Rheumatoid arthritis (RA) is a systemic chronic inflammatory disabling disorder affecting approximately 480,000 patients in the UK [1]. Treatment involves non-steroidal anti-inflammatory drugs (NSAIDs), judicious use of glucocorticoids and early use of disease-modifying anti-rheumatic drugs (DMARDs) [2]. DMARD use often begins with methotrexate. Few patients achieve remission and long-term outcomes can be poor [3], with many patients switching medication either due to lack of efficacy or adverse effects [4, 5]. Sequences of DMARD use are the common pattern of intervention as the disease progresses. Surgical interventions are frequent, including joint replacement and tendon reconstruction, to improve functional disability and chronic severe pain [6]. RA patients also have reduced life expectancy (standardized mortality ratio of between 1.5 and 3) [7, 8]. Predictors of mortality rates include Health Assessment Questionnaire (HAQ) score, functional class and joint count [9].

RA also has a substantial economic impact. A systematic review of 49 studies estimates the mean annual medical cost per RA patient at around US$5800 (£3600) using 1996 prices [10]. In 1992, UK health and social care direct costs were estimated at over £651m [11]. Higher HAQ scores have higher healthcare, carer and lost productivity costs [12–15]. Sequences of DMARD use are the common pattern of intervention as the disease progresses. Surgical interventions are frequent, including joint replacement and tendon reconstruction, to improve functional disability and chronic severe pain [6]. RA patients also have reduced life expectancy (standardized mortality ratio of between 1.5 and 3) [7, 8]. Predictors of mortality rates include Health Assessment Questionnaire (HAQ) score, functional class and joint count [9].

To assess the cost-effectiveness of etanercept therapy, a decision analytic model was developed. The central estimate cost per QALY is £16,330. Sensitivity analyses (£7800 to £42,000) showed long-term HAQ progression (etanercept, DMARDs, non-responders) as most sensitive variables. The inclusion of potential avoided nursing home admissions and indirect costs/lost employment further improves the cost-effectiveness.

Conclusions: For adults in the UK, the results suggest that etanercept is cost-effective when compared with non-biologic agents. The National Institute for Clinical Excellence has accepted that etanercept is cost-effective and recommended its availability for use in patients who have failed at least two DMARDs. This model was an important component of that decision. The model is further suitable for use for a wide range of other cost-effectiveness questions in rheumatoid arthritis.
against current standard therapy, and measuring both the health benefits of etanercept in real clinical practice and the associated resource use and costs. Such studies are simply not available at present, and policy decisions must be made based on synthesis of all currently available evidence.

Method

Model framework

The model assumes that the population concerned has already failed at least two DMARDs, one of which is methotrexate. The baseline characteristics for the population examined are based on the published etanercept monotherapy trial [17]. Mean age is 53 yr, 74% are female, baseline HAQ is 1.6, and mean previous DMARD use 3.3. The model compares two arms: a sequence with etanercept monotherapy third-line and traditional DMARDs afterwards vs a sequence of traditional DMARDs only (no biologic agents).

The model focuses on the progression of HAQ disability score for the population over time. Most of the model parameters relate to the ‘average’ population (e.g. HAQ progression, HAQ improvement following ACR 20 response). There is a large number of possible pathways through sequences of DMARDs, and an average patient history over time is produced using an individual patient-level simulation model.

The treatment pathway is described in Fig. 1. The model cycle is 6 months in line with BSR guidelines for DMARD assessment. After the first period a patient may be either an ‘initial responder’ and so remain on treatment, or a ‘non-responder’ and so be switched to the next treatment in the sequence. The mean level of HAQ improvement for responders is quantified using the clinical trial data (etanercept) and published literature (other DMARDs). Initial responders remain on treatment for several 6-monthly cycles until subsequent longer-term withdrawal owing to lack of efficacy or adverse event. At the point of longer-term withdrawal, the patient is assumed to incur a worsening in HAQ score and is switched to the next DMARD in the sequence. If failure occurs on all DMARDs in the sequence we have then assumed that best care will be provided. The model takes a lifetime perspective, incorporating probability of death, which depends on RA life-tables and epidemiology relating HAQ improvement to reduced risk of mortality.

The 6-monthly trend in HAQ disability score (following initial response, non-response, ongoing success or withdrawal) is simulated for 10 000 patients. HAQ scores are converted to QALYs using published regression of HAQ vs EuroQol (EQ-5D)-derived utility. Drug and monitoring costs are included. Other healthcare costs are estimated using published regression of HAQ scores against healthcare costs. The 10 000 simulated patient results provide the average QALYs and costs for each arm and hence an incremental cost per QALY for etanercept.

Exemplars in the DMARD sequence

To provide exemplars of the treatments in the sequence, we analysed both the UK Early Rheumatoid Arthritis Study (ERAS) data (Dr A. Young, personal communication) and a commercially available electronic general practice database (DINLINK, Compufile). The most common DMARD therapies were methotrexate (first line) and sulphasalazine (second line). Patients entering the model are assumed to have failed these previously. The most common third-line therapy was intramuscular gold, fourth line was leflunomide and fifth line was cyclosporin in combination with methotrexate. Discussion with clinical experts confirmed the use of these treatments as reasonable exemplars of DMARDs A, B and C in third, fourth and fifth line, respectively.

Fig. 1. Conceptual model of clinical pathways.
Table 1. Parameter values used in model central estimate

<table>
<thead>
<tr>
<th></th>
<th>Etanercept</th>
<th>DMARD A</th>
<th>DMARD B</th>
<th>DMARD C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial HAQ improvement (i.e. negative)c</td>
<td>-0.8421 [17]</td>
<td>-0.43 [30]</td>
<td>-0.524 [31–33]</td>
<td>-0.3531 [25]f</td>
</tr>
<tr>
<td>6-monthly HAQ progression on treatment</td>
<td>0.0075 [19,34]</td>
<td>0.017 [34]</td>
<td>0.017 [34]</td>
<td>0.017 [34]</td>
</tr>
<tr>
<td>Total yearly cost (£)</td>
<td>8594</td>
<td>553</td>
<td>666</td>
<td>2038</td>
</tr>
</tbody>
</table>

aCalculated from patient-level trial analysis. The number of responders at 3 months and 6 months.

bThe 37% estimate for the first exemplar DMARD is taken from table 4 of Anderson using mean response rate for patients with disease duration > 5 (n=473). This is in line with Anderson table 1 showing Gold ACR 20 at 44% (n = 18) [28] or 33% (n = 15) [29].

cMost trials of the traditional DMARDs simply report the mean HAQ. To estimate the HAQ change for ACR 20 responders we have used evidence on the level of HAQ change for responders in the Moreland trial (–0.8421) compared with HAQ change for non-responders (–0.3693) and the subsequent ratio (2.281) to derive the following formula: HAQ change for ACR 20 responders = mean HAQ change (% responders + % non-responders/2.28). We have also adjusted this data to take account of differences in the mix of disease durations involved in different trials. This adjustment is based on the Munro study (Table 2). Patients with duration 0–2 yr have mean HAQ change in their first year = –0.82, compared with –0.49 for patients with disease duration over 2 yr, giving a ratio of 0.598, which is used to derive the formula: mean HAQ for >2 yr = mean HAQ change (% under 2 yr × 0.589 + % over 2 yr).

dTaken from Fig. 1: group 3.

eCalculated from patient-level trial data [17]. For other DMARDs, only the mean HAQ change (including non-responders) is published, so the HAQ score improvement for ACR 20 responders must be estimated. First, the mean HAQ change is adjusted to account for patient mix because trials of other DMARDs have much higher proportions of short disease duration cases. The adjustment uses evidence on the difference between achievable HAQ improvement in patients with >2-yr disease duration compared with patients under 2 yr (mean HAQ improvement is 59.8% lower in long disease duration cases) [30]. Second, to estimate HAQ improvement for ACR 20 responders only, a formula based on the differences observed in the etanercept study was used. In that trial ACR 20 responders had an improvement in HAQ that was 2.28 times higher than non-responders [17].

Clearly, there are dozens of possible alternative sequences, and the treatments chosen in practice are determined by individual clinicians and patients, rather than strict protocol or guidelines. The use of these particular exemplars was driven partly by data on their use in practice, and partly by the availability of trial and other evidence to estimate their response rates, HAQ improvement, cost and other model parameters. Other DMARDs could be modelled and sensitivity analysis examines different response rates, the order of their use and the number of DMARDs in the sequence. Steroids are not explicitly modelled because they are very low in cost, and because normal use is alongside DMARDs rather than as alternatives (in the etanercept monotherapy trial [17] 60–80% of patients had concomitant corticosteroids). Table 1 summarizes the data used for each DMARD and the relevant evidence sources [17, 19, 21, 24–34].

Initial response to therapy

The BSR guidelines suggest withdrawal at 3 months if response to anti-TNF is not achieved. For etanercept, the model explicitly examines percentage withdrawal at 3 months and then percentage withdrawal between 4 and 6 months, providing transparent analysis of drug and monitoring costs, response and HAQ improvements. DAS 28 is considered the UK measure of response [22], but comparative data on the DAS 28 for etanercept and other DMARDs are unavailable. The phase III study of etanercept vs placebo measured the ACR responses [17]. In the model, ACR 20 data are used to estimate the response rates for each treatment. It has been shown that DAS 28 and ACR 20 are similar in etanercept recipients [35, 36]. Response data were obtained from individual patient trial data for etanercept [17] and from published literature for other DMARDs (Table 1). It is well known that disease duration significantly affects response to treatment with DMARDs [21]. Therefore, the patient characteristics from identified studies were compared with the pivotal etanercept study in an attempt to identify trials that enrolled similar patients. Where comparable studies were not available, ACR 20 response was assumed to be 35%, using published meta-analysis of patients with >10-yr disease duration [21].

Initial HAQ response

Table 1 shows the initial HAQ score improvement used for ACR 20 responders to the different DMARDs. For etanercept, the source is individual patient-level trial data [17]. For other DMARDs, only the mean HAQ change (including non-responders) is published, so the HAQ score improvement for ACR 20 responders must be estimated. First, the mean HAQ change is adjusted to account for patient mix because trials of other DMARDs have much higher proportions of short disease duration cases. The adjustment uses evidence on the difference between achievable HAQ improvement in patients with >2-yr disease duration compared with patients under 2 yr (mean HAQ improvement is 59.8% lower in long disease duration cases) [30]. Second, to estimate HAQ improvement for ACR 20 responders only, a formula based on the differences observed in the etanercept study was used. In that trial ACR 20 responders had an improvement in HAQ that was 2.28 times higher than non-responders [17].

Long-term HAQ progression

For responders to DMARDs, evidence for long-term HAQ progression, particularly in methotrexate failures receiving third- or fourth-line treatments, is scarce. A systematic review of numerous DMARD therapies in a range of functional classes and duration of treatment gave an average annual HAQ progression of 0.034 [34]. The model therefore assumes a slight progression of disability over time even whilst patients are responding to treatment (Fig. 2).

In the long-term open-label study of etanercept, the initial HAQ improvement is maintained for at least 4 yr [37]. The evidenced relationship between the HAQ score and radiological progression was used to assess variation in HAQ progression for responders to biologics [38]. Supportive evidence comes from analysis on radiological progression between etanercept and methotrexate. Radiological progression as measured by modified Sharp scores for etanercept was 44% of the rate of increase for methotrexate [19]. An annual HAQ progression rate of 0.015 (0.034 × 44%) was used in the central estimate.

For periods of non-response to treatment, the average patient’s HAQ progression is higher because the risk of irreversible radiological damage is far greater. Ideal data on this would consist of a long-term observational study on HAQ progression, in patients who had already failed at least two DMARDs and who were also resistant to other DMARDs including the new biologic therapies. Such data will take several years to obtain. As a proxy, the base case analysis uses the mean annual HAQ progression rate for patients who were functional grade III + IV in the ERAS study.
i.e. mean annual progression of 0.13 \((n = 113)\) [6]. As a sensitivity analysis, the model uses ERAS study patients who have previously failed two DMARDs no matter what their disease severity, mean annual progression of 0.065 \((n = 145)\) (A. Young, personal communication).

**Longer-term withdrawal**

Beyond the first 6 months of therapy, a longer-term withdrawal probability is estimated for each DMARD. This combines both loss of efficacy and adverse events as the reason for withdrawal because they are not reported separately in the literature. For etanercept, a study based on clinical practice in Sweden showed an annual withdrawal of 8.3\% [26]. The withdrawal rates and sources for the other DMARDs are shown (Table 1).

**HAQ score after withdrawal**

Some evidence exists to quantify the effect of withdrawal on HAQ disability score. A small study (44 successfully responding patients withdrawn from etanercept after 85 days) suggests that on average the components of the HAQ score worsen substantially after withdrawal. Figure 3 shows that typically around 85\% of the improvement in the HAQ dimensions has been lost within 28 days [Immunex Clinical study report protocol 16.0009, unpublished]. An older but much larger study of 440 patients receiving gold [30] suggests that patients withdrawing after a period of response return to their original HAQ disability. One possibility, therefore, is to assume that patients return to their original HAQ disability immediately after withdrawal, though this may be over-optimistic since it implies no underlying progression of disease during the treatment phase. We have been more conservative, assuming that...
following withdrawal, the HAQ score would immediately worsen by an exactly equivalent amount to the initial improvement. The HAQ score following withdrawal is therefore slightly higher than the baseline HAQ score because, even during the successful treatment phase, there has been a gradual progression in HAQ score (e.g. 0.034 per annum for the DMARDs). An example of a patient’s trend in HAQ is shown in Fig. 4.

**Quality of life**

A suitable single index for health-related quality of life, where 1 is equivalent to full health and 0 is equivalent to death, was not measured in the etanercept trial. This information is needed to calculate QALYs. In the model it was assumed that the HAQ disability score is directly related to quality of life. The evidence review revealed four separate studies presenting the relationship between HAQ and utility [13, 39–41]. The variation in results was small and the median relationship of \[\text{[utility change} = \frac{0.86}{0.20} \times \text{(HAQ score change)}\] was used in the primary analyses.

**Costs**

Costs of drugs/monitoring were examined for each treatment (Table 1). The drug costs were derived from current list prices [42], whilst monitoring was estimated by costing BSR guidelines [43, 44].

Other direct healthcare costs, such as general practitioner, out-patient care and hospitalization, were also examined. Evidence from the US [15] demonstrated a strong correlation \(R^2 = 0.9146\) between HAQ score and other direct healthcare costs such as physician visits, surgical procedures and hospital admissions. This was confirmed by a Swedish study [13]. We converted both studies to 2000 UK currency using the purchasing parity index and inflation [45, 46]. Both gave an almost identical linear relationship of £860 per annum increase in direct healthcare costs per point worsening in HAQ disability score. In the model, therefore, the difference between the two comparator HAQ score trends is converted into a difference in direct healthcare costs, i.e. worse HAQ scores generate pro rata higher direct costs.

The central analysis includes only drug, monitoring and other direct healthcare costs. In a sensitivity analysis, we also examined the home help, residential and nursing home care costs, which were excluded from the US study of direct costs [15]. UK evidence on RA costs showed that home help accounted for 10%, and care in nursing homes 21.8%, implying that direct costs represent around 68.2% [11]. As a broad illustration on home help, and nursing home care costs, we assumed that they were also proportional to HAQ scores and so uplifted the slope of the cost/HAQ relationship (by 100%/68.2%) to an increase of £1261 per annum per point worsened in HAQ.

A final sensitivity analysis examines improved economic productivity to society through maintained employment. Data from Sweden shows the proportion of each HAQ group in productive employment [13]. The age/sex mix of the etanercept monotherapy trial population [17] (broadly equivalent to the Kobelt cohort [13]) combined together with UK wages from the New Earning Survey 2000 from the Office of National Statistics was used to compute the average wage for the model cohort (£18 096 at year 2000 prices). If the costs of lost productivity owing to worsening HAQ scores are combined with the direct NHS healthcare costs, and the home and residential care costs, then we estimate a societal cost increase of £3434 per annum per deterioration in HAQ point.

**Mortality**

Mortality risk is higher than average for patients diagnosed with RA. We constructed age/sex mortality risks specific to RA by adjusting normal population mortality by a relative risk of 2.975 ([47], E. Yelin, personal communication). Sensitivity analysis tested lower rates. An overall life expectancy table was then calculated by combining the age/sex mix of the trial with these RA-specific mortality risks [48].

Higher mortality risk is also shown for patients with worse HAQ disability scores [9]. Most of this evidence is cross-sectional, but there is also evidence that improvement in HAQ scores over time results in mortality risk reduction. A UCSF (University of California, San Francisco) study examined 1156 patients over
TABLE 2. Summary of results for 10 000 simulated clinical pathways

| Patient pathway | 0.0–0.5 | 0.5–1.0 | 1.0–1.5 | 1.5–2.0 | 2.0–2.5 | 2.5–3.0 | 3.0–3.5 | 3.5–4.0 | 4.0–4.5 | 4.5–5.0 | 5.0–5.5 | 5.5–6.0 | 6.0–6.5 | 6.5–7.0 | 7.0–7.5 | 7.5–8.0 | 8.0–8.5 | 8.5–9.0 | 9.0–9.5 | 9.5–10.0 |
|-----------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| **EtP1** Cost  | 498.742 | 742.066 | 1085.873 | 1429.780 | 1773.686 | 2117.593 | 2461.499 | 2805.406 | 3149.313 | 3493.220 | 3837.127 | 4181.034 | 4524.941 | 4868.848 | 5212.755 | 5556.662 | 5899.269 | 6243.163 | 6587.060 | 6930.957 |
| ACR 20 response | 50.06%  | 50.06%  | 50.06%  | 50.06%  | 50.06%  | 50.06%  | 50.06%  | 50.06%  | 50.06%  | 50.06%  | 50.06%  | 50.06%  | 50.06%  | 50.06%  | 50.06%  | 50.06%  | 50.06%  | 50.06%  | 50.06%  | 50.06%  | 50.06%  |
| Mean total costs | 36 213 | 9199 | 27 014 | 1 420 | 679 | 32 335 | 25 877 | 29 199 | 32 335 | 35 473 | 38 436 | 41 400 | 44 364 | 47 328 | 50 292 | 53 256 | 56 220 | 59 184 | 62 148 | 65 112 | 68 076 |
| Costs offsets   | 1525 | 5030 | 3515 | 2990 | 2465 | 3940 | 3415 | 2890 | 2365 | 1840 | 1315 | 790 | 265 | 440 | 615 | 790 | 965 | 1140 | 1315 | 1490 | 1665 |
| Drug and monitoring costs (£) | 34 688 | 34 688 | 34 688 | 34 688 | 34 688 | 34 688 | 34 688 | 34 688 | 34 688 | 34 688 | 34 688 | 34 688 | 34 688 | 34 688 | 34 688 | 34 688 | 34 688 | 34 688 | 34 688 | 34 688 |
| Monitoring costs (£) | 2261 | 2261 | 2261 | 2261 | 2261 | 2261 | 2261 | 2261 | 2261 | 2261 | 2261 | 2261 | 2261 | 2261 | 2261 | 2261 | 2261 | 2261 | 2261 | 2261 |
| Mean QALY (10 000 patients) | 6.775 | 6.775 | 6.775 | 6.775 | 6.775 | 6.775 | 6.775 | 6.775 | 6.775 | 6.775 | 6.775 | 6.775 | 6.775 | 6.775 | 6.775 | 6.775 | 6.775 | 6.775 | 6.775 | 6.775 |
| Mean cost (10 000 patients) | 36 213 | 9199 | 27 014 | 1 420 | 679 | 32 335 | 25 877 | 29 199 | 32 335 | 35 473 | 38 436 | 41 400 | 44 364 | 47 328 | 50 292 | 53 256 | 56 220 | 59 184 | 62 148 | 65 112 | 68 076 |

Cumulative cost and QALYs discounted. HAQ at start of period. Costs include hospitalizations.

TABLE 3. Model results: central estimate

<table>
<thead>
<tr>
<th>Costs and benefits (discounted 6%, 1½%)</th>
<th>Etanercept + DMARD sequence</th>
<th>DMARD sequence</th>
<th>Incremental</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drug costs (discounted) (£)</td>
<td>32 427</td>
<td>2837</td>
<td>29 590</td>
</tr>
<tr>
<td>Monitoring costs (£)</td>
<td>2261</td>
<td>1332</td>
<td>929</td>
</tr>
<tr>
<td>Drug and monitoring costs (£)</td>
<td>34 688</td>
<td>4169</td>
<td>30 519</td>
</tr>
<tr>
<td>Costs offsets</td>
<td>1525</td>
<td>5030</td>
<td>– 3 505</td>
</tr>
<tr>
<td>Mean total costs (£)</td>
<td>36 213</td>
<td>9199</td>
<td>27 014</td>
</tr>
<tr>
<td>ACR 20 response (%)</td>
<td>50.06%</td>
<td>37.08</td>
<td></td>
</tr>
<tr>
<td>Mean QALYs</td>
<td>7.526</td>
<td>5 87 18</td>
<td>1 65 43</td>
</tr>
<tr>
<td>Cost-effectiveness (incremental cost per QALY) (£)</td>
<td>6 330</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Monte Carlo simulation samples whether the patient survives the 6-month period. If the patient dies at that point in the model, then the cumulative costs and QALYs are recorded and the model moves on to sample the lifetime of the next of the 10 000 patients (see Table 2). All costs have been discounted by 6% per annum and effectiveness by 1.5% per annum in line with NICE guidance [49].

**Results**

**Central estimate**

The results of the central estimate are shown in Table 3. Drug costs are estimated to be £30 395 higher in the etanercept sequence. These are partially offset by other healthcare cost savings (estimated at £30 395 higher in the etanercept sequence): $30 395$.
<table>
<thead>
<tr>
<th>Scenario</th>
<th>Change from central</th>
<th>Parameter change</th>
<th>Mean QALYs (disc)</th>
<th>Mean total cost (disc)</th>
<th>Cost per QALY</th>
<th>Change from central</th>
</tr>
</thead>
<tbody>
<tr>
<td>Central</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>HAQ/utility conversion</td>
<td>0.7700-0.1680×</td>
<td>1.40</td>
<td>£26 800</td>
<td>£19 126</td>
<td>+£796</td>
</tr>
<tr>
<td>2</td>
<td>HAQ/utility conversion</td>
<td>0.8500-0.2300×</td>
<td>1.77</td>
<td>£27 691</td>
<td>£15 166</td>
<td>-£1164</td>
</tr>
<tr>
<td>3</td>
<td>HAQ/utility conversion</td>
<td>0.8285-0.2344×</td>
<td>1.77</td>
<td>£27 909</td>
<td>£15 245</td>
<td>-£1085</td>
</tr>
<tr>
<td>4</td>
<td>Cost offsets per point HAQ score—no other healthcare costs</td>
<td>£0</td>
<td>1.62</td>
<td>£30 289</td>
<td>£18 670</td>
<td>+£2340</td>
</tr>
<tr>
<td>5</td>
<td>Cost offsets per point HAQ score—including nursing home care</td>
<td>£1261</td>
<td>1.66</td>
<td>£25 326</td>
<td>£15 273</td>
<td>-£1057</td>
</tr>
<tr>
<td>6</td>
<td>Cost offsets per point HAQ score—as 5 and including productivity costs</td>
<td>£4294</td>
<td>1.64</td>
<td>£12 733</td>
<td>£7787</td>
<td>-£8543</td>
</tr>
<tr>
<td>7</td>
<td>% increase in mortality per point change in HAQ</td>
<td>0.00%</td>
<td>1.26</td>
<td>£25 342</td>
<td>£20 143</td>
<td>+£3813</td>
</tr>
<tr>
<td>8</td>
<td>% increase in mortality per point change in HAQ</td>
<td>75.00%</td>
<td>1.98</td>
<td>£28 568</td>
<td>£14 439</td>
<td>-£1891</td>
</tr>
<tr>
<td>9</td>
<td>Relative risk of mortality due to RA</td>
<td>1</td>
<td>1.73</td>
<td>£27 904</td>
<td>£16 111</td>
<td>-£219</td>
</tr>
<tr>
<td>10</td>
<td>Relative risk of mortality due to RA</td>
<td>5.95</td>
<td>1.51</td>
<td>£25 353</td>
<td>£16 801</td>
<td>+£471</td>
</tr>
<tr>
<td>11</td>
<td>HAQ progression for non-responders = ERAS 2 DMARD failures</td>
<td>0.07</td>
<td>1.40</td>
<td>£26 126</td>
<td>£18 639</td>
<td>+£2309</td>
</tr>
<tr>
<td>12</td>
<td>HAQ progression equal for (a) biologics, (b) DMARDs and (c) non-responders</td>
<td>(a) = 0.03, (b) = 0.03, (c) = 0.03</td>
<td>0.75</td>
<td>£31 898</td>
<td>£42 384</td>
<td>+£26 054</td>
</tr>
<tr>
<td>13</td>
<td>HAQ Progression equivalent for biologics and DMARDs</td>
<td>(a) = 0.03, (b) = 0.03, (c) = 0.13</td>
<td>1.35</td>
<td>£26 910</td>
<td>£20 004</td>
<td>+£3674</td>
</tr>
<tr>
<td>14</td>
<td>Upper confidence interval on Enbrel Response Rate (from Moreland trial)</td>
<td>64%</td>
<td>2.11</td>
<td>£32 922</td>
<td>£15 587</td>
<td>-£743</td>
</tr>
<tr>
<td>15</td>
<td>Lower confidence interval on Enbrel Response Rate (from Moreland trial)</td>
<td>36%</td>
<td>1.14</td>
<td>£20 411</td>
<td>£17 976</td>
<td>+£1646</td>
</tr>
<tr>
<td>16</td>
<td>Upper confidence interval on Enbrel Response Rate (from Swedish observational study)</td>
<td>69%</td>
<td>2.29</td>
<td>£35 271</td>
<td>£15 408</td>
<td>-£922</td>
</tr>
<tr>
<td>17</td>
<td>Lower confidence interval on Enbrel Response Rate (from Swedish observational study)</td>
<td>51%</td>
<td>1.66</td>
<td>£27 237</td>
<td>£16 425</td>
<td>+£95</td>
</tr>
<tr>
<td>18</td>
<td>All DMARD response equivalent to highest</td>
<td>All = 48%</td>
<td>1.58</td>
<td>£26 575</td>
<td>£16 859</td>
<td>+£529</td>
</tr>
<tr>
<td>19</td>
<td>6 monthly withdrawal rates for DMARD A equal to Wolfe [62]</td>
<td>22%</td>
<td>1.60</td>
<td>£26 595</td>
<td>£16 612</td>
<td>+£282</td>
</tr>
<tr>
<td>20</td>
<td>DMARD sequence changed to DMARD B, DMARD C, DMARD A</td>
<td>All = 48%</td>
<td>1.58</td>
<td>£26 521</td>
<td>£16 774</td>
<td>+£444</td>
</tr>
<tr>
<td>21</td>
<td>Upper confidence interval on HAQ change for etanercept</td>
<td>-1.023</td>
<td>1.41</td>
<td>£26 960</td>
<td>£19 070</td>
<td>+£2740</td>
</tr>
<tr>
<td>22</td>
<td>Lower confidence interval on HAQ change for etanercept</td>
<td>-0.662</td>
<td>1.84</td>
<td>£26 692</td>
<td>£14 512</td>
<td>-£1818</td>
</tr>
<tr>
<td>23</td>
<td>Number of DMARDs in a sequence increased to 6; DMARDs A–B repeated</td>
<td>1.52</td>
<td>1.52</td>
<td>£26 390</td>
<td>£17 388</td>
<td>+£1058</td>
</tr>
</tbody>
</table>
As a result of reduced levels of disability and hence fewer events such as total joint arthroplasty. Also accounting for monitoring costs, the overall net cost difference between the two strategies is estimated at £27 014. In terms of health benefits, the etanercept strategy is estimated to provide an extra 2.73 life-years (undiscounted), which equates to 1.66 QALYs. The incremental cost per QALY gained is therefore £16 330. This result is in the region that has previously been accepted as cost-effective in the UK.

Sensitivity analyses

Scenario-based sensitivity analysis has quantified the uncertainty in the results by investigating the impact of alternative scenarios for the key parameters. Table 4 and Fig. 5 illustrate that for the majority of parameters varied, the cost per QALY is between £14 000 and £21 000, i.e. the result is relatively insensitive to changes in parameter values across their plausible range.

Importantly, changing the response rates for etanercept \{scenarios 14–17\} or HAQ changes for etanercept \{21, 22\} had little impact on the incremental cost-effectiveness ratios (ICER). Likewise, modifying the assumptions for DMARDs A–C, and the number of DMARDs in the sequence had little effect \{18–20, 23\}. Changing the estimates of mortality only changed the ICER by £4000.

The only scenario that improves the cost per QALY by an order of magnitude is the inclusion of employment productivity savings \{6\}. Using Swedish data on the rate of employment for different HAQ scores, combined with average UK wage rates (for under 65s only), we formed a broad estimate of the reduction in lost productivity. The scale of this was estimated at £10 000 to £12 000 per recipient of etanercept over their lifetime, resulting in a cost per QALY of under £10 000.

The only other parameters to affect the results concern mean 6-monthly increase in HAQ disability score. Scenarios \{11–13\} sequentially transform these variables from the base case, through more pessimistic scenarios concerning etanercept’s differential capacity to delay HAQ progression. Scenarios \{11, 13\} result in a cost per QALY in the range £19 000 to £20 000. The only scenario which produces a cost per QALY estimate over £30 000 \{12\} assumes HAQ progression on etanercept and DMARDs is no different from periods of non-response to therapy. That is, the cost per QALY estimate rises to around £42 000 only if one assumes that neither etanercept nor DMARDs have any effect on delaying underlying HAQ progression.

Good analytical practice usually involves varying all uncertain model parameters at the same time, producing a so-called probabilistic multi-way sensitivity analysis. However, such analysis is hugely computationally expensive in patient-level simulation models, requiring around 200 days running time for say 5000 multivariate simulations, each of which in its turn simulates 10 000 individual patients. Thus, the one-way sensitivity analyses, whilst not accounting for some possible interactions between parameters and therefore perhaps underestimating the overall uncertainty, remain the best description of uncertainty available at this time.

Discussion

The objective of this study was to assess the cost-effectiveness of etanercept monotherapy in accordance with BSR guidelines and, in particular, to quantify the cost per quality-adjusted life year (QALY) gained through the use of etanercept compared with current UK care. The method involved modelling clinical pathways, defining comparators and synthesis of evidence from a range of sources to quantify the model parameters. The results suggest that etanercept should be considered cost-effective when compared against traditional DMARDs in the UK setting, and the sensitivity analyses suggest that this conclusion is relatively robust to the model assumptions including alternative exemplar DMARD sequence comparators.

The main limitations of the study relate to the data available. The ideal would be a very long-term, large sample size, randomized study examining the efficacy, effectiveness and resource use associated with etanercept vs traditional care for the patient group which has failed methotrexate plus another DMARD. Such a study does not exist and the task is to assemble an evidence base for the analysis from a range of sources. First, for response rates, the review of trials and observational data has provided some evidence for the exemplar DMARDs. The alternative is to use FIG. 5. Sensitivity analysis scenarios presented on the cost-effectiveness plane.

Good analytical practice usually involves varying all uncertain model parameters at the same time, producing a so-called probabilistic multi-way sensitivity analysis. However, such analysis is hugely computationally expensive in patient-level simulation models, requiring around 200 days running time for say 5000 multivariate simulations, each of which in its turn simulates 10 000 individual patients. Thus, the one-way sensitivity analyses, whilst not accounting for some possible interactions between parameters and therefore perhaps underestimating the overall uncertainty, remain the best description of uncertainty available at this time.

The main limitations of the study relate to the data available. The ideal would be a very long-term, large sample size, randomized study examining the efficacy, effectiveness and resource use associated with etanercept vs traditional care for the patient group which has failed methotrexate plus another DMARD. Such a study does not exist and the task is to assemble an evidence base for the analysis from a range of sources. First, for response rates, the review of trials and observational data has provided some evidence for the exemplar DMARDs. Much of the evidence used on response rates is based on the ACR 20 criteria from trials. The argument that the ACR 20 criteria are not often used in clinical practice is a valid criticism, but these data provide the only evidence for response rates from relevant trials. DAS 28 measurement may become a future gold standard. The alternative is to use
observational data on patient withdrawals as the source for response data. The observational evidence so far for etanercept suggests that the trial-based ACR 20 response rates are in line with response in clinical practice [24]. The analytical solution to this question has been to undertake a series of sensitivity analyses based on response rate confidence intervals both for etanercept and the comparator exemplar DMARDs.

Second, there are several sources available for the longer-term rate of disability progression. The base case estimate using rates for patients who are functional class III and IV has been subjected to very important sensitivity analysis using alternative sources including the meta-analysis of patients on traditional DMARDs [34]. Third, there is limited evidence on HAQ score worsening following withdrawal. The etanercept data clearly show this phenomenon to be important to include in a model [ImmuneX Clinical study report protocol 16.0009, unpublished]. There are arguments that the HAQ score will probably start to worsen prior to patient withdrawal (although there is no data available with which to quantify this) and further that there may be some residual benefit for a time following withdrawal. Our simple assumption (instant worsening by the same amount as the initial improvement) may be conservative. Finally, and unavoidably, the data used in the model relating HAQ disability to healthcare costs and HAQ disability to patient utility (and hence QALYs) are sourced from observational data sets which themselves were separate from the trial data on etanercept.

The advantages of the model approach taken in this study include that it reflects the cost and benefit involved due to patients being withdrawn from etanercept therapy if they are non-responders. It models the sequential use of DMARDs as the relevant comparator and incorporates disease progression, the relevant costs and the outcome measures of interest to the NICE decision maker, i.e. costs and QALYs gained over a lifetime perspective. Technically, a patient-level simulation (rather than cohort) model was required because several variables (e.g. response rates, mortality rates) are time dependent. The probability of events depends on both time and previous events in the patient’s history and decision tree and Markov formulations soon became ‘decision bushes’ with exponentially expanding calculations. The model concentrates on HAQ change from baseline because the health economic question is to quantify the additional health benefit etanercept provides. One criticism of the model formulation is that, in clinical practice, the goal of treatment may be to achieve a particular absolute level of HAQ score (e.g. under 1.0). An alternative formulation using absolute HAQ was considered, but the level chosen would be arbitrary since no firm guidelines exist; the data themselves are very seldom reported; and most importantly, the study had to measure, not how many patients achieve a particular level of HAQ, but how much improvement in disability, and hence QALY, is gained.

Several other published models examining cost-effectiveness of RA treatments have focused on short-term costs and benefits whilst ignoring cost offsets or mortality [50], examined medium-term benefits only [13], lack the generic QALY outcome measure [51–53] and are not adequate for a UK NICE appraisal. Two recently published models address the cost-effectiveness of infliximab [54, 55]. Both consider a policy of treatment for 1 yr and then withdrawal vs traditional care and base the analysis of this short-term impact on trial evidence (ATTRACT) [56]. Following treatment withdrawal at 1 yr, observational evidence on disease progression is used to model the longer-term rate of disability progression and hence costs and QALYs. There are several similarities with the methods presented here, particularly the use of observational data on the HAQ/cost/utility relationships and use of adjusted life-tables to examine longer-term mortality. The first difference relates to the formulation of the research question—our study examines the use of etanercept over the long term with the consequent higher costs, explicitly modelling the likely withdrawal rates, rather than assuming 100% withdrawal at 1 or 2 yr.

The other main difference relates to the prediction of longer-term HAQ disability progression after withdrawal—our study assumes instantaneous HAQ worsening by the same amount as the initial improvement and for the longer term then uses evidence on annual HAQ progression based on observational data. The other two models similarly use observational data on HAQ progression long-term (one using long-term transition data on British early RA patients, the other using US data on transitions between HAQ groupings and DMARD agents). Neither explicitly models an instantaneous HAQ worsening following withdrawal from infliximab. One other published analysis has focused on both etanercept and infliximab cost-effectiveness in the UK. This model [58] is similar to our own in examining therapy switching and comparison against traditional DMARDs. The main differences were its lack of explicit modelling of response vs non-response, the examination only focusing on health benefits gained (from symptom relief) during the period of receiving etanercept treatment, hence ignoring effects on delayed disease progression and the consequent longer-term costs and benefits and hence a higher estimate for cost per QALY gained.

The analysis presented here conforms to the recently produced OMERACT discussion document on conducting economic evaluations in rheumatology [59]. In particular, outcomes are both clinically relevant (number of responders, improvement in HAQ) as well as universally comparable (e.g. QALYs via EQ-5D), the comparator is current standard care including treatment sequences, the UK decision-maker’s time horizon is beyond the trial and modelled using observational data backed up by detailed sensitivity analyses, the patient group is clearly identified and direct healthcare costs are examined separately from other societal/productivity impact.

The cost-effectiveness model was submitted to NICE in 2001, and the evidence base and analysis independently reviewed in detail. The NICE Appraisal Committee decided to rerun sensitivity analyses using ‘different estimates of clinical effectiveness and rates of disease progression’, producing ‘an incremental cost-effectiveness ratio in the region of £27 000 to £35 000 per QALY’ [59]. It is not clear exactly which variables were changed upon what evidence, but our sensitivity analyses (Table 3) show that only when it is assumed that patients who respond to treatment have the same rate of disability progression as patients who do not respond does the incremental cost-effectiveness increase above £30 000. It is clear, however, that the model was an important tool for NICE in forming its conclusion that ‘Etanercept…[59]…recommended for the treatment of adults who have continuing clinically active rheumatoid arthritis that has not responded adequately to at least two disease-modifying anti-rheumatic drugs, including methotrexate (unless contraindicated).’ [59].

Adaptation of the cost-effectiveness model to other countries is clearly possible but the particular results presented here relate to the UK context (use of etanercept monotherapy in patients who have failed at least two DMARDs, including methotrexate) and may not be generalizable. Costs may vary in other settings. The discount rates used are important as both benefits and costs are derived over a period of several years and UK guidance differs from several other countries. Although the relationship between disability and healthcare cost appeared similar from US and Swedish data, other countries may be different. The perspective required may also differ, with a lifetime perspective or the inclusion of wider analysis of cost (e.g. employment) considered unrealistic in some countries and absolutely necessary in others.

There are implications from this analysis for future research. In particular, refinement of the estimates on response rates, withdrawal rates, HAQ improvement for responders, HAQ worsening following withdrawal and long-term disability progression could be valuable. In the UK, the ‘BSR Biologics Register’ [60], an observational study with a control cohort, will record clinical outcomes, disability, quality of life (short form, SF-36), treatments, adverse events and interventions such as surgery. The registry data...
will be able to provide further evidence to refine the model and update the analysis. On the benefit measurement side, there are suggestions that EQ-5D derived utilities may better reflect severity of RA than the SF-36 instrument [41, 61]. The modelling also suggests other data collection priorities on economic-related issues, nursing home admission rates/resource use and prospective data on changes in employment status.

The general framework of the model can certainly be adapted to analysis of other health economic questions in RA. The possibility exists of examining alternatives such as sequential use of biologics, which was not recommended by NICE but may become important in future, although this would require improvement on the currently limited data on response rates in patients who have previously failed or withdrawn from a biologic therapy. The task of comparing different biologic agents against each other needs careful consideration in the absence of head to head trials. The model could be used to undertake an initial assessment with further refinements adjusting for trial differences and with assumptions on issues such as the initial HAQ improvement separately for responders, which is not usually reported in the main trial publications. The analysis can also be used and developed for other questions in other countries and settings.

In conclusion, for adults in the UK, the cost-effectiveness analysis results suggest that etanercept is cost-effective when compared against non-biologic agents. The National Institute for Clinical Excellence has accepted that etanercept is cost-effective and recommended its availability for use in patients who have failed at least two DMARDs. Our model was an important component of that decision. In future the BSR Biologics Registry should provide useful evidence for updating the model. The framework presented here is capable of making between-biologic comparisons, the impact of different countries’ costs or comparators, and a wide range of other cost-effectiveness questions in rheumatoid arthritis.

References

2. Lane S, Scott D. Drug treatment of active rheumatoid arthritis. Prescriber 5 March 1999;79–95.


60. British Society for Rheumatology Biologics Register http://www.arc.man.ac.uk/webbiologicsreg.htm


75. British Society for Rheumatology Biologics Register http://www.arc.man.ac.uk/webbiologicsreg.htm
