AbbVie's comments on the economic model produced by the School of Health and Related Research, University of Sheffield

Adalimumab, etanercept, infliximab, certolizumab pegol, golimumab, tocilizumab, and abatacept for the treatment of rheumatoid arthritis not previously treated with disease-modifying anti-rheumatic drugs and after the failure of conventional disease-modifying anti-rheumatic drugs only: systematic review and economic evaluation (review of TA guidance 130, 186, 224, 234, and part review of TA guidance 225 and 247)

AbbVie welcomes the opportunity to comment on the economic model produced by the Assessment Group, and has compiled the following points.

1 HAQ changes in the short term

1.1 ACR to EULAR mapping

In its response to the Assessment Group (AG) Report, AbbVie highlighted (section 3.3 of the document) that the Mixed Treatment Comparison (MTC) undertaken by the AG would be more transparent if it was based on ACR responses rather than on EULAR responses. This drew on the lack of clarity of the ACR to EULAR response mapping in the report which was marked academic in confidence. The parameters used by the AG for the mapping are not presented in the economic model, and AbbVie could not infer them from the information available in the model. No details are given on the estimation of the parameters for the mapping either, and therefore the issues of validity and precision of the estimates which AbbVie raised still hold at this stage of the Appraisal consultation. This compromises the ability of stakeholders to comment on the estimates of cost-effectiveness which are obtained using the EULAR treatment responses mapped from ACR responses using a US Veterans database.

1.2 Underestimate of HAQ improvement using BSRBR data

AbbVie considers that the use of BSRBR data for HAQ improvement will provide a substantial underestimation of the HAQ improvements achievable by patients who would be treated with biologic therapy in 2014 The AG report provides the following data (Table 1) on short term HAQ improvement by EULAR response category for treatment with conventional DMARDs (cDMARDs [patients treated with biologics are assumed to show the same improvement as patients treated with cDMARDs]), with an assumed baseline HAQ value of 2.03, which is in line with mean HAQ scores value in the source BSRBR data.

Table 1: HAQ improvement by EULAR response category with cDMARDs as per AG report

EULAR Response	Mean	SE
None	-0.050	0.025
Moderate	-0.509	0.035
Good	-0.650	0.043

Firstly AbbVie wishes to point out that the use of the BSRBR database will underestimate HAQ improvements for patients receiving biologic therapy, particularly patients with a DAS28 between 3.2 and 5.1. The database incorporates a large proportion of patients with high mean DAS28, long disease duration and a high baseline HAQ, taken from the years 2000 to 2008. The HAQ improvements achieved by BSRBR patients are considerably less than those achieved in all trials of biologic agents. Furthermore, the HAQ improvements achieved by patients in the BSRBR are considerably less than those achieved in the observational ReAct study of adalimumab. This highlights that underestimation of HAQ improvements by the AG is not simply attributable to differences in the HAQ improvements between those observed in randomised controlled trials versus observational studies.

AbbVie considers that there are few datasets where the mean HAQ improvements observed are as low as those achieved in the BSRBR. This is a reflection of the high disease duration, high DAS28 scores and high HAQ scores in the typical UK patient receiving biologic therapy between 2000 and 2008. The inception of the database occurred several years before the market of biologic treatments was established and as such the baseline state of the patients therein is on average very different to patients requiring treatment in the present day. The final patients were recruited to the adalimumab cohort in 2008. The implication of this is that the benefit of treatment with biologics will be underestimated using the BSRBR dataset. This will be particularly the case for patients with a DAS28 of > 3.2 with a low disease duration, as there are very few patients with both these characteristics in the BSRBR.

Data taken from recent clinical trials where the inclusion criteria focused on patients with short disease duration corroborates this point, with baseline HAQ scores being much lower than those obtained from the BSRBR. Furthermore, and perhaps more importantly, the absolute HAQ changes reported in the published literature show a much larger HAQ improvement than has been allowed for in the AG model which utilises the BSRBR data.

The other feature of this exercise has been to compare the HAQ changes generated by the AG model for cDAMRDs and placebo arms relative to the HAQ changes as reported in the trial literature. AbbVie has explored this point by presenting a comparison in Table 2. To the best of AbbVie's knowledge the examples presented therein are all those that have been published with complete six-month EULAR response data. Other trials with partially published data or different reporting time frames have not been considered here. This lack of availability of suitable trials providing fully published EULAR responses at the six month interval again casts doubt on validity of selecting the EULAR based responses as the outcome measure of choice for the AG model. The column on the left shows the mean per-treatment HAQ change predicted for a given treatment based on the AG model parameters; the column on the right shows the mean HAQ change as reported in the clinical trial, again on a per treatment basis.

Table 2: Comparison of absolute mean HAQ change derived from EULAR response data in the AG model with absolute mean HAQ improvements recorded in randomised controlled trials

Intervention	EULAR to HAQ	Actual Trial
ACT-RAY – tocilizumab + MTX	-0.51	-0.56
ACT-RAY – tocilizumab	-0.46	-0.54
van de Putte 2004 – adalimumab	-0.32	-0.38
van de Putte 2004 – placebo	-0.2	-0.07*
GO-FORTH – placebo	-0.2	-0.03*
GO-FORTH – golimumab	-0.44	-0.33

As can be seen from the two trial arms marked with an asterisk, the AG model using the assumed HAQ changes per EULAR response state (as per Table 1) predicts a much bigger change in HAQ score relative to the actual trial results in the placebo examples. In the placebo example from the GO-FORTH trial the discrepancy is close to a factor of seven.

Since the AG present the values in Table 1 as those derived for treatment with cDMARDs, the examples above support AbbVie's views that: they produce higher predicted HAQ improvements where placebo is considered, and that these assumed HAQ changes cannot, and should not, be applied to the biologics as well as the conventional treatments. In short; the HAQ improvements achieved by biologics and cDMARDs are not equivalent and should thus be modelled differently.

Exploring this point a little further, AbbVie were keen to review the per treatment HAQ changes generated in a similar manner to that detailed earlier in this section when using the AG data from the EULAR main analysis in comparison with the EULAR mapped from ACR

analysis (CODA 5 vs. CODA 1 worksheets from the AG model). The results of this comparison are presented as Table 3 below.

Table 3: Comparison of per-treatment HAQ changes generated from the EULAR main [1] and ACR to EULAR main [2] analyses from AG model with those calculated using the Roche [3] and Pfizer [4] ACR to HAQ methodology

Intervention	EULAR [1]	ACR to EULAR [2]	Roche [3]	Pfizer [4]
cDMARD	-0.28	-0.27	-0.26	-0.27
ABT iv+	-0.39	-0.38	-0.46	-0.43
ADA+	-0.42	-0.40	-0.49	-0.46
ADA mono	-0.32	-0.35	-0.40	-0.38
Int cDMARDs	-0.31	-0.35	-0.40	-0.39
ETN+	-0.49	-0.44	-0.58	-0.53
ETN mono	-0.37	-0.41	-0.52	-0.48
GOL+	-0.44	-0.41	-0.51	-0.48
IFX+	-0.36	-0.39	-0.48	-0.44
PBO	-0.20	-0.22	-0.20	-0.21
TOC+	-0.51	-0.44	-0.57	-0.52
TOC mono	-0.46	-0.44	-0.58	-0.53
CTZ+		-0.37	-0.44	-0.41
ABT sc+		-0.41	-0.51	-0.48
Grouped Biol	-0.39			

Note: *[1] HAQ changes derived from EULAR response data from the MTC results in the AG model; [2] HAQ changes derived from EULAR mapping of ACR data from the MTC results in the AG model; [3] HAQ changes derived by application of Roche ACR response group HAQ changes to the ACR MTC results in the AG model; [4] HAQ changes derived by application of Pfizer ACR response group HAQ changes to the ACR MTC results in the AG model*.

Whilst we can see from this comparison that some of the mean per-treatment HAQ changes are indeed similar when using the EULAR or ACR to EULAR mapped data, substantial differences do exist and needless to say for the last three treatments in columns [1] and [2] of the table no comparison is possible.

To complement this comparison we have also included the HAQ changes derived using the Roche [3] and Pfizer [4] per ACR response category HAQ changes to convert ACR scores directly into HAQ improvements. In doing so, not only are we able to carry out a like for like comparison, we can also add weight to the notion that the AG is underestimating the HAQ improvements achievable by biologics. All the biologic treatments in Table 3 are underestimated in terms of predicted HAQ improvement by the AG method relative to the two manufacturers' methods. Looking at etanercept, infliximab, tocilizumab, and adalimumab for example we see a 25% improvement in HAQ change from the AG ACR to HAQ method [2] and the Roche method [3]. Comparisons with the Roche and Pfizer methods are plausible since these two methods also work on the assumption that HAQ improvements are independent of treatment choice.

In absolute terms the HAQ improvements predicted by the Roche and Pfizer present a range of HAQ improvements that are in agreement with published literature and are thus in line with AbbVie's expectations (-0.44 to -0.58 and -0.41 to -0.53 respectively); whereas those predictions made by the AG model present a lower range (-0.36 to -0.51 and -0.37 to -0.44 when using AG methods [1] and [2] respectively), inconsistent with published literature.

AbbVie is keen to point out that its own method for predicting HAQ improvement is also based on ACR responses. However, it does not lend itself to this type of illustrative example, since AbbVie chose to model HAQ improvement as percentage improvements for each of the ACR response categories. Given this fact, a comparison with the AbbVie method in the context presented above would require the inclusion of baseline HAQ values for the scenarios investigated, and this was felt to be counterproductive for the purposes of illustrating this issue.

Using the ACR to HAQ improvement transformation as the method of choice AbbVie has conducted a comparison using trial data to model mean HAQ improvement using the Roche and Pfizer methods and then comparing this to the actual mean HAQ improvement recorded in the trials. We have used AbbVie trials for this comparison to investigate the predictive power of the HAQ conversions used by two other manufacturers (Table 4).

Table 4: Comparison of ACR to HAQ change using the Roche and Pfizer methods relative to trial data

Trial	Roche	Pfizer	Trial
ARMADA (ADA+MTX)	-0.59	-0.53	-0.62
ARMADA (Pbo+MTX)	-0.20	-0.21	-0.27
DE019 (ADA+MTX)	-0.51	-0.47	-0.56
DE019 (Pbo+MTX)	-0.25	-0.26	-0.24
PREMIER (ADA+MTX)	-0.66	-0.59	-0.9
OPTIMA (ADA+MTX)	-0.62	-0.56	-0.9
OPTIMA (Pbo+MTX)	-0.46	-0.43	-0.7

As Table 4 demonstrates, the HAQ predictions using the ACR to HAQ methods employed by Roche and Pfizer provide predictions more similar to the actual mean HAQ changes observed in the trials themselves than the AG method predictions, reported in Table 3. The reason for the under-predictions noted in the case of PREMIER and OPTIMA is that these trials were conducted in MTX naïve populations and as such the methods employed by Pfizer and Roche (both based on MTX experienced populations) will yield lower predicted HAQ changes than a method based on MTX naïve patients data would have done in these examples.

To conclude this section AbbVie would like to present a comparison of predicted HAQ improvements derived from the use of the AG method, the Roche method, and the Pfizer method with the actual observed HAQ improvement from the ReACT observational dataset, consisting of 5711 RA patients that have not had prior anti-TNF therapy but are cDMARD experienced receiving treatment with adalimumab.

Table 5: Comparison of predicted HAQ improvements derived from the AG, Roche, and Pfizer methods with actual HAQ improvements observed in the ReACT observational study

	AG EULAR to HAQ	ReACT publication	ACR to HAQ [Roche]	ACR to HAQ [Pfizer]
ReACT – Adalimumab	-0.32	-0.55	-0.53	-0.49

As shown in Table 5, data from the ReACT observational study is very similar indeed to the values calculated from the two methods employed by Roche and Pfizer for the response groups reported in the published study; whereas the HAQ improvement predicted by the AG

fall short of the observed value in ReACT by a long way. It should be noted that for ease of reproducibility AbbVie has used the 12 week published data as reported in the ReACT paper (Bombardieri *et al*., Rheumatology (Oxford). 2007 Jul;46(7):1191-9) to illustrate this issue.

What is clear from a comparison of the trial data, looking at overall HAQ improvement and per-response group HAQ improvements, relative to the AG modelled HAQ improvements per response group, is that published trials have higher HAQ improvements in their best response categories (be that ACR70 or EULAR good) than is assumed in the model. AbbVie would anticipate absolute mean HAQ improvements of around 1.0 as far as biologic treatments are concerned, as is seen in the Pfizer and Roche datasets. This is significantly higher than the 0.65 improvement assumed in the EULAR good group of the AG model. AbbVie considers that alternative trial datasets should be used to generate the relationship between EULAR response and HAQ improvement rather than relying on the BSRBR dataset if EULAR is to be kept as the measure of response.

1.3 Discussion of EULAR moderate response

A final point to consider given the selection of EULAR response as the outcome measure of choice is that the collective term of moderate response contains patients with a variety of DAS28 scores after treatment response has been assessed. The following matrix (Figure 1) illustrates the different ways in which a moderate EULAR response can be achieved.

Figure 1: Reproduction of the EULAR response matrix based on DAS28 improvements

	DAS28 improvement		
Present DAS28	**>1.2**	**0.6-1.2**	**<0.6**
<3.2	Good response	**Moderate response**	No response
3.2-5.1	**Moderate response**	**Moderate response**	No response
>5.1	**Moderate response**	No response	No response

AbbVie considers that network meta-analyses utilising EULAR response need to take account of the fact that overall moderate response rates from trials may vary depending on the proportion of moderate responders in each category in the trial. That is, the proportions of moderate responders cannot be appropriately compared without analysing the similarity of baseline DAS28 scores in the trials being synthesised. Furthermore, the mean HAQ improvement may vary among these different types of moderate EULAR responders but estimates are not available for whether the HAQ change would be systematically different for each of these categories of moderate responder. The BSRBR would not be the most appropriate dataset to assess whether this is the case because very few patients started biologic therapy with a DAS28 score of less than four.

2 Modelling of long term HAQ progression

AbbVie notes that an external review of the evidence for HAQ progression in RA is underway at present, as per NICE's initiative, and that it may wish to make further comments when the report is available. The inadequacy of the populations from which the data informing the AG progression model was taken has previously been highlighted in AbbVie's response to the AG report (section 3.1.1).

The lack of transparency in the AG's modelling of HAQ progression has been mentioned in section 3.1.2 of AbbVie's response to the AG report. AbbVie acknowledges that the Excel based economic model adds to the understanding of the methodology used by the AG in deriving the estimates of HAQ progression on cDMARDs, and also discloses the source data for HAQ progression on cDMARDs. However, some information obtained from the source

data for progression in the economic model still precludes discerning how HAQ scores progress in the average cohort of the model, which in turn precludes comparisons with estimates from literature, obscures the effect that HAQ progression modelling has on cost-effectiveness results, and does not allow AbbVie to incorporate what would be the AG's estimates of HAQ progression for the average cohort in its model. This makes it unfeasible to check if shifts in Incremental Cost-Effectiveness Ratios (ICERs) similar to the ones observed in the AG's model would arise from changing the HAQ progression estimates adopted by AbbVie in its base case to whole population average estimates of HAQ progression as implied by the AG model.

The HAQ progression graph in the Excel model "Progression" sheet does not match either figure 3 in reference 275 of the AG report, which gives HAQ progression observations and estimated trajectories for four different patient groups over 10 years, or either of the panels in the figure in reference 270 of the AG report, which give HAQ progression observations and estimated trajectories for four different patient groups over 15 years (both these figures are reproduced below as Figure 2 and Figure 3 respectively). Particularly noticeable is the fact that HAQ essentially tails off for any of the four groups after year 10, whereas in the work of Norton and colleagues in reference 270, in which the second panel in the figure applied a similar methodology to the ERAS database, HAQ progression is observed for all classes. It seems therefore that the source data for the graph has not got any observed data from the ERAS database, but only modelled data, whose derivation is not made clear (as would be the case if data was presented as formulas rather than values). Furthermore, there is no added information on the quantitative criteria which led to the choice of model behind the data above other alternatives, including different specifications of the Latent Class Growth Model.

Figure 2: Reproduction of Figure 3 from reference 275 in the AG report

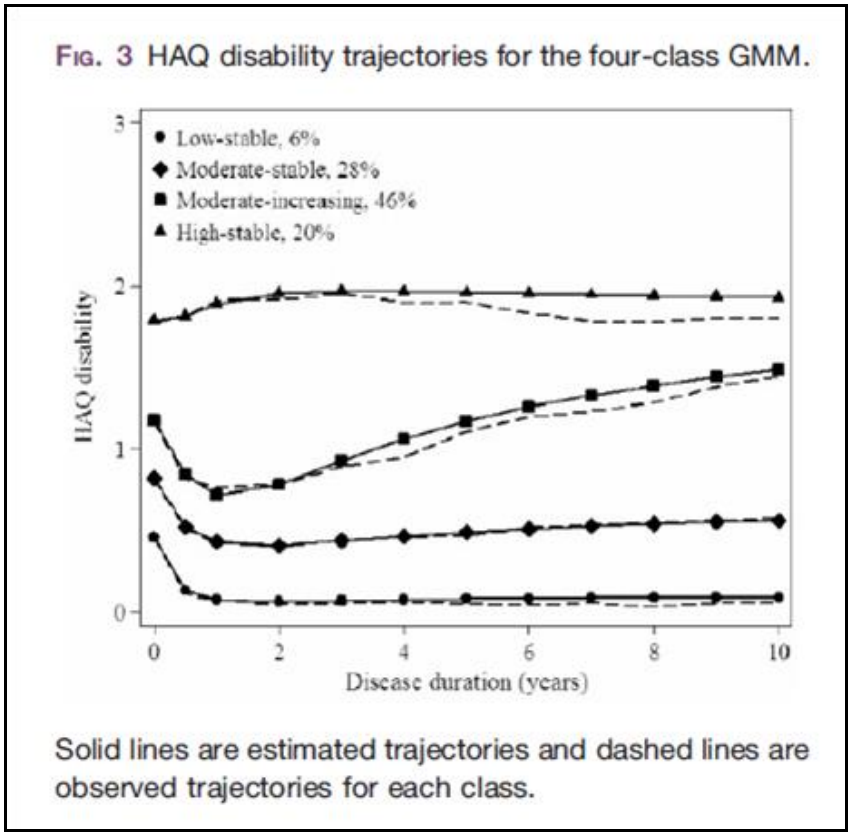
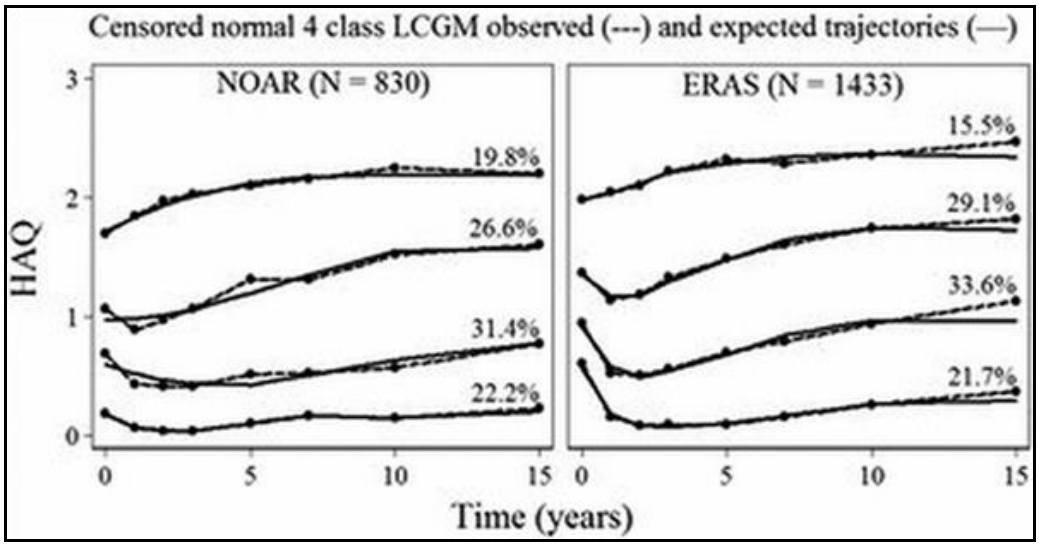


Figure 3: Reproduction of the figure from reference 270 in AG report



From the source data for the graph, it is possible to derive average annual HAQ progression rates implied for each of the four patient groups in the economic model (AbbVie adopts for simplicity and comparability the groups' designations in figure 3 of reference 275 in the AG report); presented in Table 6 below:

Table 6: Implied annual HAQ progression rates for the four groups of patients in the AG's HAQ progression model

Groups	HAQ at year 0.5	HAQ at year 14.5	Average Annual HAQ progression rate
Low-stable	0.16	0.31	0.0179
Moderate-increasing	0.53	1	0.0400
Moderate-stable	1.14	1.7	0.0360
High-stable	2.05	2.3	0.0107

An initial observation is that the annual progression rates for each group are lower, and some much lower than the annual progression rate of 0.045 used in the previous NICE appraisal of RA medicines after failure of cDMARDs. Furthermore, AbbVie could not find information in either the AG report or the economic model detailing the proportions in each of the groups in the total population. This is one of the outputs of the model (as the percentages are calculated in the Visual Basic code), but only having access to the values of these percentages for the different runs of the model would allow AbbVie to derive an estimate of the annual HAQ progression rate of the average cohort.

It should be noted that it is anticipated that very few patients with a low stable HAQ would be offered biologics in the UK. As outlined in section 3, AbbVie has analysed HAQ and pain VAS data from the ReAct (Research in Active Rheumatoid Arthritis) study. ReAct was a large, multinational open-label trial that enrolled adults with active RA who had previously been treated with traditional disease-modifying anti-rheumatic drugs or biological response modifiers with a DAS28 >3.2. A sample of 6,524 patients, the vast majority of whom were anti-TNF naïve and eligible for treatment with biologics under current NICE guidelines, had a baseline HAQ and pain VAS score. Of these only 477 (7%) had a HAQ score of 0.5 or lower at baseline. The proportion of these patients at baseline is even lower in the BSRBR registry. Out of 3856 in the registry starting treatment with adalimumab and not having had previous exposure to biologic treatments, only 97 (2.5%) had at baseline a HAQ score of 0.5 or lower. Therefore, the proportion of patients with HAQ scores of 0.5 or lower, most of which would not be eligible for treatment with biologics in the UK, should be much lower than those found in

the low-stable group of patients in the AG analyses based on either the ERAS or NOAR databases. As the average annual HAQ progression rate is quite low in the low-stable HAQ group, the fact that none of these databases provides a representative sample of the population of candidates for treatment with biologics is an additional source of underestimation of HAQ progression in the AG's economic modelling.

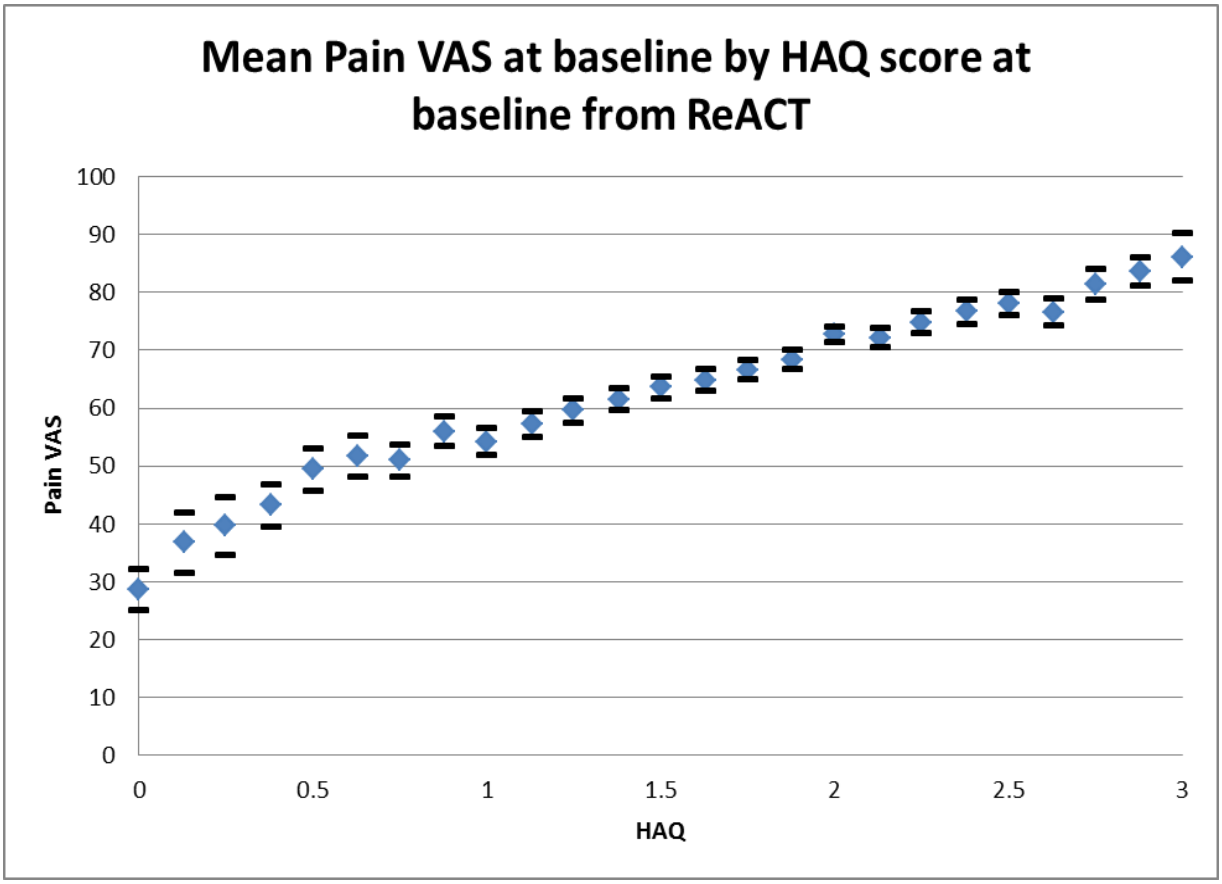
3 Relationship between pain levels and HAQ scores

As noted by the AG, pain is an important predictor of utility which is not included in estimates of utility generated using HAQ scores alone. AbbVie considers that the ERAS dataset used to generate pain VAS scores in the base case is likely to provide an underestimation of the pain levels for UK biologic eligible patients on cDMARDs, particularly at higher levels of HAQ scores.

AbbVie has analysed the relationship between HAQ score and pain VAS using the ReAct study. Patients self-administered adalimumab 40 mg subcutaneously every other week for 12 weeks and were allowed to enter an optional long-term extension phase.

At baseline there were 6,524 patients with a HAQ and pain VAS score available. Figure 4 shows the relationship between HAQ score and mean pain VAS (95% CI) at baseline and indicates that pain scores are higher for patients with a HAQ greater than 2, unlike the modelled relationship from the ERAS dataset. As these are baseline data, the majority of patients were on cDMARDs therapy but had disease that was considered severe enough (DAS28>3.2) to require adalimumab therapy. The ReAct study also supports the finding of the BSRBR that there are large numbers of patients with a HAQ score prior to starting therapy which is above 2. Therefore, it is incorrect to conclude that few patients with RA in the UK will have a HAQ score above 2. Few patients in the US in the NDB appear to have HAQ scores above 2 but this is not the case in the multinational ReACT study or the BSRBR. This may be reflective of more aggressive early diagnosis and treatment in the US.

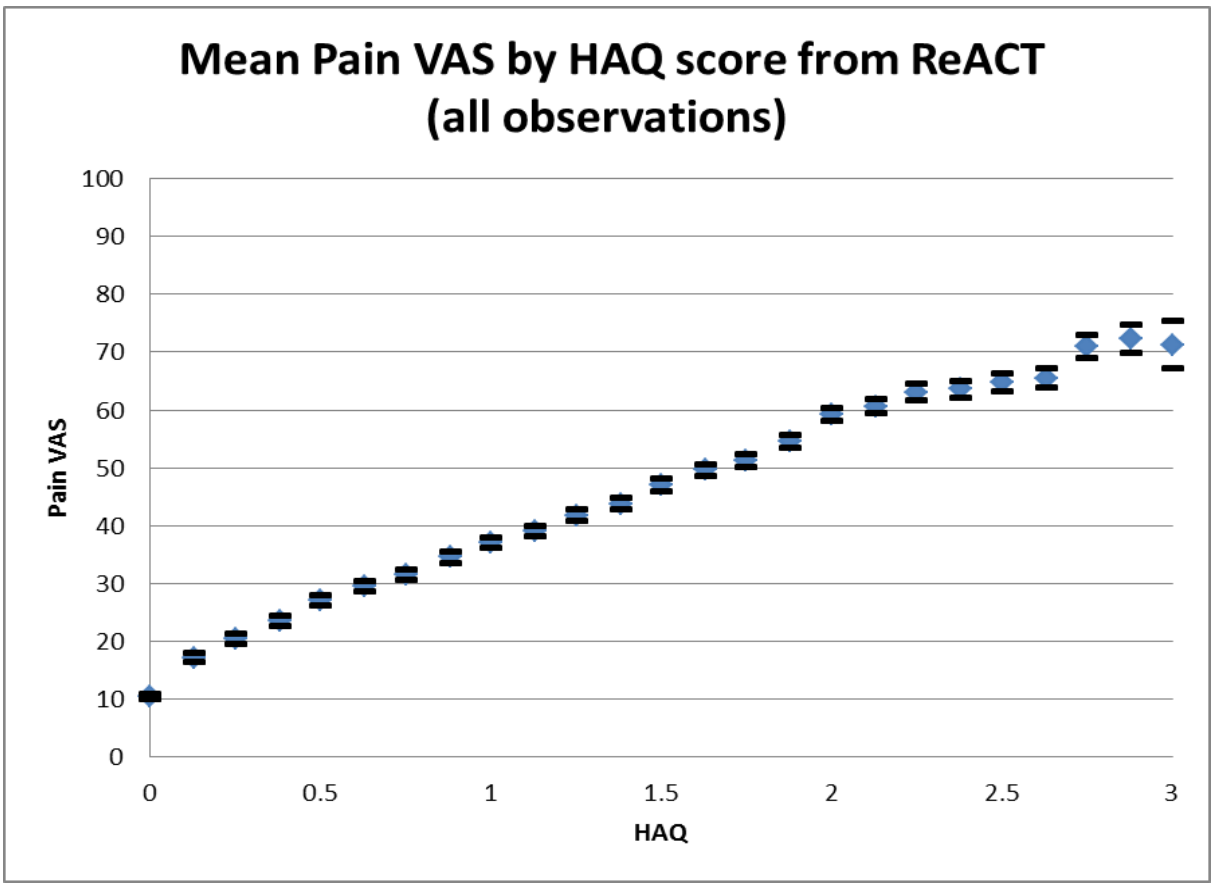
Figure 4: Mean Pain VAS relationship to HAQ score at baseline from ReAct



These data show that patients with high HAQ scores have notably higher mean pain VAS scores than patients in the ERAS or NDB dataset with high HAQ scores.

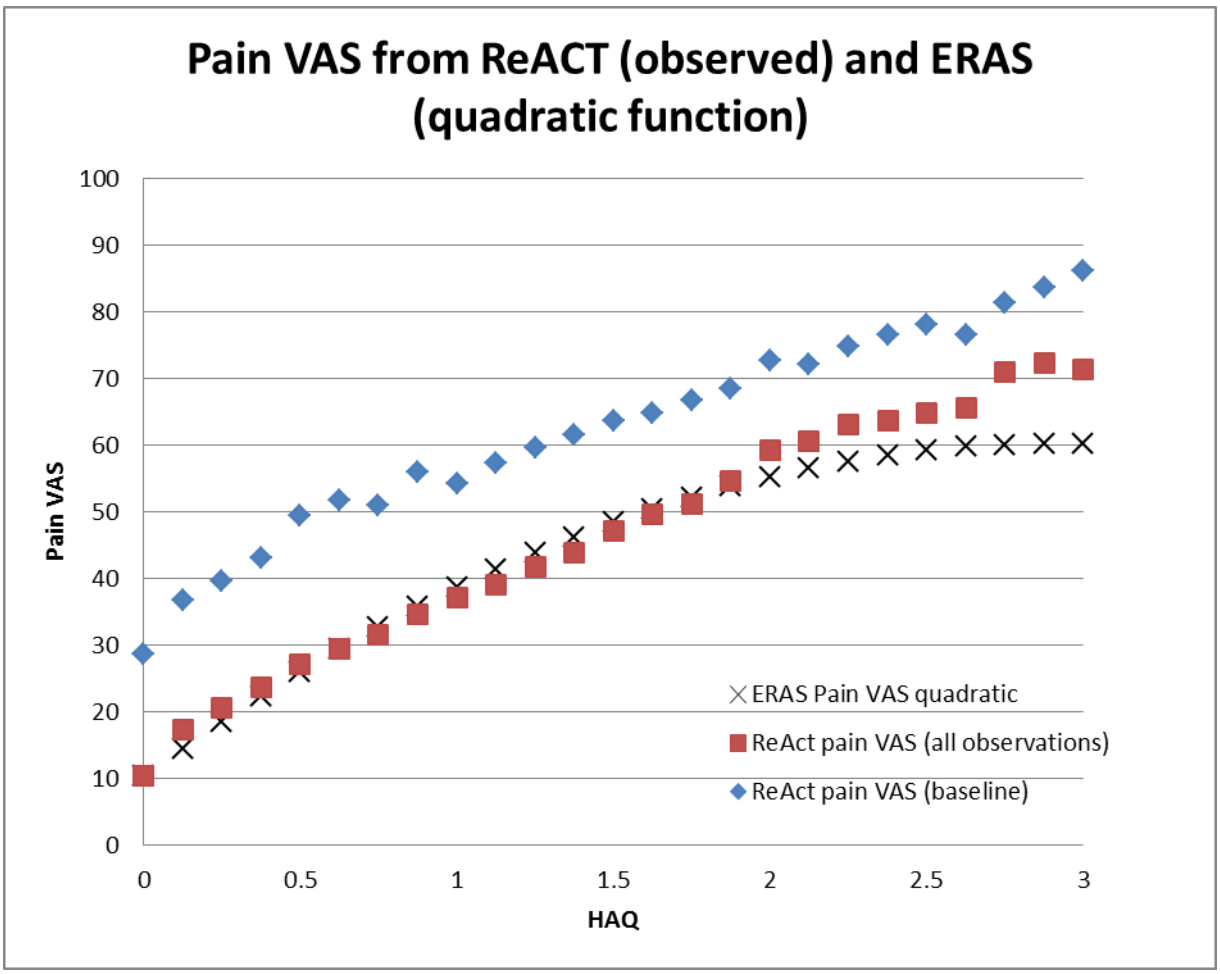
Figure 5 shows the relationship between HAQ score and pain VAS for the 41,919 paired observations available from the ReAct trial at all time points.

Figure 5: Mean Pain VAS relationship to HAQ score from ReAct (paired observations at all time points)



These data show that the mean pain VAS is systematically lower at all levels of HAQ score when all observations are used compared to baseline only scores from ReAct. It is hypothesised that this is attributable to the impact of adalimumab therapy reducing the pain associated with RA. These data highlight that pain scores increase as HAQ scores increase and do not provide evidence of a HAQ threshold above which pain scores do not increase, which is in contrast to the two datasets currently included in the AG model. Figure 6 below shows the ReAct mean pain VAS estimates from baseline and all observation estimates together with the modelled quadratic function from ERAS that is used in the AG model base case. These data suggest that the impact of anti-TNF therapy on reducing pain may be underestimated in both manufacturers and AG modelling given that pain is not modelled independently from the modelling of HAQ improvement and is not considered to differ systematically between patients on different therapies. Further analysis would be useful to consider the extent to which biologic therapies are associated with better improvements in the pain VAS compared to cDMARDs therapy when grouped by the same levels of HAQ improvement.

Figure 6: ReAct mean pain VAS estimates from baseline and all observation estimates together with the modelled quadratic function from ERAS that is used in the base case



4 Costs incorporated into the economic model

AbbVie wishes to draw attention to the following points raised in its response to the AG report (the corresponding sections of the document are further specified below), as these are still outstanding at the model consultation stage:

- (Section 3.4) The effect of considering wider costs to society of RA is not investigated in the economic model. The estimated magnitude of wider societal costs should warrant inclusion in the model to provide a more complete view of the cost-effectiveness of RA treatments.

- (Section 3.5) Each instance of infusion of the different treatment options is taken to cost the same amount (the £154 used in NICE'S TA247), regardless of anticipated mix of professionals involved, time spent and facilities used. This is a source of distortion of the cost-effectiveness of different treatments.

- (Section 3.5) AbbVie wishes to restate that the administration cost of subcutaneous (SC) adalimumab injections should be set to zero in the base case, as the product can be self-administered. In instances where guidance is required, AbbVie provides fee-free training in self-administration of SC injections.

- (Section 3.8) AbbVie wishes to restate that the cost of treating each episode of serious infections (£1749), taken from the Pfizer submission, was obtained from Swedish Healthcare Resource Groups (HRGs) and does not reflect the difference in management strategies and general cost levels between UK and Sweden. Furthermore, these costs may be underestimated, due to the Swedish HRGs not covering complications or contraindications involved in treating serious infections.