Evidence Review Group Report

Apremilast for treating active psoriatic arthritis

Erratum

Issue 1

p12 The text: “There is still uncertainty about the long-term safety of apremilast as the safety data is currently limited to one year”

Should read: “There is still uncertainty about the long-term safety of apremilast as the safety data is currently limited to two years”

p15 The text “There is no evidence that apremilast is better tolerated than biologic therapies, and longer-term safety data for apremilast are required as currently the safety data extends to only one year.”

Should read “There is no evidence that apremilast is better tolerated than biologic therapies, and longer-term safety data for apremilast are required as currently the safety data extends to only two years.”

p48 The text: “There is still uncertainty about the long-term safety of apremilast as the safety data is currently limited to one year”

Should read: “There is still uncertainty about the long-term safety of apremilast as the safety data is currently limited to two years”

p121 The text: “There is still uncertainty about the long-term safety of apremilast as the safety data is currently limited to one year”

Should read: “There is still uncertainty about the long-term safety of apremilast as the safety data is currently limited to two years”

p122 The text: “There is still uncertainty about the long-term safety of apremilast as the safety data is currently limited to one year”

Should read: “There is still uncertainty about the long-term safety of apremilast as the safety data is currently limited to two years”

Issue 2

P31 the text ““Otezla has been shown to improve physical function” was therefore removed from the indication by the manufacturer.” should be deleted.

Issue 8

p12 the text “apremilast (30mg) 13%;” has been marked AiC

p13 the text “Similarly across all other outcomes (ACR20/70, PsARC, HAQ-DI, PASI-50/75/90) apremilast showed a better response (or probability of response) than placebo, but a lower response (or probability of response) than all the other interventions.” and the text “whilst apremilast is more effective than placebo it is less effective than the biologic therapies.” has been marked AiC.

P41 the text “For all three ACR thresholds apremilast showed a higher probability of response compared with placebo, but a lower probability of response than all the other interventions.” has been marked AiC.

p42 the text “the smallest reduction was seen following treatment with apremilast” has been marked AiC

p42 the text “and the smallest reduction was for apremilast” has been marked AiC

p47 the text “outcomes apremilast showed a higher probability of response compared with placebo, but a lower probability of response than all the other interventions.” has been marked AiC

p48 the text “apremilast can be shown to improve some of the symptoms of psoriatic arthritis, although it is less effective than all of the biological therapies - sometimes considerably less effective.” has been marked AiC

p121 the text “whilst apremilast is more effective than placebo it is less effective than the biologic therapies.” has been marked AiC

Issue 12

p99 The text: “In addition to the concerns regarding positioning of apremilast, in the scenario analyses using (i) the utility model based on apremilast data and (ii) the withdrawal rates based on apremilast RCTs, the manufacturer has implemented them by applying the apremilast-specific data to all treatments.”

Should read: “In addition to the concerns regarding positioning of apremilast, in the scenario analyses using the utility model based on apremilast data, the manufacturer has implemented it by applying the apremilast-specific data to all treatments”

Issue 15

p71 The text: “Table 14 Differences in PsARC response probabilities between updated NMA and economic model”

Should read: “Table 14 Differences in PASI response probabilities between updated NMA and economic model”

Issue 17

p107 The text: “The therapies have an equivalent cost”

Should read: “The therapies have similar cost”

Issue 20

p76 The text: “In addition, the apremilast-specific withdrawal rate in the scenario analysis appeared to be applied across all active treatments (including anti-TNFs) instead of being used only for apremilast.”

Should read: “A scenario where the apremilast-specific withdrawal rates were applied to patients treated with apremilast was conducted in the manufacturer’s scenario analyses”

All three randomised trials were of a very similar design and all were well conducted. Placebo-controlled comparisons with apremilast 20mg bid and apremilast 30mg were made up to 16 weeks. Baseline characteristics were very similar across the three apremalist trials, with (treatment arm) ranges for characteristics as follows: mean age 48.7 to 51.4 years; mean PsA duration 6.8 to 8.1 years; mean SJC 9.2 to 12.8; mean BMI 29.2 to 30.1kg/m^2; mean PASI score 7.9 to 9.2; mean HAQ-DI 1.0 to 1.2; prior biologic use 14.2% to 29.6%; concomitant small molecular DMARD use 60% to 70%.

In all three trials, the primary efficacy outcome was ACR20 response at 16 weeks and the major secondary efficacy outcome was change from baseline to Week 16 in HAQ-DI score. Modified PsARC response and PASI-75 response were also included as outcomes. None of the trials included radiographic assessment of joint damage as a formal outcome.

Short-term placebo comparison results

The pooled analysis of PSA 002, 003 and 004 demonstrated that apremilast 30 mg (N=497) was statistically significantly more effective than placebo (N=496) for the outcomes: ACR 20/50/70 response; mean change in HAQ-DI; achieving a MCID of ≥0.13 on the HAQ-DI; achieving a MCID of ≥0.30 on the HAQ-DI; change from baseline in SF-36v2 PF score; change from baseline in SF-36v2 PCS score; change from baseline in FACIT-Fatigue score; achieving a modified PsARC response; change from baseline in dactylitis severity score; change from baseline in swollen joint count (0–76); change from baseline in tender joint count (0–78); and patients achieving a PASI-75. However, the proportion of apremilast patients achieving an ACR50 response is quite low (13.9% apremilast versus 6.5% placebo) and there is uncertainty about whether the improvement in function provided by apremilast reaches clinically-relevant levels.

Most adverse events reported in patients receiving apremilast were mild or moderate, with few patients reporting severe or serious adverse events during the placebo-controlled phases. Trials of apremilast in psoriasis (PSOR 008 and PSOR 009) were included as supporting evidence of apremilast's tolerability and safety; the ERG identified a three-arm trial of apremilast and etanercept versus placebo (PSOR-010), that could potentially inform the evidence base on adverse events. There is still uncertainty about the long-term safety of apremilast as the safety data is currently limited to two years.

Following an ERG request for a more comprehensive set of analyses, updated network meta-analyses were presented which compared the efficacy of apremilast with adalimumab, etanercept, infliximab, certolizumab and ustekinumab. Across the treatments the absolute probability (fixed-effect model) of ACR 50 response was: placebo 5%; apremilast (30mg) [redacted]% [AiC]; adalimumab 35%; certolizumab (200mg) 27%; etanercept 39%; golimumab (50mg) 40%; infliximab 49%; and ustekinumab (90 or

63mg) 25%. Similarly across all other outcomes (ACR20/70, PsARC, HAQ-DI, PASI-50/75/90) apremilast showed a better response (or probability of response) than placebo, but a lower response (or probability of response) than all the other interventions.

Longer term effects

The MS reported that responses beyond week 24 were maintained over those at week 16 for several outcomes including ACR20 ACR50, ACR70, PsARC and HAQ-DI. However, these results are subject to numerous methodological limitations, in particular they are based on observed data from selected patients who remain on treatment and therefore represent an optimistic estimate of treatment benefit. The ERG calculated that across the three trials the annual withdrawal rate was around 25%.

1.3 Summary of the ERG's critique of clinical effectiveness evidence submitted

The evidence for the clinical effectiveness of apremilast is based on good quality randomised trials and the results are likely to be reliable. The evidence indicates that whilst apremilast is more effective than placebo it is less effective than the biologic therapies. Furthermore, it was evident that only a small proportion of patients taking apremilast achieve the most clinically important levels of response (such as ACR 50). There is also some uncertainty about whether the improvement in function provided by apremilast reaches clinically-relevant levels. No evaluation of apremilast's efficacy relative to small molecule DMARDs was presented. Important methodological limitations of the longer -term data were not discussed by the manufacturer; the efficacy results of the long-term studies should not be considered as being reliable.

The absence of any evidence of a beneficial effect of apremilast on radiographic progression is extremely important given the importance of early management of PsA to minimise disease progression in joints, emphasised by the NICE guidance, and the evidence for anti-TNFs (which do reduce rates of radiographic progression of joint damage). Without any evidence of apremilast having an effect on radiographic progression, the validity of positioning apremilast before anti-TNFs in a treatment sequence (as proposed by the manufacturer) appears highly questionable.

There is no clear evidence to indicate that apremilast is better tolerated than biologics in patients with PsA, despite this being a key reason for the manufacturer positioning apremilast before biologics in the treatment pathway. Similarly, although the manufacturer suggested positioning apremilast before biological therapy in the treatment pathway, the MS did not present data on patients' response to biological therapies after having received apremilast; it is therefore unclear whether subsequent treatment effectiveness is affected by prior use of apremilast.

manufacturer's analysis. The ERG is very cautious of this assumption given that no long-term clinical evidence is available to support this, such as data assessing radiographic disease progression.

Equally important, the fact that not all comparators relevant to the decision problem were included in the manufacturer's model, most critical of them being ustekinumab, is a key limitation of the analysis. The inflexible economic model and the limited efficacy data for ustekinumab provided by the manufacturer in their responses did not allow the ERG to include ustekinumab as a treatment option in the model.

The ERG also had concerns about the assumptions made regarding effect degradation for subsequent line of anti-TNFs following previous anti-TNF or apremilast, the monitoring costs of apremilast, the placebo response, the application of the same withdrawal data for anti-TNFs and apremilast, disease related costs applied for HAQ and PASI and the utility algorithm used. In addition, a series of data inconsistencies were identified by the ERG between the submission document and the economic model. Updates to the NMA, in terms of excluding Phase II trial data and removing unlicensed arms of apremilast, were also not implemented within the economic model.

ERG commentary on the robustness of evidence submitted by the manufacturer

Strengths

The clinical evidence presented was appropriately based on a systematic review and the evidence for apremilast was derived from three good quality RCTs. The comparison with the biologic therapies was based on an appropriate NMA, after it was updated to address the ERGs concerns regarding missing comparators and the inclusion of unlicensed doses of apremilast.

A de novo model based on previous NICE technology appraisals was developed. The revised model submitted in response to the points for clarification allowed apremilast to be considered at different positions in the sequence.

Weaknesses and areas of uncertainty

There are no data for the effect of apremilast on the radiographic disease progression of joint damage, despite the fact that in one trial patients have been followed up for two years. This absence of any evidence that apremilast slows radiographic progression is very important, given the evidence for anti-TNFs - which do reduce rates of radiographic progression of joint damage - and the importance of early management of PsA. There is no evidence that apremilast is better tolerated than biologic therapies, and longer-term safety data for apremilast are required as currently the safety data extends to only two years.

4.2.4 Summary of the results of the included trials

Where available, pooled results from the PALACE trials are discussed here. The MS focused on efficacy results from trial PSA-002 (the reasons for this decision were not stated), and results from the pooled analysis for data on improvements in enthesitis, dactylitis, and psoriatic skin manifestations. Further result data (for PSA-003, PSA-004 and the pooled analysis) were presented in the MS appendices. The ERG received the full report of the pooled analyses following a request for this document. The ERG also requested subgroup analyses from the pooled data, relating to the following populations:

Asymmetric oligoarticular patients versus symmetric polyarticular patients
Axial disease patients versus non-axial disease patients
Dactylitis patients with versus patients without dactylitis
Enthesopathy (MASES) patients versus patients without enthesopathy

Efficacy results at 16 weeks

Table 1 shows the efficacy results for the pooled analysis of all three PALACE trials. Apremilast was associated with statistically significant improvements (compared with placebo) for the primary outcome of ACR20 response, and for most other outcomes, the exceptions being ACR70 response and enthesitis (assessed using MASES). However, size of the treatment benefit was modest with primary outcome of ACR20 being achieved by 37% of patients on apremilast compared with 19% on placebo. Also, the proportion of apremilast patients achieving an ACR50 response was quite low (13.9% apremilast versus 6.5% placebo); ACR50 is likely to be a far more clinically-important outcome for patients than ACR20.[10, 11] Similarly, the implications of apremilast achieving a PsARC response rate of 49% can only fully be realised when the placebo response rate of 30% is considered and 36.4% of patients achieving a MCID of ≥0.30 on the HAQ-DI compared with 26% on placebo. Outcomes which are prone to high placebo responses may not provide the most informative estimates of relative efficacy (see section 4.4).

The HAQ-DI outcome is important in terms of the patient's physical functioning and in assessing disease progression. The EMA assessment report commented on the HAQ-DI results, noting that the minimum clinically important difference (MCID) for HAQ-DI in psoriatic arthritis has not been clearly established. The EMA stated that improvements in the HAQ-DI score observed in the apremilast 30mg treatment groups exceeded the estimated MCID of -0.13 provided by one study, but not the estimated MCIDs of -0.3 and -0.35 provided in two other studies.[8]

- for PASI 50/75/90 data inputs were provided for the main network (13 trials) and included all treatments;
- for HAQ-DI change data inputs were provided for main network (10 trials) - golimumab and certolizumab were not included
- for PsARC data inputs were provided for main network (13 trials) - included all treatments
- for HAQ-DI change conditional on PsARC response data inputs were provided for the main network (responders 8 trials; non-responders 8 trials) – golimumab, ustekinumab and certolizumab were not included.

  In the response to an ERG request for an analysis of anti-TNF experienced and separately contraindicated patients the manufacturer conducted an analysis using the Bucher method. Whilst this simple method for indirect comparisons between two trials of different active treatments versus a common comparator (usually placebo) is acceptable, it is less so when there is more than two active comparators and when there are multiple trials for at least one comparator. In such cases more appropriate NMA methods are preferable.

4.4.2 NMA results

The NMA in the original submission did not include certolizumab or ustekinumab and also the results were provided mostly only as absolute probabilities. In response to a request from the ERG results for updated analyses were provided: absolute effect, effect relative to placebo and apremilast, and results for all active treatments relative to each other. These results were provided for the following outcomes: ACR 20/50/70; PsARC; PASI 50/75/90; HAQ-DI; HAQ conditional on PsARC response. Results from both fixed effect and random effects models were provided. As there are a very large number of results, not all are presented here.

The results of the updated network meta-analyses for the fixed effect model ACR response outcomes in the full population are presented in Tables 3 and 4; they are similar to the NMA results originally submitted by the manufacturer. The results are presented as absolute probabilities, and as odds ratios for comparisons with placebo. The highest probability of response for ACR 20, 50 and 70 outcomes was seen with infliximab. For all three ACR thresholds apremilast showed a higher probability of response compared with placebo, but a lower probability of response than all the other interventions. The results for ACR 50 – which is a more desirable clinical outcome for patients than ACR 20[10, 11] – are shown in Table 7 for all comparators against each other. The results from the random-effects model were very similar to those obtained using the fixed-effect model.

A similar pattern of results was seen for the PsARC and PASI response outcomes (Table 5). Table 6 presents changes in HAQ-DI scores for all treatment comparisons. Compared with placebo, large reductions in HAQ-DI were seen following treatment with infliximab and etanercept; the smallest reduction was seen following treatment with apremilast. Reductions in HAQ-DI were larger in patients who achieved PsARC response than in those who did not – once again the largest reductions were seen for infliximab and etanercept, and the smallest reduction was for apremilast. These responder status HAQ-DI results were only available from the original MS: the updated NMA results did not appear plausible. The ERG asked for revised results but they were not provided before the ERG report deadline.

Table 1 Updated NMA ACR response outcomes - absolute probability (fixed-effect model)

Intervention	No of trials	Mean absolute probability (95% CrI)		
		ACR 20	ACR 50	ACR 70
Placebo	12	0.16 (0.14, 0.18)	0.05 (0.04, 0.06)	0.01 (0.01, 0.02)
Apremilast 30mg	3	[redacted] [AiC]	[redacted] [AiC]	[redacted] [AiC]
Adalimumab 40mg	2	0.59 (0.5, 0.68)	0.35 (0.27, 0.45)	0.16 (0.11, 0.23)
Certolizumab 200mg	1	0.49 (0.38, 0.6)	0.27 (0.18, 0.36)	0.11 (0.06, 0.17)
Certolizumab 400mg	1	0.40 (0.3, 0.51)	0.20 (0.12, 0.28)	0.07 (0.04, 0.12)
Etanercept 25mg	2	0.62 (0.5, 0.73)	0.39 (0.27, 0.51)	0.19 (0.11, 0.28)
Golimumab 50mg	1	0.64 (0.5, 0.77)	0.40 (0.27, 0.55)	0.20 (0.11, 0.32)
Golimumab 100mg	1	0.59 (0.45, 0.73)	0.35 (0.23, 0.5)	0.17 (0.09, 0.27)
Infliximab 5mg/kg	2	0.72 (0.61, 0.81)	0.49 (0.37, 0.61)	0.27 (0.17, 0.37)
Ustekinumab 90 or 63mg	1	0.47 (0.3, 0.64)	0.25 (0.13, 0.4)	0.10 (0.04, 0.19)

Table 2 Updated NMA ACR response outcomes - odds ratios (comparisons with placebo, fixed-effect model)

Intervention	Median OR (95% CrI)		
	ACR 20	ACR 50	ACR 70
Apremilast 30mg	[redacted] [AiC]	[redacted] [AiC]	[redacted] [AiC]
Adalimumab 40mg	7.86 (5.23, 12.04)	9.88 (6.41, 15.32)	14.79 (8.99, 24.33)
Certolizumab 200mg	5.28 (3.31, 8.38)	6.55 (3.94, 10.63)	9.31 (5.13, 16.25)
Certolizumab 400mg	3.64 (2.26, 5.81)	4.399 (2.58, 7.28)	5.88 (3.11, 10.6)
Etanercept 25mg	9.05 (5.44, 15.47)	11.38 (6.72, 19.4)	17.28 (9.51, 31.31)
Golimumab 50mg	9.78 (5.22, 18.56)	12.28 (6.43, 23.07)	18.82 (9.11, 37.32)
Golimumab 100mg	7.83 (4.26, 14.86)	9.84 (5.18, 18.71)	14.72 (7.07, 29.91)
Infliximab 5mg/kg	13.86 (8.23, 24.25)	17.26 (10.22, 29.56)	27.14 (15.13, 48.63)
Ustekinumab 90 or 63mg	4.71 (2.32, 9.66)	5.80 (2.64, 12.26)	8.11 (3.21, 18.97)

Heterogeneity

The MS did not evaluate heterogeneity across the network, although the ERG acknowledges that all the other treatment data (i.e. other than apremilast) was informed by only one or two trials. For the three apremilast studies, the MS reported that statistically significant heterogeneity was observed for apremilast 30 mg for HAQ ($p < 0.001$) and PsARC ($p < 0.03$) treatment estimates. The manufacturer reported these results again in response to an ERG request for NMA method details. They also presented bar charts showing the ranges of results across the apremilast trials, including the trials which were of unlicensed doses of apremilast. These charts suggest that the PALACE 4 trial (of DMARD-naïve patients) may have been the cause of heterogeneity for the HAQ results.

Rates of placebo response

The course of psoriatic arthritis is often characterised by flares and remissions. This, together with the effect of placebo intervention associated with the treatment ritual, is likely to be important in dictating the rates of response seen in placebo groups. Across the trials in the NMA, rates of response in placebo patients had the following ranges: PsARC, 21% to 38%; ACR20, 9% to 19%; ACR 50, 0% to 8.3%; ACR 70 0% to 2.4%; PASI 50, 0% to 25%.

Outcomes which are prone to high placebo responses may not provide the most informative estimates of relative efficacy; the diluting effect of a placebo response on relative effect estimates diminishes as thresholds increase. This is illustrated by examining the increasing odds ratios as the ACR and PASI cut-offs increase (Tables 4 and 5). As the bar for response is raised, the difference in the proportion of responders between active treatment and placebo groups increases as the placebo effects outlined above become less likely to be able to achieve the higher thresholds.

4.5 Conclusions of the clinical effectiveness section

Three well-conducted double-blind RCTs of apremilast at the licensed dose showed statistically significant differences favouring treatment with apremilast over placebo for week 16 comparisons for almost all outcomes. However, the treatment effect sizes were modest: the proportion of apremilast patients achieving an ACR50 response (a clinically relevant response) was quite low (13.9% apremilast versus 6.5% placebo) and there is also some uncertainty about whether the improvement in function (HAQ-DI) provided by apremilast reaches clinically-relevant levels.

NMAs comparing the efficacy of apremilast with adalimumab, etanercept, infliximab, certolizumab and ustekinumab showed that across all outcomes apremilast showed a higher probability of response compared with placebo, but a lower probability of response than all the other interventions.

Although the manufacturer suggested positioning apremilast before biological therapy in the treatment pathway, the MS did not present data on patients' response to biological therapies after having received apremilast; it is therefore unclear whether subsequent treatment effectiveness is affected by prior use of apremilast.

The MS presented limited data that HAQ_DI benefit was maintained in patients who remain on treatment. However, important methodological limitations of the longer -term data were not discussed by the manufacturer; the efficacy results of the long-term studies should not be considered as being reliable. There are no data for the effect of apremilast on the radiographic disease progression of joint damage, despite the fact that in one trial patients have been followed up for two years. This absence of any evidence that apremilast slows radiographic progression is very important, given the evidence for anti-TNFs - which do reduce rates of radiographic progression of joint damage - and the importance of early management of PsA. Without any evidence of apremilast having an effect on radiographic progression, the validity of positioning apremilast before anti-TNFs in a treatment sequence (as proposed by the manufacturer) appears highly questionable.

Most adverse events reported in patients receiving apremilast were mild or moderate, with few patients reporting severe or serious adverse events during the placebo-controlled phases. However, there is no clear evidence to indicate that apremilast is better tolerated than biologics in patients with PsA, despite this being a key reason for the manufacturer positioning apremilast before biologics in the treatment pathway. A three arm trial of apremilast and etanercept versus placebo (PSOR-010) in psoriasis, could potentially inform the evidence base on adverse events. There is still uncertainty about the long-term safety of apremilast as the safety data is currently limited to two years.

In summary, when compared with placebo, apremilast can be shown to improve some of the symptoms of psoriatic arthritis, although it is less effective than all of the biological therapies - sometimes considerably less effective. There is no evidence that apremilast can slow radiographic disease progression of PsA, nor is there evidence that it is better tolerated than biologic therapies.

for the probability of PsARC and PASI response versus the updated NMA results in Tables 13 and 14. While there are differences in absolute probabilities for both PsARC and PASI response, these do not appear to be very big. More importantly, the fact that differences in efficacy for all treatments are moving in the same direction for all treatments and there is no re-ordering of treatments in terms of PsARC or PASI response, suggests that the impact on comparative results would not be expected to be significant. For HAQ changes, there have been several issues with the updated results that the manufacturer submitted; HAQ changes according to PsARC response status were not provided for the updated NMA. In this case, the ERG has not been able to explore the impact of the differences in HAQ score changes between the updated NMA and the economic model inputs. The fact that the appropriate efficacy inputs were not used in the manufacturer's updated model is, according to the ERG, a limitation of the submitted analysis.

Table 3 Differences in PsARC response probabilities between updated NMA and economic model

Probability of PsARC response	Updated NMA	Model
Apremilast 30mg	[redacted][AiC]	[redacted][AiC]
Adalimumab 40mg	0.6335	0.6231
Etanercept 25mg	0.7747	0.7406
Infliximab 5mg/kg	0.8016	0.7891
Golimumab 50mg	0.8013	0.7953

Table 4 Differences in PASI response probabilities between updated NMA and economic model

	PASI 50		PASI 75		PASI 90	
	Updated NMA	Model	Updated NMA	Model	Updated NMA	Model
Apremilast 30mg	[redacted][AiC]	[redacted][AiC]	[redacted][AiC]	[redacted][AiC]	[redacted][AiC]	[redacted][AiC]
Adalimumab 40mg	0.7111	0.7497	0.5076	0.5519	0.2714	0.3044
Etanercept 25mg	0.3774	0.4103	0.1938	0.22	0.06629	0.0783
Infliximab 5mg/kg	0.8889	0.9162	0.7539	0.798	0.5227	0.5751
Golimumab 50mg	0.6859	0.7261	0.4783	0.5232	0.2473	0.2797

5.2.6.3 Decline in efficacy for subsequent lines of therapy

Within the economic model, PsARC response rates for patients receiving an anti-TNF later in the treatment pathway, following primary non-response to at least one previous anti-TNF, were assumed to be reduced, based on clinical expert opinion and published studies in both PsA and RA [20 21, 34, 35]. The base case analysis included a reduction in anti-TNF efficacy rates following primary non-

The ERG notes that the assumption of zero HAQ-DI progression for apremilast is subject to considerable uncertainty. The sensitivity of the model to assumptions regarding disease progression whilst on continued treatment is explored in section 6.

5.2.6.6 Discontinuation

Rates of withdrawal from treatment in the MS were sourced from a meta-analysis reported by Rodgers et al. [20] which provides a pooled estimate of withdrawal rate for patients receiving anti-TNFα therapies based on registry data from several countries, including data for adalimumab, etanercept and infliximab. In the base case analysis, annual withdrawal rates were assumed to be the same for all anti-TNFs and apremilast (16.5%). The manufacturer justified the choice of withdrawal rate on the basis that it has been used in the recent NICE appraisals of golimumab (TA220) [31] and of etanercept, infliximab and adalimumab (TA199).[32] The withdrawal rate for apremilast was assumed to be the same as for anti-TNF therapies, in the absence of available apremilast long-term data at the time of the submission, according to the manufacturer. There is lack of clarity on whether the manufacturer conducted a search for more updated evidence on the discontinuation of the treatments included in the economic analysis, especially since the economic evaluations referenced were published prior to 2011. In addition, the assumption of equal withdrawal rate for apremilast to that for anti-TNFs is subject to significant uncertainty.

An alternative scenario was specified using withdrawal rates based on week 52–104 in the PSA-002 RCT (all cause withdrawals). This was specified as 22.1% in the original submission, based on [redacted] [AiC] patients withdrawing from continued use during the second year of the trial and later changed to 15.8% (27/171) because of a typographical error (at points of clarification). No additional information was provided in the original submission and the ERG requested additional details from the manufacturer, regarding numbers of patients discontinuing due to adverse events and loss of efficacy. The manufacturers provided the reasons for discontinuation, however the numbers of drop outs at 104 weeks do not appear to agree with the patients numbers presented elsewhere in the points for clarification (in support of continued HAQ response the tables giving the number of patients at 16,52 and 104 weeks in PSA-002). At 104 weeks the number of patients in the Apremilast 30mg arm is stated at 101. It is also unclear why only 52-104 week data is presented and not all post 16 week data (including data form the other Apremilast trials). A scenario where the apremilast-specific withdrawal rates were applied to patients treated with apremilast was conducted in the manufacturer's scenario analyses. The ERG considers that it would be more appropriate to apply the apremilast-specific withdrawal rate to apremilast only in the base case

Table 5 Scenario analyses results

Scenario	ICER for apremilast sequence (£/QALY)
Apremilast positioned post-biologic in sequence	Cost-saving but less effective £24,470 saved per QALY lost (comparison was apremilast positioned pre-biologic vs. apremilast positioned post-biologic)
1-year time horizon	Cost saving but less effective £9733 saved per QALY lost
5-year time horizon	111,552
10-year time horizon	33,442
HAQ rebound to natural history	19,114
Utility regression from apremilast trial data	30,223
Utility regression by Abbott	16,754
Utility regression by Schering-Plough	17,082
BSC and other healthcare costs by Poole et al.	7,893

Although the scenarios analyses seem to address some of the concerns raised by the ERG in previous sections, the way some scenarios have been implemented does not allow for the appropriate comparisons to be made.

The ERG is particularly concerned about the sensitivity of the model results to the position of apremilast within the proposed sequence. In addition the sequence where apremilast is positioned post anti-TNF was compared with the sequence where apremilast is positioned pre anti-TNF instead of making the comparison with the treatment sequence not including apremilast, similarly to the base case. The ERG explores the pre and post anti-TNF positioning for apremilast, along with other feasible sequences, in a fully incremental analysis in section 6.

In addition to the concerns regarding positioning of apremilast, in the scenario analyses using the utility model based on apremilast data, the manufacturer has implemented it by applying the apremilast-specific data to all treatments. In the ERG's view, a more realistic scenario would be to explore the impact on the cost-effectiveness results in these scenarios via applying the utility inputs based on apremilast data to apremilast only.

5.2.11 Subgroup analyses

As already discussed in Section 5.2.2, the manufacturer's base case considered the apremilast full trial population which included patients who are naïve to biologic therapy as well as patients who had previously received a biologic. A subgroup analysis was performed, testing the cost-effectiveness results in biologic-naïve patients, i.e. excluding patients from the apremilast full trial population who

(Table 31) the use of etanercept appears to be the most the most effective sequence (i.e. higher QALYs) and is associated with an ICER of £18,997.

Tables 32 and 33 present the scenarios where apremilast displaces an anti-TNF in a sequence of two biologics. Apremilast replaces either the first or the second biologic in: (i) the sequence of adalimumab followed by etanercept, which was used in the manufacturer's base case sequence, and (ii) the sequence reordering these two therapies i.e. etanercept followed by adalimumab. This re-ordering is considered appropriate as etanercept is associated with a higher probability of PsARC response and HAQ gain given PsARC response/no response, compared to adalimumab (see Section 5.2.6.2). The therapies have similar cost.

Table 6 Apremilast displacing a biologic in a sequence of two biologics - Pair-wise comparisons

Sequence	Mean Costs	Mean QALYs	Incremental Costs	Incremental QALYs	ICER
Apremilast replacing the 1st biologic					
apremilast → etanercept → BSC	£102,023	7.273			
adalimumab → etanercept → BSC	£104,936	7.267	£2,913	-0.0058	Dominated
apremilast → adalimumab → BSC	£96,676	6.91			
etanercept → adalimumab → BSC	£107,541	7.44	£10,865	0.5297	£20,512
Apremilast replacing the 2nd biologic					
adalimumab → **apremilast** → BSC	£97,267	6.94			
adalimumab → etanercept → BSC	£104,936	7.27	£7,669	0.3320	£23,099
etanercept → **apremilast** → BSC	£102,959	7.32			
etanercept → adalimumab → BSC	£107,541	7.44	£4,582	0.1182	£38,765

8 Overall conclusions

The evidence for the clinical effectiveness of apremilast is based on good quality randomised trials and the results are likely to be reliable. The evidence indicates that whilst apremilast is more effective than placebo it is less effective than the biologic therapies. Furthermore, it was evident that only a small proportion of patients taking apremilast achieve the most clinically important levels of response (such as ACR 50). There is also some uncertainty about whether the improvement in function provided by apremilast reaches clinically-relevant levels. No evaluation of apremilast's efficacy relative to small molecule DMARDs was presented. Important methodological limitations of the longer -term data were not discussed by the manufacturer; the efficacy results of the long-term studies should not be considered as being reliable. Importantly there is no evidence to support the impact, if any, of apremilast on radiographic progression of joint disease.

Most adverse events reported in patients receiving apremilast were mild or moderate, with few patients reporting severe or serious adverse events during the placebo-controlled phases. There is still uncertainty about the long-term safety of apremilast as the safety data is currently limited to two years. There is no clear evidence to indicate that apremilast is better tolerated than biologics in patients with PsA.

The ERG feel that the use of treatment sequences to address the decision scope as specified by NICE is not justified and the base-case sequences proposed by the manufacturer represent a limited set of potentially relevant sequences and that the manufacturer's base-case cost-effectiveness results are not a sufficient basis to inform the most efficient use and positioning of apremilast. In addition there are a number of parameter uncertainties within the manufacturer's model. The most critical of these is the assumption of zero HAQ-DI progression for PsARC responders to apremilast remaining on treatment: without radiographic data or randomised trial data to support this assumption it is not evident that such an assumption is reasonable. The ERG also has concerns about the assumptions made regarding effect degradation for subsequent line of antiTNFs following previous anti-TNF or apremilast, monitoring costs of apremilast, the placebo response application of the same withdrawal data for antiTNFs and apremilast, disease related costs applied for HAQ and PASI and the utility algorithm used.

The additional analyses undertaken by the ERG suggested that adding lines of treatments (i.e. moving from sequences including one treatment to sequences including two and three lines of treatment) results in more effective (i.e. higher QALYs) strategies. The treatment sequences where apremilast is positioned post anti-TNF were generally more effective than those where it was placed pre anti-TNF, depending however on whether the assumption anti-TNF effect degradation was applied or not. The

ICER is £24,175 for etanercept/adalimumab/apremilast/BSC compared to apremilast/etanercept/adalimumab/BSC.

In exploring a key uncertainty in the model, the assumption of zero HAQ progression for apremilast responders that remain on treatment, the ERG was extremely constrained due to the model programming. The scenario comparing apremilast with BSC and assigning the same HAQ progression to apremilast and BSC (i.e. zero progression) was shown to have a significant impact on results; the ICER of apremilast versus BSC was £66,045 versus £14,645 for the same comparison when assuming natural disease progression for BSC.

The results suggest that the cost-effectiveness results for the scenarios comparing treatment sequences would likely be significantly altered (i.e. become less favourable for the treatment sequences including apremilast) if alternative assumptions regarding disease progression for those remaining on apremilast were to be explored.

Implications for research

The absence of any evidence of a beneficial effect of apremilast on radiographic progression is extremely important given the importance of early management of PsA to minimise disease progression in joints, emphasised by the NICE guidance, and the evidence for anti-TNFs (which do reduce rates of radiographic progression of joint damage). Without any evidence of apremilast having an effect on radiographic progression, the validity of positioning apremilast before anti-TNFs in a treatment sequence (as proposed by the manufacturer) appears highly questionable.

The manufacturer suggests positioning apremilast before biological therapy in the treatment pathway. However, no data were presented on patients' response to biological therapies after having received apremilast; therefore, it is unclear whether subsequent treatment effectiveness is affected by prior use of apremilast.

There is still uncertainty about the long-term safety of apremilast as the safety data is currently limited to two years. There is no clear evidence to indicate that apremilast is better tolerated than biologics in patients with PsA.