

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/100499/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Tappenden, Paul, Carroll, Christopher, Stevens, John W., Rawdin, Andrew, Grimm, Sabine, Clowes, Mark, Kaltenthaler, Eva, Ingram, John R. , Collier, Fiona and Ghazavi, Mohammad 2017. Adalimumab for treating moderate-to-severe hidradenitis suppurativa: an evidence review group perspective of a NICE single technology appraisal. *PharmacoEconomics* 35 (8) , pp. 805-815. 10.1007/s40273-017-0488-2

Publishers page: <http://dx.doi.org/10.1007/s40273-017-0488-2>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



**Adalimumab for treating moderate-to-severe hidradenitis suppurativa: An Evidence Review  
Group perspective of a NICE Single Technology Appraisal**

**Short title: Adalimumab for treating HS – ERG perspective**

**List of authors**

Paul Tappenden, Reader in Health Economic Modelling, ScHARR, University of Sheffield, UK

Christopher Carroll, Reader in Systematic Review and Evidence Synthesis, ScHARR, University of Sheffield, UK

John W Stevens, Reader in Decision Science, ScHARR, University of Sheffield, UK

Andrew Rawdin, Research Associate, ScHARR, University of Sheffield, UK

Sabine Grimm, Research Associate in Health Economics, University of Sheffield, UK

Mark Clowes, Information Specialist, ScHARR, University of Sheffield, UK

Eva Kaltenthaler, Professor of Health Technology Assessment, ScHARR, University of Sheffield, UK

John R Ingram, Senior Lecturer and Consultant Dermatologist, Institute of Infection and Immunity, Cardiff University, UK

Fiona Collier, Specialist General Practitioner, Stirling Community Hospital, UK

Mohammad Ghazavi, Consultant Dermatologist, Nottingham NHS Treatment Centre, UK

**Corresponding author**

Dr Paul Tappenden, Reader in Health Economic Modelling, ScHARR, University of Sheffield,  
Regent Court, 30 Regent Street, Sheffield, S1 4DA, England.

Tel: +44 114 2220855

Fax: +44 114 2724095

Email: p.tappenden@sheffield.ac.uk

## **ABSTRACT**

As part of its Single Technology Appraisal (STA) process, the National Institute for Health and Care Excellence (NICE) invited the manufacturer of adalimumab (AbbVie) to submit evidence on the clinical effectiveness and cost-effectiveness of adalimumab for the treatment of moderate-to-severe hidradenitis suppurativa (HS). The appraisal assessed adalimumab as monotherapy in adult patients with an inadequate response to conventional systemic HS therapy. The School of Health and Related Research Technology Appraisal Group was commissioned to act as the independent Evidence Review Group (ERG). The ERG produced a critical review of the evidence for the clinical effectiveness and cost-effectiveness of the technology, based on the company's submission to NICE. The evidence was mainly derived from three randomised controlled trials comparing adalimumab with placebo in adults with moderate-to-severe HS. The clinical effectiveness review found that significantly more patients achieved a clinical response in the adalimumab groups than the control groups, but that the treatment effect varied between trials and there was uncertainty regarding its impact on a range of other relevant outcomes, as well as long-term efficacy. The company's submitted state transition model assessed the incremental cost-effectiveness of adalimumab versus standard care for the treatment of HS from the perspective of the National Health Service (NHS) and Personal Social Services (PSS) over a lifetime horizon. The original submitted model, which included a Patient Access Scheme (PAS), suggested that the incremental cost-effectiveness ratio (ICER) for adalimumab versus standard care is expected to be £16,443 per QALY gained. Following a critique of the model, the ERG's preferred base case, which corrected programming errors and structural problems surrounding discontinuation rules, and incorporated a lower unit cost for HS surgery, resulted in a probabilistic ICER of £29,725 per QALY gained. Based on additional analyses undertaken by the company and the ERG following the publication of the Appraisal Consultation Document (ACD), the appraisal committee concluded that the maximum possible ICER for adalimumab compared with supportive care was between £28,500 and £33,200 per QALY gained, but was likely to be lower. The appraisal committee recommended adalimumab (with the PAS) for the treatment of active moderate-to-severe HS in adults whose disease has not responded to conventional systemic therapy.

## **KEY POINTS FOR DECISION-MAKERS**

- Based on evidence provided from 3 RCTs and one OLE study, the Appraisal Committee agreed that adalimumab is efficacious and safe in producing a clinical response in adults with moderate-to-severe hidradenitis suppurativa (HS) with an inadequate response to conventional systemic therapy.
- There was uncertainty regarding the long-term effectiveness and safety of adalimumab, the cost of the surgical comparator and the definitions of “partial response” and “no response” based on the HiSCR measure.
- Based on the ERG’s exploratory analyses, the committee concluded that the maximum possible ICER for adalimumab, compared with supportive care, was between £28,500 and £33,200 per QALY gained, but was likely to be lower.

## **1. INTRODUCTION**

Health technologies must be shown to represent a clinically effective and cost-effective use of resources in order to be recommended for use within the NHS in England. The National Institute for Health and Care Excellence (NICE) is an independent organisation responsible for providing national guidance on promoting good health and preventing and treating ill health in priority areas with a significant impact. The NICE Single Technology Appraisal (STA) process usually covers new technologies within a single indication, shortly after they have received UK marketing authorisation [1]. Within this process, the company provides NICE with a written submission that summarises the company's estimates of the clinical effectiveness and cost-effectiveness of the technology, together with an executable health economic model. The company's submission (CS) is reviewed by an external organisation independent of NICE, the Evidence Review Group (ERG), which consults with clinical specialists and produces an ERG report. After consideration of the CS, the ERG report and testimony from experts and other stakeholders, the NICE Appraisal Committee formulates preliminary guidance in the form of an Appraisal Consultation Document (ACD) which indicates the Committee's initial recommendations on the use of the technology. Stakeholders are subsequently invited to comment on the submitted evidence and the ACD, after which a subsequent ACD may be produced or a Final Appraisal Determination (FAD) is issued, which is open to appeal. An ACD is not produced when the technology is recommended without restriction; in such instances, the FAD is produced directly. This paper presents a summary of the ERG report [2] and subsequent analyses [3, 4] for the STA of adalimumab for the treatment of active moderate-to-severe hidradenitis suppurativa (HS), and the subsequent development of the NICE guidance for the use of this drug in England [5]. Full details of all relevant appraisal documents can be found on the NICE website (<https://www.nice.org.uk/guidance/ta392>).

## **2. THE DECISION PROBLEM**

Hidradenitis suppurativa (HS), also known as acne inversa, is a chronic, inflammatory, recurrent, debilitating, skin follicular disease that usually presents after puberty with painful deep-seated, inflamed lesions. In patients with HS, hair follicles in the apocrine gland-bearing regions (axilla, genital area, groin, infra-mammary region, peri-anal region and buttocks) become blocked and inflamed, resulting in painful recurrent deep-seated boils and nodules. Boils and nodules may progress to abscesses, sinus tracts and scarring. In most patients, disease flares occur at varying intervals, often pre-menstrually in women. Disease flares are characterised by increased pain and suppuration with a foul smelling discharge which stains clothing. HS affects young adults, with disease onset typically between the second and fourth decades of life [6, 7]. Within the adult European population, a prevalence of 1% has been reported [8], although true prevalence is likely to be higher due to problems of under-recognition [9, 10]. Whilst there are no published data on HS prevalence in the UK, it has been suggested that this might be in the region of 1 in 600 [6]. HS has a higher prevalence in women than men and around one-third of patients have a disease in first-degree relatives. The other important known risk factors for HS

are obesity and cigarette smoking [8-10]. Studies have suggested that active disease can substantially impair patients' health-related quality of life (HRQoL), with an impact which exceeds that of other skin diseases such as alopecia, acne, mild to moderate psoriasis, vascular anomalies of the face and atopic dermatitis. Given the debilitating impact of HS, measures of pain and HRQoL, especially the Dermatology Quality of Life Index (DLQI), are recognised as being useful for clinical management of the disease [8, 10].

## **2.1 Current treatment**

There is no current standard of care for HS in England. Treatment is usually determined by the specifics of the disease in the individual patient, together with clinical and patient experience. Treatment usually aims to control the disease and to reduce the number of outbreaks; total cure is generally not expected. Alongside lifestyle changes (smoking cessation and weight loss), therapeutic options include topical antiseptics and antibiotics, systemic antibiotics (e.g. oral tetracyclines, clindamycin and rifampicin), anti-androgens, systemic retinoids, immunomodulatory agents, laser treatment, surgery and tumour necrosis factor-alpha (TNF- $\alpha$ ) inhibitors [11-13]. Treatment choices typically depend on the frequency, severity and spread of lesions and also gender in the case of the retinoid acitretin. A survey of current practice among UK dermatologists confirmed that, after topical treatments, oral antibiotics, such as lymecycline or doxycycline, represent the first-line medical treatment of choice, followed by clindamycin and rifampicin, dapsone, acitretin, ciclosporin, depending on response and gender [13]. In addition, TNF- $\alpha$  inhibitors, such as etanercept, infliximab and adalimumab are already being used for the treatment of moderate-to-severe HS in England. Surgery is usually an option after medical treatments have failed and might involve simple local incision and drainage (usually as a response to acute flares, rather than to control the disease or reduce recurrence); narrow margin excision (which might see recurrence at the edge of the excised area), and; wide margin excision for patients with advanced disease.

In October 2015, NICE issued a final scope to appraise the clinical effectiveness and cost-effectiveness of adalimumab for active moderate-to-severe HS in adult patients with an inadequate response to conventional systemic HS therapy [14].

## **3. INDEPENDENT ERG REVIEW**

The company (AbbVie) provided a submission to NICE on the clinical effectiveness and cost-effectiveness of adalimumab for the treatment of moderate-to-severe HS [15]. This submission was critically appraised by the ERG. In addition, the ERG identified areas requiring clarification, for which the company provided additional evidence prior to completion of the ERG report [16].

### **3.1.1 Clinical evidence submitted by the company**

The clinical evidence consisted of three separate reviews: (1) a review of clinical efficacy evidence from RCTs of treatments for HS, specifically trials comparing adalimumab with placebo in adults with moderate-to-severe HS: a Phase II “dosing” trial, M10-467 [17], and two Phase III trials, PIONEER I and II [18, 19]; (2) a review of evidence from a non-controlled open-label extension (OLE) study (M12-555) [20], and; (3) a review of safety evidence from the RCTs and the OLE study. The relevant efficacy data were derived from Period 1 of the M10-467 trial (up to week 16) and Periods A and B in the PIONEER trials, i.e. weeks 0-12 and weeks 12-36, respectively. The initial periods in all trials compared adalimumab 40mg every week (EW) with placebo. The second period in the PIONEER trials was initiated by re-randomisation of patients at week 12 to adalimumab 40mg EW, placebo or adalimumab 40mg every other week (EOW). The three RCTs and the OLE study were all reported by the company to be at low risk of bias following quality assessment using a range of critical appraisal tools [15].

In M10-467, significantly more patients in the adalimumab 40mg EW group achieved a clinical response (defined as achieving a Hidradenitis Suppurativa Physician’s Global Assessment [HS-PGA] score of clear, minimal or mild with at least a 2 grade improvement relative to baseline at week 16) than patients receiving placebo: 17.6% versus 3.9% ( $p<0.025$ ). Significant improvements were also observed at week 16 in individual symptoms, overall disease severity and pain scores for adalimumab 40mg EW compared with placebo. In the PIONEER trials, clinical response was evaluated by Hidradenitis Suppurativa Clinical Response (HiSCR), defined as  $\geq 50\%$  reduction in the total abscess and inflammatory nodule [AN] count with no increase in abscess count or draining fistula count [17]. In Period A of PIONEER I and II (at week 12), significantly more patients in the adalimumab 40mg EW group achieved clinical response relative to baseline compared with patients receiving placebo (PIONEER I – adalimumab 41.8%, placebo 26.0%,  $p=0.003$ ; PIONEER II – adalimumab 58.9%, placebo 27.6%,  $p<0.001$ ). Significant improvements were observed in symptoms, disease severity (according to the Modified Sartorius Severity [MSS] score) and pain in PIONEER II. However, in PIONEER I, some of the improvements with adalimumab 40mg EW were numerically but not significantly better than placebo. Subgroup analyses indicated that patients achieved benefit with adalimumab 40mg EW regardless of their baseline characteristics, although analyses were subject to small patient numbers. In PIONEER I and II, adalimumab 40mg EW significantly improved HRQoL as measured by the EQ-5D, the DLQI, and the physical components of the Short-Form 36 (SF-36) compared with placebo; improvements were not significant across all components of SF-36. The treatment effect therefore varied between the trials. The CS did not include a pairwise meta-analysis of the PIONEER trials and a network meta-analysis (NMA) was not considered feasible.

Some improvements were maintained into the second period of the PIONEER trials up to 36 weeks (Period B). During Period B, there was also a loss of effect for patients re-randomised to placebo or adalimumab 40mg EOW. The company stated that re-randomisation at week 12 and protocol-driven

discontinuation during Period B for patients with Loss of Response (LOR) or Worsening or Absence of Improvement (WOAI), led to low patient numbers in the group receiving adalimumab 40mg EW for the total study duration, meaning the analyses were underpowered.

These trials were supplemented by one unpublished, non-randomised, non-controlled, unblinded OLE study of the PIONEER trials (M12-555 OLE). The CS included an interim analysis of efficacy from this study, however patient numbers were small. Results for secondary efficacy outcomes were not reported.

The review of safety evidence included the three RCTs and the OLE study. Adalimumab 40mg EW was well-tolerated in all three RCTs. The proportion of patients experiencing serious adverse events (SAEs) or discontinuing treatment due to AEs was low and was similar in both the adalimumab and placebo arms. In an integrated summary of PIONEER I and II (n=633), six patients receiving placebo (1.9%) and three receiving adalimumab 40mg EW (0.9%) gave AEs as their primary reason for discontinuation during Period A. The most common AEs were exacerbation of HS, nasopharyngitis and headache. Rates of infectious AEs were similar for patients receiving adalimumab and for those receiving placebo. M12-555 is the only ongoing study of adalimumab for HS; final data were not available at the time of the appraisal.

### **3.1.2 Critique of clinical effectiveness evidence and interpretation**

The principal efficacy review was a poorly-reported systematic review of 3 relevant RCTs. The PIONEER trials were published only as abstracts, so clinical study reports (CSRs) provided by the company were the principal source of data and were used for quality assessment purposes. The primary outcome was clinical response, measured in the PIONEER trials using the HiSCR measure developed by the company. Clinical advice received by the ERG confirmed that HiSCR had been validated but, in terms of clinical decision-making, its findings should be viewed alongside patient-reported outcome measures (PROMs), in particular the DLQI and a pain measure [17, 21]. The ERG agreed with the company that the M10-467 trial was at low risk of bias for the relevant Period 1 (up to week 16). The ERG also conducted a critical appraisal of the RCTs using the Cochrane Risk of Bias tool [22] and the OLE study using the Critical Appraisal Skills Programme (CASP) cohort study tool [23]. The ERG considered the results from Period A (up to week 12) in PIONEER I and II to be generally at low risk of bias. However, the ERG considered there to be a moderate or unclear risk of selection, attrition and reporting bias affecting the results of Period B in the PIONEER trials, given the absence of any evaluation of blinding, the high levels of attrition, the imputation methods used to manage some of the missing data, and some differences between the outcomes reported in the protocol and those reported in the publications and study reports.



The ERG accepted that the percentage of patients achieving clinical response according to the HiSCR measure on adalimumab 40mg EW at week 12 or week 16 was significantly higher than in the placebo groups ( $p < 0.01$ ), but that the treatment effect varied between the trials. Significant or clinically relevant differences in favour of adalimumab 40mg EW that were reported for secondary outcomes in PIONEER II were not always found for those outcomes in PIONEER I, especially for AN count, MSS score, pain and some components of the SF-36. The reasons for these between-trial differences were unclear.

The company conducted an arm-based integrated summary, which breaks randomisation, for the PIONEER trials to tabulate Period B response (for all patients and for a group of HiSCR “responders” and “partial responders”). This “partial responder” group (defined as HiSCR responders with  $\geq 25\%$  AN reduction rather than  $\geq 50\%$  reduction) was not a pre-specified response category in the PIONEER trials, nor was it explained or justified in the CS, and its clinical validity had not been demonstrated. The ERG considered that findings based on this *post hoc* “partial responder” group were therefore subject to uncertainty. A small number of secondary outcomes were reported for Period B in PIONEER I and II, but only for patients who had had a clinical response at week 12, and the sample sizes in this later period were small.

The ERG considered the efficacy results from the OLE study to be uncertain because they were drawn from interim analyses of unpublished study data. This study also only offered efficacy data for up to 72 weeks for a drug that might be taken for many years by patients with moderate-to-severe HS. There were no obvious safety concerns, with most AEs being balanced between groups, and small numbers of SAEs were reported. The ERG considered that longer-term data were required to determine whether reported AE rates could be maintained for patients on long-term adalimumab maintenance treatment; whether certain subgroups of patients were at a higher risk of certain events, and; to confirm whether there were any differences between interrupted and uninterrupted regimens.

### **3.2 Cost-effectiveness evidence submitted by the company**

The company submitted a *de novo* Markov model to assess the cost-effectiveness of adalimumab versus standard care for moderate-to-severe HS. The company’s model estimates costs and health outcomes from the perspective of the National Health Service (NHS) and Personal Social Services (PSS) over a lifetime horizon (66 years). Health outcomes and costs were discounted at a rate of 3.5% per annum. Costs were valued at 2013/14 prices. All analyses relate to the full licensed population for adalimumab; no subgroup analyses were presented. Following the submission of the original ERG report, a Patient Access Scheme (PAS) in the form of a confidential price discount was agreed for adalimumab specifically in the HS indication.

The company's model includes five mutually exclusive health states, based on depth of HiSCR response: (i) high response; (ii) response; (iii) partial response; (iv) no response, and; (v) dead (see Table 1). Patients are allowed to transit between any of the living health states during each cycle. The model uses a 2-week cycle length for the first 2 cycles, and a 4-week cycle length thereafter. Health state transitions were modelled up to week 36 using pooled data from the PIONEER I/II trials, including a discontinuation rule for patients receiving adalimumab who do not achieve at least a partial response by week 12 (see Table 2). The long-term HiSCR trajectory of adalimumab responders (including partial responders) beyond 36 weeks was subsequently modelled using a time-invariant generalised logit model (GLM) fitted to last observation carried forward (LOCF)-imputed data from the OLE study. The long-term HiSCR trajectories for patients receiving standard care and for those who have previously discontinued adalimumab beyond 36 weeks were modelled using separate time-invariant GLMs fitted to data from weeks 12-36 from the PIONEER I/II trials. The CS stated that the model assumes that patients who lose response after week 36 will continue to receive adalimumab for a further 12 weeks, although this did not accurately reflect the model's implementation. Health utilities were based on depth of HiSCR response using a *post hoc* analysis of EQ-5D data collected within PIONEER II. Resource use was differentiated by depth of HiSCR response based on a survey of UK physicians undertaken by the company and included: inpatient visits due to HS surgery; outpatient visits due to HS surgery; visits to wound care due to HS surgery; non-surgical inpatient visits; non-surgical outpatient visits; visits to wound care not due to HS surgery; Accident and Emergency (A&E) visits, and; costs associated with AEs. Unit costs were taken from the British National Formulary [24], the Personal Social Services Research Unit (PSSRU) [25] and NHS Reference Costs [26]. AEs were assumed to impact only on costs.

The probabilistic version of the company's model (including the PAS) suggests that adalimumab is expected to produce an additional 1.01 QALYs at an additional cost of £16,640 compared with standard care; the incremental cost-effectiveness ratio (ICER) for adalimumab versus standard care is expected to be £16,443 per QALY gained. The deterministic results were similar (ICER=£14,952 per QALY gained). Assuming willingness-to-pay thresholds of £20,000 and £30,000 per QALY gained, the company's model suggests that the probability that adalimumab produces more net benefit than standard care is approximately 0.58 and 0.80, respectively. The ICER for adalimumab was greater than £30,000 per QALY gained in four scenario analyses: (i) time horizon=20 years; (ii) use of PIONEER II data only; (iii) use of last state carried forward imputation, and; (iv) discontinuation rate for adalimumab non-responders after week 36 based on the OLE study.

### **3.2.1 Critique of cost-effectiveness evidence and interpretation**

The ERG critically appraised the company's economic analysis and double-programmed the company's model. The main issues identified by the ERG are discussed below; the full critique can be found in the ERG report and subsequent addenda [2-4].

#### *3.2.1.1 Appropriateness of modelling according to depth of HiSCR response*

The company's model structure divides the HiSCR measure into four response categories. The CS justified this disaggregation based on a *post hoc* analysis which suggested statistically significant differences in EQ-5D between the high response and response states, and between the partial response and non-response states in PIONEER II. The ERG noted the following concerns:

- (i) Disaggregating the full HiSCR measure according to depth of response represents a *post hoc* analysis of a pre-planned endpoint.
- (ii) The HiSCR validation study reported by Kimball *et al* relates specifically to the full HiSCR threshold ( $\geq 50\%$  reduction in ANs, with no increase in abscesses or draining fistulas from baseline) [17]. Kimball *et al* reported that patients with worsening disease or minimal improvement in ANs ( $< 30\%$  reduction) did not have a meaningful improvement on the DLQI and reported some worsening in pain despite improvements in total work impairment and total activity impairment. Kimball also reported no substantial incremental benefits on patient reported outcomes beyond the  $\geq 50\%$  AN reduction threshold.
- (iii) Efficacy data from the PIONEER I/II trials are "stretched" across four rather than two states, hence, several cells in the transition matrices are populated with small patient numbers.

Based on the health state definitions and treatment continuation rules, the company's model implicitly suggests that the 50% AN reduction threshold determined in the Kimball validation study, and later pre-specified as the primary endpoint in the PIONEER trials, has been set at the wrong level for clinical practice.

#### *3.2.1.2 Disconnect between evidence used to inform efficacy and costs*

The company's modelled health gains and the resources required to generate those health gains were derived from different sources: health outcomes were modelled using observed trial data or GLMs fitted to HiSCR outcomes from the PIONEER trials, whilst resource use was based on surgery-related and non-surgery-related secondary care resource estimates from a survey of UK physicians [15]. Higher resource use was assumed for patients achieving a weaker response or no response, hence improvements in modelled HiSCR state are assumed to lead to reductions in costs. The ERG had concerns regarding whether the company's modelled predictions of overall resource use reflect the experience of patients enrolled into the PIONEER I/II trials. Whilst the CS asserted that adalimumab may delay or reduce the need for surgery, and this was reflected in the model, this potential treatment benefit had not been substantiated by evidence. As part of the clarification process, the company presented a *post hoc*

analysis of the PIONEER I/II studies which showed that at week 12, more patients who received adalimumab, compared with placebo, experienced elimination of both draining fistulas (33% vs 19%;  $p < 0.001$ ) and non-draining fistulas (15% vs 9%;  $p = 0.017$ ). These data do not however directly reflect overall reductions in surgery, particularly inpatient surgical admissions, which are a key cost driver in the model. Further, the ERG's advisors noted that whilst the adalimumab could reduce the extent to which limited surgical procedures are required for patients with previously uncontrolled disease, it may in some instances be used as a preadjuvant "bridge" to more definitive surgery, thereby increasing surgery use.

In addition, the costs of pharmacological therapies were not included in the model. Clinical advisors to the ERG were satisfied that the types of resource use included were generally relevant, but noted that some treatments (e.g. wound dressings, where needed) may be given in a primary care setting and that some patients will be prescribed antibiotics by their GPs for several years, yet these costs were not considered. Following clarification, the company provided estimates of concomitant medications used in >5% patients in Period A of the PIONEER I/II trials. These data suggested that concomitant pharmacological therapy use was broadly similar between the adalimumab and placebo groups, however this information relates only to the first 12 weeks of treatment within the RCTs and it remains unclear whether the inclusion of concomitant medication costs would substantially impact upon the cost-effectiveness of adalimumab over a lifetime horizon.

### *3.2.1.3 Treatment continuation rules*

The model assumes that patients require only a partial HiSCR response in order to continue treatment. The ERG's advisors were unclear whether patients achieving a partial HiSCR response (which could include increases in abscesses and/or draining fistulae) would obtain a clinically meaningful benefit sufficient to warrant continuing adalimumab treatment. Commentators on the validity of the HiSCR measure have highlighted the need to capture other aspects of treatment benefit such as pain and improvements on the DLQI [21].

In addition, the company's model includes an assumption whereby patients receiving adalimumab who continue to achieve no response from treatment receive an additional 12 weeks of adalimumab before discontinuing. This was applied in the model by raising the probability of remaining in the adalimumab no response state for one cycle (from the OLE GLM) to the power of 3 and adjusting all other transitions in the row accordingly. This matrix is applied from week 48 onwards. This assumption led to patients discontinuing adalimumab more quickly, thereby substantially reducing the total adalimumab treatment costs. The ERG noted that this approach is mathematically incorrect as the cubed probability reflects the 12-week probability of remaining in the no response state for three 4-week cycles. The proposed discontinuation rule should have been implemented using tunnel states.

#### *3.2.1.4 Potential overestimation of costs of surgery*

The ERG considered that the company's model overestimated the lifetime cost of surgery in both groups, and that cost savings associated with adalimumab due to surgical procedures avoided, may not be realistic. Annual surgical inpatient admission rates according to HiSCR response state were based on the company's physician survey, whilst the unit cost was derived from NHS Reference Costs 2013/14 (major skin procedures, elective inpatient code JC41Z, length of stay [LOS] = 5.1 days) [26]. The company's model predicted that the average patient receiving standard care will require 33.87 inpatient surgical admissions over their remaining lifetime, whilst patients receiving adalimumab would require 29.78 admissions. The ERG noted that the tariff cost of £5,488.32 and its associated LOS was likely to broadly reflect a wide excision procedure. Clinical advisors to the ERG suggested that excluding the management of surgical complications, the maximum number of sites which may require wide excision for a patient with very extensive disease would be 6-10 (including breasts, groin, the perineum, armpits and buttocks). Patients with less extensive disease would require fewer wide excisions than this maximum number and in some cases more than one region can be treated in the same surgical episode. The ERG's clinical advisors also suggested that patients may undergo a comparatively higher number of smaller less costly procedures such as incision and drainage and narrow margin excision.

#### **3.2.1.1 Other issues identified by the ERG**

Several further issues were identified by the ERG, although these had a less significant impact upon the ICER for adalimumab. These included: (i) the use of pooled arm-based summaries of trial data rather than a formal NMA; (ii) minor programming errors; (iii) inconsistent handling of time-dependence in transition probabilities for different time periods, and; (iv) potential bias associated with using the OLE data for adalimumab responders.

#### **3.3 Additional work undertaken by the ERG**

The ERG undertook exploratory analyses to resolve the programming errors identified and to explore alternative assumptions within the company's model. The ERG's preferred base case involved: (a) the correction of minor technical programming errors; (b) applying structural amendments to correctly reflect the company's intended adalimumab non-responder continuation rule during the maintenance phase, and; (c) re-estimation of surgery costs. The ERG's surgery cost estimates assumed that patients on average undergo 2 wide excisions over their lifetime, with the remaining procedures being intermediate day case procedures without admission or elective/non-elective intermediate skin procedures with an LOS of 2 days; this resulted in an estimated cost per procedure of £1,525.74. Further analyses were undertaken to explore uncertainty surrounding transition probabilities, the likely impact

of altering induction phase discontinuation rules and some exploration of uncertainty around the model structure.

The ERG's exploratory analyses indicated that the programming errors did not materially alter the ICER for adalimumab. The incorporation of tunnel states for adalimumab non-responders within the maintenance phase of the corrected model increased the ICER for adalimumab versus standard care (ICER=£19,551 per QALY gained). The ERG's preferred base case, which comprises a scenario whereby these two sets of corrections are combined with the lower surgery cost, resulted in a probabilistic ICER of £29,725 per QALY gained.

### **3.4 Conclusion of the ERG report**

The ERG considered the RCT evidence to be robust for the initial trial periods up to 12 or 16 weeks. However, the treatment effect varied between studies; the reasons for this were unclear. Efficacy results from Period B of the PIONEER trials were at a higher risk of bias across some domains, and were affected by the merging of "responders" with "partial responders." The safety evidence was generally at low risk of bias but was limited, and several questions remain around AE rates for patients on "continuous" or long-term adalimumab 40mg EW. The ERG's exploratory analyses suggested that the probabilistic ICER for adalimumab versus standard care is £29,725 per QALY gained.

## **4. METHODOLOGICAL ISSUES**

The principal areas of uncertainty in the clinical evidence relate to potential treatment effect modifiers and short study follow-up. These uncertainties exist due to observed differences in certain outcomes or levels of outcome between trials, differences in disease severity and other baseline characteristics between trials, and the amount of missing data and imputed results beyond 12 weeks in the PIONEER trials and the OLE study. The ERG also noted issues with respect to whether the achievement of a "partial response" according to the HiSCR measure represents a clinically meaningful treatment benefit sufficient to warrant continuing adalimumab.

The company's model was subject to several methodological issues. In particular, the ERG had concerns that the use of a 5-state model which included three responder categories may have "stretched" the available data too far and that a 3-state model (including response, no response and dead) may have represented a better use of the available evidence. The ERG also noted a selection bias in that patients who discontinued adalimumab after losing a prior response to therapy were assumed to have a different trajectory through the model (indefinitely) compared with patients receiving standard therapy alone. The joint impact of these issues on the ICER for adalimumab was unclear.

## 5. NICE GUIDANCE

The appraisal committee reviewed the data available on the clinical effectiveness and cost-effectiveness of adalimumab, having considered evidence on the nature of HS and the value placed on the benefits of adalimumab by people with the condition, those who represent them, and clinical experts. It also took into account the effective use of NHS resources.

The ACD (published February 2016) states that the committee was minded not to recommend adalimumab for the treatment of active moderate-to-severe HS. The ACD requested additional analyses including: a formal meta-analysis of the PIONEER I/II trials; the committee's preferred assumption about treatment discontinuation for non-responders at 36 weeks or later, a re-analysis of the PIONEER I/II data used in the model in which partial response is defined as a 25% to 50% reduction in AN count and no increase in abscesses and draining fistulas, and an analysis in which extrapolation of outcomes for adalimumab responders was based on the PIONEER I/II trials rather than the OLE study. The ACD also requested additional information from the company relating to resource use estimates derived from the physician survey, utility values within PIONEER II, methods for deriving transition matrices from the OLE study and clarity regarding the company's attempts to validate model predictions against the observed PIONEER I/II data.

Subsequently, the company submitted the requested analyses and additional information. The company's ACD response also included a revised model which incorporated the results of NMAs, the correction of programming errors, the committee's preferred assumption regarding treatment discontinuation in non-responders beyond 36 weeks and some structural changes. The revised model retained the original surgery cost of £5,488.32 per episode.

The ERG was broadly satisfied that the NMA had been undertaken appropriately. However, within the revised model, the NMA-derived transition matrices had been erroneously inverted (transitions to states 1, 2, 3, and 4 were inputted as transitions to states 4, 3, 2, and 1). In addition, the ERG identified a further error whereby the incorrect discontinuation rate was applied during weeks 12-36. Rectifying these errors reduced the ICER to £10,077 per QALY gained. The ERG raised concerns regarding an unwritten assumption whereby different transition matrices were applied to adalimumab discontinuers compared with the standard care group: this led to a situation whereby patients discontinuing adalimumab had a more favourable long-term prognosis compared with those who had never received adalimumab (e.g. a patient who discontinued adalimumab at 36 weeks would still be deriving benefit from therapy 20 years later). The ERG did not consider this to be clinically plausible and noted that removing this assumption increased the ICER for adalimumab. The ERG had further concerns that the company's additional analyses did not include the committee's preferred assumptions regarding surgery. The ERG also noted that the company's analyses which included the new definition of partial

response had been applied only to the transition probabilities, but should also have impacted on health state costs, discontinuation rates and utilities. The ERG undertook further exploratory analyses which included the company's NMA, the corrected discontinuation rate and alternative assumptions regarding the mean lifetime number of wide excisions. Based on the ERG's exploratory analyses of this revised model, the committee concluded that the maximum possible ICER for adalimumab compared with supportive care was between £28,500 and £33,200 per QALY gained [27]. In May 2016, NICE published its FAD which makes the following recommendations:

*“1.1 Adalimumab is recommended, within its marketing authorisation, as an option for treating active moderate to severe hidradenitis suppurativa in adults whose disease has not responded to conventional systemic therapy. The drug is recommended only if the company provides it at the price agreed in the PAS.*

*1.2 Assess the response to adalimumab after 12 weeks of treatment, and only continue if there is clear evidence of response, defined as:*

- *a reduction of 25% or more in the total AN count and*
- *no increase in abscesses and draining fistulas.”*

## **5.1 Consideration of clinical and cost-effectiveness issues**

This section discusses the key issues considered by the appraisal committee. The full list can be found in the FAD [27].

### **5.1.1 Appropriate HiSCR threshold for determining treatment response and continuation**

The committee considered how clinicians assess disease severity and response to treatment in people with HS. The clinical experts considered that the HiSCR is a reliable and reproducible tool, which has been validated and is relevant to clinical practice, but noted that the minimum clinically important difference (MCID) has not been established. Clinical experts were aware that according to the HiSCR validation study, response was defined as a 50% reduction in total AN count, with no increase in abscesses or draining fistulas from baseline. However, clinical experts considered that the 50% threshold was too high, and stated that a 25% reduction in AN count, provided there was no increase in abscesses or draining fistulas from baseline, would reflect a treatment response. Clinical experts suggested that if the reduction in AN count was between 25% and 50%, they would continue with the existing treatment but may prescribe additional concomitant treatments (e.g. anti-inflammatories, retinoids and antibiotics) to improve response. The committee heard from experts that they would stop treatment if the reduction in AN count was lower than 25%, or if there was an increase in abscesses or draining fistulas. The clinical experts stated that it was important to also use PROMs when monitoring people with HS (in particular, the DLQI, the pain visual analogue scale [VAS] and the SF-36, even



though they are not specific to this indication), because physician-reported and patient-reported scores do not always correlate. The committee concluded that it is appropriate to use HiSCR for assessing response to treatment, with supporting information provided by PROMs.

### **5.1.2 Clinical effectiveness of adalimumab for HS**

The committee discussed the clinical evidence for adalimumab and noted that people treated with adalimumab were more likely to have a clinical response than people treated with placebo and that this difference was significant. The committee was aware that the benefit with adalimumab was greater in PIONEER II than PIONEER I, possibly because PIONEER II patients appeared to have had less severe disease than those in PIONEER I, and had potentially received higher levels of systemic antibiotics. The company noted that only 19% of patients in PIONEER II took oral antibiotics during the trial. The committee noted that the company had not originally undertaken a formal meta-analysis and was concerned that they had given contradictory views on whether the PIONEER trials had similar or heterogeneous baseline characteristics, but concluded that the trials were generalisable to UK clinical practice. The committee was concerned that the OLE study only had data up to 72 weeks, given that adalimumab may be used for many years, and that full data were only available for 26% of enrolled patients. The committee concluded that adalimumab provided significant benefits compared with placebo, but these had not been shown over the long-term. The committee was also aware that adalimumab showed a beneficial effect on the SF-36 (collected in PIONEER I) and the DLQI (collected in PIONEER I and II) but noted that the difference between adalimumab and placebo was not significant for all components of the SF-36, and that the difference between arms in DLQI improvement at week 12 was not greater than the MCID. The committee considered that the DLQI may have underestimated the beneficial effects of adalimumab, based on the clinical experts' comments that people with chronic skin conditions can develop coping mechanisms, which may result in lower DLQI scores than would be expected. The committee concluded that adalimumab had a statistically significant and clinically meaningful positive effect on HRQoL.

### **5.1.3 Uncertainty surrounding the cost-effectiveness of adalimumab for HS**

The committee attempted to identify the most plausible ICER for adalimumab compared with supportive care. The committee considered that the resource use assumptions in the ERG's new exploratory analyses, provided after consultation, were more realistic than the assumptions in the company's revised model. The committee also preferred the ERG's assumption that there is no lifelong difference in prognosis between people who previously had adalimumab and then stopped treatment, and those who had never had the drug. It agreed with the ERG's corrected discontinuation rate for weeks 12-36. Based on the ERG's exploratory analyses, the committee concluded that the maximum possible probabilistic ICER for adalimumab compared with supportive care was between £28,500 and

£33,200 per QALY gained. However, the committee considered that the most plausible ICER would be lower than this for several reasons. First, the ERG's assumption of a maximum of 4 wide excisions over a patient's lifetime may be an underestimate, and the committee understood that the ICER reduced as the number of wide excisions increased. Second, the committee acknowledged that adalimumab may be associated with short-term improvements in psychological wellbeing after treatment is stopped, and so considered that the ERG's assumption about prognosis was possibly pessimistic and may have overestimated the ICER. The committee also considered that if its preferred definitions of partial response and non-response had been incorporated in the ERG's exploratory analyses, the ICER would have been reduced because continued treatment in people for whom a drug is not effective would be minimised.

## **6. APPRAISAL COMMITTEE'S KEY CONCLUSION**

The committee concluded that adalimumab provided significant benefits compared with placebo, but that these had not been shown over the long-term. The committee also concluded that the maximum possible ICER for adalimumab compared with supportive care was between £28,500 and £33,200 per QALY gained. The committee was certain that the most plausible ICER would be lower than this for several reasons, including the possibility that the ERG had underestimated the number of wide excisions and was pessimistic in its assumption about prognosis. The committee also considered that if its preferred definitions of partial response and non-response had been incorporated in the ERG's exploratory analysis, as well as the gains in HRQoL not already included in the QALY calculations, the ICER would have reduced further.

## **ACKNOWLEDGEMENTS**

This project was funded by the National Institute for Health Research (NIHR) Health Technology Assessment Programme (project no. 15/06/09). See the HTA programme website for further project information (<http://www.hta.ac.uk>). This summary of the ERG report was compiled after NICE issued the FAD. All authors have commented on the submitted manuscript and have given their approval for the final version to be published. The views expressed in this report are those of the authors and not necessarily those of the NIHR HTA Programme. Any errors are the responsibility of the authors.

## REFERENCES

1. National Institute for Health and Care Excellence. Guide to the processes of technology appraisal. London. 2014.
2. Tappenden P, Carroll C, Stevens JW, Rawdin A, Grimm S, Clowes M. Adalimumab for treating moderate to severe hidradenitis suppurativa: A Single Technology Appraisal. Sheffield. 2015.
3. Tappenden P, Carroll C, Stevens JW, Rawdin A, Grimm S, Clowes M. Adalimumab for treating moderate to severe hidradenitis suppurativa: A Single Technology Appraisal. Addendum. Sheffield. 2016.
4. Tappenden P, Carroll C, Stevens JW, Rawdin A, Grimm S, Clowes M. Adalimumab for treating moderate to severe hidradenitis suppurativa: A Single Technology Appraisal. Addendum: ERG critique of company's ACD response. Sheffield. 2016.
5. National Institute for Health and Care Excellence. Technology Appraisal Guidance: Adalimumab for treating moderate to severe hidradenitis suppurativa. London. 2016.
6. Von Der Werth JM, Jemec GBE. Morbidity in patients with hidradenitis suppurativa. *British Journal of Dermatology* 2001;144:809–13.
7. von der Werth JM, Williams HC. The natural history of hidradenitis suppurativa. *Journal of the European Academy of Dermatology and Venereology*. 2000;14:389-92.
8. Revuz J. Hidradenitis suppurativa. *Journal of the European Academy of Dermatology and Venereology* 2009;23:985-98.
9. Jemec G. Clinical practice. Hidradenitis suppurativa. *New England Journal of Medicine* 2012;366:158-64.
10. Dufour D, Emtestam L, Jemec G. Hidradenitis suppurativa: a common and burdensome, yet under-recognised, inflammatory skin disease. *Postgraduate Medical Journal* 2014;90:216-21.
11. Zouboulis C, Desai N, Emtestam L. European S1 guideline for the treatment of hidradenitis suppurativa/acne inversa. *Journal of the European Academy of Dermatology and Venereology* 2015;29:619-44.
12. Blok J, van Hattem S, Jonkman M, Horváth B. Systemic therapy with immunosuppressive agents and retinoids in hidradenitis suppurativa: a systematic review. *British Journal of Dermatology* 2013;168:243-52.
13. Ingram J, McPhee M. Management of hidradenitis suppurativa: a UK survey of current practice. *British Journal of Dermatology* 2015;173:1070-2.
14. National Institute for Health and Care Excellence. Adalimumab for treating moderate to severe hidradenitis suppurativa - Final scope. London. 2015.
15. AbbVie Ltd. Adalimumab for treating moderate to severe hidradenitis suppurativa. ID812. Company evidence submission to the National Institute for Health and Care Excellence. . Berkshire. 2015.
16. AbbVie Ltd. Adalimumab for treating moderate to severe hidradenitis suppurativa. ID812. Response to ERG's clarification questions. Berkshire. 2015.
17. Kimball A, Jemec GB, Yang M, Kageleiry A, Signorovitch JE, Okun MM, et al. Assessing the validity, responsiveness and meaningfulness of the Hidradenitis Suppurativa Clinical Response (HiSCR) as the clinical endpoint for hidradenitis suppurativa treatment. *British Journal of Dermatology* 2014;171:1434-42.

18. AbbVie Ltd. A Phase 3 multicenter study of the safety and efficacy of adalimumab in subjects with moderate to severe hidradenitis suppurativa – PIONEER I. Clinical Study Report (Interim), Adalimumab/Protocol. 2015.
19. AbbVie Ltd. A Phase 3 multicenter study of the safety and efficacy of adalimumab in subjects with moderate to severe hidradenitis suppurativa – PIONEER II (Clinical Study Report (Interim), Adalimumab/Protocol M11-810). 2015.
20. AbbVie Ltd. Phase 3 open-label study of the safety and efficacy of adalimumab in subjects with moderate to severe hidradenitis suppurativa – PIONEER (open-label extension) Clinical Study Report (Interim), Adalimumab/Protocol M12-555. 2015.
21. Ingram JR. Hidradenitis suppurativa outcome measures and treatment goals. *British Journal of Dermatology* 2014:1293-4.
22. Higgins J, Green S. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.1.0 [updated March 2011]. 2011.
23. *Critical Appraisal Skills Programme (CASP). Cohort Study Checklist*. 2013.
24. BMJ Group, RCPCH Publications Ltd and The Royal Pharmaceutical Society of Great Britain. *British National Formulary (BNF)*. 2015.
25. Curtis L. *Costs of health and social care 2014*. University of Kent, UK. 2014.
26. Department of Health. *NHS Reference Costs 2013/14*. 2014.
27. National Institute for Health and Care Excellence. *Final appraisal determination: Adalimumab for treating moderate to severe hidradenitis suppurativa*. London. 2016.

**Table 1: HiSCR response categories**

<b>HiSCR-based state definition</b>	<b>HiSCR-based state description</b>
High response	At least 75% total AN count reduction, with no increase in abscesses or draining fistulas from baseline
Response	At least 50% but less than 75% AN reduction, with no increase in abscesses or draining fistulas from baseline
Partial response	At least 25% but less than 50% AN reduction, with no increase in abscesses or draining fistulas from baseline; or at least 25% AN reductions, with an increase in abscesses and/or draining fistulas
No response	Defined as less than 25% AN reduction

*HiSCR - Hidradenitis Suppurativa Clinical Response; AN - abscess and inflammatory nodule*

**Table 2: Evidence used to inform the model transition matrices**

<b>Matrix description</b>	<b>Source</b>
<i>Standard care – induction phase</i>	
Week 0-12	Cross-tabs of outcomes based on pooling of patients initially randomised to the placebo groups within PIONEER I/II
<i>Standard care – maintenance phase</i>	
Week 12-36	Cross-tabs of outcomes for patients initially randomised to the placebo group in PIONEER II who subsequently continued on placebo during maintenance.
Week 36+	GLM based on 12-36 week data described above
<i>Adalimumab – induction phase</i>	
Week 0-12	Cross-tabs of outcomes based on pooling of patients initially randomised to adalimumab 40mg EW groups within PIONEER I/II.
<i>Maintenance phase – adalimumab 12-week responders</i>	
Week 12-36	Cross-tabs of adalimumab 40mg EW patients re-randomised to adalimumab 40mg EW after responding at 12-weeks in PIONEER I/II.
Week 36-48	GLM based on weeks 0-24 of M12-555 OLE study (including LOCF imputation as <50% patients had 24-weeks follow-up data).
Week 48+	Same as above except the probability of transiting from adalimumab no response state to standard care no response state is cubed.
<i>Maintenance phase – adalimumab 12-week non-responders and subsequent discontinuers</i>	
Week 12-36	Cross-tabs of patients randomised to adalimumab 40mg EW in PIONEER I/II who switched to placebo in the maintenance period (irrespective of whether they achieved an induction response on adalimumab).
Week 36+	GLM based on week 12-36 data described above

*GLM – generalised linear model; OLE – open-label extension; LOCF – last observation carried forward; mg – milligram; EW – every week*