

# Consensus Decision Models for Biologics in Rheumatoid and Psoriatic Arthritis: Recommendations of a Multidisciplinary Working Party

Jason Madan · Tony Ades · Pelham Barton · Laura Bojke · Ernest Choy · Philip Helliwell · Paresh Jobanputra · Ken Stein · Andrew Stevens · Jonathan Tosh · Suzanne Verstappen · Allan Wailoo

Received: October 7, 2015 / Published online: November 25, 2015  
© The Author(s) 2015. This article is published with open access at Springerlink.com

## ABSTRACT

**Introduction:** Biologic therapies are efficacious but costly. A number of health economic models have been developed to determine the most cost-effective way of using them in the treatment pathway. These models have produced conflicting results, driven by differences in assumptions, model structure,

and data, which undermine the credibility of funding decisions based on modeling studies. A Consensus Working Party met to discuss recommendations and approaches for future models of biologic therapies.

**Methods:** Our working party consisted of clinical specialists, modelers, and policy makers. Two 1-day meetings were held for members to arrive at consensus positions on model structure, assumptions, and appropriate data sources. These views were guided by clinical aspects of rheumatoid and psoriatic arthritis and the principles of evidence-based

---

**Electronic supplementary material** The online version of this article (doi:[10.1007/s40744-015-0020-0](https://doi.org/10.1007/s40744-015-0020-0)) contains supplementary material, which is available to authorized users.

---

J. Madan · T. Ades  
School of Social and Community Medicine,  
University of Bristol, Bristol, UK

J. Madan (✉)  
Warwick Medical School, University of Warwick,  
Warwick, UK  
e-mail: [j.j.madan@warwick.ac.uk](mailto:j.j.madan@warwick.ac.uk)

P. Barton · A. Stevens  
School of Health and Population Sciences,  
University of Birmingham, Birmingham, UK

L. Bojke  
Centre for Health Economics, University of York,  
York, UK

E. Choy  
Cardiff Institute of Infection and Immunity, Cardiff  
University School of Medicine, Cardiff, UK

P. Helliwell  
Leeds and St Luke's Hospital, Bradford, UK

P. Jobanputra  
Queen Elizabeth Medical Centre, Birmingham, UK

K. Stein  
Peninsular Technology Assessment Group,  
University of Exeter Medical School, Exeter, UK

J. Tosh · A. Wailoo  
School of Health and Related Research, University of  
Sheffield, Sheffield, UK

S. Verstappen  
Arthritis Research UK Centre for Epidemiology, The  
University of Manchester, Manchester, UK

medicine. Where opinions differed, we sought to identify a research agenda that would generate the evidence needed to reach consensus.

**Results:** We gained consensus in four areas of model development: initial response to treatment; long-term disease progression; lifetime costs and benefits; and model structure. Consensus was also achieved on some key parameters such as choices of outcome measures, methods for extrapolation beyond trial data, and treatment switching. A research agenda to support further consensus was also identified.

**Conclusion:** Consensus guidance that fully reflects current evidence and clinical understanding was gained successfully. In addition, research needs have been identified. Such guidance can be updated as evidence develops and policy questions change and need not be prescriptive as long as deviations from consensus are clearly explained and justified.

**Funding:** Arthritis Research UK and the UK Medical Research Council Network of Hubs for Trials Methodology Research.

**Keywords:** Arthritis; Biologics; Economic models

## INTRODUCTION

Biologic therapies represent a recent addition to treatments for inflammatory joint diseases such as rheumatoid arthritis (RA) and psoriatic arthritis (PsA). While their efficacy has been established in a number of clinical trials and cost-effectiveness demonstrated in a number of assessments [1, 2], the evidence base is still associated with substantial 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 these chronic diseases. A workshop to explore these challenges took place in 2010, the proceedings of which were disseminated in a series of papers [3–10]. A key finding of the workshop was that, despite the importance of economic models in guiding policy on the adoption of biologic therapies, 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 contradictory recommendations. 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 to inform future policy decisions. With this in mind, a Consensus Working Party on decision models for biologic therapies in RA and PsA was formed to identify the current scope for consensus, and identify gaps in the evidence base where further research is needed to support future consensus.

## METHODS

The working party was set up to bring together key expertise as comprehensively as possible. Attendees included leading clinical experts, health economists involved in the development of the main existing cost-effectiveness models that have informed policy making in the UK, and key individuals from Health Technology Assessment organizations and significant funders of

**Table 1** Members of the Consensus Working Party

Participant	Organization	Relevant expertise and experience
Professor A. E. Ades (chair)	School of Social and Community Medicine, University of Bristol	Evidence synthesis methodology, member of NICE appraisals committee since 2003
Dr. Paresh Jobanputra	Queen Elizabeth Hospital, Queen Elizabeth Medical Centre, 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, University of Exeter Medical School	Health technology assessment, vice chair of NICE appraisal committee, director of PENTAG, representative of the UK HTA program
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 modeling, developer of the BRAM
Dr. Allan Wailoo	School of Health and Related Research, University of Sheffield	Health economics/modeling, 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/modeling, 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/modeling, co-developer of the York PsA model
Dr. Jason Madan	School of Social and Community Medicine, University of Bristol and Warwick Medical School	Health economics/modeling, evidence synthesis

*BRAM* Birmingham Rheumatoid Arthritis Model, *DSU* Decision Support Unit, *EULAR* European League Against Rheumatism, *GRAPPA* Group for Research and Assessment of Psoriasis and Psoriatic Arthritis, *NICE* National Institute for Health and Care Excellence, *NOAR* Norfolk Arthritis Register, *PENTAG* Peninsular Technology Assessment Group, *PsA* Psoriatic arthritis, *RA* Rheumatoid arthritis, *ScHARR-TAG* School of Health and Related Research Technology Assessment Group

research (Table 1). Their remit was to: (1) frame and clarify the issues for which consensus needs to be sought; (2) set out, where possible, initial

recommendations for consensus approaches for models, based on sound methodology, clinical judgment, and decision-maker preferences; and

**Table 2** Overview of topics and issues for consensus

---

Topic 1: modeling 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
Modeling 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)
Modeling 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
Modeling the relationship between disease severity and mortality risk
Topic 4: structural modeling approaches:
Representing sequences of treatments
Cohort vs. individual patient models

---

(3) set out an agenda for the research needed to achieve consensus where existing evidence is inconclusive. Four main topic areas for

consensus were identified, with specific issues to address for each area (Table 2). Further details of these issues, and their representation in existing models, are presented elsewhere [8].

Two 1-day working party meetings were held at the University of Birmingham in November 2011 and March 2012. Position papers describing each of the issues above were circulated prior to each meeting, with members given time to provide feedback and suggest additional considerations. These papers defined the agenda for each meeting, where consensus among participants was sought for each aspect, 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 [11] and the UK National Institute for Health and Care Excellence (NICE) methods guide for technology appraisals [12].

Where divergent opinions remained, participants were asked to identify research programs whose results would lead to greater clarity and consensus in these areas. We present below, for each model aspect listed above, a summary of the consensus recommendations, the outcome of discussions held on this topic at the workshops, and recommendations for further research to enhance future consensus. A more detailed report of the background, process and outcomes for the consensus working party is available in online supplementary material. The report was reviewed by a separate independent panel of clinical experts, and their commentary is also available online.

This article is based on the discussions of the Consensus Working Party and does not involve any new studies of human or animal subjects performed by any of the authors.

## RESULTS

### Modeling the Initial Response to Treatment

#### *Summary of Consensus View*

- Disease Activity Score 28 (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 timing of this should follow current clinical guidelines, to aid comparison of results among models, although the impact of alternative stopping rules can be explored in sensitivity analysis.
- 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 to biologics should be based on all relevant trials and derived using formal evidence synthesis methodology that respects randomization. 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.
- Response rates to the non-biologic comparator can be based on pooling control arms from biologic trials, although the comparability of trial and decision populations should be considered.
- For PsA, PsA Response Criteria (PsARC) and Psoriasis Area Severity Index (PASI) should

be used as outcome measures, although disease-specific measures currently in development may be used once they have been validated.

#### *Outcome of Workshop Discussions*

##### *Modeling the Initial Treatment*

**Phase** Stopping rules do not fully reflect the complexity of clinical decision-making at a patient level. However, their use within models is required to synthesize 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 recognized 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 European League Against Rheumatism [EULAR] response categories depend on both absolute change in DAS and DAS at endpoint).
- It is particularly appropriate for the UK, where it has received support from clinical experts in previous NICE appraisals, and is the basis of current NICE guidance.

American College of Rheumatology (ACR) 20/50/70 was considered problematic because it is a relative measure. However, given that it is commonly reported there is a clear need for mapping functions to characterize 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 modeling the

initial treatment phase. PsA 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 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 is associated with joint damage). 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 evidence to inform modeled treatment effects, trials in biologic-naïve patients should be analyzed separately from trials in patients with prior biologic exposure, as should trials in biologics with or without concomitant disease-modifying antirheumatic drugs (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. However, in the absence of convincing evidence for effect modification, the simpler approach of using unadjusted treatment effects is preferable, particularly if

an absolute scale such as DAS28 is used for response.

**Estimating the Baseline Response in the Comparator Treatment** For modeling 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.

**Modeling 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. Information on adverse event rates for different biologics will be reported by most trials. Models should not exclude trials 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.

#### **Current Available Evidence and Further Research Needs**

**Mapping Between (Change in) DAS28 and ACR 20/50/70** While DAS28 is the preferred

measure of short-term response to treatment for the RA consensus model, many trials report ACR 20/50/70 instead. Research is required to develop mappings between ACR 20/50/70 and DAS28, so that DAS28-based models are informed by all relevant trials. Few, if any, 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. Since ACR measures are relative, while DAS28 is an absolute scale, mappings should allow for dependence on baseline disease activity. IPD from trials would be the ideal evidence for this, potentially supplemented by registry data (e.g., estimation of the DAS28/Health Assessment Questionnaire [HAQ] change relationship from the British Society for Rheumatology Biologics Register [BSRBR]).

**Mapping Between Existing PsA Outcome Measures (PsARC, PASI) and Composite Measures Currently in Development** The evaluation group for the UK NICE appraisal of biologics for PsA developed a Bayesian network meta-analysis to synthesize trial evidence on short-term response to biologics [13]. Treatment effects were estimated on four outcomes: Psoriatic Arthritis Response Criteria (PsARC), Arthritis Response Criteria (ARC; both for joint symptoms), Psoriasis Area and Severity Index (PASI; for skin symptoms), and HAQ (for functional impact). The version informing 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 organization actively engaged in the development of response measures in PsA [14]. The GRAPPA Composite Exercise (GRACE) study has collected data on multiple PsA dimensions and has recently developed novel composite responder indices [15]. 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 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 UK NICE technology appraisals can be used to identify this evidence base. There are additional reviews of adverse events in the literature [16]. Systematic reviews of sequential biologic therapy have also assessed the impact on the efficacy of a second biologic of having experienced adverse events on the first biologic [17]. This evidence base needs to be collated, updated and synthesized to inform the consensus model.

## Modeling the Long-Term Treatment Phase

### 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 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 treatment. 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.

### *Outcome of Workshop Discussions*

HAQ has been widely used in models to represent disease progression, for historical reasons. Several mapping algorithms between HAQ and quality of life (QoL) measures (e.g., EQ5D) have been developed and used in existing models [18, 19]. However, algorithms for mapping between outcome measures such as HAQ and EQ5D are an area of active research [20], 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 QoL in patients with RA, independent of HAQ [24]. Therefore, models could in future use a multidimensional (HAQ and pain) outcome measure for disease progression.

Observational data have been used to estimate the duration of treatment and the

rate of change in HAQ over time while on treatment, and sometimes support assumed differences between biologics. 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 randomized 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-randomized 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. Registries 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.

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. While 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.

#### ***Current Available Evidence and Further Research Needs***

***Estimating Duration of Treatment in Responders*** 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. Registries have several advantages as the basis for estimating this information—they are often comprehensive, provide detailed patient-level data, and are up-to-date. Registries 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-randomized data source such analyses should be interpreted with caution.

***Disease Progression on Long-Term Treatment*** The consensus group also felt existing modeling 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 QoL*** Mappings between HAQ and QoL scores have been developed using trial and/or observational data. Mappings currently used in models do not account for the independent impact of pain on QoL, 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 registries and/or IPD from trials (if available). 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. Such estimates could be derived from registry data, although follow-up visits often do not coincide with treatment withdrawal, limiting the accuracy of estimation. Elicitation techniques could be used to capture clinical judgment on rebound if empirical approaches are unsuccessful. Given the challenges of estimating rebound, the sensitivity of cost-effectiveness findings to alternative assumptions should be explored within the consensus model.

## Estimating Lifetime Costs and Benefits

### *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 utilization to be a function of disease severity.

### *Outcome of Workshop Discussions*

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 modeling 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 modeled 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.

### *Current Available Evidence and Further Research Needs*

**Arthritis Health Care Utilization** Research is required to collate diverse evidence on the relationship between disease severity in RA and PsA and healthcare utilization. This research should initially take the form of identifying current literature and appropriate

data sources. The relationship between disease severity and health care utilization 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 [21] and Sweden [22]. More recently, analysis has been published of the total costs for patients with RA and PsA, including productivity losses, using Norwegian registry data [23].

**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 analyzed mortality data in the BSRBR for this relationship [24], and their analysis included covariates such as 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.

## Structural Modeling Approaches

### *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 several advantages when representing RA and PsA patient histories, but the merits of cohort modeling approaches should also be explored.

### ***Outcome of Workshop Discussions***

Models should have the flexibility to explore alternative positions for biologics within the sequence of treatments. While there may be benefit in modeling 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 [25, 26]; as a general principle, models should be as simple as possible whilst remaining consistent with the underlying decision problem and theory of disease [27]. 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.

### ***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.

## **DISCUSSION**

Decision-analytic models have become a key resource in health technology assessment. However, models are often developed independently by manufacturers, academic groups and regulatory bodies, leading to a range of models with divergent structures and conclusions, as is the case for biologic therapies for inflammatory joint diseases [8]. This can lead to confusion over the assumptions and data selection choices driving results, and skepticism of the validity of model results. Our aim was to show how a process of bringing together independent modeling and clinical experts could lead to clear consensus guidance for future models, increasing their credibility. It may not be feasible or desirable to require manufacturers or academic experts 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 modeling approach for a specific policy question. 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.

One limitation of this consensus is that the working group consisted solely of UK-based clinicians, models and regulators. Health technology assessment clearly has aspects that are country-specific, and this may mean that certain elements of our consensus would need to be adapted to other contexts. However, the

structures we have followed, and many of our findings, are relevant internationally. Our work also provides a case study of a process that can easily be extended to support decision-making in other disease areas. The process of developing consensus, and identifying its current limits, has the added benefit of highlighting areas where further research is most needed to support reimbursement decisions.

## ACKNOWLEDGMENTS

The workshops referred to in this manuscript were jointly funded by Arthritis Research UK and the UK Medical Research Council Network of Hubs for Trials Methodology Research. Dr. Madan was supported by the MRC ConDuCT Hub for Trials Methodology Research (Grant G0800800). All named authors meet the International Committee of Medical Journal Editors (ICMJE) criteria for authorship for this manuscript, take responsibility for the integrity of the work as a whole, and have given final approval to the version to be published.

**Disclosures.** Ernest Choy has received research grants, and served as member of advisory boards and speaker bureaus of Abbott Laboratories, Allergan, AstraZeneca, Boehringer Ingelheim, Chelsea Therapeutics, Chugai Pharma, Daiichi Sankyo, Eli Lilly, Ferring Pharmaceutical, GSK, ISIS, Jazz Pharmaceuticals, MedImmune, Merrimack Pharmaceutical, MSD, Novartis, Pfizer, Pierre Fabre Medicament, Roche, Schering Plough, Synovate, and UCB. Jason Madan, Tony Ades, Pelham Barton, Laura Bojke, Philip Helliwell, Paresh Jobanputra, Ken Stein, Andrew Stevens, Jonathan Tosh, Suzanne Verstappen, and Allan Wailoo have no conflicts of interest to declare.

**Compliance with Ethics Guidelines.** This article is based on the discussions of the Consensus Working Party and does not involve any new studies of human or animal subjects performed by any of the authors.

**Open Access.** This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## REFERENCES

1. Adalimumab, etanercept and infliximab for the treatment of rheumatoid arthritis (TA130). National Institute for Health and Clinical Excellence; 2007.
2. Etanercept, infliximab and adalimumab for the treatment of psoriatic arthritis (TA199). National Institute for Health and Clinical Excellence; 2010.
3. Silman AJ. Use of biologic agents in rheumatoid arthritis: introduction. *Rheumatology*. 2011;50:iv3–4.
4. Jobanputra P. A clinician's critique of rheumatoid arthritis health economic models. *Rheumatology*. 2011;50:iv48–52.
5. Tosh J, et al. The Sheffield rheumatoid arthritis health economic model. *Rheumatology*. 2011;50:iv26–31.
6. Barton P. Development of the Birmingham Rheumatoid Arthritis Model: past, present and future plans. *Rheumatology*. 2011;50:iv32–8.
7. Bojke L, et al. Modelling the cost-effectiveness of biologic treatments for psoriatic arthritis. *Rheumatology*. 2011;50:iv39–47.
8. Madan J, Ades AE, Welton NJ. An overview of models used in economic analyses of biologic

- therapies for arthritis—from current diversity to future consensus. *Rheumatology*. 2011;50:iv10–18.
9. Ades AE, Madan J, Welton NJ. Indirect and mixed treatment comparisons in arthritis research. *Rheumatology*. 2011;50:iv5–9.
  10. Welton NJ, Madan J, Ades AE. Are head-to-head trials of biologics needed? The role of value of information methods in arthritis research. *Rheumatology*. 2011;50:iv19–25.
  11. Cochrane handbook for systematic reviews of interventions version 5.1.0. In: Higgins JPT, Green S, editors. *The Cochrane Collaboration*; 2011.
  12. *Guide to the methods of technology appraisal*. National Institute for Health and Clinical Excellence; 2008.
  13. Woolacott N, et al. Etanercept and efalizumab for the treatment of psoriasis: a systematic review. *Health technology assessment (Winchester, England)*. 2006;10(46):1–233, i–iv.
  14. Mease PJ, Gladman DD, Krueger GG. Prologue: Group for Research and Assessment of Psoriasis and Psoriatic Arthritis (GRAPPA). *Ann Rheum Dis*. 2005;64(suppl 2):ii1–2.
  15. Helliwell PS, et al. The development of candidate composite disease activity and responder indices for psoriatic arthritis (GRACE project). *Ann Rheum Dis*. 2013;72(6):986–91.
  16. Bongartz T, et al. 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 J Am Med Assoc*. 2006;295(19):2275–85.
  17. Lloyd S, et al. The effectiveness of anti-TNF- $\alpha$  therapies when used sequentially in rheumatoid arthritis patients: a systematic review and meta-analysis. *Rheumatology*. 2010;49(12):2313–21.
  18. Hurst NP, et al. Measuring health-related quality of life in rheumatoid arthritis: validity, responsiveness and reliability of EuroQol (EQ-5D). *Rheumatology*. 1997;36(5):551–9.
  19. Brennan A, et al. Modelling the cost-effectiveness of etanercept in adults with rheumatoid arthritis in the UK. *Rheumatology (Oxford, England)*. 2004;43(1):62–72.
  20. Brazier JE, et al. A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. *Eur J Health Econ HEPAC Health Econ Prev Care*. 2010;11(2):215–25.
  21. Yelin E, Wanke LA. An assessment of the annual and long-term direct costs of rheumatoid arthritis: The impact of poor function and functional decline. *Arthritis Rheum*. 1999;42(6):1209–18.
  22. Kobelt G, et al. Economic consequences of the progression of rheumatoid arthritis in Sweden. *Arthritis Rheum*. 1999;42(2):347–56.
  23. Kvamme MK, 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;51(9):1618–27.
  24. 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 Rheum*. 2010;62(11):3145–53.
  25. Barton P, Bryan S, Robinson S. Modelling in the economic evaluation of health care: selecting the appropriate approach. *J Health Services Res Policy*. 2004;9(2):110–8.
  26. Brennan A, Chick SE, Davies R. A taxonomy of model structures for economic evaluation of health technologies. *Health Econ*. 2006;15(12):1295–310.
  27. Sculpher M, Fenwick E, Claxton K. Assessing quality in decision analytic cost-effectiveness models. A suggested framework and example of application. *Pharmacoeconomics*. 2000;17(5):461–77.