Meindert Boysen
Programme Director, Centre for Health Technology Evaluation
National Institute for Health and Care Excellence
10 Spring Gardens
London
SW1A 2BU

2nd February 2015

Ref. Assessment Report: TNF-alpha inhibitors for ankylosing spondylitis and axial spondyloarthritis without radiographic evidence of ankylosing spondylitis (including a review of technology appraisal 143 and technology appraisal 233) [ID694]

Dear Meindert,

Thank you for the opportunity to comment on the Assessment Report.

The finding by the Assessment Group (AG) that structural differences between their model and the manufacturers' models did not produce significant differences across the ICERs for the respective TNF-α inhibitors underline the robustness of the cost-effectiveness of these therapies. This finding is aligned with those of previous NICE guidance that TNF-α inhibitors remain the only cost-effective treatment with disease-modifying activity for ankylosing spondylitis.

While we welcome the results of the Assessment Report, MSD has several concerns in the presentation of the results and systematic errors, as follows:

(i) The range of ICERs in the Conclusions of the Scientific Summary does not represent the most plausible scenario;
(ii) The 'rebound equal to gain' assumption is the most plausible scenario;
(iii) Underestimation of QALYs gained for TNF-α inhibitors due to misspecification of statistical model;
(iv) The NICE reference case captures only a minor component of the economic benefit of treatment with TNF-α inhibitors.

These are explored in more detail in the text overleaf.

Kind regards,

Dr. Eilish McCann
Senior HTA & EBM Manager

(i) The range of ICERs in the Conclusions of the Scientific Summary does not represent the most plausible scenario

Within the Conclusions of the Scientific Summary, the final bullet point: "*The de novo model, which had addressed many of the issues of earlier evaluations, generated ICERs ranging from £19,240 to £66,529 depending upon anti-TNF and modelling assumptions*" incorrectly implies that the estimate is highly uncertain and most likely exceeds £30,000. The range of ICERs presented reflects the results of two distinct modelling approaches employed by the AG, specifically 'rebound equal to gain' and 'rebound to conventional care'. When either scenario is considered alone, the range of ICERs decreases. The range presented should represent the most clinically plausible scenario for TNF-α inhibitors. It is the view of MSD that the most plausible scenario is 'rebound equal to gain' (as described below) and only the range from this scenario should be quoted. This scenario supports the conclusions reached in previous NICE guidance that four of the five TNF-α inhibitors are well within the established threshold for cost-effectiveness (£19,240 to £21,577). Infliximab appears to be an outlier with an ICER for the same scenario of £40,576.

(ii) The 'rebound equal to gain' assumption is the most plausible scenario

The AG model presents two rebound scenarios – 'rebound equal to gain' and 'rebound to conventional care' - in order to represent the potential limits to the ICERs presented. However, the statement in the Assessment Report (Section 7.8, p.214) "*that the reality lies somewhere between these scenarios*" may be misrepresentative and does not indicate which is the most credible scenario. MSD believes that the scenario 'rebound equal to gain' is more clinically plausible than 'rebound to conventional care' and therefore the realistic ICERs are between £19,240 and £21,577 (for four of the five TNF-α inhibitors), for the following reasons:

1. In a study[1] of patients who discontinued infliximab (patients included in this study had all been receiving infliximab for the preceding 3 years, having participated in the first published randomised clinical trial on this therapy in active AS), they experienced a worsening in BASDAI score (from a mean value of 2.5±1.8 to 6.1±1.4 [latter value recorded at the point when retreatment began]). The mean BASDAI score achieved following discontinuation was similar to the baseline BASDAI score for the original randomised controlled trial: 6.5±1.2 in the infliximab arm (Braun *et al*. 2002)[2]. This does not suggest that patients rebound to a BASDAI score worse than that at baseline i.e. supporting the 'rebound equal to gain' scenario.
2. The AG model assumes that the rate of disease progression is decreased in patients who receive a TNF-α inhibitor for more than four years and is lower than that for conventional care after this point (see below for further comment on disease progression assumption). Therefore, the 'rebound to conventional care' scenario requires the implausible assumption that the magnitude in loss of functional capacity will increase the longer a patient responds to treatment with TNF-α inhibitors.
3. As stated above, the AG model assumes that patients who respond to treatment with TNF-α inhibitors continue to experience the same rate of disease progression (0.082 BASFI units/year) for the first 4 years as those who receive conventional care. After 4 years patients who continue to respond to treatment with TNF-α inhibitors experience a reduction in disease progression (0.034

[1] Baraliakos X, Listing J, Brandt J, Rudwaleit M, Sieper J, Braun J. Clinical response to discontinuation of anti-TNF therapy in patients with ankylosing spondylitis after 3 years of continuous treatment with infliximab. (2005) Arthritis Research & Therapy 7:R439-R444 (DOI 10.1186/ar1693)

[2] Braun J, Brandt J, Listing J, Zink A, Alten R, Golder W. Treatment of active ankylosing spondylitis with infliximab: a randomised controlled multicentre trial. (2002) Lancet 359: 1187-1193

BASFI units/year). The interpretation that Haroon *et al*. (2013)[3] shows that TNF-α inhibitors do not reduce disease progression until after year 4 may be incorrect. A more credible interpretation is that disease progression in patients who respond to TNF-α inhibitors is reduced but the cumulative difference compared with conventional care is only statistically significant after 3.9 years. In addition, a reduction in disease progression from the time when a patient responds to treatment with TNF-α inhibitors is in-line with their proven disease-modifying activity. Furthermore, TNF-α inhibitors, specifically golimumab, have been shown to reduce inflammation, with lower inflammation at baseline associated with reduced disease progression[4,5]. We note that the AG has explored a reduced rate of disease progression for patients receiving a TNF-α inhibitor from the start of treatment in a sensitivity analysis and believe that this assumption is more appropriate and should be applied in the base case.

(iii) Underestimation of QALYs gained for TNF-α inhibitors due to misspecification of statistical model

There are two key sources of error in the AG model that are likely to contribute to the overestimation of the ICERs by underestimation of the QALY gain, as follows: 1. Misspecification and bias of the synthesis model used for estimation of baseline BASDAI and change from baseline in BASDAI, and 2. Underestimation of QALY gain produced by the algorithm for calculation of EQ-5D index from BASDAI and BASFI scores.

1. Misspecification and bias of the synthesis model used for estimation of baseline BASDAI and change from baseline in BASDAI. The synthesis model developed by the AG appears to incorporate a bias that produces estimates for baseline and change from baseline of BASDAI scores conditional on response that are consistently different from the observed values for the studies GO-RAISE for golimumab and ATLAS for adalimumab (Tables 77-78, p.176-177 and Tables 138-139, p.268-269; summarised in Table 1 below). Data from the study 314-EU for Enbrel were redacted and could not be compared. Discrepancies between estimated values and observed trial data along with the potential source of the errors are acknowledged in the Assessment Report "*The results obtained differ from those seen in three pivotal trials (data provided by the manufacturers upon request), probably because of distributional assumptions over the baseline and change scores*" (Section 6.4. p.180). However, misspecification of the statistical model and the potential impact of that is neither considered nor explored. Significantly, the synthesis model underestimates the change from baseline in BASDAI of responders, particularly for TNF-α inhibitor treatment, whilst the change from baseline in the observed values from GO-RAISE and ATLAS are consistent with each other. In addition, the synthesis model overestimates change from baseline in BASDAI of non-responders. The net result is to reduce the comparative benefit of TNF-α inhibitor treatment on disease activity. Since the BASDAI and BASFI values are used to calculate the EQ-5D index and thereby estimate QALYs, underestimation of comparative change from baseline in BASDAI for TNF-α inhibitors compared with conventional care will produce a higher ICER. This is reflected in the ICER results for

[3] Haroon N, Inman RD, Learch TJ, Weisman MH, Lee M, Rahbar MH. The impact of tumor necrosis factor alpha inhibitors on radiographic progression in ankylosing spondylitis. (2013) Arthritis Rheum 65: 2645-2654

[4] Braun J, Baraliakos X, Hermann KA, van der Heijde D, Inman RD, Deodhar AA, Baratelle A, Xu S, Xu W, Hsu B. Golimumab reduces spinal inflammation in ankylosing spondylitis: MRI results of the randomised, placebo-controlled GO-RAISE study. (2011) Ann Rheum Dis doi:10.1136/annrheumdis-2011-200308

[5] Braun J, Baraliakos X, Hermann KA, Deodhar AA, van der Heijde D, Inman RD, Beutler A, Zhou Y, Xu S, Hsu B. The effect of two golimumab doses on radiographic progression in ankylosing spondylitis: results through 4 years of the GO-RAISE trial. (2014) Ann Rheum Dis 73: 1107-1113 doi:10.1136/annrheumdis-2012-203075

golimumab; the ICER is £12,785 when the observed data are applied (Scenario 2 of Table 100) compared with £21,079 when estimated data from the synthesis model are used. Substitution of the corresponding observed data from the study ATLAS for adalimumab produces a similar decrease in the ICER for adalimumab versus conventional care and reinforces the likelihood that the model is misspecified.

Table 1: BASDAI scores conditional on response for the AS population. The ICERs are reproduced from Table 100, p.211 for the scenario rebound equal to gain following discontinuation.

	BASDAI					
	Conventional care			TNF-α inhibitors#		
	AG*	ADA	GLM	AG*	ADA	GLM
% Responders	0.10	0.16	0.15	0.42	0.46	0.46
Change from baseline						
Responders	-2.7	-4.5	-4.25	-3.86	-4.64	-4.74
Non-responders	-0.45	-0.2	-0.18	-1.73	-0.82	-1.22
All	-0.66	-0.9	-0.81	-2.63	-2.58	-2.84
Baseline						
Responders	3.83	6.31	6.52	4.76	6.14	6.25
Non-responders	6.31	6.37	6.63	7.03	6.35	6.69
All	6.08	6.36	6.61	6.08	6.25	6.49
ICER (£/QALY)				**ADA £21,170**	**£12,851**	**£12,785**
				GLM £21,079		

* Estimated by the synthesis model developed by the AG; ADA results provided for adalimumab from ATLAS; GLM results provided for golimumab from GO-RAISE; # ICERs for adalimumab and golimumab are reproduced from scenario 2, Table 100.

The differences between the respective estimated and observed BASFI values for golimumab and adalimumab appear smaller than for BASDAI (Table 2 below). The discrepancies in the predictive performance of the statistical model for estimation of baseline and change from baseline scores for BASFI and BASDAI indicate that the model could be inappropriate for the estimation of the cost-effectiveness of TNF-α inhibitors for ankylosing spondylitis.

Table 2: BASFI scores conditional on response for the AS population.

	BASFI					
	Conventional care			TNF-α inhibitors		
	AG*	ADA	GLM	AG*	ADA	GLM
% Responders	0.1	0.16	0.15	0.42	0.46	0.46
Change from baseline						
Responders	-1.41	-2.74	-1.8	-3.02	-2.92	-3.03
Non-responders	-0.17	-0.17	0.38	-0.63	-0.72	-0.53
All	-0.29	-0.59	0.05	-1.64	-1.73	-1.68
Baseline						
Responders	3.42	4.5	3.56	4.17	4.53	4.45
Non-responders	5.43	5.91	5.39	6.02	5.78	5.48
All	5.24	5.68	5.11	5.24	5.21	5.01

* Estimated by the synthesis model developed by the AG; ADA results provided for adalimumab from ATLAS; GLM results provided for golimumab from GO-RAISE.

While we recognise that the AG has attempted to produce a single data set for conditional baseline and change from baseline scores from the data submitted by manufacturers, it is clear that there are significant discrepancies between estimated and observed values. Data submitted by the manufacturers were obtained directly from patients and thereby should be weighted higher than estimates produced by the AG synthesis model.

2. Underestimation of QALY gain produced by the algorithm for calculation of EQ-5D index from BASDAI and BASFI scores. The AG used the multivariate, non-linear model developed by Pfizer. As highlighted in the Assessment Report (Figures 7-8, redacted): "*the models over-predicted EQ-5D at low observed EQ-5D and under-predicted at higher observed EQ-5D values*" as reported by Brazier *et al*. (2010)[6]. This is a common feature of mapping algorithms with the result that the QALY gain for TNF-α inhibitors are likely to be underestimated. Underestimation of the comparative QALY gain will produce higher ICERs for TNF-α inhibitors versus conventional care.

(iv) The NICE reference case captures only a minor component of the economic benefit of treatment with TNF-α inhibitors

The study by Rafia *et al*. (2012)[7] demonstrated that direct NHS costs contribute to only 15% of the total costs attributable to ankylosing spondylitis. The combined costs of unemployment, absenteeism from work, and reduced productivity at work accounted for the remaining costs, which are not captured in economic models based on the NICE reference case. Additional consideration of the benefits not captured by the NICE reference case would likely improve the cost-effectiveness of TNF-α inhibitors compared with conventional care.

<u>Additional issues</u>

Class effect

The AG has explored a number of different evidence synthesis models, which consider whether the TNF-α inhibitors have independent treatment effects or whether there is a common (equal) treatment effect for the class. It is not clear precisely which model was used to generate the probabilities of BASDAI50 response at week 12 (Table 82, p.186) though the narrative in Section 6 of the Assessment Report indicates that the model demonstrating a common treatment effect for all TNF-α inhibitors was preferred and adopted. MSD notes that although a number of scenario analyses were performed in the AG model, no scenario analysis was conducted to explore the impact of the choice of evidence synthesis model on the resulting cost-effectiveness estimates. It is apparent from the clinical data that the individual results for infliximab exceeded the estimates generated for the TNF-α inhibitor class. Although the AG states that: "*it is possible that infliximab has a greater effect than the other TNF-α inhibitors at least at 12 weeks, there is no strong evidence from these analyses to suggest that it does*", the supporting information for this statement (Table 11, p.65) does indicate a difference in BASDAI and BASFI scores: -2.28 for infliximab alone compared to -1.66 for all TNF-α inhibitors for BASDAI, -2.15 for

[6] Brazier JE, Yang Y, Tsuchiya A, Rowen DL. A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. (2010) Eur J Health Econ 11:215-225.
[7] Rafia R, Ara R, Packham J, Haywood K, Healey E. Healthcare costs and productivity losses directly attributable to ankylosing spondylitis. (2012) Clin Exp Rheum 30:246-253.

infliximab alone compared to -1.38 for all TNF-α inhibitors for BASFI. It is not clear why these differences have been dismissed. In summary, the high ICER for infliximab may result from the underestimation of the treatment effect for infliximab (through considering a class effect).

Quality checking of the AG model

Due to password protection of key worksheets MSD was unable to quality check the economic models provided by the AG. Furthermore, inclusion of golimumab and infliximab in the model for nr-axSpA was inappropriate since these products do not have market authorisation and are not subject to this appraisal.

Dosing of golimumab

The comment in the Assessment Report (Section 4.2.3, p.73) that: "*The results for golimumab look particularly strong with around 60% of all randomised patients achieving ASAS 40 and BASDAI 50 after 5 years. However, this is probably not reflective of clinical practice as many of the normal weight patients took the 100 mg dose of golimumab rather than the 50 mg dose: the licence only permits the use of 100 mg dose in patients with a body weight of more than 100 kg who do not achieve an adequate clinical response after 3 or 4 doses*". While we acknowledge that randomisation in the trial did not consider weight, the golimumab EPAR states that "*The impact of weight on ASAS 20 response was not apparent in subjects receiving golimumab 100 mg dose. However, in the golimumab 50 mg treatment group, the ASAS 20 response was lower among subjects in the highest quartile of weight (>87 kg).*" Therefore, any perceived benefit arising from patients receiving the higher dose would be expected to be counteracted by a lower response in those patients with greater body weight who receive the lower dose.

Data for natural history and trajectories of BASDAI/BASFI

The AG state that (Section 5.2.5, p.159): "*While such a comparison may be necessitated by the short-term nature of the double-blind periods, the lack of a more detailed consideration of the appropriateness of the comparisons being made in relation to sources of natural history data (and subsequent assumptions made concerning the BASDAI/BASFI trajectories of the different patient categories) is concerning and, hence, current ICER estimates reported by the manufacturers must be considered to be both speculative and highly uncertain*". Since the AG has also used a combination of different study types to inform their modelling approach, these issues and their possible contribution to uncertainty may apply equally to the manufacturers' models and the AG model.