Consensus Decision Models For Biologic therapies In Rheumatoid and Psoriatic Arthritis

REPORT OF WORKSHOP PROCEEDINGS

Prepared by Dr Jason Madan on Behalf of the Consensus Working Party for decision models in rheumatoid and psoriatic arthritis

March 2013

Contents

Acknowledgements

This document has been prepared by Dr Madan on behalf of the consensus working party whose members are listed in appendix 1. The workshops referred to in this report were jointly funded by Arthritis Research UK and the MRC Network of Hubs for Trials Methodology Research. Dr Madan and Professor Ades were supported by the MRC ConDuCT Hub for Trials Methodology Research (grant G0800800). The Consensus working party would like to thank the expert reviewers whose comments on their findings are documented in section 4.

1. Introduction and Background.

Biologic therapies represent a recent addition to treatments for inflammatory joint diseases. Whilst their efficacy has been established in a number of clinical trials (see Appendix B) and cost-effectiveness demonstrated in a number of assessments [1, 2], the evidence base is still associated with a high degree of uncertainty, and this poses a considerable challenge for decision-making in defining the role of different agents in the sequence of disease modifying drugs used to manage chronic disease. A particular issue of concern to many clinicians is the lack of head-to-head trials of biologic therapies, and the use of Indirect Comparisons in the decision making process by NICE, whose recommendations can effectively limit the use of biologics in the NHS. Indirect comparisons, in this context, make it possible to obtain an estimate of effectiveness of one biologic compared to another, using trials with a common comparator (usually placebo) [3, 4].

The original impetus for this report was a question from Professor Alan Silman, Medical Director of leading arthritis charity Arthritis Research UK (ARUK), to the MRC Hubs for Trials Methodology Research (HTMR) network, asking whether conclusions based on indirect comparisons should be considered to be sound, and whether head-to-head trials of biologics should be undertaken. This question eventually filtered through to the Multi-parameter Evidence Synthesis (MPES) research group, part of the ConDuCT HTMR at Bristol, who had published widely on the methodology of Indirect Comparisons and related issues in evidence synthesis [4-6]. A workshop to explore these issues took place at the Royal Institute of British Architects, London, in September 2010. This workshop was funded jointly by the MRC HTMR and by ARUK. The meeting brought together representative of the three main academic centres involved in modelling RA and PsA (York, Birmingham and Sheffield); clinical experts who had acted as advisors to the modelling groups, participated in the key biologic trials, or been involved in maintaining arthritis registries; representatives from NICE itself, and from the National Institute of Health Research (NIHR) Health Technology Assessment programme; as well as methodologists and modellers from the ConDuCT MRC HTMR.

This workshop was the precursor to the two later workshops which are the subject of this Report. The earlier workshop lead to a series of papers published as a Supplement to the journal Rheumatology [7-14]. Briefly, the meeting provided an overview of the current research issues seen from a clinical perspective[7] , followed by presentations outlining the

salient features of the rheumatoid arthritis (RA)[9, 10] and psoriatic arthritis (PsA)[11] health economic models developed by assessment groups (AGs) for Multiple Technology Assessments at NICE. Another presentation[12] set out the main "components" of modelling: the logic of the diseases and treatment allows the analyst to break down the models into a series of semi-independent components, covering initial response to treatment, treatment failure, longer term progression in responders, and so on. This work was carried over to the next stage, allowing the working group to come to the meetings with a clear idea of the issues that would need to be agreed on.

Other presentations covered evidence synthesis methodology: indirect comparisons and also mixed treatment comparisons (also known as network meta-analysis (NWMA)) in which direct and indirect evidence is pooled [13]. It was shown that although most submissions to NICE, including those from manufacturers, tended to include NWMA, as this gave manufacturers the opportunity to claim that their product was cost-effective compared to competitor products, in practice new products were approved on the basis of their cost effectiveness against placebo. To date it seems that in every case where a biologic has not been approved by NICE, this was because it was not cost-effective against placebo, not because it was inferior to another biologic.

A further methodological presentation [14] concerned the use of Expected Value of Information (EVI) methods [15-17]. These are methods that can help decision makers determine whether it is cost-effective to carry out research in order to obtain better estimates of model parameters, and hence to reduce decision uncertainty. The methods can be used to determine both *which* model parameters should be further researched (for example treatment effects or the QALY benefit of treatment), and also the optimal study design for further research, including trial sample size. These methods, it was explained, could in principle be used to determine whether head-to-head trials would be worthwhile, answering the original question that Professor Silman had brought to the MRC HTMR.

However, EVI methods require an economic model. The earlier presentations had highlighted that there were a number of modelling components where there was no clear consensus on how the models should be structured, how they should be informed from data, or even which data were the most appropriate. Moreover, the differences between the models were sufficiently substantial to lead to different predictions. This led to a realisation that a convincing answer to the question on whether head-to-head trials comparing biologics were worthwhile would require a convincing, consensus, economic model. In truth one does not need to appeal to the desirability of an EVI analysis to make a

powerful case, from an academic and regulatory perspective, for the development of a consensus model. At present, in Single Technology Assessments, each manufacturer presents their own model, and it is difficult for the AG and the Appraisals Committee to fully understand how it works within the short timeframe of the appraisal. If consensus views were available beforehand on the desirable properties of the economic model, and the data sources that should inform it, this would assist model development and review during the appraisal process, and manufacturers would be less likely to be asked to produce additional analyses. It may not be feasible or desirable to require manufacturers or AGs to follow the consensus approach in every detail. The former might view this as restricting their ability to fairly present the benefits of their product, and the latter might wish to follow their own academic opinion on the appropriate modelling approach for a specific appraisal. However, if they were encouraged to set out how their models differed from the consensus approach, and present the impact of this deviation on their results, the resulting transparency would enhance the credibility of recommendations derived from those models, and help decision-makers understand the reasons behind any differences in findings between models.

The first meeting ended, therefore, with agreement on the benefits of identifying current consensus on economic models for the evaluation of biologic therapies in RA and PsA. Work presented at that initial workshop also identified a series of issues relating to model structure and use of data, where existing models had taken divergent approaches. This provided a characterisation of models in terms of a set of modular components, which was carried forward to the later meetings. The workshop closed with a recommendation that a clinical studies group be set up to work towards consensus on model inputs and methodology.

Following the initial workshop, further funding was provided by ARUK and the MRC HTMR network for two additional workshops. The invited clinical studies group included clinical experts, health economists involved in the development of existing cost-effectiveness models, and representatives from NICE and the UK HTA (Appendix A). Many of the invitees had attended the initial workshop. The aim of the two consensus workshops was to i) identify, where possible, consensus approaches to each issue, based on sound methodology, clinical judgement and decision-maker preferences, and ii) set out an agenda for the research that needs to carried out to achieve consensus where existing evidence is insufficient.

This document is a report on the work of these two meetings. We begin by setting out the range of measures used to represent the burden of disease in RA and PsA, and the key

sources of data (trials and registry) potentially available to inform models. We then set out the aspects of treatment and natural history represented in models where a consensus approach is required. For each aspect, we describe existing modelling approaches, the views raised in the workshop, a summary of the areas of consensus, and a description of current supporting evidence and the future research required to resolve questions where that evidence is unable to support consensus.

2. Overview of information to support decision modelling in Rheumatoid and Psoriatic Arthritis

2.1. Measures of disease burden

A number of measures exist to represent the burden of disease on patients. These measures vary in the aspect of the disease they aim to capture – disease activity, functional status of patients, or quality of life. Functional status in both rheumatoid and psoriatic arthritis is commonly captured using the Health Activity Questionnaire (HAQ), of which several versions exist [18]. Disease activity is commonly reported using composite measures, changes in which allow patients to be categorised as responders or non-responders to new treatments. The ACR measure [19] has been widely used in clinical trials of biologic treatments. Patients are categorised as non-responders or ACR20/50/70 responders depending on the percentage reduction achieved in tender and swollen joints, and in a variety of other assessments which include global assessment of disease by patients as well as healthcare professionals, measures of function and systemic inflammation. An alternative measure is the Disease Activity Score (DAS), a continuous score [20] of disease status which can be converted into response categories defined by the European League against Rheumatism (EULAR)[21]. Generic quality-of-life measures, such as the SF36 and the EQ5D, can also be used to represent disease progression.

For Psoriatic Arthritis, instruments need to capture both joint and skin symptoms of the disease. The ACR measure can be used to measure joint-related disease activity in PsA, although a specific measure has been developed for PsA, the PsARC [22]. The PASI measure has been developed to measure the severity of psoriasis, and has been used to capture this aspect of PsA [23]. Work is currently in progress on developing novel composite measure of disease burden for Psoriatic Arthritis [24].

2.2. Sources of data

Trials of biologic therapies provide a key source of information on the short-term efficacy (initial response) of treatments. Appendix B lists trials that have informed existing NICE

technology appraisals. Follow-up is generally in the range 12-52 weeks, which allows the responder status of patients to be determined, although it limits the ability of trials to provide information on longer-term outcomes. Trials have differing inclusion and exclusion criteria which may pose challenges for evidence synthesis. In addition, patients entering trials differ in important ways from patients in routine care: this poses additional challenges for data interpretation and generalisability of evidence. A lack of head-to-head comparisons of biologics , and limited evidence on the use of multiple biologics in sequence, further compounds these uncertainties [25].

Since arthritis is a chronic condition, observational data sources such as patient registries provide valuable additional information to inform models, particular for parameters relating to longer-term progression. These data sources may be more representative than trial populations of the general population seen in clinics. There are a number of patient registries based in a range of countries [26], an overview of some of these registries is provided below.

- The **British Society for Rheumatology Biologics Registry (BSRBR)** was set up to register all patients with rheumatoid arthritis newly starting biologic therapy from January 2002 [27]. The registry also includes a comparison cohort of RA patients treated with standard DMARDs. The information recorded by the registry includes HAQ scores, DAS28 and associated EULAR response, adverse events, quality-of-life scores, and a number of patient characteristics such as age, gender and disease duration. Information is collected at baseline and at 6-monthly intervals for 3 years, with some further data available from annual follow-up visits. The registry also includes some patients with PsA.

- **The Norfolk Arthritis Registry (NOAR)** has been following patients in the Norfolk region of the UK with early inflammatory polyarthritis since 1989 [28]. Patients recruited between 1990 and 1994 and between 2000 and 2008 are clinically assessed at baseline and at 1,2,3,5,7,10 and 15 years by a research nurse. Patients recruited between 1994 and 2000 are followed for two years. Data is available on HAQ (at each follow-up visit) and DAS28 (at baseline and every 5 years thereafter), and there is also limited data on SF36.

- The **National Data Bank (NDB**) for rheumatic diseases is a registry of US and Canadian patients with a number of rheumatic conditions including RA[29]. Founded in 1998, it collects data via questionnaires issued to participants every 6 months. The information collected includes HAQ, pain scores, SF36 / EQ5D, adverse events, work and disability, service utilisation, mortality and adverse events.

- The **Dansk Reumatologisk Database (DANBIO)** includes information on Danish adult rheumatologic patients (including RA and PsA) receiving biologic therapies[30].

It was founded in 2000, and includes information on HAQ, DAS28, swollen and tender joints, adverse events and EQ5D.

- The **Norwegian DMARD register (NOR-DMARD)** includes information on Norwegian patients with inflammatory arthropathies followed since 2000 [31]. Information collected includes MHAQ, EQ5D/SF36, swollen and tender joint counts, and health care utilisation.

- The **Leiden early arthritis cohort** commenced in 1993 and collects demographic and clinical data including clinical and radiographic outcomes in RA patients treated according to a protocol [32].

- The **DREAM (Dutch Rheumatoid Arthritis Monitoring) r**egistry, which started in 2006, includes newly diagnosed RA patients. Data on disease outcomes including remission rates under a treat to target strategy, health care utilisation and use of biologic drugs are available [33].

In addition to these registries (and others), there are a considerable number of prospective observational studies that could also be used to supplement trial evidence when informing economic models.

2.3 Current NICE guidance on biologics

Guidance has been issued regarding the use of biologics in RA and PsA patients who have failed at least two conventional DMARDs (TA125, TA130, TA186, TA199, TA220, TA234), and in RA patients who have also failed at least one biologic (TA126, TA195, TA198). Economic models have been used to support guidance within a number of NICE technology appraisals [12]. Where a given biologic therapy is recommended, guidance states that biologic therapy should only be continued if an adequate short-term response to therapy is observed. For RA, an adequate response is defined in all guidance as an improvement in DAS28 of 1.2 points or more at 6 months [1] . For PsA, an adequate response is defined as an improvement in joint swelling /tenderness and at least one of the other three PsARC criteria at 12 weeks [2].Those who do not reach this level of response, but achieve a PASI 75 response in skin symptoms, should be assessed by a dermatologist to determine whether treatment continuation is appropriate. NICE has also issued clinical guidelines for rheumatoid arthritis [34] which incorporate the relevant guidance for biologics listed above, and also recommend their use in combination with conventional DMARDs. Clinical guidelines have not been issued in relation to psoriatic arthritis.

3. Findings of the 2011/2012 workshops

3.1. Overview of issues for consensus

At the 2010 workshop, a list of items needing consensus, relating to cost-effectiveness modelling, emerged. Four main topic areas were identified:

Topic 1: Modelling the initial response to treatment, including:

- Choice of scale to measure initial response
- Link between response level and decision to continue treatment
- Choice and use of evidence to estimate effect of treatment on initial response
- Estimating the baseline response in the comparator treatment
- Modelling adverse events in the initial treatment phase
- Influence of effect modifiers on treatment effects.

Topic 2: Longer-term disease progression in those who continue treatment, including:

- Choice of scale to measure long-term disease progression
- Rate of disease progression during long-term treatment
- Treatment duration (i.e. time to withdrawal of treatment due to lack of efficacy and/or adverse events)
- Modelling adverse events in the long-term treatment phase
- The influence of effect modifiers on treatment duration and disease progression.

Topic 3: Estimating lifetime costs and benefits of treatments, including:

- Resource use implications to include in calculations
- Modelling the relationship between disease severity and mortality risk.

Topic 4: Structural modelling approaches:

- Representing sequences of treatments
- Cohort vs. Individual Patient models

The aim of the 2011/2012 workshops was to establish, as far as possible, consensus positions on these issues, guided by an understanding of the clinical aspects of RA and PsA, and the principles of evidence-based medicine, as set out in documents such as the Cochrane handbook [35] and the NICE methods guide for technology appraisals [36] . These principles state that treatment effects should be based on randomised studies, and that all such relevant studies should inform estimates of treatment effects. This implies that models avoid the selective use of data, as occurs when models base treatment effects on single trials or exclude trials based solely on the choice of outcome measure, and should preserve

randomisation in the evidence base, rather than using absolute results from individual trial arms in isolation. The inclusion of differences between treatments based on observational data alone is not generally recommended. Models should also respect the decision context in their structure and choice of data, which implies that data sources should be relevant to the decision problem and outcome measures used in the model should reflect those used in clinical practice.

3.2. Modelling the initial response to treatment

3.2.1. Current Modelling Approaches

Choice of outcome measure

For RA, NICE multiple technology appraisals (TA130, TA195) have been based on the Birmingham Rheumatoid Arthritis Model (BRAM), which used (percentage change in) HAQ to represent initial response to treatment [37]. The model developed by Wyeth in their submission to TA130 also used HAQ, although this model assumes treatment causes an absolute change in HAQ [37]. With the HAQ, there is no set level dividing 'responders' from 'non-responders'. The BRAM therefore sources its estimate of short-term discontinuation from a Scandinavian study reporting routine data on clinical practice [38] , and assumes those who continue are those who respond most strongly to treatment. An alternative approach, taken in the Wyeth submission, is to base continuation rates on (DAS28) response rates observed in a trial [37]. Several manufacturer submissions have used ACR20/50/70 as their chosen outcome measure [37]. The outcome measure used in modelling short-term response for the BSR submission to TA130 was DAS28 and its associated EULAR response categories [39]. In the latter examples, those who fail to report an adequate response (ACR20 or EULAR moderate) are assumed to withdraw from treatment at 6 months. For PsA, NICE multiple technology appraisals (TA104, TA199) have been based on a cost-effectiveness model developed by the University of York, using PsARC (for joint symptoms) and PASI (for skin symptoms) as outcome measures [40] . Manufacturer submission models have also based on the PsARC, with or without the PASI [41].

Representing reasons for discontinuation

The BRAM is an example of a model that distinguishes between adverse events and lack of efficacy as reasons for discontinuation in the first 6 months [37]. The probability of discontinuation due to adverse events was derived from the same study used to inform the probability of discontinuation due to lack of efficacy (20). The York PsA model illustrates an

alternative approach, which is to combine these reasons in a single outcome of treatment discontinuation [40, 41].

Choice and use of data

For each NICE multiple technology appraisal mentioned above, the assessment group carried out a systematic review to identify all relevant trials. For the PsA appraisals, treatment effects in the economic model were jointly estimated using a network meta-analysis which synthesised the relative treatment effects of all identified trials [40]. For TA130, the assessment group identified 29 trials in the effectiveness review. Of the 29 trials identified, 14 inform estimates of initial relative HAQ change for the economic model – unpublished data from the treatment arm of each trial is used to derive the HAQ multiplier for that treatment (the method for combining multiple trial arms, where used, is not reported). Examples exist of models where treatment effects are based on a single trial, and where the treatment arm is used from a single trial to estimate biologic efficacy, and observational data is used to estimate the efficacy of the comparator DMARD [37]. The BSR submission to NICE TA130 is based on patients recruited into an observational study, the BSRBR [39].

Effect modification in initial response

Meta-regression of biologic trials suggests that covariates such as disease duration influence treatment effects when ACR20/50/70 is used as the outcome measure [42]. The BSR submission for NICE appraisal TA130 used an economic model in which age, baseline utility, disease duration, gender, and treatment history were allowed to modify the effect of treatment on the probability of EULAR response [39]. Data from the BSRBR was used to estimate the degree of effect modification. Neither of the assessment group economic models developed to support NICE RA and PsA technology appraisals assume effect modification, although the BRAM allows a degree of effect modification by fitting separate treatment effects for early and late RA and deriving treatment effects for biologics with or without concomitant methotrexate separately [37].

3.2.2. Views of the clinical studies group

Modelling the initial treatment phase

Stopping rules do not fully reflect the complexity of clinical decision-making at a patient level. However, the use of an explicit stopping rule within economic models is required to

synthesise trial evidence, link short-term and long-term outcomes, and explore the cost-effectiveness implications of different guidance. Therefore, models should include such stopping rules, as long as it is recognised that they do not fully specify outcomes at a patient level. Currently, for RA, the most appropriate measure to base such stopping rules on is DAS28, because:

- DAS28 most closely reflects clinical benefit of treatment in the short term.
- relatively small changes are still clinically meaningful to patients
- it is an absolute scale (although the related EULAR response categories depend on both absolute change in DAS and DAS at endpoint).
- it has received support from clinical experts in previous NICE appraisals, and is the basis of current NICE guidance.

HAQ + pain score was suggested as an alternative measure. ACR 20/50/70, whilst commonly reported in trials, was considered problematic because it is a relative measure. Given that ACR 20/50/70 is commonly reported, there is a clear need for mapping functions to characterise the relationship between the two measures, as it is not appropriate to exclude relevant studies solely because they do not report DAS28.

For PsA, both outcomes (skin and joint symptoms) need to be considered when modelling the initial treatment phase. Psoriatic arthritis is a heterogeneous condition, and there are types of PsA where DAS28 could be the most appropriate measure of response for joint symptoms. However, disease-specific measures for PsA are in development, so efforts to shift from PsARC are unlikely to be worthwhile.

Effect modification

A number of factors are potential effect modifiers for relative effects of treatment on responder status. Mechanisms for effect modification include 'treatment resistance' (failure to respond to previous drugs may indicate a lesser chance of responding to the current drug) and 'accumulated damage' (disease duration will be linked to the amount of damage that has occurred to joints). There will also be treatment-dependent modifiers (e.g. Rheumatoid factor status for rituximab). Effect modification may depend on the choice of outcome measure – in particular, effect modification may be more influential with ACR 20/50/70 response, as this is a relative response measure, sensitive to baseline disease activity, than with DAS28, which is an absolute measure.

Choice and use of evidence to estimate effect of treatment on initial response

When performing a synthesis of trial evidence to inform treatment effects in a model, trials in biologic-naive patients should be analysed separately from trials in patients with prior biologic DMARD exposure, as should trials in biologics with or without concomitant DMARDs. Formal models for effect modification could be derived from individual patient data (IPD) sourced from trials, or from observational data. A concern with the latter is potential selection bias. Where data is weak, expert elicitation could guide adjustments related to changes in position within the sequence. This approach could be used to adjust treatment effects for additional modifiers such as disease duration, disease severity, age and gender. However, in the absence of convincing evidence for effect modification, the simpler approach of using treatment effects unadjusted for these factors is preferable, particularly if an absolute scale such as DAS28 is used for response.

Estimating the baseline response in the comparator treatment

For modelling purposes, relative treatment effects need to be applied to the absolute proportion of (DAS28) responders that would be seen if a conventional DMARD was given instead of a biologic at the relevant point in the sequence. The absolute rate from the control arm of a biologic trial has often been used for this purpose, as have absolute rates from trials of conventional DMARDs. An alternative would be to use registry data. The latter would match the required patient profile most closely, but would be vulnerable to issues such as selection bias. Therefore, the approach of pooling control arms from trials with populations similar to the decision population was preferred.

Modelling adverse events in the initial treatment phase

The reason for not continuing treatment past the initial phase may have consequence for the choice and efficacy of subsequent treatments, and may also have cost implications. Models should therefore distinguish between adverse events and lack of efficacy as reasons for short-term treatment termination. This requires information on adverse event rates for different biologics, which will be reported by most trials. Models should not exclude trials completely that do not report causes for treatment discontinuation. This can be avoided by estimating the overall discontinuation rate and the split between causes, rather than estimating the absolute rate for each cause.

3.2.3. Summary of consensus view

- DAS28 should be used to represent initial response to treatment in RA.

- Models should reflect current guidelines and withdraw treatment from patients with an inadequate response. The time at which this occurs may vary from model to model, and models may also be used to explore the impact of stopping rules based on current clinical practice, rather than a fixed threshold for adequate response. However, to aid comparison of results, models should present the impact of following current guidance on this time in sensitivity analysis (for the UK, current NICE guidance states that treatment should be withdrawn from RA patients at 6 months (ref TA130) and PsA patients at 12 weeks (ref TA199).
- Currently, robust evidence for effect modification has not been identified, and effect modification should not be included in evidence synthesis of initial response treatment effects.
- Models should represent the cause for discontinuation of treatment (i.e. lack of response or adverse events).
- Estimates of short-term response should be based on all relevant trials and derived using formal evidence synthesis methodology that respects randomisation. Mapping functions should be used within the synthesis so that trials can be included even if they do not report DAS28.
- Mapping functions should also be used to relate changes in DAS28 to changes in the measure used to represent long-term disease progression (see section 3.3).
- Response rates to the non-biologic comparator can be based on pooling control arms from biologic trials, although the degree to which trial populations are similar to decision populations should be considered.
- For PsA, PsARC and PASI should be used as outcome measures, although disease-specific measures currently in development may be used once they have been validated.

3.2.4. Current available evidence and further research needs

<u>Mapping between (change in)DAS28 and ACR20/50/70</u>

While DAS28 is the preferred measure of short-term response to treatment for the RA consensus model, the relevant evidence base includes a high proportion of trials that report ACR20/50/70 instead. Currently, there are no established mappings between ACR20/50/70 and DAS28, and research is required to develop such a mapping, so that DAS28-based models are informed by all relevant trials. Few, if any trials or observational data sources collect or report both measures. Therefore, mappings will need to be constructed through indirect comparison with other outcome measures sensitive to disease activity (e.g HAQ, SF36). Since ACR measures are relative, while DAS28 is an absolute scale, mappings should allow for dependence on baseline disease activity. Individual patient data from trials would be the ideal evidence for the mapping functions, potentially supplemented by registry data (e.g. estimation of the DAS28 / HAQ change relationship from the BSRBR).

Mapping between existing PsA outcome measures (PsARC, PASI) and composite measures currently in development.

The evaluation group for the NICE multiple technology appraisal developed a Bayesian network meta-analysis to synthesise trial evidence on short-term response to etanercept, infliximab and adalimumab [40]. Treatment effects were estimated on four outcomes – PsARC, ARC (both for joint symptoms), PASI (for skin symptoms) and HAQ (for functional impact). A range of models were explored for the meta-analysis, and the reference case for the economic model involved a positive correlation between PsARC and PASI response. The analysis, once updated and extended to include newer treatments, satisfies the requirements of the workshop consensus and should inform future PsA models that are based on PsARC and PASI response. The Group for Research and Assessment of Psoriasis and Psoriatic Arthritis (GRAPPA) is an international organisation actively engaged in the development of response measures in PsA [43]. The GRAppa Composite Exercise (GRACE) study has collected data on multiple PsA dimensions and has recently developed novel composite responder indices [44]. If clinical practice changes as a result of these developments, further research will be required to develop mapping functions between new and existing response measures.

Updating reviews of short-term adverse events.

The models used to inform current NICE guidance do not fully meet the consensus view. The model for PsA does not distinguish between lack of efficacy and adverse events as reasons for discontinuation (22). The model for RA makes such a distinction, but does so without drawing fully on the available evidence. The consensus model requires estimation of the proportional split between lack of efficacy and adverse events for those who discontinue treatment at an early stage, based on comprehensive and up-to-date evidence. Systematic reviews of biologic trials undertaken to inform NICE technology appraisals can be used to identify this evidence base. There are additional reviews of adverse events in the literature[45]. Systematic reviews of sequential biologic therapy have also assessed the impact on the efficacy of a second biologic of having experiences adverse events on the first biologic [25]. This evidence base needs to be collated, updated and synthesised to inform the consensus model.

3.3. Modelling the long-term treatment phase

3.3.1. Current Modelling Approaches

Duration of successful treatment

Patients who respond satisfactorily to treatment are assumed to continue that treatment until it fails to control disease activity, an adverse event occurs, or both. The duration of successful long-term treatment is likely to be considerably longer than the follow-up time reported by trials. Models that base their estimates of treatment duration on trial data can either make the conservative assumption that treatment benefit is restricted to the follow-up time of the trial(s) used, or use some form of survival modelling to extrapolate beyond the trial follow-up time [25] . Both the BRAM and York PSA models derive treatment duration by extrapolation from routine data sources – the former using a Weibull survival model [37], and the latter an Exponential survival model[41] .

Disease progression during treatment

Whilst models exist that represent long-term changes to quality of life on treatment directly [39], a more common approach is to represent long-term changes on a disease-specific measure (e.g. HAQ), which is then mapped to quality of life scores [37],[41] Unless HAQ is also used as the measure of short-term response, models must therefore translate the initial response to treatment from the chosen outcome measure (e.g. ACR20/50/70, DAS28, PsARC) into a change in HAQ score before applying long-term disease progression. This may be derived from a meta-analysis of all available trials, as in the York PsA model [40] or from individual patient data from trials of the relevant treatment, as in the Abbott submission to TA130 [37]. Once the initial change in HAQ (or health utility) is applied, disease progression is commonly assumed to occur at a steady (linear) rate. The rate can be derived from trial or routine data, and may be assumed to be the same for all treatments, as in the BRAM [37], or to depend on the class of treatment (biologic or conventional DMARD), as in the Wyeth submission to TA130 [37].

Rebound effects on cessation of current treatment

At some point in time, a treatment may cease to be effective in controlling disease activity, or it may cause intolerable side-effects, or both. At this point it will be withdrawn, and this may be linked with a step change in HAQ (or the alternative measure used by the model). This step change may be equal to the initial step change associated with treatment response, it may be sufficient to bring HAQ in line with the value it would have reached without biologic therapy, or it may return HAQ to the baseline level at the start of treatment (fig 1). For example, the York PsA model assumes the first assumption in its reference case, and explores the second assumption in its sensitivity analysis [41]. An elicitation exercise has also been carried out to characterise expert opinion on the level of rebound [46] .

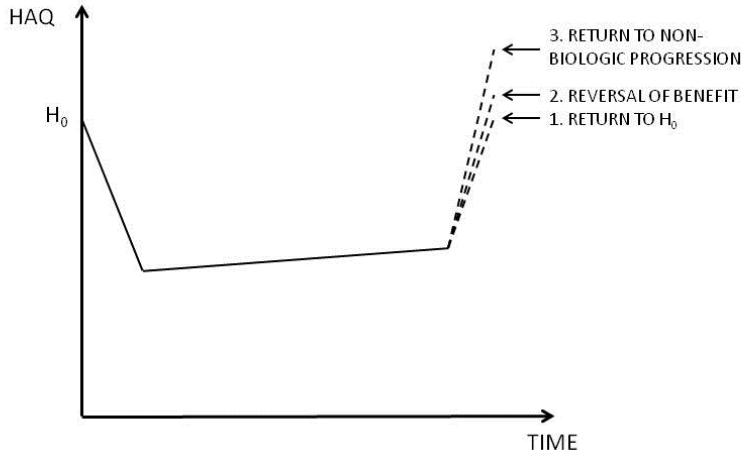

Fig 1: Alternative assumptions for rebound during treatment switch.

Mapping between disease severity and health utility

Models that represent disease progression using disease-specific measures such as HAQ need to translate this into health utility scores to estimate the QALY impact of treatments. A linear relationship between HAQ and QALY is commonly assumed, although the data sources differ. For example, early versions of the BRAM assumed that a unit change in HAQ translates into a constant change in quality of life score (0.327 in the base case analysis, based on routine data collected by Hurst et al [47]), whilst the Abbott submission to TA130 assumes a change of 0.28 in quality of life score per unit HAQ, based on patient-level data from the adalimumab trials [37]. PsA models also often assume a linear HAQ-QALY relationship, using PsA trials for estimation, but some, but not all, models take the approach of the York PsA model of including PASI when estimating health utilities [41].

Effect modification

Models do not commonly allow effect modification for parameters related to longer-term outcomes on treatment. An exception is the BSR submission to NICE TA130, which used individual patient data from the BSRBR to estimate effect modification due to age, disease duration, treatment history and initial response on treatment duration [39].

3.3.2. Views of the clinical studies group

HAQ has been widely used in models to represent disease progression, for historical reasons. Several peer-reviewed mappings between HAQ and quality of life measures (e.g. EQ5D) have been developed and used in existing models [47, 48]. However, methodological and applied research into algorithms for mapping between outcome measures such as HAQ and EQ5D is an area of active research [49], and the most appropriate algorithm for use in decision models may change over time. For example, recent research has suggested that pain has an important influence on quality of life in RA patients, independent of HAQ. Therefore, models could in future use a multidimensional (HAQ and pain) outcome measure for disease progression. Observational data has been used to estimate the rate of change in HAQ over time whilst on treatment (conventional DMARDs or biologics). This raises issues of data quality, particularly if the data are used as the basis for differences between biologics in rate of progression. Models should not be 'hard-wired' to exclude such differences, but the reference case should only allow differences between drugs of the same class if based on data from randomised studies. The impact of differences inferred from observational data could be explored in supplementary analyses, but estimates should reflect the increased risk of bias. The estimates may be more credible if based on observational data collected in a clearly relevant population, or on a synthesis of multiple sources of non-randomised evidence.

HAQ progression is sometimes assumed to be zero on biologics. This is not biologically credible in the long term in view of the effect of ageing on HAQ. Further long term data are needed in RA and PsA populations in remission. Current models for non-biologics assume linear progression at a rate which appears to result in too many people reaching the HAQ ceiling too quickly. Registry data may give some data on HAQ progression, and elicitation could also be used to incorporate expert opinion on long-term HAQ progression where existing data is insufficient. Mixture models have been fitted to registry data showing distinct sub populations with different HAQ trajectories. By averaging over these trajectories, a more realistic non-linear model could be developed for HAQ progression over time.

Differences in action between biologics may justify differences in distributions for successful treatment duration. Duration may also be age-specific and influenced by concomitant treatments, and registry data provide information for fitting treatment duration distributions. It is important to record the reason for treatment switching, as this can influence the choice and efficacy of subsequent treatment. However, there is an interaction between these factors, since adverse events are more likely to lead to treatment being withdrawn if efficacy is diminished.

Estimates of rebound on treatment termination should be based on data and assumptions avoided as far as possible. However, observations rarely coincide with treatment switching decisions. Expert elicitation may be necessary to determine the most appropriate assumption. Whilst rebound may in fact occur over a period of time, a step change is an acceptable simplifying assumption. Rebound effects are likely to differ between RA and PsA patients, and data on the former should not be used as a basis for estimating rebound in the latter.

3.3.3. Summary of consensus view

- HAQ should be used to represent disease progression, although a multidimensional measure which includes pain should be considered for mapping disease progression to health utilities.
- The source for mappings used between outcome measures (e.g. HAQ and health utilities) should be clearly stated and justified, and be consistent with current applied and methodological research.
- Survival models may be used to extrapolate beyond the follow-up period of data on the duration of successful long-term data. All relevant data should be used to fit such models, this may include open-label trial follow-up and registry data. However, treatment duration differences between biologics should not be assumed based on observational data alone.
- Assumed rates of HAQ progression should be consistent with observations from longitudinal data.
- Models should distinguish between adverse events and loss of efficacy as reasons for treatment withdrawal.
- The rebound in disease progression on treatment withdrawal should be evidence-based as far as possible. Where multiple scenarios are consistent with the available evidence, the impact of alternative plausible assumptions should be explored through sensitivity analysis.

3.3.4. Current available evidence and further research needs

<u>Estimating duration of treatment in responders</u>

Existing models use diverse data sources for estimates of biologic treatment duration, and interpret those data in different ways. None of these approaches were thought to satisfy the requirements of the consensus model, and further research is required to establish treatment duration distributions based on up-to-date and relevant data. The BSRBR has several advantages as the basis for estimating this information – it is comprehensive, provides detailed UK-specific patient-level data, and is up-to-date. It could also be used to explore the impact of effect modification and the extent to which treatment duration differs

between biologics, although as a non-randomised data source such analyses should be interpreted with caution.

Disease progression on long-term treatment

The consensus group also felt existing modelling approaches to disease progression were not appropriate for the consensus model. In particular, the assumption of linear HAQ progression leads to patients in the model reaching HAQ ceiling values earlier than is observed with real patients. Research is currently underway exploring non-linear HAQ progression models. Once this research is fully available it may prove an appropriate basis for the consensus approach. If the data available do not provide definitive evidence for long-term HAQ progression, they may be supplemented with elicitation of expert opinion.

Mappings between disease-specific severity measures and health-related quality of life

Mappings between HAQ and QoL scores have been developed using trial and/or observational data. Recent research has shown that pain has an independent effect on QoL scores when added to HAQ [50]. Mappings currently used in models do not account for this effect, and do not draw fully on all currently available evidence. Further work is required to produce definitive mapping functions between HAQ scores (with pain if appropriate) and QoL. This will first involve identifying the appropriate data sources, which may include several of the registries mentioned above, and could also involve individual patient data from trials where it can be obtained. The appropriate method for deriving mapping algorithms from this data will then need to be identified. For PsA, data collected by the GRACE study may provide information to map combined joint, skin and pain symptoms to QoL scores.

Impact of treatment switching on HAQ

Empirical estimates of HAQ rebound on treatment withdrawal are challenging to derive and lacking in existing literature. Further research could be carried out using registry data to estimate this parameter, although the fact that follow-up visits often do not coincide with treatment withdrawal may limit the accuracy of estimation. Elicitation techniques could be used to capture clinical expertise on rebound if empirical approaches are unsuccessful. Given the challenges of estimating this parameter, the sensitivity of cost-effectiveness findings to alternative assumptions should be explored within the consensus model.

3.4. Estimating Lifetime Costs and Benefits

3.4.1. Current Modelling Approaches

Potential costs of treatment include not just the direct cost of the drug itself (and costs of monitoring patients whilst receiving treatment), but also the impact of treatment on health care utilisation and social care (formal and/or informal). The base case analysis of the BRAM is an example of a restricted approach to cost inclusion, as only drug and monitoring costs are included [37]. An approach to capturing the indirect health care costs of treatment is to assume a relationship between disease severity (represented by HAQ) and health care utilisation. For example, the York PsA model assumes a relationship between HAQ and health care costs based on estimates from a study of UK and Swedish routine data [51] .

Models may include the impact of treatment on mortality as well as morbidity by assuming a relationship between disease severity and mortality. The BRAM, for example, assumes in the base case a relative risk for mortality of 1.33 per unit increase in HAQ [37] based on analysis of data from the US NDB [52].

3.4.2. Views of the clinical studies group

There is evidence to suggest disease severity has an impact on age-adjusted mortality risk, but not to suggest that choice of treatment has any additional influence on mortality. For PsA, skin symptoms may be additionally associated with mortality. The cost perspective of a model should reflect the preferences of the decision-maker involved. In the UK, for example, the reference case perspective for NICE technology appraisals is health and personal social care costs only. An acceptable approach to modelling the indirect impact of treatment on such costs is to assume a relationship between disease severity and resource use. For PsA, resource use should be modelled as a function of both joint and skin symptoms (although double counting should be avoided). Where models use discrete time-cycles, cycle duration should be short enough to accurately reflect resource use patterns.

3.4.3. Summary of consensus view

- Models should allow for an association between disease severity and mortality
- Models should adopt the decision maker's chosen perspective for costs included. This may involve assuming health care utilisation to be a function of disease severity.

3.4.4. Current available evidence and further research needs

<u>Arthritis health care utilisation</u>

Research is required to collate diverse evidence on the relationship between disease severity in RA and PsA and healthcare utilisation, as modelling for current NICE guidance either does not incorporate this relationship or bases it on selective and out-of-date

sources. This research should initially take the form of identifying current literature and appropriate data sources. The relationship between disease severity and health care utilisation has been estimated in several published analyses drawing on routine data. Work that has informed existing models includes analysis of registry data from the US [53] and Sweden [54]. More recently, analysis has been published of the total costs for patients with RA and PsA, including productivity losses, using Norwegian registry data [55]. The NOAR may also provide recent UK-specific data for estimating this relationship.

Mortality and disease severity

There are conflicting findings in the literature regarding the relationship between mortality and disease severity. Research is therefore required to establish a definitive estimate for the consensus model. Routine data may provide the most appropriate source for this relationship. For example, Lunt et al have analysed mortality data in the BSRBR for this relationship [56], and their analysis included a number of covariates including disease duration and severity. Additional research would identify the full current evidence base and use this to derive the consensus relationship, either through synthesis of multiple evidence sources or establishing clinical consensus on the most appropriate data source.

3.5. Model type and structure

3.5.1. Current Modelling Approaches

Patients receive a sequence of treatments over their lifetime, and are switched to the next treatment once their current treatment fails (due to lack of efficacy or adverse events). Models may explicitly model initial response to downstream conventional DMARD treatments. For example, the BRAM models initial response to several conventional DMARDs following biologic therapy [37]. The York PsA model illustrates an alternative approach in which, following failure of a biologic therapy, patients receive palliative care (of which a percentage may continue using a DMARD). Patients then experience steady long-term deterioration and short-term fluctuations caused by response to subsequent conventional DMARDs are ignored. Models differ in their approach to patient simulation, with some simulating patients individual patients using approaches such as Discrete Event Simulation [37, 39] and other simulating cohorts of patients, often within a Markov model structure [41, 57].

3.5.2. Views of the clinical studies group

Models should have the flexibility to explore alternative positions for biologics within the sequence of treatments. While there may be benefit in modelling specific DMARD sequences once biologic therapies have been exhausted, the group felt that treatments

have limited effects at this stage in practice. Therefore, it is preferable not to explicitly model sequences of conventional DMARDs following biologic therapy, unless data on such patients becomes available that credibly challenges this view.

The group noted that both cohort and individual sampling approaches have been adopted by previous models, and there were divergent views over the relative merits of these approaches. Guidance exists in the literature on factors which should influence the choice of model type [58, 59]; as a general principle, models should be as simple as possible whilst remaining consistent with the underlying decision problem and theory of disease [60]. However, the appropriate model structure for the evaluation of biologics in arthritis has not been definitively established in the literature, and remains a question of both practical and methodological interest.

3.5.3. Summary of consensus view

- Models should be able to represent response for each biologic therapy in a sequence, but do not need to model individual post-biologic conventional DMARDs.
- Individual patient models have a number of advantages when representing RA and PsA patient histories, but the merits of cohort modelling approaches should also be explored.

3.5.4. Current available evidence and further research needs

Given the alternative approaches to model structure in existing models, future research should involve developing models that follow the consensus approach as closely as possible whilst adopting alternative structures, to evaluate how closely each model structure is able to follow the consensus approach and the impact of structure on model results.

4. Commentary on workshop consensus statement from an independent expert panel.

In order to provide a perspective on the extent to which the workshop reflected current consensus, a separate panel of independent experts were invited to provide a commentary on the views of the clinical studies group as expressed above. These individual commentaries are presented below.

4.1. Commentary from Professor John Isaacs, Director, Institute of Clinical Medicine, University of Newcastle:

I agree broadly with the conclusions of the clinical studies group – perhaps most importantly in terms of defining the factors that should be considered in terms of cost-effectiveness. In my mind it is perverse not to include factors such as the cost of joint surgery. I also agree that it is not logical to assume that patients will return to non-biologic DMARDs following the failure of biologics – and particularly that these may maintain their pre-biologic efficacy.

I also have some specific comments on the report, which are given below.

- Measures of disease burden (section 2): I am always surprised that X-ray progression does not feature in models. While it does have a 'downstream' effect on function, it is one of the major differences between conventional DMARDs and biologics.
- Sources of data (section 2): While I realise that the provided list of registries is not meant to be exhaustive, I am surprised that the German RABBIT and Swedish registries have not been listed, as they have been particularly influential.
- Current guidance (section 2): There is an inconsistency between NICE guidance and the technology appraisals, since they emphasise the importance of treating to target (to DAS remission or LDA) but don't allow biologic initiation unless DAS > 5.1. This means clinicians cannot follow NICE guidance in a patient who has failed conventional DMARDs but still has a DAS between 3.2 and 5.1 – a large proportion of patients.
- Effect modification (Section 3.2.2). Rheumatoid factor status may not be as big an influence on response to rituximab as previously thought. Also, baseline DAS28 does have some influence on likelihood of response, since it is easier to improve from a high DAS28 than it is from a low DAS28. While the consensus

view mentions an absence of convincing evidence for effect modification, I think the evidence is pretty strong that female sex is a poor prognostic factor, both in terms of likelihood of remission and response to therapy. Smoking is also emerging as a poor prognostic factor, although the evidence is not yet as convincing.

- Mapping between DAS28 and ACR 20/50/70 (section 3.2.4): The report claims here that few, if any, trials collect or report both of these measures. However, I believe several trials have done so (for example the ph3 tofacitinib trials)

- Reviews of short-term adverse events (section 3.24): The report references the review of adverse events by Bongartz et al, the conclusions of which have been criticised and, to an extent, discredited. There is also emerging evidence that immunogenicity should also be measured and could usefully influence treatment decisions (e.g around switching within a class of drugs such as anti-TNF).

- Duration of successful treatment (section 3.3.1): I would have thought that registry data is reasonably reliable in terms of duration of treatment, and published data are quite consistent from around the world.

- Choice of long-term outcome measure (section 3.3.2): The report mentions that pain has an important influence on quality-of-life, independent of HAQ. Fatigue is also an important determinant of QoL that is poorly captured with current measures.

- Costs included in economic modelling (section 3.4.1): Restricting such costs to drug and monitoring costs only is much too narrow – one has to factor in the costs of joint surgery for RA (which has plummeted in recent years), employment prospects, cost of carers etc. to obtain a true view of cost-effectiveness in a chronic disease such as RA. I strongly agree with the statement in section 3.4.4 that research is required on the relationship between disease severity and healthcare utilisation.

- Modelling DMARD sequences (section 3.5.2): I strongly agree that models should reflect the limited impact of DMARDs once biologic therapies have been exhausted. Essentially DMARDs do not work post-biologics, or work minimally, and it is inappropriate to extrapolate pre-biologic DMARD efficacy into the post-DMARD disease stage.

4.2. Commentary from Dr Karim Raza, reader in clinical rheumatology and honorary consultant, College of Medical and Dental Sciences, University of Birmingham.

This report is well written, well informed and timely. I have only a few comments on it:

1. PATIENT INVOLVEMENT: The participants in this workshop bring highly appropriate and complementary skills but it would appear that a patient / user perspective is lacking. In developing models to help inform the use of biologic agents for rheumatoid arthritis and psoriatic arthritis it is, I think, important to reflect on whether benefit could be gained by involving those for whom these therapeutics are intended. I recognise that this is not straightforward, particularly given the very technical nature of many of the issues being discussed, but I do think that consideration should be given to this area, perhaps under the heading of "further research needs": [a] Do patients think they should be / want to be involved in discussions on decision models for biologics; what advantages do they feel their involvement would bring and what concerns would they have about being involved? [b] If patients are keen to be involved, how could involvement be facilitated to allow maximum benefit (for patients, modellers and the model)?

2. EFFECT MODIFIERS: Reference is made on several occasions to the fact that "disease duration" may be an effect modifier though a summary of the consensus view is that "Currently, robust evidence for effect modification has not been identified, and effect modification should not be included in evidence synthesis of initial response treatment effects". Table 1.2 summarises studies of biologics in biologic-naïve "early" RA patients – the group in which the impact of early vs. delayed biologic therapy is most likely to be seen. It is very important to recognise that considerable ambiguity surrounds the concept of "duration". "Duration" has to be timed from an "onset" and different researchers mean very different things by "onset" in the context of RA – ranging from the onset of symptoms of RA to the onset of patient reported joint swelling through to the time the patient first fulfilled classification criteria for RA (this concept is discussed in detail in: Raza K, Saber TP, Kvien TK, Tak PP, Gerlag DM. Timing the therapeutic window of opportunity in early rheumatoid arthritis: proposal for definitions of disease duration in clinical trials. Ann Rheum Dis. 2012 Dec;71(12):1921-3). When assessing "Disease duration" as a potential effect modifier, this issue needs to be taken into account – a mean "disease duration of 6 months" may mean completely different things in different studies.

3. MODELLING THE LONG TERM TREATMENT PHASE: The only scenarios in which treatment withdrawals appear to be considered are adverse events and loss of efficacy. In current practice, a number of Rheumatologists are reducing / withdrawing therapy (either by reducing the frequency of drug administration or discontinuing the biologic altogether and then monitoring for disease recurrence) in patients in whom they believe remission has been achieved. In addition, a number of ongoing studies are assessing factors which predict the persistence of remission

following the reduction or withdrawal of biologic therapy. Models may well need to take this approach to biologic use into account in the future.

4.3. Commentary from Dr James Galloway, clinical lecturer and honorary consultant in rheumatology, Kings College London.

I enjoyed reading the document – it is clear some very careful thought has gone into the project.

The first very general comment I would feedback relates to Professor Silman's original question, which if I am not mistaken relates to whether head to head comparisons of biologics are acceptable using indirect methods by either pooling trial data or observational data using network meta-analytic methods. It is clear that this approach is being used already, and numerous publications exist comparing efficacy of biologics in the observational setting – so almost irrespective of whether I agree with the principle of performing head-to-head comparisons in this manner, it seems wise for the consensus group to offer advice on the process… if you like, an extension to STROBE recommendations for comparative effectiveness work in rheumatology trials.

Second, there is the question of whether or not there is a case for randomised trials to compare the different biologic agents in a head to head manner. I think this particular question deserves more thought in the consensus document – partly because I was under the impression that this was a key part of the initial question. My own view is that head to head trials may not be justifiable as they would cost enormous amounts, and the currently available comparisons do not give me the impression that great differences will be apparent. Relating to this point, a key area that I would be keen to see discussed is what level of difference in efficacy between agents would be clinically important. If clinicians felt that differences of less than 5% would be clinically important, then to perform an RCT with adequate power to establish superiority would require vast numbers of participants. For example to identify a 5% difference in ACR 70 assuming 30% response in the arm A and 35% in arm B, and not allowing for any loss to follow up or dropout… sample sizes of 1800 per arm would be needed. Myself – I don't think 5% difference is worth knowing about if it requires that much effort (I am not entirely sure if it would be ethical either).

If on the other hand, we felt that we would only be interested if the differences in efficacy were much greater (20% difference between agents for example), then the numbers would be more manageable in terms of sample size – but I suspect we would have been able to detect this magnitude of difference using currently available data and the proposed meta-analytical techniques.

A further very general comment relates to the title of the document – namely using the word biologics… as many of the new drugs coming into play with be small chemical entities again – e.g. the kinase inhibitors. Perhaps incorporating the concept into the document (e.g. consensus decision models for highly targeted therapies in rheumatoid…) would not change the core content but would broaden the generalizability of the work?

I have added some more specific comments below.

- Introduction and background (section 1) : In light of the recent publication by Ben Goldacre and Fiona Goodlee in the BMJ regarding transparency in pharmaceutical drug trials, it would be nice to see a brief recommendation in the consensus model reflecting our commitment to transparency. Demonstrating a commitment to more widespread access to the decision models would be a positive step surely?
- Industry NICE submissions (section 1): It is highlighted that to date, all NICE approvals have depended upon cost-effective models versus placebo. This is unlikely to be sustainable in the future as it is becoming ethically more challenging to offer patients with active disease placebo in good faith – given the range of available options now available. I would suggest this point was highlighted.
- The EVI approach to assessing value of head to head trials is important (section 1). This section is clear and well argued. However on page 3 half way through the first paragraph I was interested to see the remarks about a consensus model resulting in more streamlined applications by industry with a reduced likelihood of requests for additional data. Am I correct in thinking that this consensus model is an independent panel without vested interests in industry?
- Measures of disease burden (section 2.1): Well written section. No obvious gaps that I spotted, and nothing controversial from my point of view.
- Sources of data (section 2.2): The highlighted examples of different data sources were appropriate. I think this section would benefit from some mention of the issue around bias and confounding in the observational setting. Confounding by indication is perhaps the elephant in the room with respect to registry data, and referring to the EULAR consensus statement on biologics registries (by Dixon / Askling et al) might be sensible. No need to reinvent a wheel.
- Effect modification (section 3.2): This is a crucial point – and I think it is important that the consensus document addresses this area and highlights the need for

comparisons between agents to incorporate this aspect. These issues are well discussed – yet in the summary statement 3.2.3 the second bullet point says that robust evidence for effect modification has not been identified! I thought this was at odds with the previous statements… and not sure I agree with the recommendation not to address effect modification in future modelling.

- DAS/ACR/HAQ mapping (section 3.2.4) Good section – and further research in this area is clearly warranted. Excellent points.
- Disease progression during treatment (section 3.3.1) There is limited mention of the issues surrounding the assumption of linear progression of HAQ over time. It is clear that modelling HAQ over time is not straight forward. Even patients in remission will progress with respect to HAQ – just due to ageing alone. HAQ trajectories have been studied in considerable detail in the ERAS / ERAN datasets by Adam Young's group. This issue was mentioned briefly in the discussion group – and alludes to the on-going work. Some of this work is now published and could be used.
- Summary of consensus view (section 3.3.3): I agree that HAQ remains the standard for mapping disease progression, but that additional thought needs to be put into addressing other outcomes – not just pain but also depression. The third bullet point I found unclear. It seems that the study group were discussing how long individuals remained on a particular drug, and how best to model this. One approach suggested was to extrapolate from trial data, while another was to use registry data. I had the feeling that the group expressed concerns about the use of observation data. There are clearly limitations to both approaches. Perhaps a sensible way forward would be to develop a study group to specifically look at the predictors of drug survival in RCT data and separately in observational data? This information could then inform subsequent strategies. There is no doubt at all that modelling drug survival needs to incorporate the reason for drug stopping (adverse event or inefficacy) - and this is highlighted by the group clearly.
- Modelling strategies (section 3.5): I must confess that I struggled to follow some of the discussion regarding modelling. I would consider myself reasonably well versed on the subject, and still got somewhat lost in the text. It strikes me that one of the great challenges around economic modelling is communication of the methodology. The paragraph at the top of page 22 for example: "Given the alternative approaches to model structure in existing models, future research should involve developing models that follow the consensus approach as closely as possible whilst adopting alternative structures, to evaluate how closely each model structure is able to follow the consensus approach and the impact of structure on model results." Is this saying that health economic modelling should involve some degree of model diagnostics?

4.4. Commentary from Professor David L Scott, Professor of clinical rheumatology, Kings College London

This is an excellent summary of a complex area from a group of leading experts. To object about their approach or conclusions would be mistaken. I think they have provided the best possible consensus of current views. However, that does not make their individual views or the overall consensus correct. My goal in this brief commentary is to summarise as concisely as I can some of the objections to this sort of economic modelling in patients with inflammatory arthritis.

One crucial limitation is the use of composite scores like DAS28. Whilst using the DAS28 has many benefits, including simplifying a complex assessment and establishing a uniform assessment method, it also has substantial limitations. DAS28 combines joint counts with an assessment of the acute phase response. This can be equated to mixing chalk with cheese; it is not logical. In addition, patients with mild rheumatoid arthritis who also have other problems which affect individual components of the DAS28 may have disproportionately high DAS28 scores. For example, patients with fibromyalgia, which often gives high tender joint counts [1], or with pneumonia, which may result in a high ESR, may be scored as having active RA when their actual disease activity is low. These objections do not invalidate the use of composite scores, but they do mean there are major limitations in generalising from changes in them. Indeed, there is a substantial body of evidence that assessing outcomes to treatment needs to consider all the components of the DAS28 as well as the overall score itself [2].

A second and more important point is the modelling often involves comparing recent trial data with historical observational data. This approach is not very convincing because the severity and impact of inflammatory arthritis appears to be lessening over time. There is substantial evidence for temporal changes in the clinical picture of rheumatoid arthritis, which appears to becoming less severe over time [3]. Whether this reflects differences in the disease, its treatment or the organisation of healthcare leading to more mild cases being seen in specialist clinics is uncertain. However, changes in the severity of present day disease make it difficult to interpret the comparison of current outcomes with historical outcomes. Any modern treatment is likely to achieve better results compared to what happened in earlier years.

A final problem is trying to interpret the value patients place upon complex cultural issues such as quality of life using simple questionnaires like HAQ and EQ5D. Whilst patients with low HAQ scores will generally have good function and vice versa using these questionnaires across time and different patient groups creates all sorts of distortions in patients with long-term inflammatory diseases like rheumatoid arthritis. One issue is that the impact of treatment on HAQ scores varies at different time-points in the course of rheumatoid arthritis [4]. An associated problem is that measures such as HAQ are influenced by

ethnicity, social class and a range of non-disease factors [5]. In addition, there are likely to be specific factors in some patients with rheumatoid arthritis which make them rank their health status as being “worse than death” [6]. Relating changes in DAS28 to changes in HAQ or EQ5D is extremely challenging and may be inappropriate.

A recent systematic review of economic studies of biologics in rheumatoid arthritis [7] concluded that “Economic evaluations of biologics are hindered by lack of data on long-term responses and consequence of responses on downstream health utilization and productivity.” I think the problem of modelling responses based on short term clinical trial data is a central issue contributing to this uncertainty. None of this means economic modelling is either inappropriate or incorrect, or that the recommendations of the working group should not be implemented. The simple point is that when evaluating the economic benefits of high cost treatments for inflammatory arthritis investigators have to rely on a very large number of assumptions. Some of these assumptions and, possibly many of them, might not prove ideal approaches. On balance some measures are better than no measures, and a flawed model is preferable to no model at all. Nonetheless the problem remains almost overwhelming difficult.

1. Pollard LC, Kingsley GH, Choy EH, Scott DL. Fibromyalgic rheumatoid arthritis and disease assessment. Rheumatology 2010; 49: 924-8.
2. Ton E, Bakker MF, Verstappen SM, Ter Borg EJ, van Albada-Kuipers IA, Schenk Y,van der Veen MJ, Bijlsma JW, Jacobs JW; Utrecht Rheumatoid Arthritis Cohort Study Group. Look beyond the disease activity score of 28 joints (DAS28): tender points influence the DAS28 in patients with rheumatoid arthritis. J Rheumatol 2012; 39: 22-7.
3. Uhlig T, Kvien TK. Is rheumatoid arthritis really getting less severe? Nat Rev Rheumatol 2009; 5: 461-4.
4. Plant MJ, O'Sullivan MM, Lewis PA, Camilleri JP, Coles EC, Jessop JD. What factors influence functional ability in patients with rheumatoid arthritis. Do they alter over time? Rheumatology 2005; 44: 1181-5.
5. Barton JL, Trupin L, Schillinger D, Gansky SA, Tonner C, Margaretten M,Chernitskiy V, Graf J, Imboden J, Yelin E. Racial and ethnic disparities in disease activity and function among persons with rheumatoid arthritis from university-affiliated clinics. Arthritis Care Res 2011; 63: 1238-46.
6. Harrison MJ, Davies LM, Bansback NJ, McCoy MJ, Farragher TM, Verstappen SM,Hassell A, Symmons DP. Why do patients with inflammatory arthritis often score states "worse than death" on the EQ-5D? An Investigation of the EQ-5D classification system. Value Health 2009; 12: 1026-34.
7. van der Velde G, Pham B, Machado M, Ieraci L, Witteman W, Bombardier C, Krahn M. Cost-effectiveness of biologic response modifiers compared to disease-modifying

antirheumatic drugs for rheumatoid arthritis: a systematic review. Arthritis Care Res 2011; 63: 65-78.

4.5. Commentary from Professor John Kirwan, Professor of Rheumatic Diseases, University of Bristol.

This report represents an impressive amount of work and thought, and is presented in a readily digestible format. Specific comments on each section are given below.

Comments on Section 1.

The case for the two consensus conferences is well made in describing the conduct and outcome of the first meeting. In effect, the main conclusion from part 1 of the report is that transparency is key to understanding the different models used and the potential for disagreements between their results given the same input data.

Comments on Section 2.

Registries report the progress of patients who were treated according to practice when they entered the registry. The treatment of rheumatoid arthritis has been improving, particularly for newly diagnosed patients. (1,2) The comparator outcomes for (at least newly diagnosed) RA patients should be in accordance with up to date treatment policies. (3)

Conventional DMARDs is an unacceptable phrase. Initial treatment for newly diagnosed RA is recommended by NICE to be a combination of DMARDs including glucocorticoids. (3) All recent evidence suggests this is a powerful method for treating newly diagnosed RA. (4) It is as good as a biologic agent in the single head-to-head study that has been conducted (the BsSt study (5)). Thus, the correct treatment for comparison in newly diagnosed RA patients in the ARC low dose glucocorticoid trial (6) treatment or the COBRA trial (7) treatment. No clear cut best comparison data are available for patients with established RA. Therefore newly diagnosed RA and established RA should be modelled separately. This is an important area which does not seem to have been considered at these meetings.

Comments on Section 3.

The issue of new and established RA is central to the whole of Section 3. This is crucial to the section on estimating baseline response in the comparator treatment. For newly diagnosed patients, using inadequate treatment in the comparator arm (for example, omitting glucocorticoids from the treatment regimen) will result in a falsely low estimate of the outcome in the comparator arm, and hence an overestimate of the comparative benefit in the treatment arm. As far as I know, the only DMARD comparator arm to include up to date

treatment of newly diagnosed RA (combination of DMARDs plus glucocorticoids) is the BeSt study. (5) Therefore, I believe the summary of the consensus view which states: “Response rates to the non-biologic comparator can be based on pooling control arms from biologic trials” is fundamentally flawed, at least as far as newly diagnosed RA is concerned.

There is a problem here with short term adverse events as well. Treatment of newly diagnosed RA using a combination of DMARDs plus glucocorticoids usually results in less adverse effects than found when treating with ‘conventional’ DMARDs. (8)

Modelling disease progression with the HAQ introduces a complication which has not been addressed in this report. The HAQ (a measure of disability) represents a combination of fluctuating joint inflammation and accumulating but irreversible joint damage. (9) Issues such as expected change in HAQ as inflammation is treated are thus quite dependant on the extent of underlying joint damage, most likely reflected in disease duration. Early control of joint damage will result in a HAQ score more likely to rebound after anti-inflammatory treatment is discontinued, but poor control of joint damage progression will mean that the HAQ is less able to reflect any subsequent change in joint inflammation when treatment is discontinued. The HAQ should be modelled with two internal components to reflect this.

Health care utilisation is likely to depend on the choice of approach to follow-up by the clinicians looking after patients at a particular department. Simply changing the approach to follow-up can result in a reduction in cost of 30%. (10) Models should be tested for sensitivity to different approaches to follow up.

References

1. Combination Therapy Including Glucocorticoids: The New Gold Standard for Early Treatment in Rheumatoid Arthritis? Annals of Internal Medicine 29012; 156: 390-91

2. Bakker MF, Jacobs JW, Welsing PM, Verstappen SM, Tekstra J, Ton E,et al. Low-dose prednisone inclusion into a methotrexate-based, tight control strategy for early rheumatoid arthritis. A randomized trial. Ann Intern Med. 2012;156:329-39.

3. Rudolph, Michael. 2009. NICE: The management of rheumatoid arthritis in adults. National Institute for Clinical Excellence.

4. Bijlsma JWJ, Marlies van der Goes M, Hoes JN, Jacobs JWG, Buttgereit F, Kirwan J. Low-dose glucocorticoid therapy in rheumatic diseases: an obligatory therapy. Ann N Y Acad Sci 2010; 1193: 123-126.

5. Goekoop-Ruiterman YP, de Vries-Bouwstra JK, Allaart CF, van Zeben D, Kerstens PJ, Hazes JM, et al. Clinical and radiographic outcomes of four different treatment strategies in patients with early rheumatoid arthritis (the BeSt study): a randomized, controlled trial. Arthritis Rheum. 2005;52:3381-90.

6. Kirwan JR. The effect of glucocorticoids on joint destruction in rheumatoid arthritis. The Arthritis and Rheumatism Council Low-Dose Glucocorticoid Study Group. N Engl J Med. 1995;333:142-6.
7. BOERS M, VERHOEVEN AC, MARKUSSE HM et al.: Randomised comparison of combined step-down prednisolone, methotrexate and sulphasalazine with sulphasalazine alone in early rheumatoid arthritis. Lancet 1997; 350: 309-18.
8. Da Silva JA, Jacobs JW, Kirwan JR, Boers M, Saag KG, Ineˆs LB, et al. Safety of low dose glucocorticoid treatment in rheumatoid arthritis: published evidence and prospective trial data. Ann Rheum Dis. 2006;65:285-93.
9. Kirwan JR. Links between radiological change, disability, and pathology in rheumatoid arthritis. J Rheumatol. 2001;28:881-6.
10. Hewlett S, Kirwan J, Pollock J, Mitchell K, Hehir M, Blair P, Memel D. Follow-up in chronic disease: A randomised controlled trial of routine rheumatologist-initiated review versus patient-initiated review in rheumatoid arthritis, extended to 6 years. British Medical Journal 2005; 330 (7484): 171-7.

4.6. Commentary from Professor Paul Emery, ARUK Professor of Rheumatology, University of Leeds

My main issue with this report is that there is insufficient discussion of issues relating to the quality-of-life of arthritis patients. Measures of function do not sufficiently capture all dimensions of quality-of-life in these patients, particularly in the case of Psoriatic arthritis. While work has been carried out on assessing quality-of-life in this group, there is clearly a need for further research in this area.

Consensus document references

1. *Adalimumab, Etanercept and Infliximab for the treatment of rheumatoid arthritis (TA130)* 2007, National Institute for Health and Clinical Excellence.
2. *Etanercept, infliximab and adalimumab for the treatment of psoriatic arthritis (TA199)*, 2010, National Institute for Health and Clinical Excellence.
3. Bucher, H.C., et al., *The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials.* Journal of Clinical Epidemiology, 1997. **50**(6): p. 683-691.
4. Caldwell, D.M., A.E. Ades, and J.P.T. Higgins, *Simultaneous comparison of multiple treatments: combining direct and indirect evidence.* British Medical Journal, 2005. **331**(7521): p. 897-900.
5. Lu, G. and A.E. Ades, *Combination of direct and indirect evidence in mixed treatment comparisons.* Statistics in Medicine, 2004. **23**(20): p. 3105-3124.
6. Lu, G.B. and A.E. Ades, *Assessing evidence inconsistency in mixed treatment comparisons.* Journal of the American Statistical Association, 2006. **101**(474): p. 447-459.
7. Silman, A.J., *Use of biologic agents in rheumatoid arthritis: introduction.* Rheumatology, 2011. **50**: p. iv3-iv4.
8. Jobanputra, P., *A clinician's critique of rheumatoid arthritis health economic models.* Rheumatology, 2011. **50**: p. iv48-iv52.
9. Tosh, J., et al., *The Sheffield rheumatoid arthritis health economic model.* Rheumatology, 2011. **50**: p. iv26-iv31.
10. Barton, P., *Development of the Birmingham Rheumatoid Arthritis Model: past, present and future plans.* Rheumatology, 2011. **50**: p. iv32-iv38.
11. Bojke, L., et al., *Modelling the cost-effectiveness of biologic treatments for psoriatic arthritis.* Rheumatology, 2011. **50**: p. iv39-iv47.
12. Madan, J., A.E. Ades, and N.J. Welton, *An overview of models used in economic analyses of biologic therapies for arthritis-from current diversity to future consensus.* Rheumatology, 2011. **50**: p. iv10-iv18.
13. Ades, A.E., J. Madan, and N.J. Welton, *Indirect and mixed treatment comparisons in arthritis research.* Rheumatology, 2011. **50**: p. iv5-iv9.
14. Welton, N.J., J. Madan, and A.E. Ades, *Are head-to-head trials of biologics needed? The role of value of information methods in arthritis research.* Rheumatology, 2011. **50**: p. iv19-iv25.
15. Felli, J.C. and G.B. Hazen, *Sensitivity analysis and the expected value of perfect information.* Medical Decision Making, 1998. **18**(1): p. 95-109.
16. Claxton, K. and J. Posnett, *An economic approach to clinical trial design and research priority-setting.* Health Economics, 1996. **5**(6): p. 513-524.
17. Claxton, K., *The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies.* Journal of Health Economics, 1999. **18**(3): p. 341-364.
18. Bruce, B. and J.F. Fries, *The Stanford Health Assessment Questionnaire: A review of its history, issues, progress, and documentation.* Journal of Rheumatology, 2003. **30**(1): p. 167-178.
19. Felson, D.T., et al., *American-College-of-Rheumatology Preliminary Definition of Improvement in Rheumatoid-Arthritis.* Arthritis and Rheumatism, 1995. **38**(6): p. 727-735.
20. van der Heijde, D.M., et al., *Development of a disease activity score based on judgment in clinical practice by rheumatologists.* The Journal of rheumatology, 1993. **20**(3): p. 579-581.
21. Fransen, J. and P.L. van Riel, *The Disease Activity Score and the EULAR response criteria.* Clin Exp Rheumatol, 2005. **23**(5 Suppl 39): p. S93-9.
22. Mease, P.J., et al., *Psoriatic arthritis assessment tools in clinical trials.* Ann Rheum Dis, 2005. **64 Suppl 2**: p. ii49-54.

23. Ashcroft, et al., *Clinical measures of disease severity and outcome in psoriasis: a critical appraisal of their quality.* British Journal of Dermatology, 1999. **141**(2): p. 185-191.
24. Coates, L.C., et al., *Development of a Disease Severity and Responder Index for Psoriatic Arthritis (PsA) — Report of the OMERACT 10 PsA Special Interest Group.* The Journal of rheumatology, 2011. **38**(7): p. 1496-1501.
25. Lloyd, S., et al., *The effectiveness of anti-TNF-α therapies when used sequentially in rheumatoid arthritis patients: a systematic review and meta-analysis.* Rheumatology, 2010. **49**(12): p. 2313-2321.
26. Zink, A., et al., *European biologicals registers: methodology, selected results and perspectives.* Ann Rheum Dis, 2009. **68**(8): p. 1240-1246.
27. Silman, A., et al., *British Society for Rheumatology Biologics Register.* Ann Rheum Dis, 2003. **62**(suppl 2): p. ii28-ii29.
28. Symmons, D.P. and A.J. Silman, *The Norfolk Arthritis Register (NOAR).* Clin Exp Rheumatol, 2003. **21**(5 Suppl 31): p. S94-9.
29. Wolfe, F. and K. Michaud, *The National Data Bank for rheumatic diseases: a multi-registry rheumatic disease data bank.* Rheumatology, 2011. **50**(1): p. 16-24.
30. Hetland, M.L., *DANBIO—powerful research database and electronic patient record.* Rheumatology, 2011. **50**(1): p. 69-77.
31. Kvien, T.K., et al., *A Norwegian DMARD register: prescriptions of DMARDs and biological agents to patients with inflammatory rheumatic diseases.* Clin Exp Rheumatol, 2005. **23**(5 Suppl 39): p. S188-94.
32. de Rooy, D.P.C., et al., *Predicting arthritis outcomes-what can be learned from the Leiden Early Arthritis Clinic?* Rheumatology, 2011. **50**(1): p. 93-100.
33. Vermeer, M., et al., *Implementation of a Treat-to-Target Strategy in Very Early Rheumatoid Arthritis Results of the Dutch Rheumatoid Arthritis Monitoring Remission Induction Cohort Study.* Arthritis and Rheumatism, 2011. **63**(10): p. 2865-2872.
34. *Rheumatoid arthritis: the management of rheumatoid arthritis in adults (CG79)* 2009, National Institute for Health and Clinical Excellence.
35. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0*, J.P.T. Higgins and S. Green, Editors. 2011, The Cochrane Collaboration.
36. *Guide to the methods of technology appraisal*, 2008, National Institute for Health and Clinical Excellence.
37. Chen, Y.F., et al., *A systematic review of the effectiveness of adalimumab, etanercept and infliximab for the treatment of rheumatoid arthritis in adults and an economic evaluation of their cost-effectiveness.* Health technology assessment (Winchester, England), 2006. **10**(42): p. iii-iv, xi-xiii, 1-229.
38. Geborek, P., et al., *Etanercept, infliximab, and leflunomide in established rheumatoid arthritis: clinical experience using a structured follow up programme in southern Sweden.* Ann Rheum Dis, 2002. **61**(9): p. 793-798.
39. Brennan, A., et al., *Modelling the cost effectiveness of TNF-α antagonists in the management of rheumatoid arthritis: results from the British Society for Rheumatology Biologics Registry.* Rheumatology, 2007. **46**(8): p. 1345-1354.
40. Woolacott, N., et al., *Etanercept and efalizumab for the treatment of psoriasis: a systematic review.* Health technology assessment (Winchester, England), 2006. **10**(46): p. 1-233, i-iv.
41. Rodgers, M., et al., *Etanercept, infliximab and adalimumab for the treatment of psoriatic arthritis: a systematic review and economic evaluation.* Health technology assessment (Winchester, England), 2011. **15**(10): p. i-xxi, 1-329.
42. Nixon, R.M., N. Bansback, and A. Brennan, *Using mixed treatment comparisons and meta-regression to perform indirect comparisons to estimate the efficacy of biologic treatments in rheumatoid arthritis.* Statistics in Medicine, 2007. **26**(6): p. 1237-1254.

43. Mease, P.J., D.D. Gladman, and G.G. Krueger, *Prologue: Group for Research and Assessment of Psoriasis and Psoriatic Arthritis (GRAPPA).* Ann Rheum Dis, 2005. **64**(suppl 2): p. ii1-ii2.
44. Helliwell, P.S., et al., *The development of candidate composite disease activity and responder indices for psoriatic arthritis (GRACE project).* Annals of the Rheumatic Diseases, 2012.
45. Bongartz T, S.A.J.S.M.J.B.I.M.E.L.M.V., *Anti-tnf antibody therapy in rheumatoid arthritis and the risk of serious infections and malignancies: Systematic review and meta-analysis of rare harmful effects in randomized controlled trials.* JAMA: The Journal of the American Medical Association, 2006. **295**(19): p. 2275-2285.
46. Bojke, L., et al., *Eliciting distributions to populate decision analytic models.* Value Health, 2010. **13**(5): p. 557-64.
47. Hurst, N.P., et al., *Measuring health-related quality of life in rheumatoid arthritis: validity, responsiveness and reliability of EuroQol (EQ-5D).* Rheumatology, 1997. **36**(5): p. 551-559.
48. Brennan, A., et al., *Modelling the cost-effectiveness of etanercept in adults with rheumatoid arthritis in the UK.* Rheumatology (Oxford, England), 2004. **43**(1): p. 62-72.
49. Brazier, J.E., et al., *A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures.* The European journal of health economics : HEPAC : health economics in prevention and care, 2010. **11**(2): p. 215-225.
50. Alava, M.H., A.J. Wailoo, and R. Ara, *Tails from the Peak District: Adjusted Limited Dependent Variable Mixture Models of EQ-5D Questionnaire Health State Utility Values.* Value in Health, 2012. **15**(3): p. 550-561.
51. Kobelt, G., et al., *Modeling the progression of rheumatoid arthritis: A two-country model to estimate costs and consequences of rheumatoid arthritis.* Arthritis & Rheumatism, 2002. **46**(9): p. 2310-2319.
52. Wolfe, F., et al., *The mortality of rheumatoid arthritis.* Arthritis & Rheumatism, 1994. **37**(4): p. 481-494.
53. Yelin, E. and L.A. Wanke, *An assessment of the annual and long-term direct costs of rheumatoid arthritis: The impact of poor function and functional decline.* Arthritis & Rheumatism, 1999. **42**(6): p. 1209-1218.
54. Kobelt, G., et al., *Economic consequences of the progression of rheumatoid arthritis in Sweden.* Arthritis & Rheumatism, 1999. **42**(2): p. 347-356.
55. Kvamme, M.K., et al., *Two-year direct and indirect costs for patients with inflammatory rheumatic joint diseases: data from real-life follow-up of patients in the NOR-DMARD registry.* Rheumatology, 2012.
56. Lunt, M., et al., *No evidence of association between anti–tumor necrosis factor treatment and mortality in patients with rheumatoid arthritis: Results from the British Society for Rheumatology Biologics Register.* Arthritis & Rheumatism, 2010. **62**(11): p. 3145-3153.
57. Kobelt, G., et al., *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**(8): p. 1174-1179.
58. Barton, P., S. Bryan, and S. Robinson, *Modelling in the economic evaluation of health care: selecting the appropriate approach.* Journal of health services research & policy, 2004. **9**(2): p. 110-118.
59. Brennan, A., S.E. Chick, and R. Davies, *A taxonomy of model structures for economic evaluation of health technologies.* Health Economics, 2006. **15**(12): p. 1295-1310.
60. Sculpher, M., E. Fenwick, and K. Claxton, *Assessing quality in decision analytic cost-effectiveness models. A suggested framework and example of application.* PharmacoEconomics, 2000. **17**(5): p. 461-477.

Appendix 1. Members of the Consensus Working Party

Participant	Organisation	Relevant expertise and experience
Professor A E Ades (chair)	School of Social & Community Medicine, University of Bristol	Evidence synthesis methodology, Member of NICE Appraisals Committee since 2003
Dr Paresh Jobanputra	Department of Rheumatology, Selly Oak Hospital, Birmingham	Clinical specialist, co-author of multiple NICE technology appraisals of biologic therapies.
Professor Ernest Choy	Cardiff University School of Medicine	Clinical specialist, Expert advisor to NICE, member of EULAR.
Dr Philip Helliwell	Chapel Allerton Hospital, Leeds and St Luke's hospital, Bradford.	Clinical specialist, Expert advisor to evaluation group for NICE PsA guidance, member of GRAPPA.
Professor Andrew Stevens	Department of Primary Care, Public and Occupational Health, University of Birmingham.	Health technology assessment, Chair of a NICE appraisal committee since 2005.
Professor Ken Stein	PenTAG, Peninsular College of Medicine & Dentistry	Health Technology assessment, member of NICE appraisal committee, Director of a Technology Appraisal Group
Dr Suzanne Verstappen	ARUK Epidemiology Unit, University of Manchester	Arthritis Epidemiology, Member of NOAR staff.
Dr Pelham Barton	School of Health and Population Sciences, University of Birmingham	Health economic modelling, developer of the BRAM.

Dr Allan Wailoo	School of Health and Related Research, University of Sheffield.	Health economics / modelling, Director of the NICE DSU, co-developer of the Sheffield RA model.
Mr Jon Tosh	School of Health and Related Research, University of Sheffield.	Health economics / modelling, member of the NICE DSU and ScHARR-TAG, co-developer of the Sheffield RA model.
Dr Laura Bojke	Centre for Health Economics, University of York	Health economics / modelling, co-developer of the York PsA model.
Dr Jason Madan	School of Social & Community Medicine, University of Bristol	Health economics / modelling, Evidence Synthesis.

.

Appendix 2: List of Biologics Trials that have informed NICE technology appraisals of biologics for rheumatoid and psoriatic arthritis

PSA: psoriatic arthritis, RA: rheumatoid arthritis, NULL: placebo, ETA: etanercept, IFX: infliximab, DMARDs: disease modifying anti-rheumatic drugs, ADL: adalimumab, MTX: methotrexate, GOL: golimumab, SUL:sulphasalazine, TOC: tocilizumab, RTX: rituximab, ABA: Abatercept CZP: certolizumab pegol.

Table A1.1: Trials of biologics in patients with PSA

Lead Author / Trial name	Year	Treatment History	Treatments	mean age	Mean Duration (years)	Mean Prior DMARDs	Mean HAQ	Outcome measures	Follow-up
Mease	2000	Biologic naive, 1+ DMARDs	ETA, NULL	43.5-46.0	9	1.5,2.0	1.3,1.2	ACR, PsARC, PASI, HAQ	12 weeks
Mease	2004	Biologic naive, 1+ DMARDs	ETA, NULL	47.3-47.6	9 - 9.2	1.6,1.7	1.1,1.1	ACR, PsARC, PASI, HAQ, Sharp	24 weeks
IMPACT	2005	Biologic naive, 1+ DMARDs	IFX, NULL	45.2-45.7	8.5 - 8.7	?	?	ACR, PASI, DAS28, HAQ, PsARC	16 weeks
IMPACT 2	2005	Biologic naive, 1+ DMARDs	IFX, NULL	46.5-47.1	7.5 - 8.4	?	1.1	ACR, PsARC, PASI, HAQ, SF36	24 weeks
ADEPT	2005	Biologic naive, 1+	ADL, NULL	48.6-49.2	9.2 - 9.8	1.5	1.5	ACR, PsARC, PASI,	24 weeks

		DMARDs						HAQ, SF36	
Genovese	2007	Biologic naive, 1+ DMARDs	ADL, NULL	47.7-50.4	7.2 - 7.5	1.7	2.1	ACR, PsARC	12 weeks
GO-REVEAL	2009	Biologic naive, DMARD / NSAID failure	GOL (+MTX), NULL (+MTX)	45.7 - 48.2	7.2 - 7.7	NR	NR	ACR, Sharp(VdH) PsARC	24 weeks

Table A1.2 Trials of biologics in biologic-naive early RA patients (mean duration < 2 years).

Lead Author / Trial name	**Year**	**Treatment History**	**Treatments**	**mean age**	**Mean Duration (years)**	**Mean Prior DMARDs**	**Mean HAQ**	**Outcome measures**	**Follow-up**
PREMIER	2004	Biologic naive, 1+ DMARDs	ADL, ADL+MTX, MTX	52	0.7-0.8	NR	1.47-1.63	ACR, Sharp. DAS28	2 years
ERA	2000	Early RA, biologic naive	ETA, MTX	49-51	0.9 - 1.0	0.5-0.6	1.4-1.5	ACR, Sharp	12 months
ASPIRE	2004	Early RA, biologic and MTX naive	IFX+MTX, MTX	50-51	0.8 - 0.9	65% - 71% DMARD naive	1.5	ACR, DAS, Sharp (VDH)	54 weeks
Taylor	2004	Early RA, MTX failure	IFX+MTX, MTX	51-55	1.3 - 1.6	NR	NR	ACR, DAS, Sharp (VDH)	54 weeks

Quinn	2005	Early RA, biologic and DMARD naive	IFX+MTX, MTX	51-53	0.5-0.6	0	1.3	DAS28, HAQ	54 weeks

Table A1.3 Trials of biologics in biologic-naive late RA patients (mean duration >4 years).

Lead Author / Trial name	Year	Treatment History	Treatments	mean age	Mean Duration (years)	Mean Prior DMARDs	Mean HAQ	Outcome measures	Follow-up
den Broeder	2002	Biologic naive, 1+ DMARDs	ADL, NULL	53-59	8.9-14.5	3.6-4.4	1.41-1.93	DAS, EULAR, ACR,	4 weeks
Weisman	2003	Biologic naive, 1+ DMARDs	ADL, NULL	50-56	13-18	NR	0.9-1.5	ACR, HAQ	4 weeks
van de Putte	2003	Biologic naive, 1+ DMARDs	ADL, NULL	50-54	9-10	3.5-4.1	1.63-1.79	ACR, HAQ, DAS28	12 weeks
ARMADA	2003	Biologic naive, 1+ DMARDs	ADL+MTX, MTX	54-57	11-13	2.9-3.1	1.52-1.64	ACR, HAQ	24 weeks
Rau	2004	Biologic naive, 1+ DMARDs	ADL+MTX, MTX	52-54	11-12	3.3-3.5	1.32-1.38	ACR, EULAR, HAQ	4 weeks

van de Putte	2004	Biologic naive, 1+ DMARDs	ADL, NULL	52-54	9-12	3.6-3.8	1.83-1.88	ACR, EULAR, HAQ, DAS	26 weeks
Keystone	2004	Biologic naive, 1+ DMARDs	ADL+MTX, MTX	56-57	11	2.4	1.44-1.48	ACR, SF36, Sharp	52 weeks
STAR	2003	Biologic naive, 1+ DMARDs	ADL+DMARD, DMARD	55-56	9 - 12	NR	1.37-1.43	ACR	24 weeks
Moreland	1996	Biologic naive, 1+ DMARDs	ETA, NULL	53-62	4.3 - 19.8	NR	NR	?	4 weeks
Moreland	1997	Biologic naive, 1+ DMARDs	ETA, NULL	52-55	71%-80% > 5years	?	?	?	3 months
Moreland	1999	Biologic naive, 1+ DMARDs	ETA, NULL	51-53	11 - 13	3.0 - 3.4	1.6-1.7	ACR, HAQ	6 months
Weinblatt	1999	Biologic naive, 1+ DMARDs	ETA+MTX, MTX	48-53	13	2.7 - 2.8	1.3 - 1.5	ACR	24 weeks

TEMPO	2004	Biologic naive, 1+ DMARDs	MTX, ETA, ETA+ MTX	53	6.3-6.8	2.3	1.7 - 1.8	ACR, SHARP,	52 weeks / 2 years
Keystone	2004	Biologic naive	ETA+ MTX, NULL	52 - 54	8.2 - 10.8	88%-90% prior use	1.4	ACR, HAQ	8 weeks
Maini	1998	Biologic naive	IFX, IFX+MTX, MTX	47-56	7.6 - 14.3	2 -3	1.4 -2.0	Paulus	26 weeks
ATTRACT	1999	Biologic naive, 1+ DMARDs	IFX+MTX, MTX	51-54	9 - 12	2.5 - 2.8	1.7 - 1.8	ACR, HAQ	54 weeks
Durez	2004	Biologic naive, MTX failure	IFX+ MTX, MPS	48 - 56 (median)	10-12 (median)	3 (median)	1.3 - 1.5 (median)	ACR, SF36, HAQ	14 weeks
GO-FORWARD	2009	Biologic naive, MTX failure	GOL+MTX, MTX	50.0 - 52.0 median	4.5 - 6.7 median	NR	1.25 - 1.375 median	ACR, HAQ, , SF36	24 weeks
Kay	2008	Biologic naive, MTX failure	GOL+MTX, MTX	48.0 - 57.5	5.6 - 9.0	NR	1.3-1.8	ACR, DAS	16 weeks

OPTION	2008	Biologic naive, MTX failure	TOC+MTX, MTX	50.6 - 51.4	7.4 - 7.8	1.5 - 1.7	1.5 - 1.6	ACR, DAS, EULAR	24 weeks
LITHE	2011	Biologic naive, DMARD failure	TOC+MTX, MTX	51.3 - 53.4	9.0 - 9.4	1.6 - 1.7	1.5	HAQ, ACR, DAS28	52 weeks
TOWARD	2008	Biologic naive, MTX failure	TOC + DMARD, DMARD	53 - 54	9.8	1.6	1.5	ACR, DAS	24 weeks
RAPID 1	2008	Biologic naive	CZP + MTX, MTX	51.4 - 52.4	6.1 - 6.2	1.3 - 1.4	1.7	ACR, HAQ	52 weeks
RAPID 2	2009	Biologic naive	CZP + MTX, MTX	51.5 - 52.2	5.6 - 6.5	1.2 - 1.3 (excluding MTX)	1.6	ACR, HAQ, DAS	24 weeks
FAST4WARD	2009	Biologic naive, DMARD failure	CZP, NULL	52.7 - 54.9	8.7 - 10.4	2	1.4 - 1.6	ACR, HAQ	24 weeks

table A1.4 Trials of biologics in patients with RA with prior exposure to biologics.

Lead Author / Trial name	Year	Disease	Treatment History	Treatments	mean age	Mean Duration (years)	Mean Prior DMARDs	Mean HAQ	Outcome measures	Follow-up
REFLEX	2006	RA	Biologic failure	RTX + MTX, MTX	52.2 - 52.8	11.7 - 12.1	2.4 - 2.6 excluding MTX, 1.5 anti-TNFs	1.9	ACR, EULAR, HAQ	24 weeks
ATTAIN	2005	RA	Biologic failure	ABA + DMARD, DMARD	52.7 - 53.4	11.4 - 12.2	NR	NR	DAS, ACR	6 months
Go-AFTER	2009	RA	Biologic experience	GOL (+MTX), NULL	54-55	8.7 -9.8	NR	1.5 - 1.8	ACR, HAQ	24 weeks

table A1.3 Trials of biologics in patients with RA and unknown prior treatment history

Lead Author / Trial name	Year	Disease	Treatment History	Treatments	mean age	Mean Duration (years)	Mean Prior DMARDs	Mean HAQ	Outcome measures	Follow-up
Wajdula	2000	RA	?	ETA, NULL	53-54	6.8 - 7.5	3.0 - 3.6	1.8 - 1.9	?	12 weeks
Codreanu	2003	RA	?	ETA, ETA+SUL, SUL	51-53	5.6 - 7.1	2.1 - 2.7	1.6 - 1.7		24 weeks
Lan	2004	RA	?	ETA+MTX, MTX	48 - 51	NR	NR	1.0 - 1.2	?	12 weeks
Baumgartner	2004	RA	?	ETA, NULL	60	10	NR	NR	?	16 / 20 weeks
Elliott	1994	RA	?	IFX, NULL	48-56	7.3-9.0	2.8-3.7	NR	Paulus	4 weeks
Kavanagh	2000	RA	?	IFX+MTX, MTX	37-53	4.9 - 7.5	NR	1.4 - 1.6	?	12 weeks
START	2004	RA	?	IFX+MTX, MTX	NR	NR	NR	NR	?	22 weeks