Critical review of ScHARR cost-effectiveness analysis of disease modifying treatments in Rheumatoid Arthritis: NICE MTA

Review report

June 2015

Purpose statement

UCB pharma UK have commissioned a review and appraisal of the independent Assessment groups economic evaluation of biologic Disease Modifying Anti-Rheumatic Drugs (DMARD) for the on-going multiple technology appraisal of adalimumab, etanercept, infliximab, certolizumab pegol, golimumab, abatacept and tocilizumab for the treatment of rheumatoid arthritis.

This review was conducted over a 3-day period and consisted of a review of the public evaluation assessment report prepared by the technology assessment group assigned to the appraisal (The School of Health and Related Research - ScHARR), and a copy of the excel file used to conduct the economic evaluation (provided by UCB).

The objectives of the review were to highlight any areas of uncertainty in the modelling work undertaken by ScHARR in response to the forthcoming appraisal committee meeting. This short document details the findings of that review.

Table of Contents

1 Objectives 2
1.1 UCB areas of review 2
1.2 Brief overview of ScHARR model 2
1.3 Error checking in assessment model 3

1 Objectives

1.1 UCB objectives

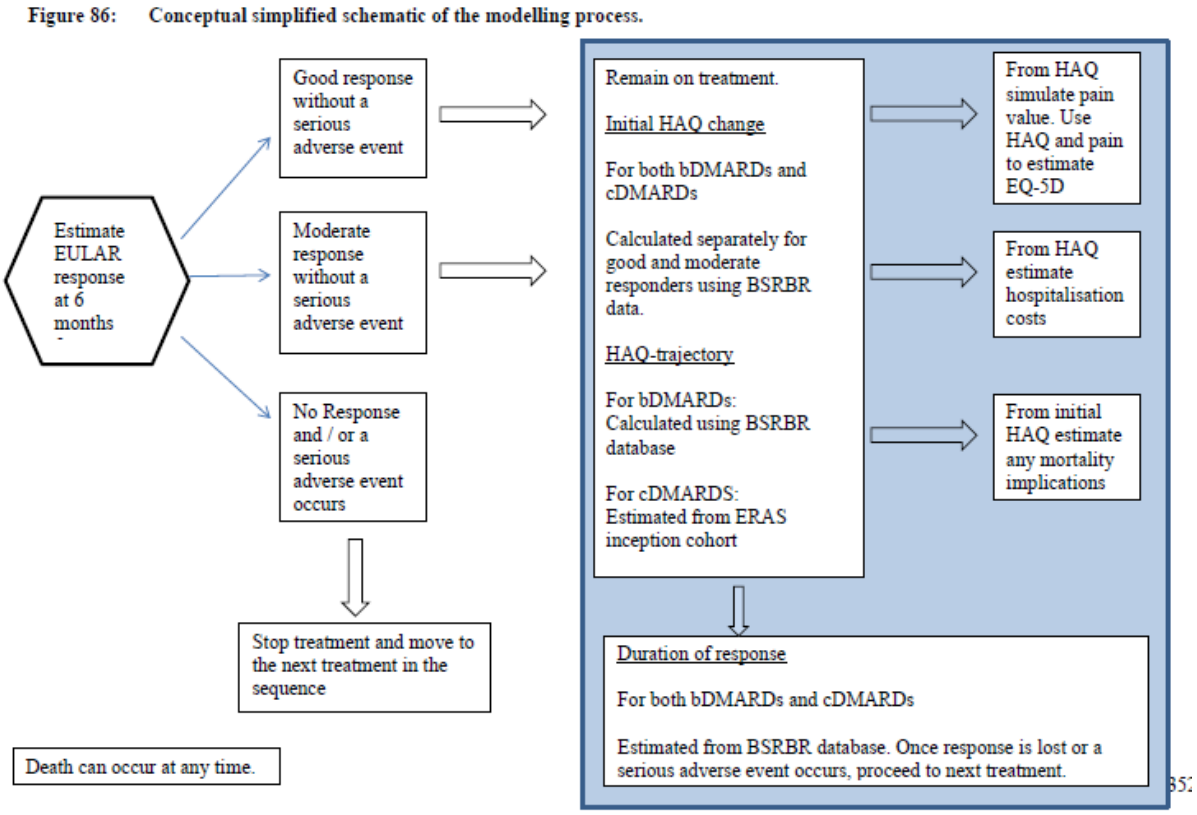
UCB Pharma would like to comment on the following elements of the economic evaluation prepared by ScHARR:

- Overall model structure (a cohort model compared with the UCB Markov model)
- Treatment of initial health gain (improvement on HAQ or ED5D when on treatment
- Cycle time for TNFs (are they all 6 months)
- Management of treatment failure – how patients are removed from treatment
- Comparator pay and HAQ change if not treated with a biologic
- Cost treatment for both active and comparator arms
- Length of time on treatment
- Treatment of the evidence base – how have the SLRs been defined and selection of evidence

1.2 Brief overview of ScHARR model

The structure of the ScHARR model is relatively complex owing to the use of multiple mapping algorithms and assumed correlations to establish relationships in the model, in addition to the use of a patient-level simulation modelling technique. The structure of the ScHARR model is depicted in Figure 1, as presented in the assessment group report.

Figure 1: Conceptual structure of ScHARR model



As explained by ScHARR, the "key parameter which affected the results [of the evaluation] is the assumed Health Assessment Questionnaire whilst on cDMARDs [conventional DMARD]; if the values used in previous National Institute of Health and Care Excellence (NICE) appraisals were instead used, the incremental cost per QALY fell to below £35,000 in some scenarios for bDMARD [biological DMARD], compared with cDMARD's alone. [page 39]". In addition, the assessment group noted that "the data source used for establishing the relationship between HAQ and pain was also seen to influence the results markedly [page 39]". These findings are not surprising when considering that HAQ score (alongside pain score) was mapped to EQ-5D weight.

Based on the key findings of the ScHARR report, part of this review has focused on the two key parameters of the ScHARR model (Table 1), as stated by the assessment group in their report, on the basis that any uncertainty in these parameters will have the largest impact on results.

1.3 Error checking in assessment model

A brief error checking review of the assessment group model was conducted. This review failed to identify any major omissions or errors in the simulation. Minor issues related to the choice of statistical distribution used to inform baseline parameters, such as HAQ score. However, as the simulation model is programmed to repeatedly sample patient characteristics to generate a sample of patients with consistent baseline characteristics, it is not clear whether this issue directly impacts on the results of the evaluation.

Table 1: Overview of review

Modelling aspect	ScHARR approach	Comments for NICE	Question raised by the ScHARR model	Priority/conclusion
Model structure - patient level versus Markov	"The model by UCB was a Markov cohort model that treated all patients as homogenous and would not have the flexibility desired for employing patient level covariates to represent the heterogeneity of patient outcomes." Page 346	It is UCB viewpoint that the technique of cohort-based Markov modelling is adequate for the purposes of evaluating the cost-effectiveness of interventions in RA. The inability to model patient heterogeneity in a cohort-based model is only considered of relevance, if the goal of the evaluation is dependent on explicitly (and parametrically) accounting for such heterogeneity. The cohort-based modelling framework is likely to implicitly account for patient heterogeneity, by modelling the outcomes expected at a cohort-level, without relying on numerous interrelated mapping algorithms to map individual patient characteristics to outcomes.	It is likely however that the patient level approach would have generated results which were more variable (and more uncertain) than the cohort-level approach. This would be due to the increased complexity of the model, and the need to parametrically accounting for patient heterogeneity. The ScHARR authors state that the Markov model approach lacks the desired flexibility for employing patient level covariates to represent the heterogeneity of patient outcomes. This statement however assumes that patient-level heterogeneity can be adequately represented by the myriad of mapping algorithms used by ScHARR in their evaluation. In the absence of any reported validation (external or internal), it is unclear whether the quest to represent heterogeneity in patient outcomes leads to a more robust simulation	The ScHARR decision to utilise a patient-level simulation approach is unlikely to influence the findings of the evaluation, however it increases the level of uncertainty in the results and introduces selective application to the patient level variation

Modelling aspect	ScHARR approach	Comments for NICE	Question raised by the ScHARR model	Priority/conclusion
Treatment of initial health gain	"In the base case, the Assessment group assume that the HAQ change, conditional on EULAR response, was the same for cDMARDs as for bDMARDs…The analyses assume that HAQ change, conditional on EULAR response is equal irrespective of the treatment (cDMARD or bDMARD)"	This assumption potentially favours cDMARD use because it assumes that the incremental benefits of bDMARD use on HAQ score is related solely to EULAR status. This assumption was not tested by the assessment group in the sensitivity analysis to their evaluation, and there was no reported attempt to validate the results of the simulation in predicting the short-term change in HAQ score reported in the RA clinical trials. It is therefore unclear whether EULAR response outcome alone can adequately predict the change in HAQ scores observed in the trials, and thus whether the direct mapping of HAQ to EULAR used by the assessment group is robust. UCB therefore invite the assessment group to provide evidence in support of this assumption, and thereby confirm whether the simulation model adequately captures the full observed benefits of treatment on HAQ score.	The UCB review has assessed the available evidence on the assumption that treatment induced improvements in HAQ score can be directly attributed to EULAR response, and attempted to assess whether there is an independent treatment effect on HAQ that is not covered by EULAR status. The available data (appendix - CERTAIN study) suggests that there may exist a small trend towards improved change in HAQ for EULAR good responders receiving certolizumab versus placebo, which may indicate an additional "biologic effect" on HAQ that is independent of EULAR response. This trend is potentially observed in the mapping analyses conducted by UCB, which showed that randomisation to placebo was associated with a numerical increase in HAQ score, which is independent of EULAR status, versus randomisation to certolizumab pegol. This data may show a trend towards a "HAQ-effect" that is independent of EULAR status for biologics.	This assumption may have some bearing on the results of the evaluation. There is significant validation work required to validate the choice of EULAR response rather than the more conventional ACR response.

Modelling aspect	ScHARR approach	Comments for NICE	Question raised by the ScHARR model	Priority/conclusion
Cycle time for TNF's	As shown in figure 1, the assessment group model considers EULAR response after 6-months of treatment.	This approach is broadly consistent with UK guidelines, however there is evidence to support the assessment of response to certolizumab pegol at the 12 week stage. This has the dual benefit of reducing the cost of patients removed from a non effective therapy and also allowing an earlier trial of a follow up treatment following the failure of the first TNF blocker There is an additional research question regarding the optimal strategy for bDMARD use in RA, which may include the use of shorter response periods for treatments, and the potential for reducing the exposure to biologic treatment in patients who are ultimately unresponsive to treatment.	The review of the patient time on treatment - and stopping rules - has highlighted significant variation in the implementation of the stopping rule. We would ask for a sensitivity analysis that places certolizumab pegol first line and implements a 12 week cycle.	Low priority

Modelling aspect	ScHARR approach	Comments for NICE	Question raised by the ScHARR model	Priority/conclusion
Management of treatment failure	As shown in figure 1, the assessment group model considers treatment failure on the basis of no EULAR response after 6-months of treatment, and/or adverse events. The management of these patients is not explicitly defined in the report, however, based on review of the Excel model, it appears that patients with no response are assumed to experience HAQ change of 0 (compared to -0.317, -0.672 for moderate and good EULAR response) prior to discontinuation of treatment On the assumed impact of treatment failure on HAQ score the assessment group state that "The change in a patient's HAQ when treatment has failed to be efficacious or is stopped due to an adverse event is not known with certainty. The Assessment Group has assumed that following cessation of treatment the initial HAQ-improvement experienced on treatment initiation would be lost. The resultant HAQ would be assumed for the subsequent six months when the next treatment in the sequence is trialled. This is similar to assumptions made within the manufacturers' models"	No comment	The approach taken by ScHARR appears reasonable in light of evidence on impact of no EULAR response. In addition, variations in the assumed HAQ score for no EULAR response is likely to have a minimal impact on the outcomes of the evaluation; given that this assumption impacts all treatments in the model (i.e. it is likely to have a minimal impact on the incremental results).	Low priority

Modelling aspect	ScHARR approach	Comments for NICE	Question raised by the ScHARR model	Priority/conclusion
HAQ change if not treated with biologic	"In order to estimate the trajectory of HAQ the BSRBR database was used. The BSRBR database measures HAQ at 6 month intervals for all registered patients for a maximum of three years. The evolution of HAQ whilst a patient remains on a biologic therapy was estimated as a function of a patient's baseline characteristics and 6-month EULAR response category." - page 365 "Norton et al. estimated270 HAQ progression in patients not receiving bDMARDs using data from patients recruited to the ERAS inception cohort study. Observations relate to patients recruited between 1986 and 1998 (n=1460), followed for 10 years, and a growth mixture model approach was taken." - page 370. As explained in the assessment report, the "key parameter which affected the results [of the evaluation] is the assumed Health Assessment Questionnaire whilst on cDMARDs [conventional DMARD]; if the values used in previous National Institute of Health and Care Excellence (NICE) appraisals were instead used, the incremental cost per QALY fell to below £35,000 in some scenarios for bDMARD [biological DMARD], compared with cDMARD's alone. [page 39]"	It is not clear from the assessment report whether the two data sources used to inform the long-term change in HAQ score for patients receiving bDMARDs (BSRBR) and cDMARDS (ERAN) are sufficiently comparable, and therefore can be used to inform future differences in the efficacy of treatment. In particular, it should be noted that the Decision Support Unit at NICE recommend that "all parameters which take on different values for different treatments should be based on randomised data…[and] the increased uncertainty and potential for bias [related to the use of non-randomised data] should be addressed". Both the manufacturers (UCB included) and assessment group models utilise non-randomised evidence to inform differences in the future HAQ score in patients on cDMARD and bDMARD. However, it should be noted that the evidence employed by the manufacturers has been subject to previous review and appraisal, and withstanding limitations, have been considered adequate in informing existing recommendations. The assessment group report does not appear to adequately address the uncertainty or the potential for bias in using two distinct sources of evidence to inform this parameter in the model. In particular, UCB would welcome/invite the assessment group to provide evidence to support the comparability of the sources used for bDMARD and cDMARD HAQ progression.	The ScHARR evaluation is sensitive to the assumptions and approach to modelling the future HAQ score in patients receiving cDMARD and bDMARD therapies. The statistical approaches used by ScHARR appear robust, so far as UCB and their advisers are able to review the material detailed in the report. There are however potential areas of uncertainty that need further explanation (or addressed should further evidence be made available from the assessment group). These include: • Confirmation that the ERAN and BSRBR populations are comparable, and that the use of these two sources of evidence in the evaluation does not bias the analysis in favour of cDMARDs (i.e. if the ERAN population are healthier, or have a more severe form of disease than BSRBR) • Confirmation that the reduction in the rate of worsening in HAQ score over time is not due to informative censoring (where patients with high HAQ scores leave the study due to the severity of their illness).	High Impact The main conclusions of the model and the subsequent ICERs are based upon sources of comparator data that are uncontrolled and potentially heavily biased. UCB does not believe that the information available to the assessment group (as opposed to their technical treatment of the information) provides a sound basis with which to make a recommendation that may lead to the withholding of a clinically effective treatment from patients with a high unmet need.

Modelling aspect	ScHARR approach	Comments for NICE	Question raised by the ScHARR model	Priority/conclusion
Cost of treatment for both active and comparator arms	The assessment group state that the "costs [of DMARDs] are similar to those used by the manufacturers however there are two comments worth noting: i) that the Assessment Group takes all patient access schemes into consideration whereas the majority of manufacturers do not and ii) that a number of manufacturers have assumed a fixed weight per person that can underestimate the costs of weight-based interventions." The assessment group however do not explicitly report the cost of bDMARDs in the assessment report, although these are now available through inspection of the economic model.	UCB would like to the assessment group to clarify whether the certolizumab pegol patient access scheme has been implemented in the evaluation model, and therefore accounted for in the deterministic analysis presented in the assessment report. In the assessment group model, the stated "responder period cost" of certolizumab pegol plus methotrexate is estimated at £8314, which is based on 7 months of fortnightly administrations of certolizumab pegol (to account for the additional loading dose of certolizumab pegol). Based on this calculation, the responder period cost of certolizumab pegol corresponds to the full initial dose schedule of certolizumab pegol administered over the initial 6-month period. Due to the complexity of the model, and the time required to fully de-bug the various routines in the simulation, we have been unable to identify if the assessment group have correctly adjusted the cost of certolizumab pegol to account for the 10 free drug vials included as part of the certolizumab pegol patient access scheme.	It is not clear from the executable model that the patient access scheme for Certolizumab pegol has been incorporated in the simulation. UCB would like to seek clarity from the assessment group on whether the results presented in the report are based on the 10-free vial patient access scheme.	Very High priority

Modelling aspect	ScHARR approach	Comments for NICE	Question raised by the ScHARR model	Priority/conclusion
Length of time on treatments	"The duration of treatment on the first biologic for adult RA patients was estimated using the BSRBR database which records the dates on which therapies are initiated and ended. Separate analyses were undertaken for those patients obtaining good and moderate EULAR responses at 6 months. … There were also little data on the duration of response for patients receiving cDMARDs. Based on the assumption that cDMARDs are not likely to be more toxic than biologics used in combination with a cDMARD, it was assumed that the survival duration for each EULAR response category for bDMARDs would be applicable for cDMARDs. " - page 373-374	The ScHARR approach to simulating duration of response is based on estimating survival curves for time to treatment switch that is dependent on EULAR response and is assumed to be independent of treatment allocation (i.e. bDMARD versus cDMARD). The assessment group consider this assumption to be valid on the basis that cDMARDs are not likely to be more toxic than bDMARDs, which in turn implies that any future decisions to switch treatment are based on toxicity, excluding other causes for discontinuation such as loss of efficacy. This assumption appears to directly contradict the assumed long-term behaviour of HAQ score in patients receiving cDMARDs, which is assumed to deteriorate over time, versus patients receiving bDMARDs, which is assumed to remain flat over time. This observation combined with the modelled relationship between HAQ and EULAR response (the stated criterion for treatment cessation) would indicate that patients receiving cDMARDs would discontinue treatment at an earlier stage than patients receiving bDMARDs (due to loss of efficacy). The impact of this assumption was not tested in the sensitivity analysis, and thus, it is not clear whether this assumption would have a material impact on the results of simulation.	It is unclear based on the complexity of the model (and need to conduct simulations) whether a shorter duration of response for cDMARDs would improve the results of the evaluation in favour of biologic therapy	Unknown impact due to absence of sensitivity analysis on time to discontinuation of therapy

Modelling aspect	ScHARR approach	Comments for NICE	Question raised by the ScHARR model	Priority/conclusion
Treatment of evidence base (construction of SLR)	The assessment group have considered a number of options in utilising data from clinical trials in the economic model: • Main analysis of mean EULAR response (excluding RAPID) • Main analysis plus RCTs with a small level of bDMARD use (including RAPID trials) • Main analysis plus RCTs with a small level of bDMARD use and a trial with low MTX-background use The above analyses were conducted using EULAR data as reported in the randomised trials, which represented a sub-set of the complete network of trials, and using data on ACR response that was mapped to EULAR response, which represented the complete network of trials in the review. The mapping was conducted using data from the US, with details of the mapping redacted from the publically available report. The statistical methods used by the assessment group to estimate the comparative efficacy of treatments in the network meta-analysis are consistent with the approaches recommended by the decision support unit at NICE. The assessment group stated that the estimates of ICERs for both EULAR data only, and when mapping ACR data to EULAR data indicate that the conclusions were not altered by restricting the selection of RCT's to only those that reported EULAR data	The assessment review group network met-analysis considered the exclusion of two key studies on the comparative efficacy and safety of certolizumab pegol plus methotrexate versus methotrexate alone (RAPID I and RAPID II), which when included in sensitivity analysis, have a marked effect on the comparative efficacy of certolizumab pegol versus comparators in the MTA. The reason for exclusion was based on the prior use of bDMARDs in less than 4% of the enrolled population. It is UCB's viewpoint that the main analysis, conducted with the exclusion of these studies, is based on an incomplete network of trial evidence, and therefore produces biased estimates of the comparative efficacy of treatment. In addition, the absence of these studies means that no data were available for EULAR response with certolizumab pegol, and thus no comparison versus certolizumab pegol was made in this analysis. As certolizumab pegol is used in practice, and licenced in this indication, the main analysis presented by the assessment group provides an incomplete picture on the cost-effectiveness of existing biologic treatments in RA. Due to the limited number of patients who received prior bDMARDs, it is unlikely that the results of RAPID I or RAPID II cannot be considered comparable to those from studies were prior bDMARD use was not permitted as part of enrolment criteria. In addition, it is not clear from the report the strength of the mapping from ACR to EULAR data. In particular, it is unclear how uncertainty in this mapping was propagated through the evaluation. UCB note however that the assessment group model was not sensitive to the choice of directly observed versus mapped trial data.	A key element of critique regarding the use of data from the SLR is the assumption that the RAPID I and RAPID II trials, which represent the bulk of evidence on the efficacy and safety of certolizumab pegol in RA, are excluded on the basis of minimal prior use of biologic therapies. The exclusion of this data has a meaningful impact on the simulated efficacy of certolizumab pegol. In particular, when comparing between main analysis and the analysis with RAPID trials, the results of the evaluation varies considerably, with Certolizumab pegol dominating other biologic therapies in the case where RAPID data is used (Table 197). It should also be considered that during the previous single technology appraisal of certolizumab pegol in RA, the evidence review group considered that "the FAST4WARD and RAPID 1 and 2 trials were of acceptable methodological quality and clearly show that therapy with Certolizumab pegol plus methotrexate was more effective than placebo plus methotrexate". In addition, there is some inconsistency in the approaches taken by the assessment group in attempting to align the network meta-analysis with UK clinical practice. In particular, the assumption to not consider trial exclusion based on geographic location. It is unclear whether restriction on geographic location would hamper claims for the efficacy/safety of Certolizumab pegol versus other comparators.	High priority. The decision to exclude the RAPID studies and the choice of EULAR rather than ACR study evidence has biased the results against certolizumab pegol whilst failing to demonstrate that the implied heterogeneity in our trials (due to prior biologics exposure) is a reason for exclusion. UCB asks that our key trials are incorporated into the main analysis