RHEUMATOLOGY

Rheumatology 2011;50:iv10–iv18
doi:10.1093/rheumatology/ker240

An overview of models used in economic analyses of biologic therapies for arthritis—from current diversity to future consensus

Jason Madan[1], Anthony E. Ades[1] and Nicky J. Welton[1]

Abstract

A number of cost-effectiveness models have been developed with the aim of providing guidance for decision making on biologic therapies for the management of inflammatory joint disease. The findings of these analyses can differ markedly, and these differences can undermine the credibility of such models if unexplained. To allow differences between models to be identified more easily, we define six components common to all models—initial response, longer term disease progression, mortality, quality-adjusted life year estimation, resource use and the selection and interpretation of data. We give examples of divergent approaches taken by model structures to the same issue, and explore the impact of divergence on model results, with particular focus on two models that have reported substantially different estimates for the cost-effectiveness of third-line etanercept *vs* conventional DMARD. The sensitivity of results to a particular assumption made in a model will depend on the decision problem and assumptions made elsewhere in the model, highlighting the importance of guidance throughout model development. To some extent, guidance from bodies such as the National Institute of Health and Clinical Excellence can be used to determine which approach should be preferred where models differ. However, there is a pressing need for clinical input and guidance before consensus can be reached on the most credible model(s) to use for decision support.

Key words: Economic model, Cost-effectiveness analysis, Biologic therapies.

Introduction

Biologic therapies have considerable impact in retarding the progression of chronic inflammatory joint diseases. However, they also add substantially to the costs of managing these diseases. Decision-analytic models have been widely used to assess whether the use of biologic therapies is cost-effective in various situations. Their use is particularly common in the UK, where manufacturers and academic evidence review groups have carried out cost-effectiveness analyses to support the production of guidance by bodies such as the National Institute of Health and Clinical Excellence (NICE). Table 1 gives a list of NICE appraisals of biologics for the treatment of RA or PsA, most of which have included at least one economic model. Decision-analytic models have also been used to assess the cost-effectiveness of biologic therapies in the USA [1], Sweden [2] and the Netherlands [3]. While decision-analytic modelling can provide useful information for policy-makers on the long-term consequences and cost-effectiveness of biologic therapies, a concern is the discrepancy between results reported in different modelling exercises. For example, NICE technology appraisal TA130 included model-based estimates of the incremental cost-effectiveness ratio (ICER) of etanercept *vs* DMARD for third-line management of RA that ranged from £18 000 to £93 000 per quality-adjusted life year (QALY) [4]. Such differences, if unexplained, reduce the credibility of modelling exercises, and thus their usefulness as an aid to decision-making.

The aim of this article is to shed light on the source of these discrepancies by identifying the differences in model structure and data sources that lead to them. To do this, we highlight key aspects of inflammatory joint disease that models are required to represent. In each area, we describe the possible assumptions that could be made when constructing models, give examples of different approaches used and explore the impact of these approaches

[1]School of Social and Community Medicine, University of Bristol, Bristol, UK.

Submitted 2 February 2011; revised version accepted 7 June 2011.

Correspondence to: Jason Madan, School of Social and Community Medicine, University of Bristol, Canynge Hall, Whatley Road, Bristol BS8 2PS, UK. E-mail: jason.madan@bristol.ac.uk

© The Author 2011. Published by Oxford University Press on behalf of the British Society for Rheumatology. All rights reserved. For Permissions, please email: journals.permissions@oup.com

Downloaded from https://academic.oup.com/rheumatology/article/50/suppl_4/iv10/1833100 by guest on 20 March 2024

Table 1 List of published NICE technology appraisals (as at January 2011) of biologic treatments for PsA or RA

Appraisal	Date	Type	Target population	Treatments and UK manufacturer(s)	Recommendation
TA199	10 August	MTA	Patients with active and progressive PsA who have not responded to at least two DMARDs	Adalimumab (Abbott), Etanercept (Wyeth), Infliximab (Schering-Plough)	Adalimumab, etanercept and infliximab are recommended as possible treatments
TA198	10 August	STA	Patients with moderate to severe RA who have failed at least one biologic	Tocilizumab (Roche)	Tocilizumab is recommended as a possible treatment where rituximab is not suitable
TA195	10 August	MTA	Patients with severe RA who have failed at least one biologic	Abatacept (Bristol-Myers-Squibb), Adalimumab (Abbott), Etanercept (Wyeth), Infliximab (Schering-Plough), Rituximab (Roche)	Rituximab is recommended as a possible treatment Adalimumab, etanercept, infliximab and abatacept are recommended as possible treatments where rituximab is not suitable
TA186	10 February	STA	Patients with severe active RA after the failure of at least two DMARDs	Certolizumab pegol (UCB)	Certolizumab pegol is recommended as a possible treatment
TA141	8 April	STA	RA patients (MTX-naive or after the failure of DMARDs)	Abatacept (Bristol-Myers-Squibb)	Abatacept is not recommended (within its marketing authorization)
TA130	7 October	MTA	RA patients who have failed at least two DMARDs	Adalimumab (Abbott), Etanercept (Wyeth), Infliximab (Schering-Plough), Anakinra (Amgen)	Adalimumab, etanercept and infliximab are recommended as possible treatments
TA126	10 August	STA	Patients who have severe RA and have failed other DMARDs including at least one other biologic	Rituximab (Roche)	Rituximab is recommended as a possible treatment
TA125	7 July	STA	Patients with active and progressive PsA who have not responded to at least two DMARDs	Adalimumab (Abbott)	Adalimumab is recommended as a possible treatment
TA104	6 July	MTA	Patients with active and progressive PsA who have not responded to at least two DMARDs	Etanercept (Wyeth), Infliximab (Schering-Plough)	Etanercept is recommended as a possible treatment. Infliximab is recommended as a possible treatment where etanercept is unsuitable
TA72	3 July	STA	RA patients who have failed MTX	Anakinra (Amgen)	Anakinra is not recommended as a possible treatment
TA36	2 March	MTA	RA patients who have failed at least two DMARDs (including MTX)	Etanercept (Wyeth), Infliximab (Schering-Plough)	Etanercept and infliximab are recommended as possible treatments

STA: single technology appraisal; MTA: multiple technology appraisal.

on model results. We then look at how guidance issued by bodies such as NICE can identify preferred modelling approaches, and where a need for consensus remains, based on guidance from those with the appropriate clinical expertise.

Components of RA economic models

The overall representation of arthritis and its treatment in cost-effectiveness models is that of a degenerative disease that persists from onset till death. Over that time, patients will receive a sequence of treatments, with clinicians switching treatments when the existing intervention loses efficacy or causes unacceptable side effects. Models estimate the incremental cost-effectiveness of introducing biologic treatments into the sequence. Modelling is a process that involves creating a simplified version of a complex real-world situation, which is amenable to analysis while maintaining the key aspects of the

Downloaded from https://academic.oup.com/rheumatology/article/50/suppl_4/iv10/1833100 by guest on 20 March 2024

situation from the point of view of the decision-maker. In the case of decision-making around biologic therapies, this involves making choices that can be organized into six categories:

- initial response to treatment;
- long-term response where treatment is initially successful;
- relationship between disease severity and mortality risk;
- translation of the health impact of treatments into a quality-of-life measure;
- resource use;
- selection and use of data to inform model parameters.

The choices involved in each category are described further below, with illustrative examples of models that make alternative assumptions for each choice. Where possible, we focus on two models—the model reported in Brennan *et al*. [5] (hereafter referred to as Sheffield 2004) and the model developed by the assessment group for NICE TA130 [4] [hereafter referred to as Birmingham Rheumatoid Arthritis Model (BRAM) 2007] These models are chosen since (i) they produce markedly different estimates of the ICER of etanercept *vs* conventional DMARD as a third-line therapy; (ii) they are reported in sufficient detail to make comparison possible; and (iii) both modelling groups present their work in greater detail elsewhere in this supplement [6, 7]. Table 2 summarizes the differences between the two model structures. Other models are used as examples where this further aids understanding of model choices and their implications.

Initial response to treatment

Clinical guidelines recommend that treatment with TNF-α inhibitors should only be continued if there is an adequate response in the short term (3–6 months) [16]. This guidance raises two issues for the structure of the model. The first relates to the choice of measure used to capture the initial impact of treatment. In RA, for example, there are several measures for (changes in) disease severity. The HAQ is a family of questionnaires widely used to measure the functional capacity of patients [17]. Several models represent the initial response to treatment on this scale, for example, BRAM 2007, and the manufacturer's submission by Wyeth for NICE TA130 [4].

Composite measures exist that combine the HAQ with clinician-reported outcomes. The ACR measure [18] has been widely used in clinical trials of biologic treatments. It combines the HAQ with a count of swollen/tender joints. There are three levels of response in common use—ACR20, ACR50 and ACR70, which represent a 20, 50 or 70% improvement in the ACR score. A number of models, including Sheffield 2004 [5], use this measure to categorize patients as responders or non-responders. An alternative measure, the DAS, is also based on a combination of swollen/tender joint counts and patient-reported outcomes [19]. It is a continuous outcome measure, although the European League against Rheumatism (EULAR) provides a scale that converts the DAS into three response levels (none, moderate or good) [20]. This measure was used by the British Society of Rheumatology submission to NICE TA130 [4]. The EULAR and ACR response levels have been shown to perform comparably in trials [20].

HAQ, ACR, DAS and other commonly used indices measure different aspects of arthritis. They are correlated, but not interchangeable. It is possible to map between measures—Sheffield 2004 translates ACR20 into change in HAQ, for example. If key studies differ in the measure used for short-term response then, without such mappings, it will not be possible to draw on the full evidence base. The choice of primary measure will only have a marked impact on model results where the correlation between measures is weak, assuming that data sources and definitions of adequate response are comparable.

The second issue relates to how decisions to maintain or switch treatment relate to the initial treatment effect. A treatment may be abandoned at an early stage because it provokes a severe adverse event, or it has not achieved a sufficient initial improvement in the disease. Models may represent both events—BRAM 2007 models adverse events and lack of efficacy as separate events [4]. Alternatively, models may take the approach of, e.g. Sheffield 2004, in which treatment withdrawal is reported without separating out reasons for withdrawal [5]. It is unlikely that pooling treatment withdrawal, rather than analysing non-response and adverse events separately, will have a dramatic effect on model results. The exception might be in estimating the incremental cost-effectiveness of one biologic against another—if costs and efficacy are similar, then different side-effect profiles may drive results.

Assumptions relating to initial response can have a marked effect on the predicted proportion of patients who will stay on a treatment beyond the short term. Sheffield 2004, for example, assumes that the percentage of patients who are withdrawn from etanercept at 6 months is 50%, while BRAM 2007 assumes this to be 7%. The difference is partly due to the data used, but different definitions of adequate response are more important (Table 2). Sheffield 2004 assumes patients continue treatment if, and only if, ACR20 is achieved, whereas BRAM 2007 uses the continuation rate observed in their chosen study, which was markedly higher than the ACR20 response rate.

Longer term response to treatment, if continued

Here, there are several questions to consider. The first relates to time to treatment withdrawal. At some point a decision will be made to cease the current treatment, due to either an adverse reaction or a loss of efficacy. While the time to this event is a parameter that will have a clear influence on cost-effectiveness, available studies may well lack the follow-up to provide conclusive evidence of its value. To bridge this gap, models may follow an approach such as that used by Kobelt *et al.* [2] in their base case, where they assume that treatment (etanercept) is withdrawn after 2 years, which is the follow-up

Downloaded from https://academic.oup.com/rheumatology/article/50/suppl_4/iv10/1833100 by guest on 20 March 2024

TABLE 2 A comparison of assumptions made by two models (Sheffield 2004 and BRAM 2007) when estimating the cost-effectiveness of etanercept *vs* conventional DMARD for third-line treatment of RA

	Sheffield 2004 [5]	BRAM 2007 (late RA case) [4]
Results (etanercept *vs* conventional DMARD after the failure of two previous DMARDs)		
Conventional DMARD	Gold	LEF
Incremental cost-effectiveness	£16 000/QALY (£30 000 and 1.65 QALYs per patient)	£93 000/QALY (£43 000 and 0.46 QALYs per patient)
Initial response to treatment		
Choice of measure	ACR20	HAQ
Threshold for efficacy	ACR20	Observed withdrawal in study
Response to etanercept		
Failure rate at 6 months	50% (lack of efficacy)	7% (6% lack of efficacy, 1% toxicity)
Source	Trial of etanercept *vs* placebo [8]	Swedish routine data [9] NB: ACR20 response on etanercept was 60% in this data set
Mean HAQ change in responders	Reduction of 0.84	Reduction of 39%
Source	Trial of etanercept *vs* placebo [5] Mean HAQ of participants at start of treatment 1.7	Three trials of etanercept *vs* placebo [4] Mean HAQ of participants at start of treatment 1.6–1.8
Response to DMARD comparator		
Failure rate at 6 months	63% (lack of efficacy) (the failure rate assumed for LEF was 63%)	43% (22% lack of efficacy, 20% drug toxicity, 1% other). [the failure rate assumed for gold was 41% (23% efficacy, 18% toxicity)]
Source	Two trials of gold *vs* MTX [10]	Swedish routine data [9]
Mean HAQ change in responders	Reduction of 0.43 (for LEF, reduction of 0.52)	Reduction of 47% (for gold, reduction of 39%)
Source	Two trials of gold *vs* MTX [10] (ratio of HAQ change in responders and non-responders assumed to be the same as for etanercept)	Trial of LEF *vs* MTX [11] Mean HAQ at start of treatment 1.03
Long-term progress if treatment continued		
Long-term response on etanercept		
Mean time-to-failure	12 years (exponential)	15 years (Weibull)
Source	Swedish routine data with 20 months follow-up [12]	Swedish routine data with 20 months follow-up [9]
HAQ change on treatment	0.034 per year	0.035 per year
Source	Trial of etanercept *vs* MTX [13]	Assumed same as conventional DMARDs
Long-term response on DMARD comparator		
Mean time-to-failure	5 years (exponential)	NA (redacted in evaluation report)
Source	Meta-analysis of 56 treatment arms [14]	UK routine data with up to 15-year follow-up [15]
HAQ change on treatment	0.034 per year	0.035 per year
Source	Routine data on 48 patients starting DMARD therapy	Routine data on 48 patients starting DMARD therapy
Rebound on treatment failure	Equal to initial improvement	Equal to initial improvement
Other model assumptions		
Mortality	Relative risk related to disease severity (1.375 per unit HAQ)	Relative risk related to disease severity (1.33 per unit HAQ)
Resource use	Health-care utilization related to disease severity (£800 per unit HAQ)	Drug and monitoring costs only
QALY estimation	$Q = 0.86 - 0.2*HAQ$	$Q = 0.86 - 0.33*HAQ$

BRAM 2007 gives separate estimates for early and late RA—details presented here are those for late RA.

period of the trial used to inform treatment effects [Trial of Etanercept and Methotrexate with Radiographic Patient Outcomes (TEMPO)]. Alternatively, models may extrapolate beyond the follow-up period of the available data. In their submission to TA130, Wyeth use data from TEMPO to extrapolate treatment withdrawal beyond 2 years [4].

Where models choose to assume that the benefits of biologics extend beyond the period for which data are available, there are a number of ways in which extrapolations can be made. Sheffield 2004 assumes a constant hazard rate (exponential model) for the time on etanercept, whereas BRAM 2007 assumes a varying hazard

Downloaded from https://academic.oup.com/rheumatology/article/50/suppl_4/iv10/1833100 by guest on 20 March 2024

rate (Weibull model). Models may also treat biologics as a class to derive estimates of time to withdrawal (e.g. the assessment group model from NICE appraisal TA104 [21]), or allow differences between them (e.g. BRAM 2007). Models also vary in how they represent disease progression while on treatment. They may assume that the disease does not progress during biologic treatment [e.g. the British Society for Rheumatology (BSR) submission to NICE appraisal TA130 [4]], that progression occurs at a rate independent of treatment (BRAM 2007) or that it occurs at a treatment-dependent rate (Sheffield 2004).

The decision to extrapolate beyond the follow-up period of a study, and the method used for extrapolation, can have a dramatic impact on model estimates of the time until a biologic treatment is withdrawn. Both Sheffield 2004 and BRAM 2007 extrapolate from Swedish routine data [9, 12], with a follow-up period of 20 months, to predict mean etanercept treatment durations of 12 and 15 years, respectively. Sheffield 2004 assumes exponential distributions when extrapolating, while BRAM 2007 assumes Weibull distributions, and this may be the reason for the difference. When extrapolating so far beyond available data, the choice of method can lead to even more dramatic differences. However, extending the time for which biologic treatment is received would increase costs as well as benefits, so the impact of this assumption on cost-effectiveness need not be dramatic. Kobelt *et al*. [2] found that extrapolating treatment from 2 years to a maximum of 10 years changed the ICER of etanercept plus MTX *vs* MTX alone from 37 000 to 46 000 per QALY (still below their stated willingness-to-pay threshold of 50 000 per QALY).

There may be a worsening of disease around the time at which treatment is switched, reflecting either the loss of efficacy that leads to the switch, the impact of ceasing treatment or both. Models may assume that this is equal to the initial short-term improvement (both Sheffield 2004 and BRAM 2007 assume this), that it is high enough to raise disease severity to where it would be in the absence of biologic treatment [2] or somewhere in between (Wyeth submission to NICE appraisal TA130 [4]). The impact of assumptions about disease progression during treatment, and at the point of switching treatment, is difficult to assess because their impact will interact with each other, and with assumptions made about time to treatment failure. BRAM 2007 assumes that HAQ progression on treatment was the same for all (biologic and conventional) DMARDs. Their sensitivity analysis showed that, if HAQ is assumed not to progress on biologic therapies, the cost-effectiveness of etanercept *vs* conventional DMARD as third-line therapy falls from £93 000/QALY to £30 000/QALY. The sensitivity of their results to this assumption is related to two other assumptions made—that the duration of treatment can be extrapolated well beyond the available data, and that the rebound is equal to the initial gain. The former extends the time period over which differences in progression accrue, and the latter assumption allows for that difference to persist beyond the treatment period.

Translation of the health impact of treatments into a quality-of-life measure

If model results are to be broadly comparable to other economic evaluations of health technologies, the health benefits of treatments must be reported using a generic measure such as the QALY. This requirement applies to both short-term changes when treatments are initiated or withdrawn and disease progression while on treatment. If the appropriate data are available, a model could allocate the QALY impact of treatment directly—Brennan *et al.* [22], for example, use data from the British Society of Rheumatology Biologics Registry (BSRBR) to estimate the short-term impact of treatment in QALY terms. More commonly, HAQ is used as a proxy for disease progression, and then a HAQ–QALY mapping is used to estimate the impact of treatment in QALY terms; both Sheffield 2004 and BRAM 2007 do this. While there may be differences between the mappings used to derive QALY changes, these are likely to reflect different data sources rather than fundamental differences in model structure.

RA mortality

Where models have a lifetime horizon, they may adjust mortality for RA. They may assume that this adjustment is independent of disease severity (BSR submission to NICE appraisal TA130) or that there is a relationship between disease severity and mortality risk. The assessment group model from NICE appraisal TA130, for example, assumes that the relative risk of mortality increases by 1.33 per unit increase in HAQ [4]. Sensitivity analysis found that removing the relationship between HAQ and mortality had little impact—the cost-effectiveness of etanercept *vs* conventional DMARD as third-line therapy, for example, fell from £93 000/QALY to £86 000/QALY. It seems intuitively plausible that this assumption will have little effect on estimates of cost-effectiveness given that absolute differences in mortality are unlikely to be marked until late in a model's time horizon, at which point their impact will be reduced by discounting.

Resource use

As well as direct drug costs, models may include the impact of treatment on the direct and indirect economic costs of RA, such as routine health-care use, surgery, lost productivity and social care (formal or informal). The assessment group model from NICE appraisal TA130 is an example of a relatively narrow approach to cost inclusion, in that only drug and monitoring costs are included. In contrast, Kobelt *et al.* [2] include the full range of costs listed above, and relate these costs to disease severity. They explored the impact of excluding indirect costs, and found that the increment cost of etanercept plus MTX *vs* MTX alone increased from 14 000 to 20 000 per patient, and this changed the ICER from 37 000 to 53 000 per QALY. These estimates still include health-care costs beyond drug and monitoring costs. The incremental direct costs of etanercept *vs* DMARD in Sheffield 2004 were £27 000 per patient if differences in health-care utilization were included, or £31 000 if costs were restricted to drugs

Downloaded from https://academic.oup.com/rheumatology/article/50/suppl_4/iv10/1833100 by guest on 20 March 2024

and monitoring only. The ICER increased from £16 000 to £18 000 per QALY. Conversely, including lost productivity and social care reduced incremental costs to £13 000 and the ICER to £8000/QALY. These results are consistent with an expectation that relative to the high cost of biologic therapies, changes in health-care utilization are not significant, but costs associated with lost productivity and the need for social care have a greater potential to influence results.

Selection and use of data to inform model parameters

Issues around the selection and interpretation of data arise with all of the areas of model structure choice described above. These issues are as follows.

Choice of data

Models may use different data sources for the same parameter. For example, BRAM 2007 derives the mean HAQ change achieved by etanercept given as a first-line treatment from the Early Rheumatoid Arthritis (ERA) trial, whereas Wyeth base their estimate for this parameter in their submission to NICE on the TEMPO trial [4].

Method of synthesis

Where several treatments are being compared, models may need to draw on a group of studies to estimate parameters such as the initial response rate. Models have used a range of methods to achieve this. Some use absolute response rates from different trials. Both Sheffield 2004 and BRAM 2007, for example, use absolute responder rates for etanercept and conventional DMARDs from separate studies (Table 2). An alternative approach is to carry out a formal evidence synthesis of all relevant studies—the assessment group model for NICE appraisal TA104, for example, carry out a Bayesian mixed treatment comparison of all trials identified through their systematic review [21].

Interpretation of data

Even when models use the same study, they may use it in different ways. For example, in NICE appraisal TA130, the manufacturer submission from Wyeth represents initial response in terms of an absolute change in HAQ, whereas BRAM 2007 represents initial response in terms of percentage change in HAQ [4].

Influence of patient heterogeneity

Some models attempt to adjust key parameters for patient characteristics such as age, sex, disease duration or treatment history. Brennan *et al*. [22], for example, fit a predictive model for EULAR response to data from the BSRBR that include as covariates age, disease duration, disease severity and previous number of DMARDs. Their model for time to treatment failure includes these covariates, and adds EULAR response as a predictive factor. An issue of particular interest here is the efficacy of a second biologic after one has already been given, often referred to as sequential use. This issue was analysed specifically in NICE technology appraisal TA195, where the assessment group noted that 'any evidence suggesting statistically significant differences (between the clinical effectiveness of biologics used after the failure of one biologic) come from uncontrolled studies' [23].

The ability to influence the results of a model through selective use of data is a concern, more so if data sources vary substantially in their estimates of the same parameter. If factors such as disease severity have a marked impact on relative treatment effect, adjusting for them becomes more important, and the scope to influence results through selective use of data becomes greater—even more so if models take absolute results from different trials, breaking randomization. Nixon *et al*. [24] constructed a meta-regression of biologic trials, which suggested that, after accounting for differences in disease duration, there was little difference between them in their effect on short-term response.

Methods used to interpret data can have a significant impact on model results. Sheffield 2004 assumes that, among those continuing to long-term treatment, the reduction in HAQ is greater with etanercept than with conventional DMARD, while BRAM 2007 assumes the opposite (Table 2). While the two groups have chosen different data to inform this parameter, this is not the reason for the divergence. With both models, the data sources for etanercept report a greater absolute reduction in mean HAQ after treatment than the data sources for the conventional DMARD. However, the results come from separate studies with different patient populations, and the patients in the studies informing etanercept response have more severe disease than patients informing response to conventional DMARD (Table 2). In percentage terms, the change in HAQ is greater on conventional DMARD. The decision to interpret treatment effects in terms of absolute (Sheffield 2004) or relative (BRAM 2007) HAQ leads to a marked difference in the assumed comparative benefit of response to etanercept or conventional DMARD.

Identifying the assumptions that drive differences in model results

There are clearly a number of choices faced by modellers when constructing and populating their models, and at each step there are examples of models that take divergent approaches. Which of these have the most impact on results? It is not easy to give a definitive answer to this question. The importance of an assumption will be linked to the decision problem—the willingness-to-pay threshold, the decision population, and the decision question (for example, first-line biologic *vs* DMARD, third-line biologic *vs* DMARD or sequential use of biologics). Furthermore, several examples have been presented of interactions between assumptions. A number of sensitivity analyses have been quoted to help indicate key assumptions. However, most analyses reported are one-way, which fails to account for the interactions described. Sensitivity analyses are also vulnerable to bias by omission, or through arbitrary choice of range over which to vary parameters.

Downloaded from https://academic.oup.com/rheumatology/article/50/suppl_4/iv10/1833100 by guest on 20 March 2024

Table 2 identifies many differences between Sheffield 2004 and BRAM 2007, some of which appear to bear particular responsibility for their divergent outputs. The first is the choice of threshold used to determine adequate response. The estimated mean incremental cost per person of etanercept *vs* conventional DMARD was £30 000 in the former model and £43 000 in the latter, which must be almost completely due to different assumed response rates, and these largely follow from different thresholds for response. As well as lower costs, Sheffield 2004 predicts higher QALY gains from etanercept over conventional DMARD (1.65 *vs* 0.46). There are two main reasons for this. While both models predict that patients who have failed two conventional DMARDs are more likely to respond to etanercept than a third conventional DMARD, Sheffield 2004 also predicts that responders to etanercept see greater improvement than responders to conventional DMARD, while BRAM 2007 assumes the opposite, based on a different reading of the evidence. Secondly, Sheffield 2004 assumes that disease progresses more rapidly on etanercept than conventional biologic, whereas BRAM 2007 assumes progression is independent of treatment. The latter assumption becomes more important in explaining divergent results because both models extrapolate etanercept treatment duration well beyond the data available, and neither assumes that any long-term slowing of disease progression is reversed when treatment is withdrawn.

The comparison described above is based on an interpretation of published descriptions of the models, rather than direct access to the models themselves, which would permit a more rigorous analysis of the key factors driving differences in their results. It also relates to a specific decision problem analysed at a specific time—both groups have updated the assumptions and data sources of their models since the versions described here were created [6, 7]. However, it does highlight more general issues. Results are likely to be sensitive to assumptions around the short-term response required to justify continuing with a treatment. Also, a key challenge when estimating the cost-effectiveness of biologic therapies is determining their long-term impact from short-term data, and results are likely to be sensitive to the assumptions used in doing so. There were also assumptions where, in this specific case, results were less sensitive to choices made. They may have been more influential in a different decision context, however, and there is clearly a need for wide-ranging guidance around preferred modelling approaches and data selection.

Methods guidance—implications for preferred modelling approaches

Guidance issued by bodies such as NICE provide some support for judging alternative approaches used in decision models. The most recent methods guide suggests that the evidence base for modelling should be pre-specified and the result of a systematic review with explicit selection criteria [25]. Treatment effects should be based on randomized controlled trials, preserving randomization. This guidance suggests that models based on specific trials in isolation should be avoided unless no other studies relevant to the decision have been identified during the systematic review. It also suggests that it is inappropriate to select single arms from different trials to get absolute event rates, as this breaks randomization. Instead, a formal synthesis of relative treatment effects, using methods such as Bayesian mixed treatment comparison, should be preferred.

The methods guide also suggests that the choice outcome measure should be relevant to clinical decision making, have some form of link to generic quality-of-life measures, and be based on studies identified in the systematic review. This may be inconsistent advice if the DAS is more commonly used in clinical practice, as trials most commonly report treatment effects in terms of ACR response. Models may need to use both measures, with some form of mapping used to translate between them. Mapping functions are particularly important where studies include a range of outcome measures—it is not appropriate to exclude studies purely on the grounds that they report an outcome measure different from that favoured by the model.

While stressing the importance of basing models on data of the highest quality available, the methods guide does allow for the use of observational data for estimating baseline event rates and extrapolating beyond the follow-up period of the available trials. To avoid selection bias, the data used should represent all available relevant information, identified through a systematic review with pre-specified selection criteria (as with included trials). Thus, the *ad hoc* and selective use of observational data should be avoided. Furthermore, the reference case for the model should not assume any differences between treatments unless they are supported by randomized evidence. This has particular relevance for any extrapolation of treatment effects beyond the follow-up period observed in biologic trials. Such extrapolation can be explored in supplemental sensitivity analyses. This is particularly relevant to the issues discussed above that relate to longer term response, where a number of models assume differences between treatments based on observational data, or extrapolation beyond the follow-up period of trials.

Areas for consensus

There is considerable diversity in modelling approaches that have been used to date in assessing the cost-effectiveness of biologic therapies. Recent methods guidance from the modelling community, as produced by bodies such as NICE, sheds light on which approaches should be preferred, particularly regarding the choice of data and the synthesis of evidence. However, it may be that further input from clinical experts can lead to greater consensus, and this would improve the credibility of decisions made on the basis of model results.

In our view, the need for consensus can be organized around similar themes to those used here to compare

Downloaded from https://academic.oup.com/rheumatology/article/50/suppl_4/iv10/1833100 by guest on 20 March 2024

modelling approaches. Regarding initial response to treatment, questions remain over the choice of measure. Advice is needed on the measures seen as most useful in clinical practice. If these are not the same as those most commonly used in trial reporting, guidance is required on the appropriate way to construct mappings between measures, and consensus is required on the appropriate data to use. Guidance is then required on how to represent the options available to clinicians once initial response has been observed—this includes the timing of any decision to switch treatment, and whether it is essential to represent switching due to adverse events and lack of response separately.

Specific guidance is needed on the issue of sequential use of biologics in particular. It may be that, at an individual level, each patient will respond best to a particular molecule. If this is true, we would expect (when trying a second biologic) a higher rate of response compared with biologic-naive patients in those who are non-responders to their first biologic, and a lower rate of response in those whose disease was controlled by their first biologic for a significant length of time. An alternative assumption might be that some patients are better candidates than others for biologics as a class, in which case we would expect a lower rate of response to the second biologic than the first, no matter what the result of that first treatment was. Guidance is needed on the plausibility of these alternative assumptions, and the appropriate sources of data against which to test them.

For modelling longer term response, similar guidance is needed on the criteria used in clinical practice to determine treatment failure and prompt switching. While data can provide some insight into the appropriate model for representing time to treatment failure, it would be useful to reach consensus on the clinical plausibility of alternative models used (i.e. whether the hazard of treatment failure is likely to vary over time). Clinical views on the side-effect profile of different biologics, and guidance on data sources to support clinical judgement, would also benefit model development. The issue of progression during treatment (plausible assumptions and supporting data sources) also requires clinical input.

Additional questions remain beyond those relating directly to treatment effects, where consensus is required on the plausibility of alternative assumptions and appropriate sources of evidence. These include the extent to which the ability to better control disease progression can extend as well as improve lives and/or reduce the use of health-care resources during the long-term management of these chronic conditions.

Conclusion

There are a number of independently constructed models that have been used to assess the cost-effectiveness of biologic therapies for RA and other inflammatory diseases. Some of these models yield conflicting recommendations for decision-makers. These contradictory findings are unsurprising given differences between the models in their structure, data sources and methods of interpretation. Recent guidance from the modelling community, as published by bodies such as NICE in the UK, can resolve many of these differences. However, a number of questions still remain, and the input of clinical experts to achieve further consensus is to be welcomed.

A Rheumatology key messages

- Economic models of biologics vary in their assumptions and data sources, leading occasionally to markedly different results.
- The impact of a model assumption will depend on the decision context and assumptions made elsewhere.
- For consensus on credible models to be achieved, both technical guidance and clinical judgment are essential.

Supplement: This paper forms part of the supplement 'Biologic therapies in inflammatory joint diseases: models, evidence and decision making'. This supplement was supported by unrestrictive funding from Arthritis Research UK.

Disclosure statement: The authors have declared no conflicts of interest.

References

1 Chiou CF, Choi J, Reyes CM. Cost-effectiveness analysis of biological treatments for rheumatoid arthritis. Expert Rev Pharmacoecon Outcomes Res 2004;4:307–15.

2 Kobelt G, Lindgren P, Singh A, Klareskog L. Cost effectiveness of etanercept (Enbrel) in combination with methotrexate in the treatment of active rheumatoid arthritis based on the TEMPO trial. Ann Rheum Dis 2005;64:1174–9.

3 Welsing PM, Severens JL, Hartman M, van Riel PL, Laan RF. Modeling the 5-year cost effectiveness of treatment strategies including tumor necrosis factor-blocking agents and leflunomide for treating rheumatoid arthritis in the Netherlands. Arthritis Rheum 2004;51:964–73.

4 National Institue for Health and Clinical Excellence. Adalimumab, etanercept and infliximab for the treatment of rheumatoid arthritis - Technology Appraisal TA130. London: NICE, 2007.

5 Brennan A, Bansback N, Reynolds A, Conway P. Modelling the cost-effectiveness of etanercept in adults with rheumatoid arthritis in the UK. Rheumatology 2004;43:62–72.

6 Tosh J, Brennan A, Wailoo A, Bansback N. The Sheffield rheumatoid arthritis health ecomomic model. Rheumatology 2011;50(Suppl 4):iv26–31.

7 Barton P. Development of the Birmingham Rheumatoid Arthritis Model: past, present and future plans. Rheumatology 2011;50(Suppl 4):iv32–8.

8 Moreland LW, Schiff MH, Baumgartner SW *et al*. Etanercept therapy in rheumatoid arthritis. A randomized, controlled trial. Ann Intern Med 1999;130:478–86.

Downloaded from https://academic.oup.com/rheumatology/article/50/suppl_4/iv10/1833100 by guest on 20 March 2024

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

10 Anderson JJ, Wells G, Verhoeven AC, Felson DT. Factors predicting response to treatment in rheumatoid arthritis: the importance of disease duration. Arthritis Rheum 2000; 43:22–9.

11 Emery P, Breedveld FC, Lemmel EM *et al*. A comparison of the efficacy and safety of leflunomide and methotrexate for the treatment of rheumatoid arthritis. Rheumatology 2000;39:655–65.

12 Crnkic M, Teleman A, Saxne T, Geborek P. Survival on drug as a tool for the evaluation of drug tolerability. Initial experience in southern Sweden of infliximab, etanercept and leflunomide in rheumatoid arthritis. Rheumatology 2001;40(Suppl.1 abstract 231b):82–3.

13 Genovese MC, Bathon JM, Martin RW *et al*. Etanercept versus methotrexate in patients with early rheumatoid arthritis: two-year radiographic and clinical outcomes. Arthritis Rheum 2002;46:1443–50.

14 Maetzel A, Wong A, Strand V, Tugwell P, Wells G, Bombardier C. Meta-analysis of treatment termination rates among rheumatoid arthritis patients receiving disease-modifying anti-rheumatic drugs. Rheumatology 2000;39:975–81.

15 Scott DL, Garrood T. Quality of life measures: use and abuse. Baillieres Best Pract Res Clin Rheumatol 2000;14: 663–87.

16 Jerram S, Butt S, Gadsby K, Deighton C. Discrepancies between the EULAR response criteria and the NICE guidelines for continuation of anti-TNF therapy in RA: a cause for concern? Rheumatology 2008;47: 180–2.

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

18 Felson DT, Anderson JJ, Boers M *et al*. American College of Rheumatology. Preliminary definition of improvement in rheumatoid arthritis. Arthritis Rheum 1995; 38:727–35.

19 van der Heijde DM, van't HM, van Riel PL, van de Putte LB. Development of a disease activity score based on judgment in clinical practice by rheumatologists. J Rheumatol 1993;20:579–81.

20 Fransen J, van Riel PL. The Disease Activity Score and the EULAR response criteria. Clin Exp Rheumatol 2005;23: S93–9.

21 National Institue for Health and Clinical Excellence. Etanercept and infliximab for the treatment of adults with psoriatic arthritis - Technology Appraisal TA104. London: NICE, 2006.

22 Brennan A, Bansback N, Nixon R *et al*. Modelling the cost effectiveness of TNF-alpha antagonists in the management of rheumatoid arthritis: results from the British Society for Rheumatology Biologics Registry. Rheumatology 2007;46:1345–54.

23 National Institute for Health and Clinical Excellence. Adalimumab, etanercept, infliximab, rituximab and abatercept for the treatment of rheumatoid arthritis after the failure of a TNF inhibitor - Technology Appraisal TA195. London: NICE, 2010.

24 Nixon R, Bansback N, Brennan A. The efficacy of inhibiting tumour necrosis factor alpha and interleukin 1 in patients with rheumatoid arthritis: a meta-analysis and adjusted indirect comparisons. Rheumatology 2007;46:1140–7.

25 National Institute for Health and Clinical Excellence. Guide to the methods of technology appraisal. London: NICE, 2008:1–76.

Downloaded from https://academic.oup.com/rheumatology/article/50/suppl_4/iv10/1833100 by guest on 20 March 2024