exercise was undertaken, which used health utility index (HUI) scores collected in the Canadian AS and ATLAS studies. The increased risk of death due to AS was modelled by incorporating a standard mortality ratio of 1.5.

The base-case results are shown in Table 6. Univariate sensitivity analyses on a number of parameters including annual discontinuation rates were undertaken: incremental cost-effectiveness ratios (ICERs) varied from £18,000 to around £27,000 (over 30 years). In a scenario where the BASFI-based cost model was used, ICERs increased to £29,000 at 5 years and £25,300 at 10 years. However, the Assessment Group noted that both models yielded predictor coefficients with very wide confidence intervals. Consequently the Assessment Group argued that the predictive accuracy of these regression models was limited.

Table 6  Base-case economic results from the manufacturer submissions and Assessment Group’s independent modelling

<table>
<thead>
<tr>
<th>Model</th>
<th>Time period</th>
<th>Incremental cost per patient</th>
<th>Incremental utility per patient (QALYs)</th>
<th>Incremental cost per QALY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abbott - adalimumab</td>
<td>48 weeks</td>
<td>£5,025</td>
<td>0.107</td>
<td>£47,083</td>
</tr>
<tr>
<td></td>
<td>5 years</td>
<td>£13,273</td>
<td>0.504</td>
<td>£26,332</td>
</tr>
<tr>
<td></td>
<td>30 years</td>
<td>£23,857</td>
<td>1.033</td>
<td>£23,097</td>
</tr>
<tr>
<td>Schering-Plough – infliximab</td>
<td>Braun lifetime (70 years)</td>
<td>£29,399</td>
<td>1.62</td>
<td>£18,192</td>
</tr>
<tr>
<td></td>
<td>ASSERT (70 years)</td>
<td>£30,326</td>
<td>1.58</td>
<td>£19,169</td>
</tr>
<tr>
<td>Wyeth - etanercept</td>
<td>1 year</td>
<td>£6,174</td>
<td>0.14</td>
<td>£44,684</td>
</tr>
<tr>
<td></td>
<td>2 years</td>
<td>£10,298</td>
<td>0.33</td>
<td>£30,754</td>
</tr>
<tr>
<td></td>
<td>5 years</td>
<td>£19,136</td>
<td>0.84</td>
<td>£22,844</td>
</tr>
<tr>
<td></td>
<td>10 years</td>
<td>£26,940</td>
<td>1.50</td>
<td>£18,002</td>
</tr>
<tr>
<td></td>
<td>15 years</td>
<td>£29,782</td>
<td>1.99</td>
<td>£14,978</td>
</tr>
<tr>
<td></td>
<td>20 years</td>
<td>£30,880</td>
<td>2.26</td>
<td>£13,694</td>
</tr>
<tr>
<td></td>
<td>25 years</td>
<td>£31,365</td>
<td>2.37</td>
<td>£13,201</td>
</tr>
<tr>
<td>Assessment Group</td>
<td>12 months</td>
<td>Adalimumab and etanercept: £5,650</td>
<td>0.099</td>
<td>£57,260</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Infliximab: £11,850</td>
<td>0.099</td>
<td>£120,110</td>
</tr>
<tr>
<td></td>
<td>2 years</td>
<td>£9,000*</td>
<td>0.171</td>
<td>£52,534*</td>
</tr>
<tr>
<td></td>
<td>10 years</td>
<td>£23,713*</td>
<td>0.332</td>
<td>£71,454*</td>
</tr>
<tr>
<td></td>
<td>20 years</td>
<td>£33,159*</td>
<td>0.335</td>
<td>£98,910*</td>
</tr>
</tbody>
</table>

* Results for adalimumab and etanercept only. Infliximab ICERs ≥ £100K in all instances.

QALY, quality-adjusted life year.

3.2.2.2  Wyeth model – etanercept

The Wyeth model compares the use of etanercept plus NSAIDs with NSAIDs alone. The model generates a hypothetical patient population based on patient-level data from two RCTs and an open-label extension. The principal RCT evidence used in the model is drawn from the unpublished Wyeth study. The time horizon is up to 25 years.

Response to treatment is determined by BSR criteria, and is evaluated at 12 and 24 weeks within the model. Response rates in the placebo arms of the trials have been used to model outcomes for the comparator in the economic evaluation. Responders obtain an initial health gain in BASDAI/BASFI scores. Patients are assumed to remain at the new BASDAI/BASFI levels during the period they respond to treatment.

It was assumed in the base case that responders to etanercept do not progress in terms of BASDAI and BASFI scores while on treatment, and that the annual disease progression...
(BASDAI and BASFI) in non-responders is 0.3 units. In terms of the rate of withdrawal from etanercept treatment, it was assumed that initial responders to treatment would discontinue etanercept at a rate of 10% per year. In the model, patients who withdraw from etanercept would continue to receive NSAIDs, and therefore it was assumed that disease progression would mirror that of the comparator arm.

On treatment withdrawal, assumptions were made with respect to the impact on BASDAI and BASFI scores. Two scenarios were modelled:

- On withdrawal, the BASDAI and BASFI scores return to baseline levels (base case).
- The disability of the person is assumed to have progressed while on treatment. On withdrawal from etanercept scores rebound by the same magnitude as the initial improvement to a state worse than baseline.

Changes in BASDAI and BASFI scores drive changes in both predicted disease costs and utility. Regression analyses of EQ-5D (n = 356) and SF-36 (n = 511) data collected from AS patients were used to establish a relationship between changes in BASDAI and BASFI and utility. In the base-case analysis, the output from the EQ-5D regression was used to estimate quality-adjusted life years (QALYs).

Disease-related costs were based on a retrospective analysis of resource use by 147 AS patients attending the Staffordshire Rheumatology Centre. The Assessment Group noted that patients in this costing study had lower BASDAI and BASFI scores compared with those recruited in the RCTs and were approximately 10 years older. Adverse events were not considered in the modelling because the trials did not report statistically significant differences in adverse events between the two arms.

The obtained incremental cost-effectiveness ratios are presented in Table 6. A number of univariate sensitivity analyses were undertaken. When a utility model based on SF-36 data was used, ICERs were found to vary between £17,000 and £70,000 per QALY. Probabilistic sensitivity analysis indicated that over a 25-year time period, etanercept has an 88% probability of being cost effective at a threshold willingness to pay of £15,000.

The Assessment Group raised a number of concerns relating to the technical execution of the Wyeth model. The group noted, for example, that it predominantly generates female patients,
and also a number of hypothetical paediatric patients. (Etanercept is licensed for use in patients of 18 years and above.) In addition, the age and duration of disease for hypothetical patients are generated independently in the model.

3.2.2.3  **Schering-Plough model – infliximab**

The Schering-Plough model is based on a combined decision tree and Markov chain structure, and compares infliximab versus ‘standard therapy’. Two analyses are described, one based on the 24-week outcomes of the ASSERT trial and another based on those reported by Braun and coworkers up to 12 weeks. The placebo groups in the Braun registration trial and ASSERT were assumed to have received standard therapy as these studies allowed the concomitant used of NSAIDs.

In the base case, infliximab dosing was modelled as per the administration schedule in the clinical trials, that is, 5 mg/kg every 6 weeks, with an additional infusion after 2 weeks. Note that the SPC allows dosing to take place every 6–8 weeks. The drug dose was calculated based on the mean body weights reported in the Braun and ASSERT studies: 73.6 kg and 77.5 kg, respectively.

Patient-level data from the two RCTs informed the decision-tree component of the model. At 12 or 24 weeks (depending on the randomised period of the trial), continuation on infliximab treatment was determined by BSR criteria relating to BASDAI response at 12 weeks. (In the case of the ASSERT data, 24-week data were used.) All patients then enter the Markov model, which has three health states (on treatment, off-treatment, and dead) and an annual cycle.

The age at the start of the economic evaluation was fixed at 40 years and the analysis adopted a time horizon of 70 years. In addition, it was assumed that there would be no effect on mortality, so differences between the comparator and infliximab were based solely on issues of morbidity, the extent infliximab improved disease and functional activity, and delayed disease progression. Costs and utilities were assigned based on regression models controlling for BASDAI, BASFI, age and gender, and bootstrapping for the probabilistic analysis.

It was assumed that for patients, off-treatment in the model (infliximab, non-responders and placebo patients) would have a natural disease progression of 0.07 BASFI units/year. In
terms of withdrawal rate from treatment with infliximab, the analysis assumed an annual
drop-out rate of 15%, based on 2-year follow-up data on 18 patients who were eligible for
treatment continuation beyond 12 weeks in the Braun study. Patients who withdrew from
infliximab treatment were assigned the BASFI and BASDAI scores of the no-treatment
group, reduced by the underlying natural progression of BASFI during the years of treatment.
In other words, the rebounded BASFI score remains lower than for those patients in the
comparator arm.

Table 6 presents the results of the Schering-Plough model. In the base case, the estimated
ICERs were under £20,000. A number of sensitivity analyses were undertaken. One explored
varying the assumption that there is no disease progression rate while patients remain on
infliximab treatment in the model. If disease progression while on treatment is assumed to be
50% of natural history (that is, 0.035 units/year), the ICER rises to £31,000. If progression
while on treatment is the same as natural history (0.07 units/year), the ICERs increase to
between £50,000 and £54,000. Reducing the time horizon to 20 years also increases the
ICERs to between £32,000 and £34,000. The Assessment Group found that in the submitted
model, patients who withdrew from infliximab treatment appeared not to be assigned an ‘off-
treatment’ disease progression. On correcting this apparent error, the lifetime ICERs were
found to be between £41,000 and £50,000.

3.2.2.4 Assessment Group model
The Assessment Group examined the use of adalimumab, etanercept and infliximab versus
‘conventional treatment’ in a cohort of 1000 patients. ‘Conventional treatment’ was defined
in terms of the placebo arms of two adalimumab RCTs. The group explored cost
effectiveness of these interventions over the short term (1 year) and over a time horizon of up
to 20 years. The short-term model used weekly cycles for its Markov-like transitions, and the
long-term model (2–20 years) was based on quarterly cycles.

It is notable that the Assessment Group assumed that all three interventions were of equal
effectiveness. Differences in the cost effectiveness between these drugs were therefore driven
by issues related to their relative acquisition and administration costs. Short-term
effectiveness is modelled using response rates (based on BSR criteria) from the pooled week-
12 data from the adalimumab, etanercept and infliximab trials, and week-24 response rates after pro-rata imputation of the missing data from the Wyeth study.

Under base-case assumptions, from week 30 onwards it was assumed that spontaneous recovery without treatment would occur at a rate of 17.1% as identified in the patient-level analysis supplied in the Abbott submission.

The Assessment Group model assumes that patients withdraw from TNF inhibitor treatment at a rate of 15% per annum, although in sensitivity analyses, rates of 6% and 24% were also explored. The annual withdrawal rate (after 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 (see page 118 of the assessment report).

The Assessment Group model took into account the cost of adverse events and used the Abbott estimate of £95.29 in the first year of treatment, reducing by 50% per patient–year thereafter.

In terms of estimating disease related costs, the Assessment Group utilised OASIS data as per the Abbott submission, fitting an exponential cost model to the weighted aggregate OASIS data. BASFI was used as the major predictor of costs since it was considered to reflect long-term disease progression. An exponential model was used because evidence noted by the Assessment Group appears to suggest that cost increases in a steep, non-linear way for the highest BASDAI/BASFI scores.

Health-related quality of life was estimated by using the utility model provided by Schering-Plough on the grounds that it used a comparatively larger sample of UK AS patients, and also because it incorporates age and gender variables.

The Assessment Group adopted a long-term increase of BASFI scores of 0.07 units/year for the conventional treatment comparator arm of the model. This progression rate is applied for all periods after week 20 in the model. The model explored two alternative scenarios for BASFI progression on treatment:

- BASFI progression continues while on TNF inhibitor treatment.
- No increase in BASFI score while on treatment.
Results for the short- and long-term analyses are presented in Table 6. In the first year of treatment, costs are high because many patients are treated for a minimum of 12 weeks. ICERs for adalimumab and etanercept were essentially the same (around £57,000). In contrast, the ICER for infliximab was over £120,000.

With respect to the long-term modelling, the results for adalimumab were considered as representative of etanercept, and only the former were provided. It is notable that in contrast with the manufacturer models, ICERs increase steadily from year 2 onwards (see Table 6). This is because the costs of treatment accrue at a greater rate than the accumulation of QALYs for patients continuing on treatment with a TNF inhibitor. This is due primarily to assumptions made by the Assessment Group regarding spontaneous recovery of patients given conventional therapy (see figure 7.9 of the assessment report).

Univariate and multi-way sensitivity analyses were undertaken. Probabilistic sensitivity analysis was not performed because the Assessment Group considered that the results from such an analysis would be misleading. Notably it was found that increasing the baseline disease activity score from a BASDAI score of 4 to 8 worsened the cost-effectiveness profile of these agents, suggesting that treatment may be less cost effective for patients with the most severe symptoms at initiation and reflecting the data obtained from the patient-level analysis. Multi-way sensitivity analysis identified scenarios in which adalimumab/etanercept could be considered cost effective, with ICERs dropping to around £20,000. Important factors influencing the long-term cost effectiveness of these two agents included assumptions about spontaneous recovery, withdrawal rate from treatment and the BASFI progression rate. Multi-way sensitivity analysis on the infliximab results identified no scenario, assuming a discount rate of 3.5% for both costs and benefits, in which the ICER dropped below £37,000.

4 Issues for consideration

- Effectiveness and cost effectiveness of the TNF inhibitors have been considered in terms of AS-specific outcomes. This was the approach adopted by the manufacturers and the Assessment Group. The outcomes used may not fully capture any possible effects of treatment on non-skeletal manifestations of AS.
A critical issue in the cost effectiveness of these interventions is their long-term effectiveness and how that interacts with the natural progression of the disease. In the economic analyses undertaken by the Assessment Group, a number of factors influence the cost effectiveness of these agents, such as:

- whether or not disease progression continued while on treatment (no progression decreases the ICER)
- the withdrawal rate from treatment (lower withdrawal rates decrease the ICER)
- natural history, that is, the level of spontaneous recovery without treatment (no spontaneous recovery reduces the ICER substantially).

Analysis of patient-level data from two trials of adalimumab alone, led the Assessment Group to discuss the appropriateness of available guidelines for the use of TNF inhibitors in AS. The Assessment Group identified a set of factors that could be useful in deciding whether continued treatment with any particular intervention is appropriate (see section 7.2.4, page 111 of the assessment report).

On the basis of an uncertain natural history, the Assessment Group argued that episodic treatment with TNF inhibitors may be appropriate, for example targeting flare ups of the condition. Moreover, the group considered there to be no evidence supporting the continuous use of these interventions as implied by the economic evaluations produced by the manufacturers. However, BSR guidelines state that cessation of treatment with etanercept and infliximab (adalimumab was not considered at the time) usually results in recrudescence of symptoms. The guidelines state that response should be monitored regularly and that failure to maintain response should lead to a cessation or change in treatment.

The BSR guidelines state that treatment with TNF inhibitors should not be withdrawn within 12 weeks because of lack of effectiveness. However, the SPC for infliximab states that if there is a lack of response by 6 weeks (that is, after 2 doses), no additional treatment with infliximab should be given.

As noted in section 1 of this document, the patient-completed BASDAI and BASFI measurement tools are based on VAS scoring and there have been reservations about the use of such approaches. In addition, it’s not clear if such tools, most commonly used to
measure outcomes in clinical trials, will help inform clinical decision-making at the individual level. Nevertheless, BSR guidelines from 2005 on the use of TNF inhibitors in AS partly define eligibility and response in terms of BASDAI scores and spinal pain VAS scores.

- None of the economic evaluations considered the consecutive use of the TNF inhibitors in AS, although it is likely that a lack of response to one TNF inhibitor may lead a prescriber to trial another. The consecutive use of these interventions has been examined principally in patients with RA. There is limited observational evidence available suggesting that, for these patients, a second TNF inhibitor might be effective.

- The Assessment Group noted that an individual patient may be exposed to several trials of TNF inhibitors lasting at least 12 weeks (at least 36 weeks in total) before it can be determined if therapy with any of these agents is beneficial.

- NICE is carrying out a technology appraisal on etanercept and infliximab for psoriatic arthritis. In the final appraisal determination, etanercept was recommended for use in this indication. Infliximab had a poorer cost-effectiveness profile, but was recommended for use in patients with contraindications to/intolerant of etanercept.

- In the appraisal of TNF inhibitors in RA the Assessment Group analysis found that in the base case, all three TNF inhibitors (etanercept, adalimumab and infliximab) had poor cost-effectiveness profiles, although etanercept was found to be comparatively more cost effective than the other two. The appraisal consultation document for this appraisal has not recommended one drug above the others.
5 Ongoing research

The are a number of ongoing RCTs examining the use of the TNF inhibitors in AS. One RCT aims to compare etanercept with sulfasalazine in AS. Another aims to assess the efficacy of infliximab administered at a dose of 3 mg/kg.

6 Authors

Janet Robertson and Francis Ruiz
Health Technology Analysts, NICE Appraisal Team
July 2006
Appendix A. Sources of evidence considered in the preparation of the overview


B  Submissions from the following organisations:
   
   I  Manufacturer/sponsors:
      - Abbott Laboratories Ltd
      - Wyeth Pharmaceuticals
      - Schering-Plough Ltd
   
   II  Professional/specialist and patient/carer groups:
      - Royal College of Nursing
      - British Society for Rheumatology
      - ARMA – The Arthritis & Musculoskeletal Alliance
      - The National Ankylosing Spondylitis Society
      - BackCare
      - British Health Professionals in Rheumatology
      - Royal College of Physicians of Edinburgh