Etanercept, Infliximab and Adalimumab for the treatment of psoriatic arthritis: a Systematic Review and Economic Evaluation

The Assessment Group's Responses to the Comments from NICE Consultees

This document contains responses to specific comments from consultees on the assessment group report. An accompanying document provides details of the additional analyses undertaken in response to these comments.

	CLINICAL REVIEW		
	Consultee	**Comments**	**The Assessment Group Response**
1	Hull Teaching PCT	Section 2.5 lists a large number of uncertainties about the comparisons used, all of which raise fundamental questions about how the trials relate to everyday practice. *("The patients in most trials are not precisely representative of the population recommended for biologics in current guidelines. It is unclear whether the beneficial effects are similar in those treated in routine clinical practice." p24)* Were the study groups comparable with general practice populations or even the groups currently prescribed these drugs? Indeed, section 5.2.2.1 *states "the populations in these trials of etanercept are not representative of the patients for whom etanercept is licenced for use….."(p48)* and section 5.2.2.2 states *"Relative to the patients for whom infliximab treatment is recommended in practice, these trial populations may be less severely affected, with only around half in IMPACT and possibly even* fewer *in IMPACT 2 having failed to respond to two or more DMARDs.."(p54)* It is not clear how this flaw may influence the usefulness of the treatment or the cost-effectiveness for use by a PCT. Further clarification of this point would be useful.	In the page 95 of the Assessment report, we have acknowledged that the majority of patients in the trials had previously received at least one DMARD, and no trial specified the failure to respond to at least two DMARDs (patients whom the current BSR guidelines consider eligible for biologic treatment) as a recruitment criterion. Therefore, trial participants were likely to have had less severe disease compared to those patients receiving agents in practice. Despite this fact, it should also be important to note that trial participants were generally likely to represent the population with moderate to severe PsA requiring further treatment in practice, as we have clearly indicated in the report. The model base-case population has the mean HAQ score of patients in the RCTs and mild-to-moderate psoriasis. We conduct subgroup analyses for more severe baseline arthritis and psoriasis. We assume that absolute treatment effects are the same for different baseline severity of arthritis, and proportional treatment effects are the same for different baseline severity of psoriasis.
2	Hull Teaching PCT	Tables 5.3 (p.49), 5.7 (p.55) and 5.11 (p.61) containing RCT data on efficacy outcomes: although p-values for the 95% CIs are quoted for the majority of outcome data, there are several instances where p-values are absent and would have been helpful to produce a better overall picture of efficacy in terms of statistical significance.	Where reported, we extracted the p-values and 95% CIs from the primary studies. It should be noted that presenting 95% CIs for outcomes solely is sufficient in terms of statistical significance; we can infer whether the result is statistically significant on the basis of 95% CIs.
3	British Society	We question the feasibility of pooling the results of trials for all three biologics as it is	Ideally, the assessment group would have been able to derive its findings

	for Rheumatology	clear that each trial included patients of different severity, and recommend that it is treated with considerable caution. There are differences in baseline characteristics - particularly previous DMARDs, baseline joint counts and psoriasis severity. We have tabulated the differences between the trials in terms of three important variables: sub-group based on number of joints (the figures differ a little from the report), number of joints involved at baseline, and number of previous DMARDs used. These differences may well have influenced the outcome (i.e. rates of PsARC response, ACR20 and change in HAQ).	from a direct head-to-head comparison of the agents of interest. However, in the absence of such evidence, efficacy data from randomised placebo controlled trials for each agent were included in the review. Based on the characteristics shown in table 5.1, studies were considered sufficiently clinically similar to allow pooling. Estimates of uncertainty are provided and the findings were cautiously interpreted. The validity of indirect comparison (mixed treatment comparisons (MTC)) meta-analysis is built on the assumptions that no important differences exist between trials in terms of baseline characteristics such as disease severity. If the two sets of trials differ with respect to a clinical feature that influences the outcome, then the results of indirect comparisons could be confounded. Nevertheless, it should be noted that the differences between the mean baseline variables of RCTs that evaluate the same drug are as great as differences between trials that evaluate different drugs. Therefore these limitations and cautions would apply to any meta-analysis, not just an indirect comparison.
4	British Society for Rheumatology	There are no head to head trials and using this approach to try to compare the agents may well give misleading results. The effect of the agents on the skin cannot be compared again due to differences in baseline PASI and relatively small numbers involved.	Please see response to item 3.
5	British Society for Rheumatology	It should be noted that all three trial populations contain patients who are either DMARD naive or have yet to fail 2 or more DMARDs and are therefore not necessarily representative of the true clinical setting. The recommendations from the comparison analysis need to take this into account. For example the etanercept data from 2004 utilise a population where only 20% have failed 2 or more DMARDs (the population NICE are recommending we use these agents in). Comparing HAQ and joint count responses is likely to be different in a population that has already failed 2 or more drugs to one where a drug is being used de novo. This is another example where attempts to compare results from different trials can be fraught with difficulty. It is worth noting that the ACR20 response depends in part on the number of joints involved at outset, so that a person is less likely to achieve ACR20 with fewer joints at onset.	We have acknowledged in our report that the patients in most trials are not precisely representative of the population recommended for biologics in current guideline. We therefore concluded that it is unclear whether the beneficial effects are similar in those treated in routine clinical practice in the report. For the ACR20 response, it should be noted that the number of joints involved at outset was generally similar across the trials.
6	Abbott	Given that the Assessment Group has only considered the 12 week RCT evidence when determining the effectiveness of the agents, effectively disregarding any open-label extension data and observational data from routine clinical practice in the UK (BSRBR data); Abbott considers it necessary to examine the RCTs included in the effectiveness analysis and	All available placebo-controlled data were included in the review of clinical efficacy, including both 12- and (where reported) 24-week outcomes. Longer term randomised data would be desirable but were unavailable. The use of standard meta-analytic techniques meant that

		highlight the impact the smaller RCTs have on the estimates of efficacy for the different anti-TNFs when the data are pooled.	small trials were appropriately weighted in any pooled analysis. We have acknowledged the limitation (p. 154) that our findings were based on a limited amount efficacy data, though these limited data were of good quality.
7	Abbott	Adalimumab long term radiographic data demonstrates that it reduces the progression of peripheral joint damage in PsA patients: Abbott can understand why the Assessment group has only used results from the randomised controlled parts of the anti-TNF trials to avoid any potential biases arising from either open-label data or observational studies, although Abbott considers that these data are important in the clinical effectiveness analyses. This is particularly true of outcome measures that evaluate radiographic progression. As the TAR acknowledges, radiographic measures are more objective and are therefore a better reflection of the estimates of biologic treatment effect. Given that the radiographs are blinded, then the results should be measured consistently and objectively regardless of treatment arm. This is supported by evidence submitted to the EMEA in 2008, which led to a change in the wording of the adalimumab licence to include the following: "Humira has been shown to reduce the rate of progression of peripheral joint damage as measured by X-ray in patients with polyarticular symmetrical subtypes of the disease and to improve physical function." The evidence to support this change was based on the 24-week randomised double-blinded data and also on the 2 year open-label data. Therefore, Abbott believes that the data for 1-2 years follow-up are adequate to support the premise that adalimumab inhibits long-term radiographic progression.	The radiographic data from a single controlled trial for adalimumab in PsA demonstrated a beneficial effect on progression of joint disease at 24 weeks. This is a short time over which to identify a clinically significant effect of therapy: at least one-year follow-up is considered to be required to measure the radiographic changes in responding to the treatment. However, the 24 week data do indicate a rapid onset of action of adalimumab. The two-year follow-up data report a mean change in TSS from baseline, which is difficult to interpret because of lack of a control group; the value of 0.5 (SD 4.20) does not seem particularly low compared to the placebo rate at 24 weeks (0.1). Therefore, we believe that the limited available data do not adequately support that adalimumab inhibits long-term radiographic progression.
8	Abbott	Skin improvements were assessed in the Genovese study using Target Lesion Score (TLS) On page 65 of the TAR, the Assessment Group state that: "*There is limited evidence from a single RCT that adalimumab treatment has a beneficial effect on the psoriasis component of the disease in patients with PsA.*" This statement is incorrect as the Genovese study did examine the psoriasis component of the disease using the Physicians Global Assessment of disease and also the Target Lesion Score. Unfortunately the PASI was not used as an outcome measure so it is not possible to include the data in the modelling, however the data do show statistically significant improvements in both PGA and TLS in patients receiving adalimumab compared to patients receiving placebo. At Week 12, the mean target lesion score had decreased from baseline by 3.7 units for adalimumab patients compared with 0.3 units for placebo patients ($p \leq 0.001$). At Week 12, the physician global assessment for psoriasis was "Clear" or "Minimal" for significantly more adalimumab patients (40.6%, 13/32) than placebo patients (6.7%, 2/30) ($p = 0.002$)[i]. Furthermore, from Week 12 to Week 24, target lesion	As stated on p.34, PASI was chosen as the primary measure of skin response, as this is an outcome recommended in the BAD guidelines, and was measured in all RCTs, unlike other indicators of severity of psoriasis,. In addition, the AG received specific clinical advice that TLS would not provide a useful measure. However, the TLS data is available in the data extraction table 10.3.3. The statement on p. 65 will be amended to clarify it refers to skin disease as measured by PASI.

		scores decreased by 4.4 and 0.8 for patients from the placebo and adalimumab arms, respectively, resulting in total improvements from baseline of 4.7 and 4.5. From Week 12 to Week 24, the percentages of patients who had achieved physician global assessments of "Clear" or "Minimal" increased by 43% (from 6.7% to 50.0%) for placebo patients treated with open-label adalimumab, and by 16% (from 40.6% to 56.3%) for patients in the adalimumab arm.	
9	Wyeth	Mode of action of biologics (Section 2.1, Page 20): The report refers to biologics targeting pathologic T-cell activity, whereas it is customary to refer to the three biologics licensed for psoriatic arthritis as targeting tumour necrosis factor (TNF).	The AG acknowledges this point and the sentence will be amended accordingly.
10	Wyeth	Suggestion of higher incidence of TB reactivation with etanercept than adalimumab (Section 2.4, Page 22): Data from the BSRBR registry suggests that the risk of TB reactivation is higher with the monoclonal antibodies (adalimumab and infliximab) than with etanercept, which is not acknowledged within the Assessment Report. There are data from a number of European registries that suggest rates of TB reactivation are lower with etanercept than with the monoclonal antibodies : Dixon et al, (2009), examining data from rheumatology patients in the BSRBR, come to the conclusion that the rate of TB in patients with RA treated with anti-TNF therapy was 3-4 fold higher in patients receiving infliximab and adalimumab compared to etanercept. Data from the French RATIO database suggests that exposure to infliximab or adalimumab versus etanercept was an independent risk factor for TB (Tubach 2009) in their cohort of patients on anti-TNF treatment. BIOBADASER data (Gomez-Reino, 2007) suggests that, although the number of cases was very small for all three anti-TNFs, the incidence ratio was higher for both infliximab and adalimumab compared to etanercept. The Swedish ARTIS data also shows the same trends, with the risk of TB being higher with infliximab and adalimumab than etanercept. In addition, the recent Cochrane Review (Singh 2009) of biologics for rheumatoid arthritis, refers to a report from the FDA (2008) that states: "Data from clinical trials and preclinical	Data from the Gomez-Reino 2007 report are summarised in tables 5.26 to 5.29. The Dixon et al 2009, Tubach 2009, and ARTIS data do not appear to be based on summary reports that were published in time to be identified in the updated AG search. The Cochrane review summarises FDA warnings for each of the biologics whereas the AG report aimed to obtain direct empirical evidence on adverse events as outlined by the inclusion criteria.

		studies suggest that the risk of reactivation of latent tuberculosis infection is lower with Enbrel than with TNF-blocking monoclonal antibodies."	
11	Wyeth	Exclusion of the PRESTA data (Section 5.2.2.1, Page 47): Given the need to consider a comprehensive evidence base reference should have been made to the randomised, multi-centre outpatient study conducted in 752 subjects with PsA (PRESTA), which supports and extends the information of the efficacy and safety of etanercept available from controlled trials.	The PRESTA trial was a 12-week RCT (with open-label follow-up to 24 weeks) comparing two different doses of etanercept (one of which is not licensed) in patients with psoriasis and psoriatic arthritis. As PRESTA did not include any of the comparators specified on p.40 of the AG report, it did not meet inclusion criteria for the review of clinical effectiveness.
12	Wyeth	Safety of biologics (Section 5.2.3.1, Page 78): The focus in this section are data observed from RCTs and some observational studies, which are within the boundaries of the inclusion criteria for the systematic review, however, there are data published from European registries that we believe should also be included in the review of adverse events with biologics: There is some reference to data from the Spanish BIOBADASER and German RABBIT registries already in this section, however, there is also pertinent data published from the BSRBR on serious infections and latent tuberculosis reactivation, and also data from the Swedish ARTIS registry and the French RATIO database. There has also recently been data presented on skin cancer from the BSRBR, at the ACR conference 2009. In addition, there is no reference to the recent Cochrane review (Singh et al, 2009) of biologics in rheumatoid arthritis, in which the authors conclude that there is less withdrawal due to adverse events with etanercept, compared with anakinra, infliximab and adalimumab.	Please see response to comment no. 10
	Evidence synthesis (MTC)		
13	Abbott	Baseline HAQ: In the trials of PsA included in the MTC, it is clear there are differences in baseline HAQ. Surprisingly the Assessment Group did not adjust for baseline HAQ when considering HAQ change in the MTC. A failure to adjust for baseline HAQ when considering HAQ change in the MTC will therefore bias the results of the analysis.	The Assessment Group is aware of the importance of baseline adjustments with respect to the analysis of outcome measures and that a relationship between baseline HAQ and HAQ change may be present. However, the fact that the are only 6 trials and that the Assessment Group did not have access to Individual Patient Data makes it very difficult to evaluate such a correlation. Also within a meta-analysis it is the relative treatment effects that are used.

		Abbott requests that the Assessment Group re-run the analyses adjusting for baseline HAQ as it appears that these changes will have a significant impact on the results.	
14	Abbott	Abbott requests that the Assessment Group re-run the analyses using more precise input values for each of the anti-TNF therapies	Please see response to comment no. 34
15	Abbott	Different types of patient are being compared in the smaller RCTs: Abbott considers that the differences in joint efficacy reported in the Assessment Group's MTC between adalimumab and etanercept are based solely on differences arising from the smaller RCTs. However, an important point to note when comparing these smaller trials is that two <u>different</u> patient populations are being compared indirectly. This could be another contributing factor to the notable differences in arthritis efficacy, in addition to the effect of chance due to the small sample size. Abbott understands that it is very difficult to account for these differences in sub-types in the modelling, particularly when the studies are so small. However, it is important to highlight that the differences in joint efficacy between adalimumab and etanercept that drive the Assessment Group's conclusions are based on pooled response rates from different sub-types of PsA patient. Therefore, the Assessment Group's conclusions that etanercept is more efficacious in treating the arthritic component of PsA than adalimumab should be treated with caution.	Please see response to comment no. 3

16	Abbott	Importance of registry data mimicking routine clinical practice to the clinical evidence base: Given that there are no head to head trials of the three anti-TNFs, and the efficacy evidence from randomised controlled trial data for each anti-TNF agent is limited to two studies for each drug, it is also important to consider effectiveness data available from observational data sources. There is limited discussion in the assessment report of the larger evidence base for effectiveness of the three drugs based on observational data. The conclusions regarding comparative effectiveness of the drugs are based on the mixed treatment comparison (MTC) data, which are also used to populate the economic model. Given the magnitude of the differences in HAQ improvement calculated from the MTC of trial data, it would be expected that these differences would also be apparent in the data for effectiveness in clinical practice. The registry data sources indicate that adalimumab is not associated with a lower effectiveness in terms of the arthritis component of the disease. (see details in the report). Furthermore, evidence of comparable joint effectiveness amongst the three anti-TNFs has been found in rheumatoid arthritis. Nixon *et al*. evaluated the use of mixed treatment comparisons and meta-regression to perform indirect comparisons to estimate the efficacy of biologic treatments in rheumatoid arthritis[ii]. The authors found that including study level characteristics of mean baseline disease duration and mean baseline HAQ had a substantial effect on the estimated log odds ratio of an ACR50 event. Results showed that the three TNF antagonists (adalimumab, etanercept and infliximab) appeared to have comparable effectiveness.	Please see response to comment no. 3
17	Abbott	Limitations of the MTC in both the Abbott and Assessment Group MTC: In the MTC for both the Abbott and Assessment Group model, the estimated probability of an ACR70 response for adalimumab is lower than etanercept. Yet, crude comparison of the reported ACR70 data from the trials show that adalimumab has a better ACR70 response than etanercept (see Table 3.6.1 in the report). Given that the ACR70 response level is a much harder level of response to achieve than either the ACR20 or ACR50, it supports the premise that the joint efficacy data for adalimumab and etanercept are similar. Both the Assessment Group and Abbott models link the probability of achieving an ACR70 response to the probability of first achieving an ACR20 response, then achieving an ACR50 response, etc which is a logical approach. However, due to the unexpectedly low ACR20 response rates for adalimumab as a result of the Genovese study, the probability of achieving an ACR70 response is predicted to be lower than etanercept in the MTC because of the hurdle-like approach used to construct the MTC. This low ACR20 response for adalimumab effectively	As Abbott state, the AG followed the same logical approach as the manufacturer took in their own model. It should also be noted that ARC20 response data for adalimumab was not taken from only the Genovese study, but was in fact pooled with the larger dataset from the ADEPT study.

		caps the proportion of patients able to achieve an ACR70 response in the MTC.	
18	Abbott	<u>*Executable Model*</u>: Issue 1 - Mixed Treatment Comparison Inputs provided to different levels of precision Since the results of the MTC are so sensitive to the number of decimal places reported for the change in HAQ, Abbott has provided these data to 4 decimal places in Appendix 1 of our response to the Assessment Group report. Abbott suggests that the Assessment Group request the data to this level of detail from the other manufacturers and uses these data to re-run the MTC.	Please response to comment no. 34
	ECONOMIC EVALUATION		
	Consultee	**Comments**	**The Assessment Group Response**
19	Hull Teaching PCT	The Technology Assessment Report uses different scoring systems for psoriasis severity (PsARC and ACR for arthritis component and PASI for the skin component) and 4 different economic analyses were used (3 pharmaceutical industry and 1 developed by York); are these reflective of everyday practice for initiation, monitoring progress and discontinuation of therapy? (p112-118 in the Assessment Report)	Initiation: we are assessing the licensed indication for PsA. Monitoring: we assume assessment of progress takes place at 3 months after initiation. The base-case analysis assumes clinicians follow the BSR guidelines for patients with PsA and mild-to-moderate psoriasis. Patients withdraw to no therapy if they do not achieve PSARC response. The model does not consider ACR, though chapter 5 estimates the response rates. Discontinuation: the rate of discontinuation after the first 3 months is estimated from general practice (biologics registers). Everyday practice might be more variable in terms of initiation criteria including use in unlicensed indications, and clinicians in practice might use other measures of treatment efficacy such as ACR for arthritis and DLQI for skin response. The current BSR/BAD guidelines recommend PSARC and PASI, among other measures. The model considers alternative scenarios for whether the RCT results reflect general practice, or whether there may be some treatment effect that is specific to patients enrolled in an RCT. These alternative scenarios do not materially affect the results.
20	Hull Teaching PCT	The PASI scoring system has a number of deficiencies (described on p33-4) such as the criteria that 3% of the body surface area has to be affected for the PASI score to be used. The clinical impression from GP representatives in Hull is that patients with less than 3% could	The AG and our clinical advisors also felt that this was an important issue. However, most of the RCTs measured skin involvement using PASI, therefore we lack an evidence base to assess other measures of

		have psoriasis and joint involvement as well – it would have been helpful to have had more discussion on this and the effect on the generalisability of the study findings in the light of this, in the main discussion on p147-151 in the Assessment Report.	psoriasis and its impact on HRQOL and costs. We could recommend that future trials assess a wider set of measures of psoriasis.
21	Hull Teaching PCT	The York model adopts a NHS perspective. We would like to see a supplementary societal perspective discussion or at least an explanation of why this was not undertaken.	Current NICE guidelines (2008) require an NHS perspective as part of its reference case. It is likely that the non-NHS costs and productivity losses of arthritis and psoriasis are very substantial, but we have not modelled the impact of drugs.
22	British Society for Rheumatology	The estimated rate of progression of HAQ is based on poor quality data from the NOAR database where the diagnosis was unclear and it is likely that the more severe cases were excluded.	This is a valid criticism. However, the BSR will be aware how difficult it is to obtain accurate estimates of the natural history of the disease, given that severe patients are usually offered biologics. Previous models used data provided by Wyeth from a very small number of patients in Leeds, and the inclusion criteria were not clear. We also asked experts about natural history in the elicitation exercise. These 3 sources of data gave similar results.
23	British Society for Rheumatology	We also have serious doubts about the elicitation exercise and the assumptions and model parameters used as a result of the exercise. The response rate of experts was poor (5/16) and the reasons for this are not clear. Further, the report is somewhat patronising about the experts, claiming that they did not understand the exercise (despite talking some of the respondents through the process). The results obtained were thought to be unreliable so that the York group chose, as we understand, to dismiss the estimates of progression after discontinuing the drug.	An expert elicitation is by definition an uncertain exercise, as we are asking experts to fill in gaps in the evidence base where we have little or no data from other sources. We made extensive use of the results of the elicitation. The results were used in the base-case for the rate of progression of HAQ on drug, and for long term progression of HAQ after withdrawal. We were very surprised by the results of the exercise for the estimates of the immediate change in HAQ at the time of discontinuation of the drug. The experts appeared to be telling us that patients who withdraw from the drug maintain much of the initial benefit, in addition to maintaining the delay in progression relative to natural history obtained while on the drug. All previous modelling work in PsA and RA, including the industry submissions, maintained that even in the most optimistic scenario patients would 'rebound' and lose all the initial HAQ gain on withdrawal (though patients would maintain the delay in progression obtained while on drug). The base-case model maintains this conventional scenario, while Scenario 2 uses the estimate from the elicitation exercise of the mean HAQ change at the time of withdrawal. This scenario does not materially change the conclusions of the base-case analysis, apart from improving etanercept' cost-effectiveness We also noted that the response rate was poor, however the reason for

			this is unclear and to suggest reasons would only be conjecture. The report did not intent to be patronising to the experts claiming that they did not understand the exercise, indeed it is the analyst's responsibility to ensure that the exercise is designed in such a way to enable experts to complete without difficulty. The AG was very appreciative of the time and effort provided by the experts. We aimed to give a balanced view of the strengths and weaknesses of the exercise. We will review the wording of Appendix 10.11 to ensure the tone is appropriate.
24	Schering-Plough	The cycle length in the model (3months) is inconsistent with the calculation of the cost of the drugs(12 week periods)	This is an error in the AG report and model. The previous version calculated the costs of etanercept and adalimumab over 12 week periods but the cost of infliximab over 3 month periods. We have corrected it and rerun the analysis. All costs are now calculated for 3 month periods for all drugs. This increases the cost of etanercept and adalimumab, but not infliximab. The revised cost calculations are shown in a new version of Appendix 10.13 and the revised model results are shown in a separate paper. Overall, conclusions about the relative cost-effectiveness of the drugs are unchanged.
25	Schering-Plough	The TAG assumed an administration cost of £144 per infusion in their calculations. Schering-Plough would also like to point out that an infusion cost of £124 per infusion is the upper limit of £65.02-£124 range of plausible administration costs for infliximab accepted by the Committee in a recent appraisal of infliximab in psoriasis (TAG 134; Section 4.11, page 14) and should therefore be used in the calculations of ICERs.	The administration costs were taken from the most recent version of the NHS reference costs (April 2007 to March 2008) and are for an elective excess bed day for inflammatory Spine, Joint or Connective Tissue Disorders without complications. TAG 134 was published in Jan 2008 and so the costs relate to an earlier year.
	Schering-Plough	No consideration of vial sharing: Vial optimisation with infliximab has implications on the cost-effectiveness argument currently being appraised. In conjunction with point 1.1, this may significantly reduce the drug acquisition and administration cost of infliximab thereby affecting final ICERs. The TAG did not consider vial optimisation in their analysis. Schering-Plough believes that vial optimisation is a widespread practice in the UK, acknowledged by NICE (Technology Appraisal 133) and should therefore be considered as part of evidence presented to the Committee. Schering-Plough thus recommends: • Incorporation of vial optimisation when calculating the cost of infliximab	The AG were advised by NICE not to consider vial sharing.
26	Wyeth	Extent of redacted information:	The AG will clarify redacted information for the HTA report

		Many of the references to the inputs, assumptions, model structure and results from the economic evaluation submitted by Wyeth have been unnecessarily redacted in the published Assessment Report. Appendix 2 of the Wyeth submission was marked as AIC to prevent the disclosure of the report in its entirety ahead of publication in a peer reviewed journal. It was not intended to restrict the abstraction and reporting of the requisite data to enable a comparison between the evaluation undertaken by the manufacturers and the Assessment Group. Apologies for any ambiguity.	
27	Wyeth	Anticipated costs of biological interventions (Section 3.3, Page 36) The differences in cost between etanercept and adalimumab are due to rounding errors in (hypothetical) cost of a single 25mg vial of etanercept (see Table 10.13.2). It is unclear how the 'annual costs thereafter' have been derived.	The list prices are £357.50 for 40mg of adalimumab and £89.38 for 25mg of etanercept. These account for the very small differences in the acquisition costs of the drugs, of a magnitude of less than £1 per year. Appendix 10.13 shows the calculation in detail.
28	Wyeth	Doses of etanercept (Section 10.13,Page 327): There is no evidence to suggest that the posology of etanercept in practice exceeds the licensed recommendation of 50mg per week in PsA patients. Thus it is unclear why it is assumed that 26 rather than 24 vials are used in subsequent 3 month (12 week) cycles.	(See above for a similar comment from Schering-Plough). This is an error in the AG report and model. The previous version calculated the costs of etanercept and adalimumab over 12 week periods but the cost of infliximab over 3 month periods. We have corrected it and rerun the analysis. All costs are now calculated for 3 month periods for all drugs. This increases the cost of etanercept and adalimumab, but not infliximab. The revised cost calculations are shown in a new version of Appendix 10.13 and the revised model results are shown in a separate paper. Overall, conclusions about the relative cost-effectiveness of the drugs are unchanged.
29	Wyeth	Cost of infliximab infusion (Section 10.13, Page 329, Table 10.13.2): This report sources the cost of infusion as an elective inpatient excess bed day @ £144 whilst a recent NICE costing template (TA 126) utilised HRG H26 @ £793 per day. It would seem appropriate to use a common infusion cost across all appraisals.	(See comment 25 from Schering Plough on the same issue, although Schering Plough recommended using a lower cost for infusion than that used by the AG). The administration costs were taken from the recent version of the NHS reference costs (2007/08) for an elective excess bed day for inflammatory Spine, Joint or Connective Tissue Disorders without complications. The cost of £793 is the average cost for a daycase admission which includes drug costs. The AG has estimated the acquisition costs of drugs separately. Using the cost of £793 for administration would double count the drug costs.
30	Abbott	Abbott is surprised that the report draws such strong inferences from the base case analysis, in particular that adalimumab is extendedly dominated by etanercept, without emphasising that this conclusion is very sensitive to a number of assumptions enumerated elsewhere in the	The AG carried out incremental cost-effectiveness analysis, comparing each drug with the next best alternative. The NICE methods guidance 2008 does not require comparison of treatments with the least effective

		report. The cost-effectiveness of adalimumab and etanercept versus palliative care are very similar in the base case and the report indicates that there is considerable uncertainty around many of the model inputs. In light of these uncertainties and the very similar cost-effectiveness results for adalimumab and etanercept, we would suggest that a strong conclusion of extended dominance of one treatment over the other is misplaced and would ask that this be drawn to the attention of the Appraisal Committee when they consider the report. Abbott also considers that the results of the probabilistic sensitivity analyses do not appropriately reflect the uncertainty that etanercept will be a more cost effective therapy option than adalimumab. This is partly because the mixed treatment comparison results generated by the Assessment Group give a much greater HAQ improvement for etanercept than adalimumab. Furthermore, Abbott considers that the low ICERs for infliximab versus palliative care are highly questionable as they are predicated on both a greater HAQ response for patients receiving infliximab and an average patient weight of 70kg. Neither of these assumptions seems to be easily supported. The base case analysis indicates that the cost-effectiveness of adalimumab and etanercept versus palliative care are very similar, with ICERs of £17,274 and £15,990 respectively. Furthermore, the results of the mixed treatment comparison indicate that there is significant overlap in the credible intervals for response – in particular for ACR and PsARC response rates. It is therefore surprising that the probabilistic sensitivity analysis shows that there is very little uncertainty in which is the most cost-effective of these two treatments (p=0.524 for etanercept and p=0.044 for adalimumab).	alternative, as Abbott seems to be proposing; indeed this would contrary to standard cost-effectiveness decision rules. The probabilistic sensitivity analysis fully propagates the uncertainty in the HAQ change from the MTC into the economic model. According to the base-case MTC results, the mean change in HAQ for responders is considerably lower for adalimumab than the other drugs. We recognise the modelling uncertainty in the base-case and we run a sensitivity analysis (#22) assuming that the mean change in HAQ for responders is the same for all drugs. Adalimumab has an ICER of less than £20,000 per QALY in this scenario, though with a low probability of being the most cost-effective drug for patients with PsA and mild-to-moderate psoriasis.
31	Abbott	There is a high degree of uncertainty in the Assessment Report conclusion that etanercept is the most cost-effective anti-TNF for the treatment of PsA. The effectiveness estimate is driven by two 12-week trials, one containing 30 patients receiving etanercept vs. another trial of 51 patients receiving adalimumab. This uncertainty is compounded by the possibility that the patient populations included in each of these two trials are not the same. Furthermore, data from the BSRBR mimicking routine clinical practice in the UK in a much greater number of patients, suggest that the three anti-TNFs are similarly effective in treating the arthritis component of the disease.	Please see the AG response to comments number 3 and 6 on the clinical review.
32	Abbott	Abbott welcomes the inclusion of the benefits of the different treatments on the skin component of the disease in the modelling conducted for this appraisal. The Assessment Group state that "*the assessment of effectiveness in Section 5.2.2 did not find any appreciable differences in the biologics' response rates for joint disease or psoriasis between approximately 12 weeks compared with 24 weeks*." As a result of this, the Assessment Group used 12 week efficacy data to inform the clinical-effectiveness estimates in their model. Given the strong inference the Assessment Group make in their conclusions about the most cost-effective drug, it is worth noting that there are appreciable differences in the PASI response	There are 3 reasons why the AG used 12 weeks to assess psoriasis response in the model. (i) The RCT evidence base is reduced at 24 weeks (some studies go to open label at 12 weeks). (ii) The clinical review found limited differences in treatment effects between the time points. (iii) Current NICE/BAD guidelines for psoriasis recommend assessment between 10-16 weeks

		rates between weeks 12 and 24 for adalimumab. Further, it is important to recognise that improvements in psoriasis with adalimumab, when 12 week data are used, have been underestimated in the Assessment Group model. Therefore, given that the ICERs for adalimumab and etanercept vs. palliative care are similar, the improved PASI data at week 24 for adalimumab could have an impact on the conclusion made by the Assessment Group that etanercept is the most cost effective treatment option.	
33	Abbott	On page 21 of the Technology Assessment Report (TAR), it states that: "*The response in joint disease (PsARC and ACR) is greater with etanercept than with adalimumab, whereas the response in skin disease (PASI) is greater with adalimumab than with etanercept, though these differences are not statistically significant.*" This statement is based solely on 12 week pooled data from the randomised controlled trials (RCTs) for the two anti-TNFs. Although there were no statistically significant differences between adalimumab and etanercept with regards to joint response, the conclusions from the Assessment group are worded such that etanercept is considered to be the most cost-effective anti-TNF for patients with PsA. Based on the RCT evidence, and given that the annual drug cost of adalimumab and etanercept is equivalent, Abbott considers that this conclusion cannot be robustly supported by the data.	The probabilistic sensitivity analysis propagates the uncertainty in the measures of response from the MTC into the economic model. We conduct sensitivity analyses (#18 and #19) assuming no differences between drugs in PSARC and PASI response respectively, holding the values of all other variables as in the base case.
34	Abbott	HAQ change by PsARC responder/non-responder – MTC inputs: In order to assess the impact of rounding to different decimal places, the MTC was re-run with the HAQ changes rounded to 2 decimal places for each drug. Additionally, the standard errors for adalimumab appear to have been incorrectly calculated. It seems that instead of dividing SD by the square root of the number of patients in each cell, SD was divided by the square root of the total number of patient in each treatment arm. This error was also corrected when re-running the MTC. A comparison of the results shows that the apparently inconsequential issue of rounding has a large impact on the mean HAQ improvements for both adalimumab responders and non-responders, with an increase of 0.0506, and 0.0573 respectively. Given that the differences in the change in HAQ between treatments is a key driver of the results, this improvement in HAQ will result in an increase in QALYs for adalimumab thus changing the cost-effectiveness results. Since the results of the MTC are so sensitive to the number of decimal places reported for the change in HAQ, Abbott has provided these data to 4 decimal places in Appendix 1. Since Abbott does not have access to the response rates from the etanercept and infliximab clinical trials to this level of accuracy, we were unable to determine the exact impact this change will have on the cost-effectiveness results. Abbott therefore suggests that the Assessment Group	The MTC has been re-run using the data provided to 4 decimal places by Abbott. It should be noted that the original data provided by Abbott to the AG was to 1 decimal place. The AG has also corrected an error it made in calculating the standard error. This gives revised estimates of the change in HAQ for responders and non responders by drug. These revised values (and revised costs) have been used as inputs to rerun the cost-effectiveness model base-case and sensitivity analyses. Results are shown in a separate paper. Overall, conclusions about the relative cost-effectiveness of the drugs are not materially changed.

		request the data to this level of detail from the other manufacturers and uses these data to re-run the MTC and the cost-effectiveness analysis.	
35	Abbott	Assumption of different change in HAQ by treatment: On page 114 of the Assessment Report, it states that: *"It is uncertain whether the change in HAQ is the same for all PsARC treatment responders, or depends on the particular biologic treatment followed. In the opinion of our clinical advisor, either scenario could be plausible (Ian Bruce, personal communication) In the base-case model, we allow the change in HAQ for treatment responders to depend on PsARC response and the biologic treatment, and consider the alternative scenario as a sensitivity analysis".* No justification for the decision to allow the change in HAQ to depend on both PsARC response and the biologic treatment is provided. However, this is a key model assumption with sensitivity analyses indicating that adalimumab is no longer dominated by etanercept when the same change in HAQ is applied for all PsARC responders, regardless of treatment. This analysis is presented as scenario 22 in the Assessment Group report, with the results provided in Table 6.6 of the Assessment Group Report. Comparing the results of this sensitivity analysis against the base case analysis, it is clear that the decision to allow the change in HAQ to depend on both PsARC response and the biologic treatment rather than just PsARC response gives lower effectiveness estimates for adalimumab. As outlined in section 2, Abbott considers that it is highly unlikely that adalimumab would have a lower effectiveness on treating the arthritis component of the disease than etanercept or infliximab. The total costs and QALYs for each strategy are shown in Table 2.3.1 (see the report). It can be seen that while the assumption of different HAQ change by PsARC response for each treatment increases the QALYs and reduces the costs for both etanercept and infliximab, the opposite is true for adalimumab. It therefore appears that had the alternative assumption that all biologics have the same change in HAQ at 3 months for a PsARC responder been made, the conclusion of the Assessment Report may have been different as adalimumab would no longer have been dominated by etanercept in the base case analysis, and many sensitivity analyses would have indicated that adalimumab is in fact the most cost-effective strategy. Such a situation is highly likely since this clinical assumption appears to have been entirely arbitrary.	It is unlikely that PSARC response at 12 weeks alone is a perfect predictor of future HAQ. This is clearly seen in the much lower mean HAQ change for PSARC responders on placebo, compared to HAQ changes for PSARC responders on biologic drugs. However, according to our clinical advisor, experts are unsure whether HAQ for responders differs by the type of biologic used. The base-case model assumes that there is a different effect for each biologic. This assumption is uncertain but not arbitrary. First, the MTC shows that the mean HAQ change for responders on adalimumab is considerably lower than for infliximab and etanercept, and the 95% CIs only just overlap. Second, the ACR responses show that the probability of ACR 70, given an ACR 20 response has been gained, does seem to differ by drug. From Table 5.18, in etanercept, 25.9% (= 0.158/0.609) of ACR20 responders have ACR 70, in infliximab it is 29.9% and in adalimumab it is 23.4%. This ranking is consistent with the MTC results for mean change in HAQ by responders for each drug. Therefore in the base case we allowed HAQ for responders to differ by drug. We recognise the modelling uncertainty in the base-case and we run a sensitivity analysis (#22) assuming that the mean change in HAQ for responders is the same for all drugs. Adalimumab has an ICER of less than £20,000 per QALY in this scenario, though with a low probability of being the most cost-effective drug for patients with PsA and mild-to-moderate psoriasis.
36	Abbott	Abbott considers that there is a high degree of uncertainty in the Assessment Group's conclusions about the most cost-effective anti-TNF for the treatment of PsA based on	Please see comments 3 and 6

		differences in joint efficacy arising from a trial containing 30 patients receiving etanercept vs. 51 patients receiving adalimumab for only 12 weeks. Particularly when the patient populations included in each of these two trials are likely not the same (see section 3.3 in the report); and when there are data from the BSRBR mimicking routine clinical practice in the UK in a much greater number of patients, which suggest that the three anti-TNFs have similar efficacy in treating the arthritis component of the disease	
37	Abbott	Impact of the psoriasis component of PsA on quality of life: In the Assessment Group model, the psoriasis component of PsA is not given as much weight in patients with moderate to severe skin disease as the arthritis component. Abbott understands that a proportion of PsA patients will not have moderate or severe psoriasis with their arthritic symptoms. However, it is important to acknowledge the impact psoriasis has on quality of life of those patients who do have moderate-to-severe skin disease. Symptoms of the skin component of PsA occur as visible manifestations that can also cause physical discomfort. The circumscribed, thickened, scaly plaques often cause itching, irritation, and redness, or more severely, physical pain, skin soreness, bleeding from lesions, fatigue and insomnia[iii]. Furthermore, the impact of severe psoriasis on health-related quality of life is considered to be similar to that of other major medical conditions including diabetes, heart disease, and cancer[iv,v]. Compared to placebo-treated patients in psoriasis clinical trials, adalimumab-treated patients demonstrated significant improvements not only in dermatology-specific quality of life measures (DLQI), but also in general health-related quality of life measures (SF-36) and work productivity measures (WPAI-SHP)[vi]. The utility of interrogating the psoriasis rather than the PsA database is that these changes can be ascribed primarily to the effect of adalimumab on skin disease, so these data reinforce that (a) psoriatic skin disease [in PsA or psoriasis] is associated with impairment in general health-related quality of life and work productivity, and (b) that adalimumab is efficacious at mitigating these skin-associated impairments. Abbott considers that improvements in the skin manifestations of PsA should be given greater weight in the consideration of the cost-effectiveness for each intervention in PsA patients with moderate-to-severe psoriasis.	The impact of psoriasis on HRQoL (and costs) has been fully taken into account in the model. The data provided to us by all 3 manufacturers for EQ5D clearly shows that changes in PASI have a much lower impact on HRQoL than changes in HAQ, within the magnitude of the treatment effect expected from biologic drugs. For example, using this HRQoL function, a 0.5 point change in HAQ is associated with a 0.5*0.298 = 0.15 change in EQ5D, while a 10 point change in PASI leads to a 0.04 change in EQ5D. Using the Abbott (log-linear) utility function gave similar conclusions.
38	Abbott	Subgroup analyses: The Assessment Group model is a cohort model, and therefore assumes a homogeneous mix of patients. Although the base case patient characteristics were selected based on expert opinion as to the most common patient type observed in clinical practice, these characteristics by definition represent only a subgroup of the PsA patient population and are not reflective of the mix of PsA patients. In recognition of this limitation, the Assessment Group conducted some subgroup analyses using alternative patient.	The basecase assumes that the BSR continuation rule will be used for patients with PsA and mild to moderate psoriasis. We have now run another sensitivity analysis where the BAD continuation rule is applied to patients with PsA and mild to moderate psoriasis. This does not materially change the results For patients with PsA and moderate to severe psoriasis the report

		The Assessment Group has shown that the patient characteristics have a significant impact on the results of the cost-effectiveness analysis. In order to obtain an accurate picture of the expected cost-effectiveness of each of the treatments, it is therefore important to consider the results of these subgroup analyses alongside the base case results. Furthermore, the base case analysis assumes that patients continue to receive treatment only if a PsARC response is achieved at 3 months in line with the BSR guidelines. However, the BAD guidelines state that a patient should also continue to receive treatment if a PASI 75 response is achieved. Abbott considers that it is reasonable to assume that patients with both skin and joint involvement would be managed by both a rheumatologist and a dermatologist, and that both of these guidelines would therefore apply. Scenario analyses conducted by the Assessment Group indicate that when using either the BSR or the BAD stopping rule, adalimumab is the most cost-effective treatment when using a threshold of £20,000/QALY.	considered an analysis where the BSR rules were applied and another analysis where the BAD continuation rules were applied. Abbott is correct that if BAD rules are used for patients in this group, the model shows adalimumab is cost-effective at a threshold of £20,000/QALY, though all biologics have a similar probability of being cost-effective (that is, it is the most effective drug with an ICER less than £20,000 per QALY, though all drugs have a similar probability of being cost-effective at this threshold).
39	Abbott	Sensitivity analysis varying baseline HAQ: The change in HAQ score is a key model input which is modelled using a random-effects meta-analysis. The code for this analysis indicates that the change in HAQ score depends on the baseline HAQ. However, when the Assessment Group conducted the sensitivity analysis in which the baseline HAQ was increased from 1.05 as per the base case analysis to 1.8 (analysis 10), it appears that they failed to alter the change in HAQ score simultaneously. This sensitivity analysis is therefore incorrect and does not accurately reflect the expected cost-effectiveness in this population.	The clinical review did not explore baseline-treatment effect interactions (subgroup effects), as proposed by Abbott. Therefore we do not have data to examine whether the treatment effect varies for different baseline HAQ. We therefore assume that relative treatment effects between drugs are independent of baseline HAQ. In the model, a greater baseline HAQ only means that a greater proportion of patients will reach a ceiling effect on the progression of arthritis without biologics, as the maximum HAQ score is 3.
40	Abbott	Infliximab Costs: Abbott notes that the Assessment Group assumes an average patient weight of 70kg based on the average weight of the UK population. In order to determine whether this weight is representative of the psoriatic arthritis population, Abbott used data from M02-570 and ADEPT trials for adalimumab, and the smaller etanercept trial (Mease, 2000) to calculate the average weight of moderate to severe PsA patients enrolled in clinical trials. Patient weight was not reported for either of the infliximab clinical trials, nor in the larger etanercept trial (Mease, 2004). The average weight from these three trials was calculated to be 87kg. Since infliximab has a weight-based dosing schedule, patients weighing over 80kg would require one additional vial than patients weighing 70kg which will increase the costs associated with infliximab.	The base-case assumes 4 vials of infliximab (100mg/vial) per administration, assuming no vial sharing, 5mg/kg and patient weight of 60-80kg We believed this would be the most common modality based on the weight of men and women in the general population, though a minority would require 3 or 5 vials per administration. However, our clinical advisor suggested that patients with psoriasis tend to be heavier than the general population. Mean patient weight was 82 to 88kg in the IMPACT trials. We have run another sensitivity analysis for 5 vials of infliximab. Infliximab would be very unlikely to be cost-effective if 5 vials were required per administration.

41	Abbott	The Assessment Group assume a ½ day in-patient hospital cost for each infusion of infliximab at a cost of £144 per infusion. However, since an infliximab infusion is more likely to be a day case rather than an in-patient procedure, this would be a more appropriate cost to use. The NHS reference costs (2007/08) indicate that the cost would therefore be £462[vii].	(See comment no. 25 on the same issue, although Schering Plough recommended using a lower cost for infusion than that used by the AG). The administration costs were taken from the recent version of the NHS reference costs (2007/08) for an elective excess bed day for inflammatory Spine, Joint or Connective Tissue Disorders without complications. The cost of £462 is the average cost for an admission which includes drug costs. The AG have estimated the acquisition costs of drugs separately. Using the cost of £462 for administration would double count the drug costs.

[i] Genovese M, Mease P, Thomson G, Kivitz A, Perdok R, Weinberg M, Medich J, Sasso E. Safety and efficacy of adalimumab in treatment of patients with psoriatic arthritis who had failed disease modifying anti-rheumatic drug therapy. *The Journal of Rheumatology* 2007;**34**(5)1-11.

[ii] Nixon RM, Bansback N, Brennan A. Using mixed treatment comparisons and meta-regression to perform indirect comparisons to estimate the efficacy of biologic treatments in rheumatoid arthritis. *Stat Med* 2007;**26**(6):1237-54.

[iii] Mease PJ, Menter MA. Quality-of-life issues in psoriasis and psoriatic arthritis: Outcome measures and therapies from a dermatological perspective. *J Am Acad Dermatol.* 2006;**54**:685-704.

[iv] Rapp SR, Feldman SR, Exum ML, Fleischer AB,Jr, Reboussin DM. Psoriasis causes as much disability as other major medical diseases. *J Am Acad Dermatol.* 1999;**41**:401-407.

[v] Krueger G, Koo J, Lebwohl M, et al. The impact of psoriasis on quality of life: results of a 1998 National Psoriasis Foundation patient-membership survey. *Arch Dermatol.* 2001;**137**(3):280-284.

[vi] Revicki DA, Willian MK, Menter A, Gordon KB, Kimball AB, Leonardi CL, Langley RG, Kimel M, Okun M. Impact of adalimumab treatment on patient-reported outcomes: results from a Phase III clinical trial in patients with moderate to severe plaque psoriasis. J. Dermatolog Treat. 2007; 18(6): 341-350.

[vii] NHS Reference Costs 2007/08 Day Case HRG code: HD23C (Inflammatory Spine, Joint or Connective Tissue Disorders without cc)