CONFIDENTIAL UNTIL PUBLISHED

Evidence Review Group Report commissioned by the NIHR HTA Programme on behalf of NICE

Afamelanotide for treating erythropoietic protoporphyria

Produced by Southampton Health Technology Assessments Centre (SHTAC)

Authors Joanne Lord, Professorial Fellow in Health Economics
Olu Onyimadu, Research Fellow (Health Economics)
Wendy Gaisford, Research Fellow
Micah Rose, Research Fellow (Health Economics)
Geoff Frampton, Senior Research Fellow
Jonathan Shepherd, Principal Research Fellow

Correspondence to Dr Jonathan Shepherd
Southampton Health Technology Assessments Centre (SHTAC)
Wessex Institute
Alpha House
Enterprise Road, University of Southampton Science Park
Southampton SO16 7NS
www.southampton.ac.uk/shtac

Date completed 26th October 2017

Source of funding: This report was commissioned by the NIHR HTA Programme as project number 16/27/02.

Declared competing interests of the authors

The authors declare no competing interests. Dr Sarkany declares no competing interests. Dr Badminton has attended a European EPP expert meeting in June 2016 with funding for travel and accommodation provided by CLINUVEL.

Acknowledgements

We would like to thank the following for providing expert clinical advice in the preparation of this report:

Dr Robert Sarkany, Consultant Dermatologist, Guys & St Thomas' Hospital, London.

Dr Mike Badminton, Honorary Consultant and Clinical Lead, National Acute Porphyria Service (Cardiff), Medical Biochemistry & Immunology, University Hospital of Wales, Cardiff.

We also thank Dr Jo Picot, Senior Research Fellow at SHTAC, for acting as an internal editor for the ERG report, and Karen Welch, Information Specialist, SHTAC, for appraising the literature search strategies in the company submission.

Rider on responsibility for report

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.

This report should be referenced as follows:

Lord, J., Onyimadu, O., Gaisford, W., Rose, M., Frampton, G., Shepherd, J. Afamelanotide for treating erythropoietic protoporphyria: A Highly Specialised Technology Appraisal. Southampton Health Technology Assessments Centre (SHTAC), 2017.

Contributions of authors

Joanne Lord critically appraised the health economic review, critically appraised the economic evaluation, and drafted the report; Olu Onyimadu critically appraised the health economic review, critically appraised the economic evaluation, and drafted the report; Wendy Gaisford critically appraised the clinical effectiveness review and drafted the report; Micah Rose critically appraised the health economic review, critically appraised the economic evaluation, and drafted

the report; Geoff Frampton critically appraised the clinical effectiveness review and drafted the report; Jonathan Shepherd critically appraised the clinical effectiveness review, drafted the report, project managed the assessment and is the project guarantor.

Word count:

44,029

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AE	Adverse effect
BIM	Budget impact model
BAD	British Association of Dermatologists
BPA	British Porphyria Association
CBC	Complete blood count
CHMP	Committee for Medicinal Products for Human Use
CI	Confidence interval
CS	Company submission
CSR	Clinical study report
DALY	Disability Adjusted Life Year
DLQI	Dermatology Quality Life Quality Index
EAR	European Assessment Report
EPAR	European Public Assessment Report
EPP	Erythropoietic protoporphyria
EPP-QoL	Erythropoietic protoporphyria Quality of life questionnaire
EMA	European Medicines Agency
EQ-5D	EuroQoL 5 Dimensions questionnaire
ERG	Evidence review group
FECH	Ferrochelatase
GBD	Global Burden of Disease
GBP	Great Britain pounds
GCP	Good clinical practice
HRQoL	Health related Quality of life
HST	Highly specialised technology
HTA	Health technology assessment
ICER	Incremental cost effectiveness ratio
IMP	Implant
ITT	Intention-to-treat
MD	Melanin density
MSH	Melanocyte stimulating hormone
NHS	National Health Service

NICE	National Institute for Health and Care Excellence
PASS	Post authorisation Safety Study
PPIX	Protoporphyrin IX
PSA	Probabilistic sensitivity analysis
QALY	Quality adjusted life year
RCP	Royal College of Pathologists
RCT	Randomised controlled trial
SD	Standard deviation
SF-36	Short Form survey-36
SmPC	Summary of Product Characteristics
UK	United Kingdom
USA	United States of America
UVB	Ultraviolet B

SUMMARY

Scope of the company submission

The company submission (CS) presents evidence of the clinical and cost effectiveness of afamelanotide (SCENESSE®) for adult patients with erythropoietic protoporphyria (EPP) above the age of 18 years old compared to best supportive care. In all studies afamelanotide (16 mg) was given as a subcutaneous implant. The main outcomes measured were duration of tolerance to sunlight and other forms of visible light, phototoxic reactions, health related quality of life (HRQoL) and adverse effects (AEs) of treatment.

Summary of submitted clinical effectiveness evidence

The CS presents evidence for the clinical effectiveness of afamelanotide based on a small open label phase II study (CUV010; five patients); four phase III RCTs (CUV017; CUV029: CUV030 and CUV039) comparing afamelanotide to placebo, and two observational studies on the safety and efficacy of long-term afamelanotide (one a retrospective assessment of up to eight years of the treatment of Italian and Swiss patients, the other an on-going post authorisation safety study).

All but one of the studies were sponsored by the company. Study CUV017 was based in eight EPP expert centres within Australia and Europe and included 100 patients (including three patients from the UK). Study CUV029 was based in eight EPP expert centres within Europe (including the UK) and included 74 patients (16 from the UK); study CUV030 was based in six EPP expert centres within the USA and included 77 patients and study CUV039 was based in seven EPP expert centres within in the USA and included 94 patients. Study CUV039 was the study that the European Medicines Agency (EMA) considered methodologically adequate enough to based it's licensing approval on.

Due to the lack of detail provided, the ERG is unable to make a fully informed judgement on the methodological quality of the RCTs. The methods used to generate random allocation sequences of patients to study groups were sufficient. However, it was not possible to determine from the information given whether study groups were comparable at baseline; or whether concealment of allocation was adequate; or whether there was selective reporting of outcome measures. Furthermore, although trials were double-blinded the increased skin pigmentation in participants who received afamelanotide was acknowledged to reveal treatment

allocation in some patients. The impact of this on patients' sun exposure behaviour and hence the effectiveness of afamelanotide is uncertain. The company's statistical analyses appear generally appropriate but information is lacking on how sample sizes and statistical power were estimated and on how missing data were handled. The level of patient drop-out, where reported, was low.

The company's evidence review included a narrative synthesis of the results of the studies, but no meta-analysis. The ERG considers meta-analysis would not be meaningful due to heterogeneity between the studies. Results from study CUV029 revealed a significant difference in the number of hours over the nine month study period in direct sunlight (measured between 10.00 - 15.00 hours) with no pain between patients receiving afamelanotide (median number of hours per patient, 6.0 (range 0-193)) compared to the placebo group (median number of hours per patient 0.8 (range 0-35)) p = 0.005 (primary outcome). In study CUV039 there was a significant difference in number of hours over the six month study period per patient in direct sunlight (measured between 10.00 - 20.00 hours) with no pain between study groups (afamelanotide median no of hours per patient 69.4 (range 0-651) vs placebo median number of hours per patient 40.8 (range 0-224)) p = 0.044 (primary outcome).

There was a higher number of phototoxic reactions observed in patients receiving the placebo in studies CUV029 and CUV039 though the difference between study groups was only statistically significant in study CUV029. In the phase II study (CUV010) there was a change in melanin density during the first 30 days after administration of afamelanotide, with a mean melanin density change of 124% above baseline and a small increase of 6% to 130% above baseline, following the second implantation at 90 days. The long-term retrospective observational study of Swiss and Italian patients reported an increase in melanin density that was maintained over the six year treatment assessment period.

Adverse events were mild to moderate in severity and the most common events reported in the studies included headache, nausea, gastrointestinal discomfort and migraine. Mortality was not reported in the CS; however, publications indicated that four deaths occurred during these trials (which had approximately 340 patients in total). The deaths were regarded by the investigators as definitely not related to the study treatment.

The impact of treatment on HRQoL was measured using the disease specific EPP-QoL instrument devised by the company (scores measured from 0-100, with higher scores indicating better HRQoL). [redacted], these data were used to inform the company's assessment of cost-effectiveness (see below). Quantitative results are available for studies CUV029 and CUV039. In CUV029 the scores increased over time in both study groups, although the increase was higher in the afamelanotide group at all assessment time points, with the highest score around 85 points. The differences between the groups were statistically significant at days 120, 180, and at day 240. In study CUV039 scores increased over time from baseline in both groups with larger increases in the afamelanotide group. The highest score was 77.7 points for the afamelanotide group at day 180 (scoring range 0-100, higher scores mean better HRQoL). Differences between the groups in the change from baseline were statistically significant at day 60, day 120, and day 180. By day 360 (240 days after the last implant) scores had fallen in both study groups illustrating a reduction in HRQoL, though they remained above baseline levels. The retrospective observational study of Swiss and Italian patients showed an increase in HRQoL after afamelanotide administration which was maintained up to six years of treatment observation, though HRQoL was shown to be higher in winter months than summer during this period indicating seasonal variation. The clinical significance of the changes in EPP-QoL results was unclear as minimal important differences have not been established.

HRQoL was also measured using the Dermatology Life Quality Index (DLQI) in studies CUV029, CUV030, and CUV039. Results available for study CUV039 showed that scores declined over time (thus showing an improvement in HRQoL) for both afamelanotide and placebo: 2.4 (± 4.2) and 3.1(± 4.1) respectively at day 180 compared to 10.7 (± 6.3) vs 10.4 (± 5.7) at baseline (N.B. a score of between 2 to 5 indicates a small effect on a patient's life). The decline in scores was larger in the afamelanotide group, though differences between the groups in the change from baseline were not statistically significant.

Summary of submitted cost effectiveness evidence

An evidence review was conducted by the company to identify economic evaluations of afamelanotide in adult patients with EPP. They reported that no relevant economic evaluations were identified. The ERG's search, however, identified a 2016 conference abstract reporting a relevant cost effectiveness analysis of afamelanotide for EPP. The ERG noted that the model

which was used for this study appeared to be similar to that of the model submitted by the company to NICE, and it included an exploratory sensitivity analysis using QALYs derived from SF-36 data from early clinical trials and for other 'similar' conditions. The estimated Incremental cost effectiveness ratios (ICERs) ranged from £208,000 to £1.1 million per Quality Adjusted Life Year (QALY).

The company's submitted cost effectiveness evaluation comprised a model to estimate the cost-effectiveness of treatment with afamelanotide compared with a standard treatment control for adult patients with EPP. This addressed the decision problem specified in the scope, with the exception of the measure of value for money: the model estimates incremental cost per DALY avoided, rather than the incremental cost per QALY gained expected by NICE. The company's rationale for this approach (which the ERG disagrees with – see below) is due to the lack of available robust utility data, and their view that a cost per DALY framework is more appropriate for this condition.

[redacted]

[redacted] The ERG has not identified any evidence to

contradict this. Non-compliance or discontinuation of treatment is not explicitly modelled. The model assumes that treatment continues throughout the modelled time horizon, with the same mean number of implants per patient and the same effectiveness estimates every year over the [redacted] year time horizon. The model does not include any additional disability, mortality risk or healthcare cost to reflect the impact of adverse reactions to afamelanotide. This is reasonable given the generally low incidence and mild severity of adverse events observed in the clinical effectiveness studies.

The company used individual EPP-QOL data from studies CUV029, CUV030 and CUV39 to estimate the proportions of patients in the intervention and control groups with mild, moderate and severe disease at baseline and at 120 days (assuming that the 120 day values apply for the whole year). The base case analysis uses disability weights from the World Health Organisation Global Burden of Disease (GBD) study conducted in 2010. The survey did not include EPP, or the company's preferred proxy of [redacted]. Instead, the company used a proxy of [redacted] in their base case analysis, and an alternative proxy of [redacted] in a scenario analysis. The ERG questions the relevance of these proxy conditions for EPP.

The cost per implant is reported as £12,020. This equates to [redacted] per year assuming a mean number of implants of [redacted] per year. The company estimates the administration cost of afamelanotide at [redacted] per patient per year.

The company's base case cost per DALY averted was £278,471 (see table).

Base case cost effectiveness results

	Discounted costs	Discounted DALYs
Afamelanotide	[redacted]	[redacted]
Standard care	[redacted]	[redacted]
Incremental	[redacted]	[redacted]
ICER	**£278,471 per DALY averted**	

The company conducted deterministic sensitivity analyses to explore variations in estimates of disability weights, starting age and time horizon, number of implants per year, and societal costs. The ICERs varied between £97,624 and £727,143 in these sensitivity analyses. No

probabilistic sensitivity analysis is reported. This represents a very limited exploration of uncertainty. In particular, the CS does not present any sensitivity analysis over the parameters that reflect treatment effectiveness in the model or the methods and assumptions used to derive them.

Commentary on the robustness of submitted evidence

Strengths

- The clinical effectiveness evidence base comprises four multi-centre double-blind RCTs including approximately 340 patients in total, plus a long-term retrospective observational study of 115 patients providing data on safety and efficacy up to eight years of afamelanotide use. Two of the RCTs included a small number patients from UK expert porphyria treatment centres (amongst other countries). The ERG believes that all relevant clinical effectiveness studies have been included in the CS.
- The clinical effectiveness studies measured a range of outcome measures of relevance to patients and clinicians, including: time patients are able to spend in sunlight without experiencing pain or with only mild pain; phototoxic reactions; adverse events and HRQoL (though not HRQoL of carers and family members). There do not appear to be any clinically important outcome measures that have not been included in the study programme.
- Recorded adverse events were mild to moderate in severity and the level of patient drop-out from treatment (where data are reported) was low (less than 10%).
- The company's economic model, though simplistic, is appropriate for the condition, and some, though not all, of the assumptions are reasonable.

Weaknesses and areas of uncertainty

- Full methodological details of the included clinical effectiveness studies are lacking and this prevents a full assessment of quality by the ERG. In particular, it isn't clear whether randomised study groups were comparable at baseline in all studies, or whether concealment of random allocation to study groups was adequate, indicating the potential for selection bias. It is also unclear whether there is selective reporting of outcome measures, as for most studies, protocols and clinical study reports were not supplied to the ERG (though requested). The influence on the study results of apparent unblinding

due to increased skin pigmentation in some patients who received afamelanotide is not entirely clear.

- Information is lacking on how sample sizes were estimated and on how missing data were handled in the trials.
- Meta-analysis of the studies was not conducted in the CS (though pooling of EPP-QoL results was done to inform the economic model – see below), rather, a narrative summary of the individual studies was presented. The ERG considers that meta-analysis would not be advisable given clinical and methodological heterogeneity between the studies.
- Due to concerns by the EMA about the methodological conduct of two of the RCTs (studies CUV030 and CUV029), the sole pivotal RCT to inform the decision to grant a marketing application was the CUV039 trial. The CUV039 trial was conducted in seven expert centres in the USA and therefore it does not include patients taking afamelanotide in the UK. There are differences in latitude and hence potential exposure to sunlight over the course of a year between the USA and Europe which is likely, amongst other things, to influence the amount of time patients can spend outdoors during the day (the European centres were at higher latitudes). The mean and median time that patients in the CUV039 trial were able to spend in sunlight with no or mild pain cannot therefore necessarily be generalised to England and the UK as a whole.
- Although an improvement in HRQoL was reported in the studies, the interpretation of the clinical significance of this is unclear. The EPP-QoL instrument was devised specifically for the afamelanotide study programme [redacted]. HRQoL, as assessed by EPP-QoL with results pooled for studies CUV029, CUV030, and CUV039, is the clinical outcome effectiveness measure that informs the company's cost-effectiveness analysis.
- The ERG has insufficient information about how the EPP-QoL results from the three trials, CUV029, CUV030 and CUV039 were analysed and pooled for use in economic evaluation. There is a lack of clarity over whether intention to treat (ITT) datasets were used, the number of patients included from each trial and whether the method of pooling accounted for clustering.
- The company's economic model relies on a definition of mild, moderate and severe EPP by division of the EPP-QOL scale into thirds. This is arbitrary and we cannot assess if it is consistent with the disability weights attached to these levels of severity in the DALY calculations.

- The company's use of a single time point (120 days) to represent disease severity over a whole year is simplistic and is likely to have biased DALY estimates in favour of afamelanotide. It does not account for baseline imbalance in trial arms in EPP-QoL estimates (which are amplified when extrapolated over time). In addition, we note that data at 180 days were collected in the three included trials, but not used for the economic evaluation (the largest between-arm difference in mean EPP-QOL was observed at 120 days in CUV039 and CUV029).
- The ERG notes that the analysis of uncertainty presented in the CS was inadequate. No probabilistic sensitivity analysis was reported and there was no attempt to estimate the extent or consequences of uncertainty over the effectiveness parameters and assumptions.
- Contrary to the company, the ERG believes that QALYs are a conceptually appropriate metric for quantifying the value of health effects of afamelanotide for patients with EPP, as they are for other lifelong and chronic disabling conditions and that satisfactory methods for estimating QALY gain are available. It is considered that these methods, although not perfect, are superior to the methods used by the company to estimate DALYs averted. A QALY based analysis is presented by the ERG (see Summary of additional work undertaken by the ERG below).

Summary of additional work undertaken by the ERG

The ERG made adjustments to the company's model to estimate cost-utility, generating costs per QALY. Two alternative analyses have been conducted:

- A simple QALY version of the company model by assuming utility values for mild, moderate and severe disease equal to 1 minus the disability weights used in the company's basecase proxy of [redacted]
- An ERG base case analysis, in which we estimate QALYs from mean DLQI results at 0, 60, 120 and 180 days from study CUV039 mapped to EQ-5D scores.

The simple QALY model was intended as a platform to investigate alternative scenarios and sensitivity around the company's base case. This demonstrated that the company's incremental cost per DALY averted of £278,471 (£278,386 per QALY gained after a small correction by the ERG) is likely to be an underestimate. With correction for baseline differences in EPP-QOL, the ICER rose to £454,800 per QALY gained. It rose further, to £779,657 per QALY gained, when

we assumed that treatment benefits would gradually decline over a 2 month period from month 6. Use of utility estimates from the literature for the same proxy condition as in the company base case, further increased the estimated ICER to over £1.7 million per QALY gained.

We conducted a 'best case' analysis, which combined the most favourable scenario that we had tested (our simple QALY conversion of the company's base case model), with the most favourable sensitivity analysis limits for treatment effects, disability weights and mean number of implants used for costing. This brought the ICER down to £151,212 per QALY gained. The ERG does not believe that this or any of the other ICER estimates based on our simple adaptation of the company model are plausible.

Our preferred set of analyses were based on mean DLQI data from the pivotal study (CUV039) mapped to EQ-5D utility values using a published algorithm. Results from this model were less favourable, and did not fall below £1.1 million per QALY gained in any of the scenarios that we tested. The ERG believes that this set of estimates is more plausible than the company's approach.

Budget impact in the first year varied between [redacted] and [redacted] depending on variations in the estimate of EPP prevalence in England.

1 Introduction to ERG Report

This report is a critique of the company's submission (CS) to NICE from CLINUVEL UK on the clinical effectiveness and cost effectiveness of afamelanotide for erythropoietic protoporphyria (EPP). It identifies the strengths and weakness of the CS. Clinical experts were consulted to advise the ERG and to help inform this review.

Clarification on some aspects of the CS was requested from the manufacturer by the ERG via NICE on 1st September 2017 (early clarification questions) and on 12th September. Sets of responses from the company via NICE were received by the ERG on 12th September, 26th September and 2nd October 2017, and these can be seen in the NICE HST committee papers for this appraisal.

2 BACKGROUND

2.1 Critique of company's description of underlying health problem

The ERG considers that the CS provides a clear and accurate overview of the genetic disorder, erythropoietic protoporphyria (EPP, CS, pp15-17). The disease is caused by impaired function of the enzyme ferrochelatase (FECH) which disrupts the haem biosynthesis pathway, resulting in the accumulation and storage of protoporphyrin IX (PPIX), predominantly in patients' skin and liver. PPIX is a phototoxic molecule, which reacts after brief exposure to visible light (the most reactive wavelength being at 408 nm; CS, p 9). Upon exposure to light, PPIX in the capillaries underneath the skin reacts to create oxygen radicals which attack capillary walls, causing onset of erythema, oedema and an intense burning sensation which can last for days or weeks.[1] This can also lead to second degree burns (CS, p 9). During a reaction, any subsequent exposure to light, as well as heat variation, pressure and air movement, can exacerbate and prolong symptoms. Cumulative exposure to light has a 'priming' effect and after only a few minutes of daily light exposure severe phototoxicity may be triggered (CS, p 15).

EPP is described as a disease that requires lifelong and cyclical management. Phototoxicity is most predominant in the UK, from February to November each year, during the period of highest light intensity (CS, p 66). Phototoxic reactions are unresponsive to regular analgesics or

any other medication and require the recovery of damaged tissue (i.e. time) prior to their subsidence.

The CS states that “both environmental and artificial light sources (particularly modern ‘energy saving’ globes) can cause anaphylactoid and phototoxic reactions” (CS, p 9). Clinical experts advising the ERG commented that only a minority of patients experience phototoxic reactions resulting from exposure to artificial light sources. The clinical experts also highlighted that there is a variation in severity of disease amongst patients, where some are able to cope with light exposure for longer periods (e.g. up to an hour) before suffering any reaction. On average, however, the majority of UK patients will start to experience pain within 15-20 minutes of light exposure outdoors between early March and October.

2.2 Critique of company’s overview of current service provision

The company correctly state that no NHS guidance has ever been issued for EPP and suggest that current standard care is limited to patients avoiding sunlight. Upon discussing treatment options with the ERG’s clinical advisors it was noted that beta-carotene compounds (taken orally, on average eight tablets daily) seem to provide some protection for a minority of people. However, it can sometimes be hard to obtain beta-carotene in the UK and it has to be sourced from overseas (e.g. the USA). The ERG’s clinical advisors also described the use of narrow-band ultraviolet beta (UVB) phototherapy (e.g. 3 x weekly for 4-6 weeks or variations of), which has, according to clinical experience and a few case reports, been shown to marginally increase patients time of exposure to sunlight. Although the ERG’s clinical advisors did mention that few patients choose this option due to the practical issues and impact on lifestyle and work routine. The ERG experts state that the use of Dundee cream can also slightly increase the time patients can be exposed to sunlight. However, it tends to be reserved for particular outdoor occasions rather than being used daily. This is because large volumes need to be applied, and it can adhere to clothing. In addition, these creams have an appearance similar to cosmetic make-up and are therefore not always acceptable to some patients (e.g. younger males). They can also be difficult to get from general practitioners on prescription. Vitamin D and calcium are recommended (though patients may not always take them regularly) and this would not change if afamelanotide is prescribed.

The current treatment options discussed by the ERG experts above were not mentioned in the CS (apart from a brief reference considering beta-carotene as part of the cost effectiveness model (CS Table D3, p74)).

The ERG's clinical advisors state that there is little evidence for the above current treatments and that helping patients to manage their exposure to light is a key part of management. Patient experience of the currently available treatments is discussed in the consultee submissions to NICE (described in section 7 of this report).

2.3 Critique of company's definition of decision problem

Population

The population described in the company's decision problem is adults (CS Table A1, p. 11) which matches that specified in the final scope issued by NICE. The age range of adults is not mentioned in the decision problem section of the CS but the inclusion/exclusion criteria in the summary of methodology for the RCTs (CS table 5, pp 48 24) state adults aged between 18-70 years.

The CS states that there are 394 known patients in the UK with EPP based on published estimates (CS p 9). The CS also states separately that there are [redacted] and an estimated current total of 513 patients in England based on disease prevalence (CS, p 9). Furthermore, it is suggested that there are [redacted] patients eligible for treatment (CS, p 91), though it does not mention the proportion of patients in whom afamelanotide may be contraindicated (such those over the age of 70 years or below 18 years old, pregnant women or those with liver disease). Although this figure is higher than that previously cited, the ERG clinical experts consider that this figure is generally correct and would probably not vary by around 100 patients either way.

Intervention

The intervention in the decision problem (CS Table A1, p 11) is stated as afamelanotide (16mg), delivered as a controlled release injectable implant. Afamelanotide has a European marketing authorisation from the European Medicines Agency (EMA), granted in December 2014 under "exceptional circumstances" (CS, p 53). The European Public Assessment Report (EPAR, p 89)

[2] describes the discussions between the company and the Committee for Medicinal Products for Human Use (CHMP) regarding these circumstances, namely the fact that EPP is a rare condition and that comprehensive data on the efficacy and safety under normal conditions of use could not be generated, resulting in the granting of a marketing authorisation under exceptional circumstances.

The afamelanotide Summary of Product Characteristics (SmPC) states that the recommended dose of afamelanotide is 16mg, delivered as a subcutaneous implant (1.7 cm in length x 1.5 mm in diameter), administered every two months prior to expected and during increased sunlight exposure e.g. spring to early autumn). Three implants a year are recommended with a maximum of four per year [CS table A2, p 12]. The SmPC states that the safety and efficacy of afamelanotide has not been established for patients under 18 or over 70 years of age, or during pregnancy or lactation (SmPC, pp 3-5). It also states that long term safety data (after two years) have not been evaluated (SmPC, pp 3-4).

Comparators

The only comparator included in the scope and the decision problem is best supportive care. The CS does not explicitly define best supportive care within the decision problem (CS Table A1, p 11), but the ERG assumes that it would include the various current management options that are described above (section 2.2). The CS states that there are no alternative treatments or comparators used or in development at present (CS, p 10).

Outcomes

The outcomes specified in the NICE scope are duration of tolerance to sunlight and other forms of visible light; phototoxic reactions; change in melanin density; adverse effects of treatment; health-related quality of life (HRQoL) (for patients and carers); and mortality. These outcomes are included in the company's decision problem (CS Table A1, p11) although the CS does not explicitly report mortality and does not report HRQoL for carers of people with EPP (due to lack of relevant information). Section 3.1.5 of this report provides a description and critique of the company's assessment of the outcome measures.

3 CLINICAL EFFECTIVENESS

3.1 Critique of company's approach to systematic review

3.1.1 Description of company's search strategy

The company reported a single search for clinical effectiveness evidence, economic evidence, and resource identification and valuation (CS section 9.1 and CS Appendix 1, Appendix 3, and Appendix 4). PubMed was the sole external database searched, with the date of the search up to 15th July 2017. The company justifies only searching this database and not Embase, Medline In-Process and the Cochrane Library (as required by NICE) as it is the sole supplier of afamelanotide and is aware of all clinical research undertaken on it. The ERG acknowledges that an orphan drug/first in class product is unlikely to have been evaluated outside of the company, however the expectations of a systematic literature review have not been fully met. The ERG considers that free text search terms used in the search strategy are appropriate. The quantity of references identified from the search was not recorded nor tabulated into a PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow-chart as is customary in health technology assessment reports. The ERG requested this flow-chart for transparency but the company declined due to the burden of the administrative request (clarification response question A11, 26/09/17)). The company did not conduct separate searches for literature on adverse events, however, it is likely that any available evidence on adverse events would have been identified by the company's main search and from their in-house pharmacovigilance database (CS, Appendix 2).

The company cross-checked their internal reference library against their PubMed search results. The ERG considers it would have been informative as a minimum to quote which sources were used in the weekly current awareness alerts that feed the in-house company database. The company also reported searching for ongoing trials on the National Institute of Health clinicaltrials.gov and Eudract (European Clinical Trials Database).

The ERG elected to search Embase, Web of Science, The Cochrane Library, Econlit, and the NHS economic evaluation database (NHS EED) for any additional references relating to afamelanotide. In addition, the ERG searched the following additional databases: clinicaltrials.gov, UK Clinical Trials Gateway (UKCTG), ISRCTN, and the WHO International Clinical Trials Registry Platform (WHOICTRP). The 2017 proceedings of the International

Congress on Porphyrins and Porphyrias was also checked by the ERG. The results of the ERG searches were screened to identify any additional relevant data. Only one relevant publication was identified, a conference abstract of a cost effectiveness analysis of afamelanotide [3]. The ERG discusses this study further in section 4.2 of this report.

3.1.2 Statement of the inclusion/exclusion criteria used in the study selection.

The inclusion criteria for the company's systematic review of both published and unpublished studies are clearly stated in the CS, (tables C1 and C2, p 22-23). No exclusion criteria were stated. The inclusion criteria stated reflect the decision problem for population and intervention.

As stated above (section 3.1.1) a PRISMA flow diagram to show the numbers of records retrieved, included or excluded at each stage of the literature review was not included. It was stated that a total of 18 peer-reviewed journal articles were identified. However, there were only four citations to these retrieved articles in the subsequent paragraphs of the CS. The ERG requested a full reference list for these 18 articles, with stated reasons for any exclusions from the submission. These have now been provided (clarification response question A12, 26/09/17). However reasons for the omissions were not stated. The ERG notes that an additional three of these 18 references were cited in later sections of the CS, however the remaining 11 do not appear to have been cited anywhere in the CS.

3.1.3 Identified studies

The CS included seven relevant studies: CUV010, CUV017, CUV029, CUV030, CUV039 (see Table 1) a long-term treatment observational study, and a post authorisation safety study CUV-PASS-001 (Table 5). Some of these studies (CUV017 and CUV030), are currently unpublished although data were presented for these studies at the International Congress of Porphyrins and Porphyria 2013[4] and the 19th European Association of Dermatology and Venerology Congress, 2010 respectively.[5]

Table 1 Overview of clinical effectiveness studies in the company submission

Trial	CUV010 (Harms et al. 2009)[6]	CUV017 (unpublished)	CUV029 (Langendonk et al. 2015)[7]	CUV030 (unpublished)	CUV039 Langendonk et al. (2015)[7]
Trial design	Phase II, open label, single arm	Phase III, double blind RCT, alternating cross-over every 60 days	Phase III, double-blind RCT	Phase III, double-blind RCT	Phase III, double-blind RCT
Location	Switzerland	Europe/ Australia	Europe	USA	USA
Study duration	4 months	12 months	9 months	6 months	6 months
Number of patients	N=5 No withdrawals/drop outs	N=100 (93 treated) Withdrawal/drop outs unclear	N=76 (74 treated) Withdrawal/drop outs =5	N=77 (77 treated) Withdrawal/drop outs =5	N=94 (93 treated) Withdrawal/drop outs =6
Intervention (n in arm)	Afamelanotide (20 mg) (n=5)	Afamelanotide (16 mg) (n=93)	Afamelanotide (16 mg) (n=38)	Afamelanotide (16 mg) (n=39)	Afamelanotide (16 mg) (n=48)
Comparator (n in arm)	NA	Placebo (all patients received both treatments)	Placebo (n=36)	Placebo (n=38)	Placebo (n=45)
Primary outcome measured	• Provocation response time (PRT) under standardised laboratory controlled conditions	• Frequency of days of pain (by severity)	• Hours of direct sun exposure on days with no pain / mild pain (10:00 to 15:00 hrs) per subject (median)	• Hours of direct sun exposure on days with no pain / mild pain (10:00 to 15:00 hrs) per subject (median)	• Hours of direct sun exposure on total no of pain free days (10:00 to 18:00 hrs) per subject (median/mean)
Secondary outcomes measured	• Melanin density • HRQoL (SF-36 form) • Phototoxic reactions • Safety	• Hours per day of sunlight exposure • HRQoL (SF-36 form)	• Hours of direct sun exposure on days with no pain / mild pain (10:00 to 20:00 hours). • Mean number of phototoxic episodes per subject (+ mean	• Hours of direct sun exposure on days with no pain / mild pain (10:00 to 20:00 hours). • Time in direct sunlight when no or mild pain.	• Hours of direct sun exposure on days with no pain / mild pain (10:00 to 18:00 hours). • Days of 'some' sun exposure on days with no pain / mild pain (10:00 to 18:00 hrs) per subject (median/mean)

Trial	CUV010 (Harms et al. 2009)[6]	CUV017 (unpublished)	CUV029 (Langendonk et al. 2015)[7]	CUV030 (unpublished)	CUV039 Langendonk et al. (2015)[7]
			severity of episodes per subject) • Duration of phototoxicity (days) • HRQoL (EPP-QoL 15 form).	• HRQoL (EPP-QoL 15 form),	• HRQoL using -(EPP-QoL 15 form and revised 12 question form (post hoc)) and generic DLQI

NA = Not applicable; Dermatology Life Quality Index (DLQI)

Table 2 Overview of observational studies in the company submission

Name	Design	Number of patients	Intervention	Duration of study	Country/region	Outcomes measured
Biolcati et al. (2015)[8]	Longitudinal observation study	115	Afamelanotide 16mg	Up to 8 years	Italy, Switzerland	Primary Outcome • HRQoL (EPP-Qol) • Compliance + dropout Secondary outcome • Safety
Langendonk (2017) [9] CUV-PASS-001	Post-Authorisation Disease Registry Safety Study (incorporates European EPP Disease Registry (EEDR))	150 (as of Aug 2017)	Afamelanotide 16mg	On-going	International	Primary Outcome • Safety Secondary Outcome • HRQoL (EPP-QoL) (18 question form), • Length of severity of phototoxicity reported

- CUV010 was a four month phase II, open label, single am study carried out in Switzerland on five patients. This study compared afamelanotide (20 mg) versus placebo on the time to appearance of provoked symptoms; melanin density, phototoxic reactions and safety.[10]
- CUV017 was a 12 month phase III, crossover RCT, carried out in Europe and Australia, on 100 patients (93 treated). The study compared the effect of afamelanotide (16 mg) versus placebo on the frequency of days of pain (by severity); number of hours per day of sunlight exposure, melanin density, and HRQoL using the short form survey-36 (SF-36). The EPAR states that this trial was originally intended to be submitted as a pivotal study for marketing authorisation in 2009. However, the CHMP deemed that the crossover design was unsuitable and that pivotal, confirmatory parallel group studies should be run.[2] This study is unpublished.
- CUV029 was a nine month phase III, double blind RCT, carried out in Europe on 76 patients (74 treated). The study compared the effect of afamelanotide (16 mg) versus placebo on the number of hours of direct sun exposure on days with no pain and on days with no pain or mild pain (between 10.00-15.00 hours or 10.00-20.00 hours; number of phototoxic episodes; duration of phototoxicity, HRQoL using the Erythropoietic protoporphyria questionnaire (EPP-QoL), and adverse events.[7] HRQoL results from this study are used to inform the company's cost-effectiveness analysis. The trial was conducted between January 2010 and May 2011.
- CUV030 was a six month phase III, double blind RCT, carried out in the USA on 77 patients. The study compared the effect of afamelanotide (16 mg) versus placebo on the number of hours of direct sun exposure on days with no pain / mild pain (between 10.00-15.00 hours or 10.00-20 hours); number of phototoxic episodes; duration of phototoxicity and HRQoL using the EPP-QoL questionnaire. HRQoL results from this study are used to inform the company's cost-effectiveness analysis. This study is unpublished.
- CUV039 was a six month phase III, double blind RCT, carried out in the USA on 94 patients (93 treated). The study compared the effect of afamelanotide (16 mg) versus placebo on the number of hours of direct sun exposure on days with no pain / mild pain (between 10.00-15.00 hours or 10.00-20 hours); number of days of "some" sun exposure on days without pain or with no pain / mild pain (between 10.00-20.00 hours) and HRQoL using a 12 item revised version of the EPP-QoL and the Dermatology Quality

Life Quality Index (DLQI).[7] Some of the design characteristics of this trial (e.g. the length of treatment) were informed by experience gained from earlier trials, including CUV029. The trial was conducted between May 2012 and July 2013 with inclusion restricted to two months to allow the trial to be performed mainly during the summer months[7]). The EMA considered this trial to provide pivotal data for the assessment of efficacy of afamelanotide and was robust enough to support the marketing authorisation (however, they did not consider that studies CUV029 or CUV030 were pivotal due to concerns about their conduct; see section 3.1.6.5 of this report for further details).[2] HRQoL results from this study are used to inform the company's cost-effectiveness analysis.

The two long-term observational studies included are:

- Biolcati et al.[11] followed up 115 patients (retrospectively) treated in Italy and Switzerland who had been treated for up to eight years between 2006 and 2014, to assess HRQoL (EPP-QoL), melanin density, adverse events and compliance and dropout.
- Langendonk et al.[9] describes the post authorisation disease registry safety study (PASS) which was set up as a condition of the European licensing authorisation. Afamelanotide can only be prescribed by designated and trained porphyria centres according to a protocol (supplied as an appendix to the CS). Centres are required to monitor patients and the company to submit yearly reports. As of May 2017 104 Dutch patients have been included in the treatment programme where patients have received up to five implants (CS p 39, [9]). The European EPP Disease Registry (EEDR) collects safety and effectiveness data from European Centres in the PASS. The first safety data from the EEDR have been reported to the EMA, with subsequent annual reports to be submitted in December each year.

The CS reports details of the included studies including the location, study design, study duration, sample size, inclusion/exclusion criteria, method of randomisation and blinding, intervention and comparator, statistical tests and outcomes. The numbers of patients discontinuing treatment are reported in most studies, although it was stated to be not applicable in CUV017. The CS stated that statistical tests were reported, however the summary tables contain no details of power/sample size calculations. Participant characteristics at baseline are not given for all studies and the ERG requested clinical study reports and trial protocols from the company, though the company chose not to provide these.

No ongoing studies have been listed in the CS apart from the ongoing PASS study mentioned (CUV-PASS-001).

The ERG believes that all relevant studies have been included in the CS and all of those that have been included meet the stated inclusion criteria.

3.1.4 Description and critique of the company's approach to validity assessment

The company assessed the quality (using the NICE recommended criteria) of studies CUV029, CUV030, and CUV039 (CS, Table 7) but not studies CUV010 or CUV017. The company provided a brief critical appraisal of the long-term observational study by Biolcati et al.[11] (CS, Table 8). Table 3 below provides the company's quality assessment judgements for studies CUV029, CUV030, and CUV039 and the ERG's quality assessment judgements for these three studies, plus study CUV017 (the ERG requested the company to provide a critical appraisal of this study but the company said that this was not appropriate as it was a cross-over trial. The ERG contends that critical appraisal criteria are applicable to cross-over RCTs as well as parallel-group RCTs and has conducted a critical appraisal of this study based on the information given in the CS).

Table 3 Company and ERG assessment of trial quality

Study Name	CUV017		CUV029		CUV030		CUV039	
Critical appraisal criterion	**Judgement**							
1. Was the method used to generate random allocations adequate?	**CS:**	Not stated	**CS:**	Yes	**CS:**	Yes	**CS:**	Yes
	ERG:	Yes	**ERG:**	Yes	**ERG:**	Yes	**ERG:**	Yes
	ERG comment: The CUV017 trial was not included in critical appraisal table C7 p 41. However, in the CS [table C5 p 28], it is stated that "each patient was assigned to a treatment arm according to a computer generated randomisation list. For each study site, patients who satisfied the inclusion/exclusion criteria were allocated patient randomisation numbers sequentially and chronologically, based on the timing of their attendance at the clinic for the first study implant". This method was also used for CUV029 and CUV030 [CS page 31 and page 34], however for CUV039, the randomisation process differed. Here subjects were randomised on a site basis to maintain a geographic/climatic balance between treatment arms [CS p 36]. The randomisation method used a small block size (four) to ensure that treatment was balanced within study sites. Five individually sealed sets of computer-generated randomisation codes (each set containing 48 randomised numbers) were provided to the pharmacy. The study pharmacist chose one of the five sealed envelopes and the selected randomisation list was used to randomise the subjects in this study.							
2. Was the allocation	**CS:**	Not stated	**CS:**	Yes	**CS:**	Yes	**CS:**	Yes

Study Name	CUV017		CUV029		CUV030		CUV039	
Critical appraisal criterion	**Judgement**							
adequately concealed?	**ERG:**	Unclear	**ERG:**	Unclear	**ERG:**	Unclear	**ERG:**	Unclear
	ERG comment: The CS does not explicitly state whether procedures were followed for concealment of random allocation. The randomisation procedures used in CUV039 suggest that allocation may have been concealed, but it is not completely clear: "Five individually sealed sets of computer-generated randomisation codes (each set containing 48 randomised numbers) were provided to the pharmacy. The study pharmacist chose one of the five sealed envelopes and the selected randomisation list was used to randomise the subjects in this study" (CS page 36). For study CUV017 Table C5 [page 28] does not state whether any procedures to conceal allocation were used. It is stated however, that all sponsor, investigator site and monitor staff were blinded to the treatment code except the unblinded pharmacy monitor; pharmacy staff preparing treatments and statistician preparing randomisation. In studies CUV029 and CUV030 it is stated that the randomisaton code was kept in a sealed code break envelope. However, this appears to be reserved for emergencies in the event that the blinding needed to be broken, rather than a process for concealing the allocation of patients during enrolment.							
3. Were the groups similar at the outset of the study in terms of prognostic factors, e.g. severity of disease?	**CS:**	Not stated	**CS:**	Yes	**CS:**	Yes	**CS:**	Yes
	ERG:	Unclear	**ERG:**	Unclear	**ERG:**	Unclear	**ERG:**	Unclear
	ERG comment: Some baseline data are provided in the journal article for CUV029 and CUV039.[7] The data in this article shows a difference in the percentage of patients at baseline with Fitzpatrick type 1 skin (never tans, always burns) between the afamelanotide and placebo groups (16% vs 33%) in study CUV029. A similar difference is not observed in study CUV039 (27% vs 22%, respectively). The CS states that "Due to the limited potential sample size (i.e. orphan indication), it was not possible to actively control groups at baseline. At no point in the evaluations of these studies (including by EMA) was concern raised on this issue" [Table C7, page 41]. It is not clear to the ERG exactly what is meant by "actively control" in this context. In principle, adequate randomisation should ensure an even distribution of patient characteristics between trial arms, with any notable differences occurring due to chance. These can be adjusted for in statistical analysis of the outcome variables. The CS does not state if any adjustment was made for any instances of imbalance. Full baseline data for the other trials are not given in the CS and the ERG therefore requested these from the company (clarification response question A5, 02/10/17). The company did not supply these data but commented that there is no evidence that gender, age, skin type or the concept of race have any impact upon the safety or efficacy of afamelanotide. Expert clinical advice to the ERG suggested that it is reasonable to assume that skin type does not necessarily influence the effects of afamelanotide since the effectiveness of the treatment is unlikely to rely only on increases in melanin density.							
4. Were the care providers, participants and outcome assessors blind to treatment allocation? If any of these people were not	**CS:**	Not stated	**CS:**	Yes	**CS:**	Yes	**CS:**	Yes
	ERG:	Yes	**ERG:**	Yes	**ERG:**	Yes	**ERG:**	Yes
	ERG comment: All the trials are described as being double-blind. However, in the journal publication of the CUV029 and the CUV039 trials[7] it is stated that "the increased skin pigmentation in participants who received afamelanotide partially unblinded the trial" (p 53). This is not mentioned in the CS and presumably it was encountered in the other trials. The risk of patients being unblinded to the treatment due to the tanning effect of afamelanotide was acknowledged by the							

Study Name	CUV017		CUV029		CUV030		CUV039	
Critical appraisal criterion	**Judgement**							
blinded, what might be the likely impact on the risk of bias (for each outcome)?	company (clarification question response A2, 02/10/17). They stated that this issue had been addressed by the CHMP. It was considered that the beta-carotene that was evaluated in EPP patients causes tanning with no treatment effect and therefore does not translate in a change in the EPP patient's behaviour (in terms of their willingness to expose themselves to sunlight). It is not clear to the ERG if the statement is referring to beta-carotene taken by patients in the afamelanotide trials, or patients more generally. The point seems to be that tanning effects, whether caused by afamelanotide or beta-carotene, do not necessarily influence patient sun exposure behaviour.							
5. Were there any unexpected imbalances in drop-outs between groups? If so, were they explained or adjusted for?	**CS:**	Not stated	**CS:**	No	**CS:**	No	**CS:**	No
	ERG:	Unclear	**ERG:**	No	**ERG:**	No	**ERG:**	No
	ERG comment: Information on patient drop-out between study groups CUV017 was unavailable from CS table C5 (p 27-30). In the remaining three trials, all patients lost to follow up were explained in the CS (table 5 pp 30-38). The CS states that study drops outs were minimal and generally balanced between the active and placebo groups. The ERG notes that the number of patients discontinuing early was twice that in the afamelanotide group than the placebo group in study CUV029 (n=4 vs n=2), but these were small proportions of the study sample (10% vs 5% respectively).							
6. Is there any evidence to suggest that the authors measured more outcomes than they reported?	**CS:**	Not stated	**CS:**	No	**CS:**	No	**CS:**	No
	ERG:	Unclear	**ERG:**	Unclear	**ERG:**	Unclear	**ERG:**	Unclear
	ERG comment: Due to the absence of detailed study protocols (apart from CUV039 which was available as an appendix to the journal publication [7]), it was not possible to fully assess whether additional outcomes were measured in these studies. The ERG requested detailed study protocols from the company for the studies but these were not provided.							
7. Did the analysis include an ITT analysis? If so, was this appropriate and were appropriate methods used to account for missing data?*	**CS:**	Not stated	**CS:**	Yes	**CS:**	Yes	**CS:**	Yes
	ERG:	Unclear	**ERG:**	Unclear	**ERG:**	Unclear	**ERG:**	Unclear
	ERG comment: For study CUV017 CS table 5 states that the ITT population included all treated subjects who provided at least one post-dose efficacy assessment. The protocol for trial CUV039 [7] defines ITT in the same way. The ITT definition given by the company is effectively that of a "modified ITT" analysis rather than a true ITT analysis (which would require all randomised patients to be analysed). For the other studies the analysis is described as ITT but no definition is given to enable the ERG to determine whether it was a true ITT analysis. The CS highlighted in table C7 (pp 41-42) that although ITT was used for CUV029,030 and 039, it was stated in the critical appraisal section (CS p 42) that the principle of last value carried forward was not considered appropriate to the assessment of the chosen endpoints in this indication. Due to the variable nature of sun exposure and phototoxicity from day to day, using these as endpoints where the last value carried forward would not result in meaningful results. The CS reasoned that if a patient experienced a severe phototoxic reaction and dropped out with a pain scale score of 10 then that value would							

Study Name	CUV017	CUV029	CUV030	CUV039
Critical appraisal criterion	**Judgement**			
	need to be imputed for all future assessment points - would be nonsensical. The ERG agrees with this assertion.			

The ERG's judgement concur with that of the company for some of the quality assessment criteria, namely the adequacy of randomisation procedures and the procedures for ensuring blinding of patients, care providers and outcome assessors. However, the ERG notes that afamelanotide is associated with a tanning effect and that this is likely to have led to unblinding in many patients. The company state in their clarification response that, based on experience with beta-carotene in EPP patients, skin tanning does not appear to affect patients' behaviour in relation to exposure to sunlight. The ERG agrees that unblinding due to a tanning effect might not necessarily lead to systematic differences in patients' behaviour between the study groups, although it is unclear whether study investigators would be influenced by such unblinding. The ERG also agrees with the company that there were no unexpected imbalances in drop-outs between study groups (though this information is not available for study CUV017).

The ERG disagrees with the company's quality assessment for allocation concealment, similarity of the study groups at baseline, and use of an ITT analysis. It is unclear to the ERG whether random allocation was adequately concealed in the studies as the descriptions given did not explicitly mention concealment procedures. Also, due to the absence of detailed patient baseline information in the CS it is not possible to determine whether the randomised study groups were similar at the outset of study, and there was one notable imbalance in Fitzpatrick type 1 skin between the afamelanotide and placebo groups (16% vs 33%) in study CUV029 (the company asserts that skin type does not modify the effects of afamelanotide). Furthermore, although the studies were described as using ITT analyses the precise definition of ITT is not given for all studies. The ERG notes that there is much variation in definition of ITT analyses in descriptions of clinical trials and that they do not always describe a "true" ITT analysis (i.e. all randomised patients within the groups to which they were allocated) (see section 3.1.6 of this report for description and critique of the statistical procedures in the studies).

3.1.5 Description and critique of company's outcome selection

The CS states that the company proposes no variations in outcomes to the NICE scope (CS p 11). The ERG agrees that the outcomes selected by the company match the NICE scope, apart from two exceptions:

- For the NICE scope outcome "HRQoL (patients and carers)" the CS has only provided HRQoL data for patients. Following a clarification question (clarification response question A14, 26/09/17) the company confirmed that they are not aware of any published data on the impact of EPP on the quality of life of carers, though anecdotal evidence is available with reference to Food and Drug Agency Scientific Workshop transcripts (see section 7 of this report for the ERG's summary of the consultee submissions to NICE, which includes patient perspectives).
- The CS does not report mortality, which is an outcome specified in the scope.

3.1.5.1 Outcomes specified in the NICE scope

The CS reports data for the outcomes in the NICE scope as follows.

Duration of tolerance to sunlight and other forms of visible light

Outcomes reported in the CS refer to two types of light exposure among EPP patients: voluntary exposure to natural light, including sun exposure; and exposure to artificial light under standardised laboratory test conditions (in the form of a 300W Xenon Arc Lamp), which in the CS is termed "photoprovocation". The majority of light exposure outcomes reported in the CS relate to EPP patients' voluntary exposure to natural light.

Exposure to natural light

The voluntary light exposure outcomes reported by the company are shown in Table 4. In addition to the outcomes shown in Table 4, study CUV017 assessed patients' voluntary sun exposure but the CS does not specify during which hours of the day assessments were made and only brief descriptive results are given (see section 3.3 of this report). The CS also states that in the small study CUV010 (n=5), "sun exposure" was a secondary outcome, but no further information defining this, or results, are presented in the CS.

Table 4 Voluntary light exposure outcomes assessed in the studies

Outcome	Study CUV029	Study CUV030	Study CUV039
Total hours in study in direct sunlight with no pain	Assessed 10:00 – 15:00 (5h) per day (co-primary outcome) and 10:00-20:00 (10h) per day (secondary outcome)	Assessed 10:00 – 18:00 (8h) per day (primary outcome) and 10:00-15:00 (5h) per day (secondary outcome)	Assessed 10:00 – 18:00 (8h) per day (primary outcome) and 10:00-15:00 (5h) per day (secondary outcome)
Total hours in study in direct sunlight with no pain or mild pain	Assessed 10:00 – 15:00 (5h) per day (co-primary outcome) and 10:00-20:00 (10h) per day (secondary outcome)	Not assessed	Assessed 10:00 – 18:00 (8h) per day (secondary outcome)
Total hours in study in direct sunlight regardless of pain score	Not assessed	Not assessed	Assessed 10:00 – 18:00 (8h) per day (secondary outcome) (data in EPAR only)
Total days in study "in some direct sunlight" on days with no pain [a]	Not assessed	Not assessed	Assessed 10:00 – 18:00 (8h) per day (secondary outcome; also referred to in the CS as an exploratory outcome)
Total days in study "with some sunlight" on days with no pain or mild pain [a]	Not assessed	Not assessed	Assessed 10:00 – 18:00 (8h) per day (secondary outcome; also referred to in the CS as an exploratory outcome)

[a] Phrasing of outcomes (indicated here in quotation marks) as reported in the study publication[7] is inconsistent between these two outcomes – unclear whether this is a typographic error or reflective of a material difference in how the outcomes were assessed; the ERG assumes these outcomes differ only in the degree of pain experienced, not in sunlight exposure

The majority of results relating to voluntary light exposure behaviour of EPP patients are from the CUV039 study which was conducted in the USA, and from CUV029 which was conducted in Europe. As well as being of different duration, the studies differed according to the daily times when outcomes were assessed, which were 10:00-15:00, 10:00-18:00, and 10:00-20:00. The CS does not explain these differences in the timing of exposure assessments between the trials. Although the CS designates the different sunlight exposure outcomes as "primary" and "secondary" within each study, insufficient information is reported in the CS to determine whether the primary outcomes would be any more reliable than secondary outcomes in terms of their statistical power (see section 3.1.6).

The CS provides varying descriptions of light exposure, including (amongst others) "direct sunlight exposure" (e.g. CS P 32), "direct light/sunlight exposure" (e.g. CS p 32), "light/sun

exposure" (e.g. CS p 36) or "direct light/sunlight exposure" (e.g. CS p 36). The ERG requested clarification of the exposure definitions from the company via NICE (clarification response question A6, 26/09/17). The company responded stating that the studies "evaluated the excitation of protoporphyrin IX by "visible light (>408 nm)" and that "patients were asked to expose themselves to conditions of direct light/sunlight exposure, which was the best approximation that was possible at the time of the clinical programme".

Duration of tolerance to sunlight is dependent on the amount of pain caused by light exposure. For this reason, in trials CUV029, CUV030 and CUV039 the company assessed duration of direct sunlight exposure for subgroups of patients who experienced "no pain" and "no pain or mild pain". The intensity and duration of pain and exposure to sunlight and shade were recorded daily by the patients in a diary, with the time spent outdoors being recorded in 15-minute intervals. Pain was scored on a 0-10 Likert scale. The CS describes the scale only for trial CUV017, stating score 0 was used for no pain, scores of 1 to 3 for mild pain, scores of 4 to 6 for moderate pain, scores of 7 to 9 for severe pain and 10 for worst imaginable pain. The ERG notes that the cut-off for mild and moderate pain is arbitrary, not explained by the company, and differed between the trials (CUV017, CUV029, CUV039 defined mild pain as 1-3 whilst CUV030 defined mild pain as 1-4). Full details of the Likert scale used in each trial and an explanation for the cut-off discrepancy between trials were requested by the ERG from the company via NICE (clarification response question A10, 26/09/17). The company responded that a panel of biostatisticians were consulted about defining anaphylactoid reactions and phototoxic episodes. The Likert scale was "a near approximation since EPP patients describe their ordeal as "pain" while a proper medical lexicon is lacking". However, no justification was given regarding the scoring threshold used.

The company presents sunlight exposure outcomes in terms of the total hours of exposure to sunlight during the study (i.e. the first three outcomes listed in Table 1) and the days with sunlight exposure (i.e. the last two outcomes listed in Table 1). The company calculated the light exposure outcomes based on the patients' diary records of light exposure and pain scores. The CS, study publication,[7] SmPC, EPAR[2] and company's clarification response do not clearly explain how the outcomes were calculated from the diary card data.

Total hours in direct sunlight

The ERG assumes that to obtain the first outcome listed in Table 4 the company summed the patient's sunlight exposure time in each of the 15-minute study intervals that had a maximum pain score of zero, to give the total time of sunlight exposure per patient during the study with "no pain". Similarly, for the second outcome listed in Table 4 we assume that the company summed the sunlight exposure time in each of the 15-minute study intervals that had a maximum pain score of 3 (or 4), to give the total time of sunlight exposure per patient during the study with "no pain or mild pain". The third outcome listed in Table 4 would have been calculated similarly, by summing sunlight exposure time across all 15-minute intervals irrespective of the pain score of each interval. Results of these outcomes (see section 3.3) are presented as the mean and median duration of sunlight exposure per patient.

Total days in direct sunlight

The method of calculating the final two outcomes listed in Table 4 for study CUV039 is not reported in the CS or study publication[7] and is not clear to the ERG. The EPAR[2] (p. 50) implies that these outcomes were calculated for each subject by dividing the total time in the study spent in direct sunlight (without or with mild pain) by the number of days each subject was in the study. This would result in fractional outcomes <1.0 since the denominator would be larger than the numerator, but this does not agree with the format of the reported outcomes, which are expressed in days (section 3.3). The wording of these outcomes is inconsistent in the study publication (see footnote to Table 4) which adds ambiguity to the interpretation.

Photoprovocation

Photoprovocation is a test of the duration of tolerance of artificial light under standardised laboratory test conditions (in the form of a 300W Xenon Arc Lamp) in which the time taken to provoke minimal symptoms is recorded. Advantages of photoprovocation testing are that exposure conditions can be clearly controlled (which is not possible with patients' voluntary outdoor exposure behaviour and heterogeneous weather conditions), and patient exposure to the light stimulus can be ensured (i.e. behavioural avoidance of light exposure does not occur). Disadvantages of photoprovocation testing are that it is unclear how generalisable the exposure conditions are (specific areas of the body are assessed rather than all exposed skin areas); and photoprovocation does not capture patients' behavioural response to light exposure which could be an important determinant of compliance with therapy.

Photoprovocation is reported in the CS only for the small (n=5) study CUV010 (CS p 26), in which photoprovocation, carried out on the dorsal surface of the hands, was the primary outcome, but no further information defining this, or results, are given in the CS, although results are reported in more detail in a publication by Harms et al.[12]

According to publications, photoprovocation was also tested in small subgroups of patients in study CUV030[13] and study CUV039.[2 7] The photoprovocation tests were conducted on subsets of patients: n=15 in CUV030 (but only six completed testing); and n=21 in CUV039 (number completing testing not reported). However, no rationale is given in the CS or study publications for the patient subgroup selection.

The CS does not provide any explanation of why photoprovocation was conducted and the very limited descriptive results given suggest that the company does not view this as being an important outcome for the current appraisal.

Phototoxic reactions

Phototoxic reactions are reported in the CS for five studies (CUV010, CUV017, CUV029, CUV030, and the ongoing study CUV-PASS-001). Phototoxic reactions are reported for the CUV039 trial in a publication by Langendonk et al.[7] but these data are not mentioned in the CS.

The outcome relating to phototoxicity, as reported in the CS, is “pain”. The company specifies “pain” within quotation marks without defining explicitly what they mean by “pain”. However, in the NICE “Response to consultee and commentator comments on the draft remit and scope (pre-referral)” the company had stated in a comment to NICE that “…patients have an ingrained fear for an episode of anaphylactoid reaction, burns, oedema and scarring, causing an unspeakable internal ordeal often poorly – and by lack of a better word – expressed as “pain”…”. This statement suggests that the “pain” outcome reported in the CS somehow captures other aspects of phototoxicity such as burns and oedema.

The CS does not report any specific outcomes for non-pain aspects of phototoxicity (e.g. burns, oedema, rash, scarring). The ERG understands from clinical experts that pain is a significant burden to patients but it is unclear to us whether the other aspects of phototoxicity are also

important to patients relative to the pain and, if so, whether the “pain” outcome adequately captures the full burden of phototoxic effects.

Pain was assessed on the 11-point Likert scale described above (see Exposure to natural light above). In reference to trial CUV039, the afamelanotide EPAR[2] states that “the **number** of phototoxic reactions was determined by counting the number of episodes on which patients report a Likert score of 4 or more for 1 or more consecutive days. The **total severity** of an individual phototoxic reaction was determined by adding the Likert scale severity scores for all days in an individual phototoxic reaction. The **maximum severity** of a phototoxic reaction was determined by the highest daily Likert scale score that occurred during that phototoxic reaction” (p 51).

The CS states for study CUV017 that “the primary efficacy objectives were to determine whether afamelanotide could reduce the number and severity of phototoxic reactions in patients with EPP” (CS p 27). For study CUV029 the CS states that the primary objective was modified when preparing the statistical analysis plan, with the modified objective being “to determine whether afamelanotide can enable patients to expose themselves to direct sunlight during the most intense periods of sunlight during the day in spring and summer” (CS p 30). This outcome recognises that without sunlight as a causative factor no pain or phototoxic reaction is possible. For study CUV030 the CS states that “during the initial stages of analysis, and in order to determine the clinically relevant impact of afamelanotide treatment, the sequence of the study objectives was adapted to assess whether the study subjects are able to modify their lifelong conditioned behaviour. This was assessed by evaluating time spent in direct sunlight while remaining pain free or experiencing only mild pain, during spring and summer months” (CS p 34). For study CUV039 the stated objective was “To determine whether afamelanotide can enable EPP patients to expose themselves to light/ sunlight without incurring phototoxic reactions and pain” (CS p 36).

Change in melanin density

The CS states that melanin density was measured by spectrophotometry (reflectometry according to Harms et al.[12]) but no technical details of the method are reported. The change in melanin density is mentioned briefly in the CS only for the small (n=5) study CUV010. A journal publication by Harms et al.[10] gives further melanin density results for study CUV010. Change in melanin density was also assessed in the crossover study CUV017 (according to the EPAR)

and in the long-term observational study (Biolcati et al.[11] supplemental appendix) but these assessments of melanin density are not mentioned in the CS. The CS does not explain the mode of action of afamelanotide, other than that it is a melanocortin-1 receptor (MC1R) agonist, and the reliability of melanin density as a clinical effectiveness outcome is not discussed in the CS. The ERG notes that EPP can occur in some people who have dark skin[14] and that melanin density is cited in the afamelanotide EPAR as an indicator of pharmacodynamics, rather an effectiveness outcome (EPAR section 2.4.3).

Adverse events

The adverse events section (CS Table C10, pp 48-50) reproduces the list of adverse events given in the SmPC which is a summary list and is not explicit about which of the studies provided source data. The CS also provides limited information on adverse events for studies CUV010, the long-term observational study (safety reported anecdotally from the study publications[11 15]) and the ongoing study CUV-PASS-001. Detailed information on adverse events in trials CUV029 and CUV039 is available in a journal publication (Langendonk et al.[7]) but is not reported in the CS. Brief information on adverse events in trial CUV030 is given in a document by CLINUVEL 2010[16] but this is also not mentioned in the CS.

HRQoL

HRQoL was measured in all seven of the included studies, but the information provided in the CS is descriptive and very brief for most of the studies. Three HRQoL instruments were employed. These were the Short-Form 36 (SF-36) in studies CUV010 and CUV017; and the Dermatology Quality Life Quality Index (DLQI) and an EPP-specific questionnaire (EPP-QoL) in studies CUV029, CUV030 and CUV039 (CS p 71). The EPP-QoL was also employed in the long-term observational study (Biolcati et al.[11]) and in the monitoring study CUV-PASS-001 (CS Table C5). However, the CS states that the SF-36 “did not prove to be useful for the assessment because most patients reported a very high quality of life from baseline assessments onwards, a finding contrary to the published literature” (CS p 71). The CS also states that that the SF-36 and the DLQI are not suitable for the quantification of the humanistic burden of EPP and hence a new disease-specific questionnaire, the EPP-QoL, was designed by expert porphyria physicians globally together with the sponsor (CS pp 9 & 71).

According to the CS, a 15-question version of EPP-QoL was developed (CS p 71), but the publication reporting results for studies CUV029 and CUV039 (Langendonk et al.[7]) presents a

12-question version of EPP-QoL. It is not clear in the CS which version of EPP-QoL was used in each study. In several places in the CS the company mentions that the [redacted]

The CS appears inconsistent in its criticism of the DLQI, since a survey of EPP patients by Holme et al.[17] which utilised the DLQI, is cited as evidence that EPP has a marked impact on patients' quality of life (CS pp 15-16). Although DLQI is a generic instrument for assessing HRQoL impacts of skin conditions, we note that it includes a question about pain whereas the EPP-QoL does not directly (it does include a question asking patients how often they feel they are risk of developing EPP symptoms. The ERG notes that this could therefore include pain). The wording of the DQLI pain question is "Over the last week, how itchy, sore, painful or stinging has your skin been?" This appears pertinent to the nature of pain experienced by EPP patients, since the survey by Holme et al.[17] indicated that patients found the cutaneous sensation following sunlight exposure difficult to describe, with the most frequent responses being burning (85%), tingling (33%), prickling (4%) and stinging (3%). The Holme et al.[17] survey is the largest survey conducted in EPP patients and demonstrated that DLQI scores in EPP patients are higher than in other skin conditions and indicative that EPP has a substantial impact on patients' quality of life. The DLQI has been widely used and subjected to validation in a number of studies.[18] It has also been used to measure quality of life in EPP patients in other studies.[19] The ERG therefore disagrees with the company's assertion that DLQI is not necessarily suitable as a measure of HRQoL in EPP. The CS does not report any DLQI scores, although we note that DLQI scores from study CUV039 are given in the afamelanotide EPAR.[2] (we have reported these in section 3.3.5 of this report). The ERG requested standardised DLQI scores from the company via NICE (clarification response question A2, 12/09/17) but the company declined to provide these. Further discussion of the use of DLQI to inform cost effectiveness of afamelanotide for EPP is provided in section 4.3.2.2 of this report.

Mortality

Mortality is not reported in the CS but is mentioned in the journal publications[7] and the EPAR[2] for studies CUV029, CUV039 and the long-term observational study[11] (see section 3.3.7 of this report).

3.1.5.2 Outcomes not specified in the NICE scope

According to the publication, in study CUV029 only, the levels of protoporphyrin IX (in erythrocytes) were assessed at baseline and follow-up.[7] Levels of protoporphyrin IX may indicate disease severity but are not influenced by afamelanotide therapy, so this is a prognostic factor rather than an efficacy or effectiveness outcome.

Summary

The company's outcomes are appropriate for the health condition and match the NICE scope, apart from no data being provided for the HRQoL of carers and for mortality. Not all of the information regarding outcome measures is provided in the CS, with additional information being sought by the ERG from journal publications and the EPAR.

3.1.6 Description and critique of the company's approach to trial statistics

The CS does not report trial results for all of the outcomes specified in the NICE scope. Where results are presented they are often descriptive only (CS Tables C5 and C9) and do not reflect all relevant results that are available elsewhere in trial publications (e.g. Langendonk et al. report relevant outcomes for trials CUV029 and CUV039 in more detail than the CS[7]).

3.1.6.1 Overall analytical approach

For six of the studies (not including the CUV-PASS-001 monitoring study) the CS states that analysis was by ITT. However, the CS only defines ITT for the crossover trial CUV017, stating that the ITT population included all treated subjects who provided at least one post-dose efficacy assessment, and that this was planned to be the main population for all efficacy analyses (CS p 42). The protocol for trial CUV039 (not initially provided by the company but available in a supplement to a journal publication[7] defines ITT in the same way. The ITT definition given by the company is effectively that of a "modified ITT" analysis rather than a true ITT analysis (which would require all randomised patients to be analysed).

The afamelanotide EPAR (p 52)[2] notes that for study CUV039 there are three "ITT" populations, reflecting the availability of post-dose effectiveness data for different data types, i.e. diary card, photoprovocation subset and HRQoL. In CUV039 the "study completers" population included subjects who received all doses of study treatment and returned adequately completed diary card entries ("diary card population"), completed all HRQoL assessments ("HRQoL population) or had the required number of photoprovocation tests ("photoprovocation subset"). The safety

population included all enrolled subjects who were randomised and received at least one dose of study medication (afamelanotde EPAR[2]). The company did not provide clinical study reports or protocols for any studies, but we assume that the population definitions for CUV039 apply also to the other studies, CUV017, CUV029 and CUV030 (clinical study reports and protocols were requested by the ERG from the company via NICE but these were not supplied; clarification response question A3, 26/09/17).

For study CUV039 the updated (June 2013) Statistical Analysis Plan (available in an appendix to Langendonk et al.[7]) does not name the specific statistical tests that would be employed in analyses, but it states that descriptive statistics would be provided in summary tables. According to the CS and journal publication (Langendonk et al.[7]), differences between the study-drug groups were assessed with the use of the Kruskal–Wallis test with Hodges-Lehmann shift estimate of difference for primary outcomes; chi-square tests for proportions; and a Wilcoxon rank-sum test for changes in HRQoL. The Hodges-Lehmann shift estimate of the difference between two groups uses the information contained in all pairwise differences between the groups and can provide a robust estimate of the median difference between groups when the underlying distributions for the groups are symmetric about their respective medians.[20] However, the CS does not provide any explanation of the rationale for using this statistical test and whether the distributions of data were symmetric. In cases of non-symmetry the reliability of the Hodges-Lehmann shift estimate is less clear.[20] In study CUV017 a Cochran-Mantel Haenszel test for two categorical datasets obtained in a crossover design was employed (CS p 29), but the CS does not specify whether a treatment-by-period interaction was tested and if a washout period between observations from alternating afamelanotide and placebo treatments was necessary (each patient alternated between an afamelanotide or placebo implant every 60 days. The duration of the effect of an afamelanotide implant, and hence the appropriate washout period, is not clear). The ERG agrees that the tests employed by the company appear generally appropriate, but few details are reported, and the descriptive statistics provided in the CS are incomplete and inconsistent across studies and outcomes (in some cases only qualitative narrative statements of results, sometimes with p-values, are reported; in other cases mean ± SD, 95% confidence intervals (CIs) and/or median and range are reported). The ERG has obtained missing descriptive statistics that were available from the study journal publications and the EPAR[2]) (see section 3.3).

3.1.6.2 Sample size

The CS, study publications and EPAR[2] do not provide any justifications for the sample size or statistical power of the studies. The statistical analysis plan for study CUV039 (provided in a supplementary appendix to the publication by Langendonk et al.[7]) states that analysis of data from the prior CUV029 and CUV030 studies demonstrated that a significant difference in the primary endpoint could be detected with "approximately 75-100 patients", but the variance, detectable difference and statistical power values used in the sample size calculation are not specified. The eventual number of patients randomised in study CUV039 was 94 which is at the upper end of the range specified. The EPAR reports that basing the sample size on a previous phase III trial was considered acceptable by the CHMP[2].

3.1.6.3 Attrition

According to CS section 9.4.6, overall patient withdrawal rates were low across the clinical trial programme. Across the three late stage studies (CUV029, CUV030 and CUV039), 17 patients did not complete the full protocol, including three who were lost to follow up but received all study medication.

The CS does not provide any Consolidated Standards of Reporting Trials (CONSORT) charts to show patient flow through the studies although the afamelanotide EPAR [2] provides a flow chart for study CUV039. (The ERG requested charts for all the studies from the company via NICE, but the company did not provide them - clarification response question A6, 02/10/17). The CS does not mention any patient attrition for studies CUV010 (which only included five patients), CUV017, or the long-term observational study[11]. Patient discontinuations in the remaining three core studies are reported in the CS as follows (CS Table C5):

- CUV029: Four subjects discontinued from the afamelanotide arm and two from the placebo arm, with reasons reported separately by study arm.
- CUV030: The CS states "[redacted]". It is unclear whether all patients who discontinued are accounted for by this statement.
- CUV039: According to the CS, 3 subjects in each arm discontinued. Reasons for discontinuation are given, but not separately by study arm. The afamelanotide EPAR[2] reports that reasons for discontinuation from the afamelanotide arm were withdrawal of

consent (no reasons given) (n=2) and a physician's decision (n=1) (clinical reasons not related to implant); whilst 2 patients from the placebo arm were lost to follow up and 1 discontinued due to a physician's decision (serious adverse event, clinical reasons not related to implant).

For studies CUV029 and CUV039 although attrition rates per arm ranged from 5.5% to 10.5% the reasons for discontinuation do not suggest that the discontinuations would have led to systematic imbalances in prognostic characteristics of the study arms (i.e. bias). For study CUV030 it is unclear whether all the discontinuations have been reported.

The CS states that given the low numbers and the reasonably even distribution of withdrawals, these withdrawals were not considered to have had an impact on the outcome of the overall assessment of the study endpoints. The ERG agrees that the company's assertion is reasonable for studies CUV029 and CUV030 but there is uncertainty as to whether all discontinuations in CUV030 have been reported, and no information on discontinuations is available for study CUV017.

3.1.6.4 Handling missing data

The CS states that for studies CUV029, CUV030 and CUV039, ITT was used but "the principle of last value carried forward was not considered appropriate to the assessment of the chosen endpoints in this indication" (CS Table C7, p 42). The company's rationale is that "Sun exposure and phototoxicity are not endpoints where the last value carried forward would give meaningful results because both are quite variable day to day. As an example, if a patient dropped out because they experienced a severe phototoxic reaction with a pain scale score of 10, then that values [sic] would need to be imputed for all future assessment points – this would be nonsensical" (CS p 42). The ERG agrees with this assertion.

The CS, in describing study CUV039 (CS p 37), states that analyses were therefore performed on a best and worst cases imputation, as described in the statistical analysis plan. The ERG agrees that this imputation approach is appropriate. However, the company does not report for any studies or for any individual study arms whether the results presented in the CS and in the journal publication (Langendonk et al.[7]) are for the best-case or the worst-case imputation.

According to the CS, in study CUV039 "compliance of diary completion was very high. There were 185 out of 15608 diary days (1.2%) with missing Likert pain scores, and 296 diary days

(1.9%) with missing information about time outdoors. Last observation carried forward for missing phototoxicity or "pain" scores on days after a "pain" score of greater than 2 was applicable to only four subjects, for a total of 6 diary days" (CS p 37). Although not explicit, this appears to suggest that relatively few imputations would have been necessary, affecting 4/93 of the randomised subjects (4%) in study CUV039 (data are not reported by study arm). Corresponding information for the other studies is not given in the CS. The EPAR (p 71) states that (for post-hoc analyses of secondary outcomes) sensitivity analyses using the ITT diary card population produced similar results to the study completers diary card population.

In summary, the company's approaches to statistical analyses appear generally appropriate but information is lacking on how sample sizes and statistical power were estimated and on how missing data were handled. However, rates of attrition appear low for patients and for diary card data and it appears unlikely that attrition would have led to bias.

3.1.6.5 Additional criticisms by the European Medicines Agency

The EPAR[2] reports that a Good Clinical Practice (GCP) inspection was conducted of studies CUV029 and CUV030 as a result of changes to their analysis plans and the lack of clarity regarding sample size. The conclusion of the inspection was that the main efficacy data from these two studies were not considered robust and they could not be used to inform the marketing authorisation of afamelanotide. The key criticisms were: (1) that the design of the patient diary for capturing the data as needed for the analysis of endpoints related to duration of sun exposure was not suitable; (2) there was a change to the statistical analysis plan of study CUV030 after data had been analysed; (3) improper statistical planning and data handling for both trials; and (4) verification of the databases and of relevant events such as database lock / unlock was not possible. The inspection of study CUV039 concluded that it was compliant with the GCP hence its status as the sole pivotal study informing the marketing authorisation.

The GCP inspection and it's results are not mentioned in the CS though the company did acknowledge in their response to a clarification question that studies CUV029 and CUV030 were not used within the CHMP's efficacy assessment for the reasons explained within the EPAR (clarification response question A3, 02/10/17).

The ERG considers that criticisms of the EMA need to be taken into account in the interpretation of the results of these studies. This is particularly pertinent given that EPP-QoL results from

study CUV029 and CUV030 (pooled with those of CUV039) are used in the company's assessment of cost-effectiveness (discussed further in section 4.3.2.2 of this report).

3.1.7 Description and critique of the company's approach to the evidence synthesis

A narrative review is provided, with results of the included studies provided individually in tables, though the level of detail given is superficial and inconsistent across the studies. The interim NICE highly specialised technology (HST) company submission template states that the review should summarise the overall results of the individual studies with reference to their critical appraisal. However, there is no structured critical summary or comparison of the results across the trials in the CS.

A meta-analysis was not provided, with the authors stating that "it is not considered appropriate for the appraisal of SCENESSE®, due to the lack of scientific tools, alternative therapies and the extensive evaluation of the product in clinical trials compared to placebo (standard of care)" (CS p 52). It is not clear what the company means in this statement and the ERG does not agree that there is a lack of scientific tools for meta-analysis, since the outcomes analysed by the company would in principle be amenable to statistical pooling using orthodox methods. The ERG's view is that, in principle, a meta-analysis comparing afamelanotide with placebo plus standard of care (thus in keeping with the NICE scope) could be possible. However, due to clinical heterogeneity between the trials (e.g. duration of treatment; country/region and associated differences in outside light exposure) a meta-analysis would not be meaningful. Further, there are differences in the definitions of outcomes between trials which would which make meta-analysis potentially inappropriate (e.g. Hours of direct sun exposure on days with no pain / mild pain (10:00 to 15:00 hours) per subject in one study (CUV029), versus the same outcome with a time period of 10:00 to 18:00 hours in another study (CUV039)).

3.2 Summary statement of company's approach to systematic review

Table 5 provides the ERG's quality assessment of the company's review of clinical effectiveness.

Table 5 Quality assessment of CS review (Centre for Reviews and Dissemination (CRD) criteria)

CRD Quality Item; score Yes/No/Uncertain with comments
1. Are any inclusion/exclusion criteria reported relating to the primary studies which address the review question? Yes – brief criteria are reported in CS Tables C1, p22 (published studies) and C2, p23 (unpublished studies). Only criteria for inclusion are given, with no specific exclusion criteria reported. The criteria do not conflict with the decision problem and the NICE scope. The company state that the literature searches were conducted by a single author, and that the clinical sections of the CS were written by a single author and reviewed by others (clarification response question A8, 02/10/17). It is not clear whether inclusion criteria were applied by a single reviewer or by more than one reviewer.
2. Is there evidence of a substantial effort to search for all relevant research? No – only PubMed and a couple of trials registers were searched, and cross-referenced to the company's internal literature library. However, given the orphan nature of the drug indication and it being a first-in-class drug it is unlikely that there would any other relevant studies that the company would not be aware of (see ERG report section 3.1.1).
3. Is the validity of included studies adequately assessed? Yes – the criteria in the NICE HST template for company submissions is used, but only for some of the studies. The RCTs CUV029, CUV030 and CUV039 are appraised jointly in a single table (CS Table C7, p41). However, the RCT CUV017 and small study CUV010 are not appraised at all (CUV010 is wrongly included in the summary table for RCTs (Table C5, p 24) - it is a single arm study with a very small number of patients and it does not contribute data to the economic model). A critical appraisal of the observational study by Biolcati et al[11] is given in Table C8 (p42) but the level of detail is very superficial and many of the items declared as not applicable. The ERG asked the company to provide full quality assessments of these studies (clarification response question A1, 02/10/17) but the company declined to do so, stating that this was not appropriate as they were not traditional RCT design studies. The ERG considers that all studies should undergo critical appraisal, regardless of design, using appropriate criteria.
4. Is sufficient detail of the individual studies presented? No – Full results for the clinical studies are not given – only selected outcomes. For example, EPP-QoI scores, which were collected in trials CUV029, CUV030, CUV039 are not given in the CS. The CS does state in various places that due to the requested format of the data, effect size information (contained in study report tables) cannot be provided, and that further information can be provided on request. The ERG requested the full clinical study reports but the company did not supply these stating that they had submitted all data from the studies to the CHMP (clarification response question A3, 26/09/17). However, the ERG does not have access to such data.
5. Are the primary studies summarised appropriately? No – The CS provides a study-by-study description of study characteristics and results, but does not provide a critical summary of the results across the studies (e.g. what the collective evidence is for each outcome in turn). The justification for not doing meta-analysis given is not very clear (see section 3.1.7 above).

The CS does not state that the review of the clinical effectiveness literature was systematic (there are no instances of the term 'systematic review'). The review stated the inclusion criteria and undertook critical appraisal of some but not all of the included studies. As stated earlier, a limited number of databases were searched for clinical effectiveness studies, however, it is unlikely that there would be any studies that the company is not aware of. The level of detail provided on the characteristics and results of the studies provided is limited, and there is no overall systematic critical summary of the clinical effectiveness of afamelanotide for EPP.

3.3 Summary of submitted evidence

The following sub-sections provide the results of the clinical effectiveness review for each of the outcomes included in the decision problem, as collated by the ERG from the CS and, where necessary, from the study journal publications and the EPAR.

3.3.1 Voluntary natural light exposure results

Outcomes relating to the duration of tolerance to light exposure are reported in the CS for four studies (CUV017, CUV029, CUV030 and CUV039), with the most detailed data being reported for study CUV039. Further detailed results for light tolerance outcomes are given by Langendonk et al.[7] for studies CUV029 and CUV039 and in the afamelanotide EPAR[2] for study CUV039. The results for studies CUV029 and CUV039 drawn together from the CS, journal publication and EPAR are shown in Table 6. Only brief results for studies CUV017 and CUV030 are available and these are summarised in the text below and in Table 7. These results include the primary outcomes of the trials, though light exposure data is not used as an input parameter in the company's economic model.

Table 6 Duration of tolerance to sunlight in studies CUV029 and CUV039 (Diary Card population)

Outcome[a]	**Study CUV029 (Europe)**		**Study CUV039 (USA)**	
	Afamelanotide N=38	**Placebo N=36**	**Afamelanotide N=46**	**Placebo N=43**
Total hours in study in direct sunlight with no pain, mean per patient ± SD; median per patient (range)	**Daily assessment 10:00-15:00 (5h) (co-primary outcome)**		**Daily assessment 10:00-15:00 (5h) (secondary outcome)**	
	20.4 ± 40.5; 6.0 (0-193)	5.6 ± 9.3; 0.8 (0-35)	71.2 ± 89.2; 39.6 (0-419)	41.6 ± 45.3; 31.8 (0-199)
	Difference between groups p=0.005 CS states p=0.006		*Difference between groups 13.1 hours (95% CI -1.3 to 28.0); p=0.092*[b]	

	Daily assessment 10:00-20:00 (10h) (secondary outcome)		**Daily assessment 10:00-18:00 (8h) (primary outcome)**	
	Not reported; [redacted]	Not reported; [redacted]	115.6 ± 140.6; 69.4 (0-651)	60.6 ± 60.6; 40.8 (0-224)
	Difference between groups p=0.007		*Median difference between groups 24 hours (95% CI 0.3 to 50.3); p=0.044*	
Total hours in study in direct sunlight with no pain or with mild pain, mean ± SD; median (range)	**Daily assessment 10:00-15:00 (5h) (co-primary outcome)**		**Daily assessment 10:00-18:00 (8h) (secondary outcome)**	
	Not reported; [redacted]	Not reported; [redacted]	141.1 ± 165.1; 80.0 (0.5-825)	74.6 ± 67.5; 51.0 (1.25-251)
	Difference between groups p=0.043			
	Daily assessment 10:00-20:00 (10h) (secondary outcome)			
	Not reported; [redacted]	Not reported; [redacted]		
	Difference between groups p=0.026		*Median difference between groups 26.8 hours (95% CI -0.3 to 57.5); p=0.053*	
Total hours in study in direct sunlight regardless of pain score, mean ± SD; median (range)	Not reported	Not reported	**Daily assessment 10:00-18:00 (8h) (secondary outcome)**	
			145.0 ± 164.1; 83.5 (0.5-825)[b]	81.8 ± 71.2; 65.3 (3.5-278.5)[b]
			Difference between groups 26.1 hours (95% CI -2.3 to 57.3); p=0.066[b]	
Total days in study "in some direct sunlight" on days with no pain, mean ± SD; median (range)	Not reported	Not reported	**Daily assessment 10:00-18:00 (8h) (secondary outcome)[c]**	
			80.5 ± 48.9; 85.5 (0-167)	51 0.7 ± 37.3;[d] 54.0 (0-124)
			Difference between groups 29 days p=0.005[e]	
Total days in study "with some sunlight" on days with no pain or mild pain, mean ± SD; median (range)	Not reported	Not reported	**Daily assessment 10:00-18:00 (8h) (secondary outcome)[c]**	
			93.9 ± 51.0; 97.0 (2-185)	64.0 ± 40.6; 61.0 (3-145)
			Mean difference between groups 32.0 days (95% CI 9.0 to 54.0); p=0.004	

[a] The CS and journal study publication are not explicit that the reported sunlight exposure times are cumulative over the full study period; this is clarified in the afamelanotide SmPC (p. 9) and EPAR (p. 58). Time differences between groups are as reported in the CS, journal publication and EPAR and according to the EPAR are based on the Hodges-Lehmann shift estimate. Unless stated, the CS does not specify whether stated differences between groups are medians or means.
[b] sourced from the EPAR (not reported in the CS or study journal publication)
[c] The CS (p. 36) states this was an exploratory analysis
[d] data for mean ± SD are as written in the study journal publication (typographic error)
[e] source: EPAR (p. 70)

In both CUV029 and CUV039 studies patients in the afamelanotide group experienced a greater mean and median total number of hours in direct sunlight with no pain (Likert scale score of 0).

In CUV029 for the primary outcome (sunlight exposure between 10:00 and 15:00 hours) the median number of total hours per patient in direct sunlight with no pain was 6 (range 0-193) compared to 0.8 (0-35) for afamelanotide and placebo groups respectively (p=0.006) after nine months. As a secondary outcome (sunlight exposure between 10:00 and 20:00 hours) the median number of hours in direct sunlight with no pain was [redacted] versus [redacted] respectively (p=0.007). In study CUV039 for the primary outcome (sunlight exposure between 10:00 and 18:00 hours) the median number of hours per patient in direct sunlight with no pain was 69.4 (range 0-651) versus 40.8 (0-224), p=0.04 after six months. For the secondary outcome of sunlight exposure between 10:00 and 15:00 hours the median number of hours in direct sunlight with no pain was 39.6 (range 0-419) versus 31.8 (range 0-199), p=0.09 after six months. The journal publication suggests that the difference between the two trials in sunlight exposure without pain may be in part due to higher latitudes of the European centres compared with the US centres. Thus patients in the US would, on average, have greater potential for sunlight exposure during the year.[7]

Results for the outcome of total hours per patient in direct sunlight with no pain or with mild pain (Likert scale score of 0-3) were also more favourable for afamelanotide than placebo patients, with statistically significant differences between study groups in both studies CUV029 and CUV039 (Table 6). In study CUV039, additional outcomes for sunlight exposure per patient expressed in terms of the total days in sunlight with no pain, or with no pain or mild pain, also favoured afamelanotide over placebo, with the differences being statistically significant, although the ERG is unsure how these outcomes were calculated (see section 3.1.5.1).

The EPAR (p 72)[2] states there were 15 patients, in trial CUV039, who experienced more than 60 minutes of direct sunlight exposure per day, of which 12 were receiving afamelanotide and 3 receiving placebo (i.e. 26% of the afamelanotide group and 7% of the placebo group).

Duration of tolerance of sunlight was a secondary outcome in the crossover study CUV017. The CS states that significantly more sun exposure occurred in patients receiving afamelanotide (p=0.0136), suggesting that afamelanotide facilitated more outdoor activity compared to placebo (CS p 29). The CS mentions (CS p 44) that this analysis refers to the number of days of exposure categorised as <1 hour, 1 to 3 hours, 3 to 6 hours and >6 hours per day, but no further information is given so it is unclear which data comparison the p-value refers to. The ERG requested the CSR for study CUV017 from the company but this was not provided.

According to a company announcement (CLINUVEL 2010[16]) for study CUV017, “Clinically relevant daily exposure of longer than one hour per day symptom-free was recorded by the trial physicians (CRFs) at the end of each 60 day treatment. In assessing the duration of sunlight exposure per patient, there was significantly more sun exposure in patients receiving SCENESSE® ($p<0.0001$).” However, no outcome data are provided and it is unclear which analysis this p-value refers to.

Duration of tolerance to sunlight was a primary outcome in study CUV030. The results as presented in the CS are shown in

Table 7. Patients receiving afamelanotide achieved a significantly greater duration of exposure to direct sunlight during the study without incurring pain than those receiving placebo.

Table 7 Duration of tolerance to sunlight in study CUV030

Outcome	Afamelanotide N=39	Placebo N=38
Total hours of direct sunlight exposure per patient on pain-free days, median (range)	**Daily assessment 10:00 to 15:00 (5h)**	
	8.88 (0-48.3)	0.75 (0-70.3)
	Difference between groups p=0.011	
	Daily assessment 10:00 to 20:00 (10h)	
	16.0 (0-126.3)	1.25 (0-106.3)
	Difference between groups p=0.006	

In summary, the available evidence for EPP patients’ tolerance to direct sunlight based on voluntary exposure in studies CUV017, CUV029, CUV030 and CUV039 consistently demonstrates a favourable effect of afamelanotide over placebo in prolonging patients’ duration of sun exposure. The clinical significance of these findings is difficult to ascertain since there is no universally accepted measure of how much additional sunlight tolerance is beneficial to patients; this is likely to vary on a patient-by-patient basis given the heterogeneous nature of EPP in which some patients are affected more profoundly than others, and patients vary in the extent to which they may need to be outdoors where they are exposed to sunlight.

3.3.2 Photoprovocation results

In study CUV010 (n=5), photoprovocation was carried out before afamelanotide treatment and repeated at days 30, 60, 90 and 120 on the dorsal surface of the hands. The mean photoprovocation response time increased at day 30 to 347%, day 60 to 595%, day 90 to 663% and day 120 to 1077% of that recorded at baseline (CS, p 26). The CS states that except for the most sensitive individual, all patients reached the maximum photoprovocation response time of 15 minutes during some point of the study. These results indicate that afamelanotide improved the patients' tolerance of the artificial light stimulus. However, the CS does not discuss the clinical interpretation of these findings or their generalisability or limitations. A graph of photoprovocation times reported by Harms et al. indicates there was considerable heterogeneity of responses even within the small sample of five patients.[12]

In study CUV029, photoprovocation was assessed in a small subset of patients, however the exact number of patients and the results were not reported.[7] In study CUV030, 15 patients were given provocation on the dorsal surface of the hands and lower back but only six (40%) completed testing which was "attributed to the rigors of the phototesting protocol".[13] Only descriptive results are reported, stating a "positive trend" (not explained) in the first 60 days but lack of a detectable effect at days 90 or 120 when fewer patients were available for testing.[13] For study CUV039, the EPAR notes that the photoprovocation testing subset of patients (n=21) was located at one of the USA study centres.

The study publication by Langendonk et al. provides a table of results for photoprovocation to the dorsum of the hand and the lower back in study CUV039.[7] The results are presented as the change from baseline in minimum symptom dose, expressed in J/cm^2 of light energy and they show that higher doses were tolerated by afamelanotide patients than placebo patients, both on the hand and back, with the differences being statistically significant from 90 days after baseline onwards. However, limitations of these results are that tolerance appeared to be higher in the afamelanotide group than the placebo group at baseline; sample sizes were small (dorsum of hand n=10; lower back n=11) and only limited clinical interpretation of the findings is provided by the study authors.[7] The EPAR notes that due to an error several patients received a lower light exposure dose than intended and this was corrected for using an unexplained 'mathematical adaptation' (not mentioned in the study publication). According to the EPAR, the company observed that the median response to photoprovocation in the afamelanotide group appeared to

follow a cyclical pattern which would be consistent with the expected pattern of change in melanin density, although melanin density was not measured in the study.

Overall, the limited evidence available on photoprovocation indicates that afamelanotide improves patients' tolerance to artificial light in controlled settings but the wider clinical significance of these findings is unclear, and the data are heterogeneous and of uncertain generalisability due to the small sample sizes tested. Photoprovocation data is not used as an input parameter in the company's economic model.

3.3.3 Phototoxic reactions

The CS reports information on phototoxic reactions in two studies (CUV017 and CUV029) principally referring to the frequency or severity of pain experienced. More extensive results for phototoxic outcomes in study CUV029 and also in study CUV039 are reported in the study publication by Langendonk et al.[7] and in the EPAR.[2] Phototoxicity was specified as a secondary outcome in each study. It is not used as an input parameter in the company's economic model.

For the cross-over study CUV017 (CS, pp 29-30) the CS states "the distribution of frequency of days on which patients experienced pain in the various pain severity categories is consistent with the mean scores and was different between the active and placebo groups (p=0.0042)". In CUV017, placebo patients experienced "more moderate and severe pain (p=0.0009)" and "individual daily pain scores" (p=0.0017) were significantly lower following afamelanotide treatment than when patients were receiving placebo. A publication referring to study CUV017 (CLINUVEL 2010[16]) states that "pain scores in patients willing to modify behaviour by continuous exposure to daily (sun)light showed a positive trend toward a reduction in average pain score following active drug treatment (p=0.1654)". These statements are the only information available to the ERG on phototoxicity outcomes in study CUV017.

For study CUV029 the CS tabulates quantitative results for three phototoxicity outcomes (number of phototoxic episodes per subject, overall sum of the severity score per patient, and the overall maximum severity per subject) (CS, p 33). These data are included below in Table 8 and Table 9, together with other phototoxicity outcomes results which are reported by Langendonk et al.[7] and the afamelanotide EPAR.[2]

Table 8 Phototoxic reactions in studies CUV029 and CUV039

Outcome	Study CUV029 (Europe)		Study CUV039 (USA)	
	Afamelanotide N=38	**Placebo N=36**	**Afamelanotide N=46**	**Placebo N=43**
Number of phototoxic episodes per subject, mean ± SD; median (range)	2.0 ± 2.8; 1.0 (0-11)	4.1 ± 5.1; 2.0 (0-20)	2.0 ± 3.3; 1.0 (0-15)	3.3 ± 6.8; 1.0 (0-35)
	Difference p=0.04		*Difference p=0.602*	
Number of phototoxic reactions during study	77	146	Not reported	Not reported
	Difference p=0.04			
Duration of photo-toxic reactions, days, mean ± SD; median (range)	Not reported	Not reported	3.2 ± 6.0; 1.0 (0-34)	6.6 ± 16.8; 1.0 (0-98)
			Difference p=0.50	
Duration of longest phototoxic reaction, days, mean ± SD; median (range)	1.5 ± 1.8; 1.0 (0-7)	3.8 ± 7.4; 2.0 (0-37)	1.3 ± 1.9; 1.0 (0-12) [a]	1.7 ± 2.1; 1.0 (0-10) [a]
	Difference p=0.08		*Difference p=0.519* [a]	
Duration of photo-toxicity, days, mean per patient ± SD; median per patient (range)	3.7 ± 5.6; 1.0 (0-23)	10.0 ± 18.3; 3.0 (0-90)	Not reported	Not reported
	Difference p=0.04			
Sum of Likert score for severity of phototoxic reactions during study, mean per patient ± SD; median per patient (range) [b]	[redacted]	[redacted]	16.3 ± 33.2; 4.0 (0-196)	34.1 ± 86.7; 6.0 (0-507)
	Difference p=0.020		*Difference p=0.44*	
Overall maximum severity per subject (Likert score) across all phototoxic episodes, mean ± SD; median (range)	[redacted]	[redacted]	3.5 ± 3.1; 4.0 (0-8) [a]	3.9 ± 3.3; 5.0 (0-9) [a]
	Difference p=0.010		*Difference p=0.544* [a]	
Patients with severe phototoxic reactions, n (%)	25 (66)	28 (78)	Not reported	Not reported
	Difference p=0.25			

[a] Sourced from the EPAR (not reported in the CS or publication)
[b] The Likert scale ranged from 0 (no pain) to 10 (worst imaginable pain)

Table 9 Pain severity in studies CUV029 and CUV039

Outcome (n= total days recorded in patient diaries)	Study CUV029 (Europe)		Study CUV039 (USA)	
	Afamelanotide n=9742	Placebo n=9601	Afamelanotide n=8055	Placebo n=7368
Number (%) of diary days with no pain (Likert score 0)	8914 (92) [a]	8463 (88)	7156 (89)	6245 (85)
Number (%) of diary days with mild pain (Likert score 1-3)	687 (7)	777 (8)	753 (9)	840 (11)
Number (%) of diary days with moderate pain (Likert score 4-6)	124 (1)	298 (3)	127 (2)	293 (3)
Number (%) of diary days with severe pain (Likert score 7-10)	17 (<1)	63 (<1)	19 (<1)	44 (<1)

[a] p<0.001 for comparison with placebo – other comparisons in the table were not statistically significant

Overall, patients in both study arms had infrequent phototoxic reactions during the studies. In the European study (CUV029), however, the number of phototoxic reactions recorded during the study for those receiving afamelanotide was approximately half that recorded compared to the placebo group (77 vs 146; mean per patient 2.0 ±2.8 vs 4.1 ± 5.1, respectively, p=0.04).[7] In the US study (CUV039), although not statistically significant, phototoxic reactions were slightly higher in the placebo group (46 vs 43; mean per patient 2.0 ± 3.3 vs 3.3 ± 6.8 p=0.60).[7] The company suggested that sun avoidance behaviour in the US trial may have been a contributory factor to the lack of difference in phototoxic reactions between treatment groups in this study (EPAR, pp 68-69[2]).

In addition to the phototoxic reactions reported above, the EPAR provides tables showing the distribution of daily and maximum pain scores calculated post hoc (EPAR, p 73). Given that the planned analyses on phototoxicity outcomes did not identify statistically significant differences, these post hoc data have not been reproduced here.

3.3.4 Melanin density

A change in melanin density (MD) following administration of afamelanotide, although only reported in the CS for study CUV010, was stated to be a secondary endpoint for CUV017.[2] In addition, the observational study by Biolcati et al.[11] also reported this outcome.

The EPAR highlighted that early pharmacokinetic studies demonstrated that both 16 mg and 20 mg doses increased MD (quantified by spectrophotometry) by 33%.[2] However in the crossover trial, CUV017, it was demonstrated that the increase in MD in clinically relevant skin areas was smaller, ranging between 15-20% on the forehead and 6-12% on the cheeks' skin, which indicated a non-homogeneous pigment distribution.[2]

In CUV010, MD measured as a secondary outcome, was seen to increase during the first 30 days after administration at all tested anatomical sites with one exception in one patient (CS, p 26). Further data from this study showed a mean melanin density increase of 124% of the baseline level at day 30, which slightly decreased by day 60 (121%).[10] This study also showed that a rise in MD after the second implant (at day 90) to 130% of initial MD was only slightly higher than at Day 30. The absolute difference in MD between treatment days (measured on days 30, 60, 90 and 120 at 6 anatomical sites) was stated to be significantly different to baseline (P = 0.004) (CS p 26). In addition, three patients with high sunlight exposure had a stronger MD increase at day 120 (1.084–1.824 MD units) than the other two patients (0.085 and 0.765 MD units).[10]

In the long term observational study, MD, measured in the Swiss cohort only, was reported in units (where one MD unit corresponds roughly to the difference in skin colour between two skin types in the Fitzpatrick scale of skin types).[11] The increase in MD is compared to MD before the first exposure to afamelanotide. It was reported that MD rose by about 0.4 units during months 1 and 2 and by about 0.7 units during months 3 and 4. Between the fifth month and the sixth year, MD remained stable between 0.7 and 1.0 units.

Melanin density was not used as an input parameter in the company's economic model.

3.3.5 Health Related Quality of Life (HRQoL)

3.3.5.1 EPP-QoL results

As mentioned earlier, the EPP-QoL instrument was used in studies CUV029, CUV030, CUV039, the long-term observational study by Biolcati et al.[11] and the on-going post authorisation safety study CUV-PASS-001. Pooled EPP-QoL data from studies CUV029, CUV030, CUV039 are used by the company to inform their assessment of cost-effectiveness in their model (discussed further in section 4.3.1.1 of this report). Limited quantitative EPP-QoL data for the respective studies are reported in the CS and the company declined to supply further data requested by the ERG (clarification response question A1, 12/09/17). Quantitative results are available for two of the studies (CUV029, CUV039), which were reported in the trial publication.[7] These are reproduced in Table 10.

The EPP-QoL score ranges from 0 to 100 (transformed from the original scoring scale), with higher scores indicating a better quality of life. The results for study CUV029 are reported as absolute scores at study visits (up to day 270), whilst in study CUV039 they are reported as change from baseline up to day 180, with absolute scores given for day 360 (240 days after the last dose). The baseline EPP-QoL scores differed between the two trials, with lower scores in study CUV039 indicating a study population with a lower HRQoL.

In study CUV029 there was a minor imbalance in scores at baseline between study groups (mean difference of 3.70). In this study the scores increased over time in both study groups, though the increase was higher in the afamelanotide group at all assessment time points, with the highest score around 85 points and with mean differences between groups ranging from around 7.9-15.2 points across the time points. The differences between the groups were statistically significant at days 120, 180, and 240.

Table 10 EPP-QoL results

Trial and questionnaire score	Afamelanotide			Placebo				Difference[a]					
Study CUV029 (Europe)	*Mean*	*SD*	*n*	*Mean*	*SD*	*n*	*P value*	*Mean*	*SD*	*SMD*	*SE*	*95% CI*	
Baseline score at day 0, before dose 1	39.00	25.80	37	35.30	23.70	34	0.39	3.70	24.82	0.15	0.24	-0.32	0.62
Score at day 60, before dose 2	68.00	19.10	37	60.10	22.00	35	0.09	7.90	20.56	0.38	0.24	-0.08	0.85
Score at day 120, before dose 3	78.80	16.20	37	63.60	23.90	35	0.005	15.20	20.31	0.75	0.24	0.27	1.23
Score at day 180, before dose 4	84.60	12.60	35	73.50	24.30	35	0.03	11.10	19.36	0.57	0.24	0.10	1.05
Score at day 240, before dose 5	84.80	10.70	34	73.10	24.10	34	0.01	11.70	18.65	0.63	0.25	0.14	1.11
Score at day 270, final visit	79.70	16.10	32	67.20	25.70	34	0.06	12.50	21.59	0.58	0.25	0.09	1.07
Study CUV039 (USA)[b]	*Mean*	*SD*	*n*	*Mean*	*SD*	*n*	*P value*	*Mean*	*SD*	*SMD*	*SE*	*95% CI*	
Baseline score at day 0, before dose 1	26.6	19.9	47	26.2	19.4	43	NR	0.40	19.66	0.02	0.21	-0.39	0.43
Score at day 60, before dose 2	70.6	24.2	47	49.6	29.8	43	NR	21.00	27.02	0.78	0.22	0.35	1.21
Score at day 120, before dose 3	76.9	22.0	46	55.8	30.2	42	NR	21.10	26.23	0.80	0.22	0.37	1.24
Score at day 180	78.1	24.9	46	63.0	26.2	43	NR	15.10	25.54	0.59	0.22	0.17	1.02
Score at day 360 (follow up visit)	38.4	27.0	44	45.4	29.6	40	NR	-7.00	28.27	-0.25	0.22	-0.68	0.18

[a] Descriptive statistics for the difference between study groups were calculated by the ERG using a published method,[21] [b] results reproduced from the EPAR.[2] SE = standard error; SD = standard deviation; SMD = standardised mean difference; NR = not reported.

Scores in both groups reduced slightly between day 240 and the final visit at day 270. The score improvements observed over time in both the afamelanotide group and placebo groups of study CUV029 would indicate a change from moderate to mild EPP according to the company's EPP-QoL score thresholds (whereby for the purposes of economic modelling the EPP scores are stratified as 'mild' – 66.7 to 100; 'moderate' – 33.4 to 66.6, and severe' – 0 to 33.3 – see section of the CS 10.1.9, p 59). However, caution is advised in this interpretation as these thresholds and any minimal important clinical differences have not been clinically justified by the company.

[redacted]

In study CUV039 scores increased over time from baseline in both groups with larger increases in the afamelanotide group. The highest score was 51.1 points for the afamelanotide group at day 180. Differences between the groups in the change from baseline were statistically significant at day 60, day 120, and day 180. By day 360 (240 days after the last implant) scores had fallen in both study groups illustrating a worsening of HRQoL, though they remained above baseline levels. The score at this time point was slightly higher in the placebo group (mean difference -7 points) suggesting better HRQoL than for afamelanotide patients. This observation is not discussed in the CS or the journal publication.[7]

The CS reports brief results for the untransformed EPP-QoL scores from study CUV039 (CS Table C5, p 38). The total score range is from -10 (best possible HRQoL) to 35 (worst imaginable HRQoL) and therefore the desired scoring direction is the opposite of the transformed scoring version. Median change from baseline for the afamelanotide group was between 1.6 and 1.9 times that of the placebo group, with statistically significant differences in favour of afamelanotide at days 60, 120 and 180 (p values not provided).

Overall the results from studies CUV029 and CUV039 show that HRQoL increases following implant and is maintained over time as implants are replaced every 60 days. However, the clinical significance of the increases observed is unclear as no clinically justified interpretation of changes in EPP-QoL scores is available. Once implants have been withdrawn there is deterioration in HRQoL over time; however, the rate at which HRQoL reduces following implant removal is uncertain and is an issue explored in the ERG cost-effectiveness analysis (section 4.4).

The CS reports brief narrative EPP-QoL results for study CUV030, stating that at each time point the mean change from baseline for the afamelanotide group was approximately twice that of the placebo group ($p<0.05$) (CS, p 35). The ERG notes that it is not possible to know how comparable the study groups were at baseline as the baseline values are not reported.

Long-term EPP-QoL results

The EPP-QoL instrument was also administered to patients in the long-term observational study of 115 patients who received afamelanotide for up to eight years.[11] Patients in the Swiss cohort of this study completed the original version of the questionnaire containing 18 questions (n=161 questionnaires completed). In the Italian cohort patients completed a version with three questions removed (n=460 questionnaires completed). For both cohorts data from the original and revised questionnaires were presented.[11] The mean number of implants per year was 4.4 ± 1.6 in the Swiss cohort and 2.6 ± 1.6 in the Italian cohort. In the Swiss cohort prior to afamelanotide the mean HRQoL score was 32 ± 22% of maximum (revised questionnaire 31 ± 24%). In the first six months of treatment, it rose to 74 ± 17% (74 ± 17%) and remained between 69% and 91% (66% and 84%) of maximum during the HRQoL observation period of six years.

In Italy, questionnaires were not given before afamelanotide was administered; data were available for assessment time points between the second month and the fifth year of treatment. The mean HRQoL score remained stable at between 73% and 80% (revised questionnaire 74% and 80%) of maximum with a slight increase in year five, to 85% (83%). The mean HRQoL treatment scores were stated to be similar between the two cohorts, with larger variation between assessment time points observed in the Swiss cohort. Seasonal variations in EPP QoL scores were also reported. The mean HRQoL score in winter (December to February) was higher (approximately 84%) than during summer (June to August), where it dropped to 75% in July. The difference between the months was statistically significant (P=0.037). It is mentioned that more questionnaires were available for the summer period than the winter period due to more patients requesting implants at that time of year. However, the number of questionnaires analysed from each season is not given. Also, the publication does not state which of the two cohorts these data apply to. Overall, data from this study show that HRQoL increases markedly following afamelanotide administration (as observed from the Swiss cohort) and is maintained over time (observed in both cohorts). However, there were seasonal variations, with HRQoL higher during winter months.

The only EPP-QoL information available for the PASS study is a statement that there was a trend towards improved patient quality of life (CS Table C6, p 40).

3.3.5.2 DLQI results

As stated earlier, the DLQI was administered to patients in the CUV029, CUV030, and CUV039 studies. However, the CS does not report any results for these studies. The ERG requested these data but the company declined to provide them citing their perceived inappropriateness of the DLQI for assessing quality of life in EPP (clarification response question A2, 12/09/17) (see section 3.1.5 of this report for the ERG discussion of the DLQI). The ERG was able to identify DLQI data from the EPAR for study CUV039 (Table 11).

Table 11 DLQI results in study CUV039

		Afamelanotide	**Placebo**
DLQI total score at visit 1 (Day 0)	N	47	43
	Mean (SD)	10.7 (6.3)	10.4 (5.7)
DLQI total score at visit 2 (Day 60)	N	47	43
	Mean (SD)	4.7 (5.7)	6.4 (6.0)
DLQI total score change from baseline at visit 2 (Day 60)	Mean (SD)	-6 (5.9)	-4 (5.5)
	P value	0.214	
DLQI total score at visit 3 (Day 120)	N	46	42
	Mean (SD)	2.8 (4.2)	4.1 (4.8)
DLQI total score change from baseline at visit 3 (Day 120)	Mean (SD)	-7.8 (6)	-6.5 (6.2)
	P value	0.589	
DLQI total score at visit 4 (Day 180)	N	46	43
	Mean (SD)	2.4 (4.2)	3.1 (4.1)
DLQI total score change from baseline at visit 4 (Day 180)	Mean (SD)	-8.1 (6.2)	-7.3 (5.6)
	P value	0.799	

Scale 0 = no effect on QoL, >20 = extremely large effect on QoL.

The DLQI scoring range is 0-30 with a score of 0 indicating no effect on QoL, and a score of 30 indicating an extremely large effect on QoL. DLQI scores between the study groups were comparable at baseline at the mid-point in the scale at around 10.4 to 10.7 out of 30 (scores of 6-10 indicate a moderate effect on a patient's life and scores of 11-20 indicate a very large

effect on a patient's life[22]). Scores declined over time in both groups to a nadir of 2.4 to 3.1 for afamelanotide and placebo respectively at day 180 (a score of between 2 to 5 indicates a small effect on a patient's life[22]). The decline in scores was larger in the afamelanotide group, though differences between the groups in the change from baseline were not statistically significant. The EPAR states that "there were no clinically relevant or statistically significant differences between groups in quality of life at any time point when assessed by the DLQI questionnaire" (p 60). The ERG notes that for general inflammatory skin conditions (e.g. psoriasis, eczema) a change in DLQI score of at least four points is considered clinically important.[23] The largest change observed for afamelanotide was around eight points which is double the recognised minimal clinically important difference for general skin conditions.

3.3.5.3 SF-36 results

The CS reports that the SF-36 instrument was used in study CUV017 but does not provide any quantitative results. The CS states that the baseline SF-36 results were "higher than expected, with the mean across all patients of the eight quality of life scales and the physical and mental component scores being above the population average score of 50" (CS p 29). The suggested explanation in the CS is that patients are likely to have adapted their lives to live with the condition without significantly affecting their HRQoL. The ERG requested SF-36 results from the company but they declined to provide them (clarification response question B1, 12/09/17). The EPAR states that in study CUV017 results "showed no improvement in QoL during and after treatment with Scenesse" (CS p 85) but no further detail is presented.

3.3.6 Adverse events

An overall list of adverse events (AE) that occurred in afamelanotide patients is provided in the CS, as reproduced from the SmPC (CS Table 10, p 48-49). However, the CS does not identify which AE arose in each of the individual included studies, except for providing a list (without numbers) of the most frequent AE that occurred in the small (n=5) study CUV010 (these were: nausea, tiredness and headache within the first 24 hours after the first implantation; CS, p 27). Details of the AE that occurred in studies CUV029 and CUV039 are reported in the study journal publication[7] and are summarised in Table 12.

Table 12 Adverse events for trials CUV029 and CUV039

Type of AE (according to MeDRA (v14.0) Preferred term)	**EU trial (CUV029)**		**USA trial (CUV039)**	
	Afamelanotide N=38	Placebo N=36	Afamelanotide N=48	Placebo N=45
Adverse events that occurred during the study period, n	189	166	272	216
Patients with any adverse event that occurred during study period, n (%)	34 (89)	32 (89)	45 (94)	39 (87)
Serious adverse events, n	1	0	3	2
	Severity of adverse events that occurred during the study period, n (%)			
Mild	19 (50)	17 (47)	17 (35)	14 (31)
Moderate	12 (32)	14 (39)	25 (52)	23 (51)
Severe	3 (8)	1 (3)	3 (6)	2 (4)
	Most frequent adverse events that occurred during the study period, n (%)			
Headache	13 (34)	14 (39)	19 (40)	13 (29)
Neopharyngitis	8 (21)	8 (22)	6 (12)	10 (22)
Nausea	7(18)	6 (17)	9 (19)	8 (18)

MeDRA = Medical Dictionary for Regulatory Activities

Adverse events collated from studies CUV010, CUV017, CUV029, CUV030, and CUV039, are presented in the EPAR (Table 8, p 92)[2] and are summarised in Table 13. The combined study results (which include 231 patients) reveal that the five most common AE were nausea, headache, migraine, nasopharyngitis and back pain.

Table 13 The five most common adverse events in studies CUV010, CUV017, CUV029, CUV030, and CUV039 (reproduced from EPAR Table 8, p 92)

Type of AE (according to MeDRA (v14.0) Preferred term)	MeDRA (v14.0) System/Organ Class	Number of patients (number of events)					
		Afamelanotide n=231			Placebo n=220		
		Total	Related to drug study	Not related to drug study	Total	Related to drug study	Not related to drug study
Headache	Nervous system disorders	87[a] (259)	54 (161)	46 (98)	75[a] (251)	39 (116)	50 (135)
Nausea	Gastrointestinal disorders	60[a] (106)	53 (93)	11 (13)	36[a] (54)	25 (31)	17 (23)
Nasopharyngitis	Infections and disorders	41 (46)	0 (0)	41 (46)	36 (43)	0 (0)	36 (43)
Back pain	Musculoskeletal and connective tissue disorders	23 (34)	4 (4)	19 (30)	21[a] (43)	4 (7)	18 (36)
Migraine	Nervous system disorders	13[a] (38)	6 (22)	8 (16)	15 (32)	4 (8)	11 (24)

[a] numbers do not sum to the specified total number of patients (data are as reported in the EPAR)
MeDRA = Medical Dictionary for Regulatory Activities

The ERG notes that slightly different adverse events are listed in the CS under “Interpretation of clinical evidence” (CS, p 52). The CS states “The adverse events occurring in 12% of all EPP patients consist mostly of 1. transient headaches (first 48 hours); 2: nausea; 3: gastrointestinal discomfort (infrequent); 4: transient darkening of the epidermis” (CS, p 52).

Longer term data on adverse events

Although the first dataset from the ongoing safety PASS study is still to be reported, longer-term data from the two longest treatment programmes (8 years), operating at EPP expert centres in Switzerland and Italy, have been presented by Biolcati et al.[11] This study, which reports on a total of 115 EPP patients (treated with 1023 implants) revealed that the most frequent adverse events (treatment related and unrelated) were nausea, headache, administration site conditions and fatigue (CS, p 51). Within this study, it was highlighted that two patients noted the appearance of new melanocytic naevus, appearing 2.5 and 5 years after the first dose of afamelanotide. One of them was removed and showed no signs of malignancy.[11]

Serious adverse events

In total, 31 serious adverse events were reported with afamelanotide in the clinical trial programme (EPAR summary of the five clinical trials as above), all of which were considered unlikely or definitely not related to study drug (EPAR, p 93).[2] Early data from the PASS study (23 June 2016 – 31 May 2017) identified four serious adverse events, of which three were unrelated to treatment (CS, p 40).

3.3.7 Mortality

The CS does not report mortality. For trials CUV029 and CUV039 the study journal publication states there was no mortality.[7] The EPAR (p 93) states "Four deaths were reported during clinical studies with the afamelanotide implant, all of which were regarded as definitely not related to study treatment by the investigators," although the EPAR is not explicit about which studies are being referred to.[2] For the long-term observational study the publication by Biolcati et al. [11] states that one patient died of heart failure, but does not specify whether this was treatment-related.

3.3.8 Sub-group analyses results

The NICE scope and company's decision problem do not specify any subgroups to be included. Some of the company's analyses involved subgroups of the randomised population (e.g. where tolerance to light exposure was analysed according to different pain severity subgroups) and these are considered above.

3.3.9 Mixed treatment comparison results

The company did not conduct a mixed treatment comparison. The ERG considers this appropriate, given that the NICE scope specifies the comparison should be between afamelanotide and best supportive care. Insufficient evidence is available to form a network to support such a comparison. Accordingly, the CS focuses on studies that directly compared afamelanotide against placebo (which is a proxy for best supportive care).

3.4 Summary of clinical effectiveness

The CS presents an evidence review of four RCTs and three observational studies of afamelanotide, most of which were sponsored by the company. The decision problem as defined by the company is consistent with the NICE scope of the appraisal. Although the searches for evidence were limited to a small number of databases, due to the orphan nature of the drug and the rarity of the condition it is unlikely that any additional relevant studies have not been included.

Some of the afamelanotide clinical effectiveness studies remain unpublished and limited detail on these and also on the published studies is provided in the CS. Clinical study reports and study protocols for all studies have not been made available to the ERG and therefore a full independent assessment of the methodological characteristics and results of the studies has not been possible for this appraisal. Although the company has conducted placebo-controlled RCTs, in such a poorly understood rare condition the ERG has concerns about the methodological quality and potential risk of bias of the studies. It is not possible to ascertain whether randomisation was adequately concealed and whether study arms in all trials were balanced at baseline. In one of the studies (CUV029) there were twice the number of patients with Fitzpatrick skin type 1 in the placebo group compared to the treatment group. The significance of this is not discussed in the CS. Unblinding is known to have occurred in some patients, yet the impact of this on the results is uncertain. The ERG also notes that the EMA expressed concerns about the conduct of two of the RCTs and only one of them (CUV039 conducted in the USA) was considered of sufficient validity to support the marketing authorisation. The ERG's quality assessment of this RCT identified potential risks of bias in this study (as in the other studies), but given its status as the pivotal trial the ERG has used it to inform it's cost-effectiveness analysis (see section 4 of this report).

The available evidence shows that afamelanotide is associated with clinical effectiveness benefits, in terms of increasing the amount of time patients can spend in sunlight without incurring pain, or incurring only mild pain; a reduction in phototoxic episodes; and a statistically significant reduction in duration of phototoxic episodes (the latter observed in CUV029 but not in CUV039). Adverse events were generally mild in severity. Statistically significant improvements are reported in the HRQoL measurements, although the clinical significance of this is unclear. The instrument used (EPP-QoL) has been designed specifically to measure the impact on EPP,

with highly specific questions about impact of the condition on ability to undertake daily activities inside, around and outside the home, choice of clothing, and mode of transport outside. [redacted] and it does not include a question about pain, which is one of the most debilitating aspects of the condition. This is an important consideration as EPP-QoL results are the sole outcome from the clinical effectiveness studies that directly informs the company's cost-effectiveness analysis.

The ERG suggest that caution is exercised in the interpretation of the results of the clinical effectiveness studies for the reasons stated above.

4 COST EFFECTIVENESS

4.1 Overview of company's approach to economic evaluation

The company conducted a review of published economic evaluations (CS section 11, pp 62 to 63), but did not find any relevant studies. However, the ERG search identified one relevant study in a published conference abstract,[3] which we describe in section 4.2 below (p 66). The company produced a model-based economic evaluation comparing afamelanotide to standard care in adults with EPP, using Disability Adjusted Life Years (DALYs) as the measure of benefit (CS section 12, pp 64 - 80). They argued that a Quality Adjusted Life Years (QALY)-based model would be inappropriate for EPP. We describe and critique the company's approach to economic evaluation in section 4.3 below (p 69). Additional ERG analyses are presented in section 4.4 (p 91), including: a simple adapted version of the company's base case model with QALYs as the measure of benefit; an ERG base case analysis with QALYs; and exploration of uncertainty around the company and ERG base cases, with probabilistic sensitivity analysis (PSA) and deterministic sensitivity and scenario analyses.

4.2 Description and critique of company review of economic evaluations

The company identified a published systematic review of economic evaluations of ultra-orphan drugs with marketing authorisation in Europe, published in 2015 by Schuller et al.[24] This review did not identify any economic evaluations of EPP. The company included terms to identify economic evaluations in their PubMed search (see section 3.1.1 above, p 22), but reported that this did not identify any economic evaluations. However, the ERG search for additional evidence found an abstract published in 2016 by Thompson et al.[3] which we consider relevant to this appraisal.

The abstract by Thompson and colleagues reported a cost-effectiveness analysis of afamelanotide for EPP that was presented at the ISPOR 21st Annual International Meeting, held in Washington in May 2016, with authors from ICON, a consultancy based in the UK and an author from CLINUVEL.

*[redacted]

The abstract reported on an economic model that appears to be very similar to the model submitted to NICE, with both sharing the following characteristics:

[redacted]

- Levels of EPP symptoms categorised as mild, moderate or severe
- Proportion of patients by level of severity based on trial quality of life scores
- Disability Adjusted Life Years (DALYs) were the primary measure of benefit

But, unlike the company submission, Thompson et al. also presented a sensitivity analysis using QALYs derived from 'preliminary SF-36 data from early clinical trials' and from other 'similar' conditions.

Broadly, one might think of one DALY averted (a year of life adjusted for the level of disability experienced during that year) as similar to a one QALY gained (a year of life adjusted for the level of quality of life experienced during that year). QALYs are calculated as the area under a weighted survival curve and DALYs as the area above a similar curve. Thus, one wants to maximise QALYs and minimise DALYs. There are, however, differences in the conceptualisation of the weighting factors (disability versus health-related quality of life) and in the methods by which these weights are obtained. See section 4.4.1.1 for more formal definitions and discussion of the differences and relative merits of QALYs and DALYs.

Results from the Thompson et al. DALY model are summarised in Table 14. They reported a base case estimate of 1.87 DALYs averted over a lifetime (discounted) with afamelanotide compared with standard care, with a range from 0.72 to 2.50 in sensitivity analysis with alternative sources for DALY weights.

[redacted] The Thompson et al. base case incremental cost-effectiveness ratio (ICER) was £373,000 per DALY averted, which was higher than that reported in the company submission: £278,471 per DALY averted (CS Table D9, p 82).

[redacted]. Thompson et al.[3] stated that "the model showed sensitivity to the number and cost of each dose". We further note that the ICER in the Thompson et al.[3] model must also have been sensitive to the source of disability weights, as illustrated in Table 14.

14 Base case DALY model results from Thompson et al. abstract

Afamelanotide vs. standard care	Base case	Lower limit for DALYs	Upper limit for DALYs
DALYs averted	1.87	0.72	2.50
Incremental cost *	£697,510 *	£697,510 *	£697,510 *
ICER: £ per DALY averted	£373,000	£968,764 *	£279,004 *

* Figures inferred by ERG from results reported by Thompson et al.[3]

Thompson et al. cited an ICER of £401,000 per QALY gained from a sensitivity analysis using the condition hereditary angioedema (swelling under the skin) as a proxy for EPP, and a range from £208,000 to £1.1 million per QALY in sensitivity analyses using alternative sources for utility weights. We note that, assuming the same incremental cost as in the Thompson et al.[3] DALY analysis, these cited ICERs suggest a base case discounted lifetime gain of 1.7 QALYs, with a range from 0.6 to 3.4 QALYs. This illustrates that DALYs averted are of a similar magnitude to QALYs gained, but that they cannot be assumed to be equal.

At the clarification stage of the HST appraisal process, additional information was requested on the methods, parameters and results of the Thompson et al. model. The company declined to provide this information, arguing that:

[redacted]

[redacted]"

(Company response to clarification question B1, p 6, 12/9/17)

The ERG disagrees with this position. We believe that QALYs are a conceptually appropriate metric for quantifying the value of health effects of afamelanotide for patients with EPP, as for other lifelong and chronic disabling conditions; that satisfactory methods for estimating QALY gain are available; and that these methods, though not perfect, are superior to the methods used by the company to estimate DALYs averted. We present this case in section 4.4.1.1. Further, we note that the HST committee does need to make a judgement about the plausible range of incremental cost per QALY gained to assess whether afamelanotide for EPP represents good value to the NHS, in relation to other uses of NHS funds and measured in a way that is consistent with other NICE health technology assessments. We therefore highlight the above QALY-based ICER estimates and present our own estimates and exploration of uncertainty around them in section 4.4.

4.3 Description and critique of the company's economic evaluation

4.3.1 NICE reference case

The ERG assessment of whether the submitted economic evaluation met the NICE Reference Case requirements is presented in Table 15. As the company did not present cost effectiveness using incremental cost per QALY, they failed to comply with the NICE Reference Case,[25] the interim methods guide for HSTs,[26] or the final scope for this appraisal.[27]

Table 15 NICE reference case requirements

NICE reference case requirements:	Included in submission	Comment
Decision problem: As per the scope developed by NICE	Yes	
Comparator: As listed in the scope developed by NICE	Yes	
Perspective on costs: NHS and PSS	Yes	
Evidence on resource use and costs: Costs should relate to NHS and PSS resources and should be valued using the prices relevant to the NHS and PSS	Yes	Estimate of societal costs presented as sensitivity analysis.
Perspective on outcomes: All direct health effects, whether for patients or, when relevant, carers	No	The outcome measure used in model (12 item version of EPP-QOL) does not include all direct health effects for patients (no direct questions on distress, anxiety or impact on work).

Type of economic evaluation: Cost utility analysis with fully incremental analysis	No	The economic evaluation uses DALYs, which are not utilities.
Synthesis of evidence on outcomes: Based on a systematic review	No	The review reported in the CS is not described as a systematic review. However, it is unlikely that there would be any studies that the company is not aware of.
Time horizon: Long enough to reflect all important differences in costs or outcomes between the technologies being compared	Yes	Whilst a full lifetime horizon is not adopted, sensitivity analyses extending the horizon have no effect on ICERs.
Measuring and valuing health effects: Health effect should be expressed in QALYs. The EQ-5D is the preferred measure of health related quality of life.	No	The company used DALYs as the primary measure of benefit.
Source of data for measurement of health related quality of life: Reported directly by patients and/or carers.	Yes	EPP-QOL used in the submission to define severity of disease was derived from patients
Source of preference data: Representative sample of the UK population	No	DALY weights not derived from a representative UK sample.
Equity considerations: An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit.	Yes	There are no QALYs, but the DALYs are the same weight regardless of other characteristics.
Discount rate: 3.5% pa for costs and health effects	Yes	
Notes: ? = uncertain; N/A=not applicable		

4.3.2 Model structure and assumptions

The company model is described on pages 64 to 81 of the CS, with further discussion of how health effects were measured and valued on pages 57 to 61 of the CS. The model was designed to estimate the cost-effectiveness of treatment with afamelanotide compared with a standard treatment control for adult patients with EPP. It addresses the decision problem specified in the scope, with the exception of the measure of value for money: the model estimates incremental cost per DALY avoided, rather than the incremental cost per QALY gained expected by NICE.[27] The company stated that the rationale for this decision was **"**the extreme paucity of robust utility data**"** and **"**the fact that a cost per DALY framework provides a better fit for the condition and treatment provided**"** (CS, pp 64-65). As stated above, we disagree with this conclusion.

[redacted]

The model entails a number of key assumptions:

Survival

[redacted]. The company states that with the exception of liver failure (estimated to affect 2-5% of patients), EPP has no known effect on life expectancy (CS, p 66). The ERG has not identified any evidence to contradict this claim. Available evidence from afamelanotide trials and observational studies does not suggest any impact on mortality (section 3.3.7 of this report, p 64).

Starting age and time horizon

[redacted] Although not a lifetime horizon, the company correctly explains that the ICER is independent of the time horizon, given their assumptions that cost and disability effects are constant over time and that treatment does not affect survival (CS, p 69). The company demonstrated this by conducting a scenario analysis with a starting age of 18 and a time horizon of 60 years (CS, p 80).

Change in costs or effects with age

[redacted] This might seem strong, but we have not identified any evidence of changes in quality of life, effectiveness or costs with age or years of treatment. Holme et al. did not find a relationship between quality of life (measured by DLQI) and age in 176 adults with EPP.[17] In the longitudinal study of 115 EPP patients in Italy and Switzerland treated with afamelanotide for up to eight years, it was found that mean quality of life (measured by EPP-QoL) was stable after the first year of treatment.[11]

Treatment compliance and continuation

Non-compliance or discontinuation of treatment is not explicitly modelled. It is not clear whether the effectiveness estimates used in the model (CS Table C12, p 59) implicitly account for non-compliance in the clinical trials by including all randomised participants, regardless of whether or not they had an implant or, if so, how many (see discussion of ITT analysis in section 3.1.6) above). The mean number of implants per patient per year assumed for costing purposes does seem to allow for 'real-life' non-compliance, as it is based on an average of expanded access and commercial distribution (CS, p 66).

Looking over a longer period, the model assumes that treatment continues throughout the modelled time horizon, with the same mean number of implants per patient and the same effectiveness estimates every year over the [redacted] time horizon. Evidence on long-term trends is inevitably limited, but what there is suggests that most patients will continue to ask for implants as reported within the observational data on 115 patients in Italy and Switzerland, of whom around three quarters were continuously treated for 6 to 8 years.[11] Of those who discontinued, half stopped in the first year and 90% within three years. A more interesting issue from an economic perspective is whether patients who experience limited benefit from implants stop having them. If so, this would suggest that the real-life cost-effectiveness might be better than that estimated by the model. One of the clinical experts who we consulted has suggested that patients who do not feel that they are benefiting from afamelanotide might well decide not to continue, due to the need for travel and discomfort and inconvenience of having the implants. However, in the absence of an objective measure of response it would be difficult to define an explicit stopping rule.

Adverse effects

The company model does not include any additional disability, mortality risk or healthcare cost to reflect the impact of adverse reactions to afamelanotide (CS, p 70 and p 78). This is reasonable, given the current evidence from the clinical trial programme and observational cohorts, where reported adverse events were mild in severity (transient nausea, headache, administration site conditions and fatigue (CS, p 51. See section 3.3.6 of this report for a summary of adverse events).

In summary, the ERG agrees that the basic model structure, although simple, is appropriate for evaluation of afamelanotide in DALY terms. With simple adjustments, it can be adapted to estimate QALYs (see section 4.4). However, the robustness of both DALY and QALY versions of the model depends on how the average annual disability/ utility losses and net healthcare costs are estimated.

4.3.3 Model parameters

The company model has four sets of input parameters, described in the following sections:

- **Disability weights**: 0 to 1 index for mild, moderate and severe disease (a higher number represents greater disability). The weights were assumed to be equal with and without treatment and constant over time.
- **Disease severity:** proportions of patients with mild, moderate and severe EPP. This distribution differed between treatment arms – reflecting the effectiveness of afamelanotide at reducing severity compared with usual care - but was assumed to be constant over time.
- **Mortality rates**: annual probabilities of death by age, [redacted]
- **Resource use and costs**: healthcare costs calculated from the mean number of implants per year and drug acquisition, administration and monitoring costs. Costs of EPP-related productivity were also estimated and included in a scenario analysis.

4.3.3.1 Disability weights

The company's base case analysis uses disability weights from the World Health Organisation Global Burden of Disease (GBD) study conducted in 2010, reported by Salomon et al.[28]. This was a large international survey to elicit judgments from the general public about health losses

associated with multiple causes of disease and injury. The survey did not include EPP, or the company's preferred proxy condition of [redacted]. So instead, the company used a proxy of [redacted] in their base case analysis (Table 16).

Table 16 Disability weights (from CS Table C11, p 58)

Severity	[redacted]	[redacted]
Mild	[redacted]	[redacted]
Moderate	[redacted]	[redacted]
Severe	[redacted]	[redacted]

1 Salomon et al. (2012)[28]
[redacted]

The company explained their choice of proxy on page 58 of the CS. They stated that the effects of EPP are similar to [redacted], one of the DSM-IV class of [redacted], defined as: [redacted]". The company argued that, although the reasons differ, the behavioural impact of EPP can be likened to that of [redacted]. Fear of painful phototoxic reactions has a psychological impact and people with EPP learn to avoid high light exposure at an early age, hence restricting their ability to participate in a variety of social, occupational and other activities. We recognise this as a description of the psychological and functional impacts of EPP,[17 30-32] and acknowledge some similarity with the functional impacts of [redacted]. However, an obvious difference is the direct effect of phototoxic reactions for people with EPP (section 3.3.3, p 52).

However, even if the nature of the effects of a proxy condition were broadly similar to those of EPP, it does not mean that the magnitude of effects or definitions of severity are comparable. The GBD 2010 disability weights for mild, moderate and severe [redacted] were elicited using short lay descriptions provided to respondents (Table 17). We consider below whether these are descriptions are compatible with the definitions of severity used to analyse the afamelanotide clinical trial data.

Table 17 Lay descriptions of [redacted] (GBD 2010)[28]

[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]

The company used an alternative proxy of [redacted] in a scenario analysis, with disability weights reported by [redacted] (Table 16).[redacted] The company did not explain the reasons for choosing [redacted] as an alternative proxy, although they stated that in their clinical research, people with EPP had been likened to people suffering with [redacted] (CLINUVEL data on file). The ERG cannot judge the validity of this claim. We note, however, that the 'disability weights' in the [redacted] are actually utility decrements that could have been used to calculate QALYs: they were derived from SF-6D scores (utilities) from a sample of 71 adult men with [redacted] minus the mean SF-6D score for males in the general population (population norms) in the same age group.

4.3.3.2 Treatment effects

The company used individual EPP-QoL data from studies CUV029, CUV030 and CUV39 to estimate the proportions of patients in the intervention and control groups with mild, moderate and severe disease at baseline and at 120 days (CS Table C12, p 59). Levels of severity were defined by an equal division of the 0 to 100 EPP-QoL scale: 'severe' (0 to 33.3); 'moderate' (33.4 to 66.6); and 'mild' (66.7 to 100). The EPP-QoL severity distributions and mean disability weights by treatment and time point (for [redacted] proxies) are shown in Table 18 below. The model actually only makes use of the 120 day results and assumes that these values apply for the whole year. Thus in the base case model (with the [redacted] proxy),

[redacted] DALYs are lost per year of life under standard care and [redacted] with afamelanotide: [redacted] DALYs are assumed to be avoided per person per year of treatment.

Table 18 EPP-QoL categories and disability weights by treatment

		Baseline		120 days	
		Afamelanotide	**Standard care**	**Afamelanotide**	**Standard care**
Proportions by EPP-QoL severity[a]					
'Mild'	0 to 33.3	[redacted]	[redacted]	[redacted]	[redacted]
'Moderate'	33.4 to 66.6	[redacted]	[redacted]	[redacted]	[redacted]
'Severe'	66.7 to 100	[redacted]	[redacted]	[redacted]	[redacted]
Mean disability weights[b]					
[redacted] proxy		[redacted]	[redacted]	[redacted]	[redacted]
[redacted] proxy		[redacted]	[redacted]	[redacted]	[redacted]

[a] Distribution of EPP-QOL scores by thirds of scale, CUV029, CUV030 & CUV039 (CS Table C12 p 59)
[b] Mean disability weights calculated from EPP-QOL distribution and proxy weights (CS Table C11 p 58)

The ERG has the following serious concerns about the source of these effectiveness estimates and the way in which they are used in the model:

Choice of outcome measure: 12 item version of EPP-QoL

The company describe their rationale for developing the EPP-QoL on pages 58-59 of the CS. They argue that other quality of life measures in their trial programme (the generic SF-36 and dermatology-specific DLQI) had proved inadequate to reflect the "humanistic burden of EPP", and so they undertook development of a new EPP-specific measure, in consultation with a number of clinical experts (CS, p 58 and p 71). Methods used in this development process are not reported: for example, it is unclear how items were generated, tested and selected for inclusion in the questionnaire. Biolcati et al. mention three versions of the EPP-QoL, containing 18, 15 and 12 questions.[11] They state that the latter was developed following a psychometric validation study by Oxford Outcomes. The results of this validation study have not been reported. The EMA stated that the clinical research organisation "were not able to fully validate the questionnaire but did review the scoring algorithm" (EPAR, p 64 [2]). In response to a clarification question, the company [redacted] (clarification response 12/9/17, question A1). The clinical trials used to inform the

model (CUV029, CUV030 and CUV039)

[redacted]

(CS, p 71).

Given the lack of information about the development and validation of the EPP-QoL, [redacted], the ERG has serious concerns about use of the EPP-QoL to drive the economic model. The DLQI has undergone extensive validation, we believe that it has face validity for use in EPP and that it has been shown to reflect marked impairment in quality of life for people with EPP.[17] See section 4.4.1.2 for further discussion about the relative merits of the EPP-QoL and DLQI for use in the economic model.

Definition of disease severity

The CS did not address the clinical relevance of defining disease severity by thirds of the EPP-QoL scale. We acknowledge the lack of accepted definitions of disease severity for the EPP-QoL, but note that the validity of the DALY estimates does depend on using compatible definitions of severity for the disability weights (Table 17) and for clinical outcome data (Table 18). Thus, for example, we need to know whether the scale of psychological and functional impact for patients scoring between 66.7 and 100 on the EPP-QoL scale is similar in severity to the GBD description

[redacted]

". The company has stated that [redacted] (CLINUVEL clarification response 12/09/17, question B2). But this contention is not supported by evidence. The ERG therefore concludes that it is uncertain if the disability weights used in the company model are consistent with the outcome data used in the model.

Use of data from CUV029 and CUV030

The company has provided limited information about the methods and results of studies CUV029 and CUV030. We have not had access to their study protocols, statistical analysis plans or clinical study reports (section 3.1.3, p 23 above). Although selected results from CUV029 were reported by Langendonk et al.[7], the protocol and analysis plan were not included in the online appendices. We also note that following GCP inspection of CUV029 and CUV030, the EMA concluded that they could not be relied on for the benefit-risk assessment (EPAR, p 39 and pp 83-84,[2]). The EMA used CUV039 as the pivotal study, to provide evidence of efficacy

and detailed methods and results are available for that trial in the EPAR document in addition to the 2015 Langendonk publication.[2 7] We therefore believe that it would be more robust to use results from CUV039 alone to inform the cost-effectiveness model.

Statistical analysis

The CS provides little information about the methods of statistical analysis used to derive the effectiveness estimates for the model (CS Table C12, p 59). It is simply stated that "The individual patient data for EPP-QoL scores was provided and the baseline/120-day data were used to stratify the results into three EPP-QoL groups". Thus we do not know whether ITT datasets were used, and if so what definition of ITT was employed (see section 3.1.6, p 40). We do not know the number of patients from each of the three studies included in the analysis and so it is not possible to estimate confidence intervals around the proportions cited. It is also unclear how the data from the three trials were pooled. In particular, it is unclear whether the method of analysis correctly reflected clustering of patients within trials, using a two-step or one-step approach suitable for ordinal data.[33 34] This is potentially important, given heterogeneity in study location and possibly patient characteristics (section 3.1.3 above). The company did not explain these issues in response to clarification questions (response to clarification questions 26/09/17 Question A5).

Extrapolation over time

Finally, we believe that the company's use of a single time point (120 days) to represent disease severity over a whole year is simplistic and likely to have biased DALY estimates. It ignores the following features of the data:

- There was a degree of imbalance between the trial arms at baseline, with a greater proportion of control patients in the severe EPP-QoL state at the start of the trials than afamelanotide patients: [redacted] vs [redacted] respectively (Table 18). We cannot assess whether this difference was statistically significant, but note that a small imbalance in disability at baseline can be amplified as DALYs are extrapolated over a long time horizon. As there was no correction for baseline severity in the model, this may have introduced bias in favour of afamelanotide.
- The company stated that they used day 120 as the follow up point because this was the longest follow-up interval available in all trials. However, it appears from the summary of included studies in section 9.4 of the CS that EPP-QoL was also collected at 180 days in all three trials; CUV029 (p 33), CUV030 (p 35) and CUV039 (p 38). We cannot assess

the effect of using 120 days rather than 180 days for the economic analysis, as we do not have the 180 day results. However, we note that for both CUV029 and CUV039, the largest between-arm difference in mean EPP-QoL was observed at four months.[2 7] This can be seen in Figure 1 below (top panel).

- A large reduction in severity was evident between baseline and 120 days in the control group as well as in the intervention group (see Table 18). For comparison, the mean EPP-QoL results for all time points in studies CUV029 and CUV039 are shown in Figure 1 below.[2 7] This shows a pattern of improvement in both groups over the first 6-8 months, followed by a return close to baseline by 12 months in CUV039. The reasons for the initial improvement in the control group might be related to a placebo effect (although some degree of unblinding was likely in these studies); improved monitoring and standard treatment for all trial participants; seasonal effects (recruitment occurred in May and June in the US CUV039 study); and/or a 'regression to the mean' effect (if patients were more likely to consult a specialist and hence be recruited to a trial, at times when their quality of life was worse than usual).
- Whatever the cause of these trends, it does not appear that the four-month snapshot of quality of life is representative of the whole year. We conclude that the company's analysis is likely to have overestimated the benefit of treatment whether quantified in DALY or QALY terms.

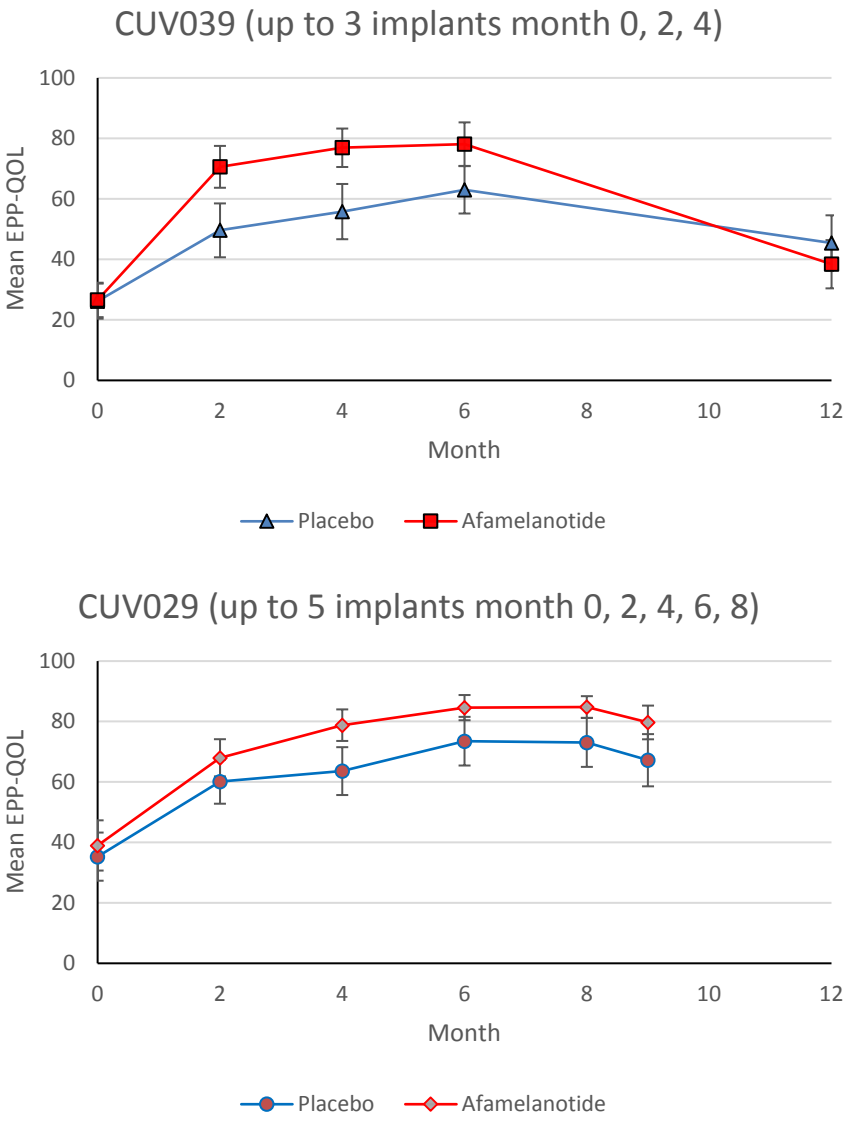

Figure 1 Mean EPP-QoL for studies CUV029 and CUV039

Source: CUV029 data from Langendonk et al. 2015 (Table 4, p 56)[7]. CUV039 data from EPAR (Table 23, p 64)[2], Error bars show 95% confidence interval estimated by ERG, using large sample method based on reported numbers of observations and standard deviations.

4.3.3.3 Mortality rates

Annual probability of death by age for both treatment groups was taken from UK National Life Tables (ONS) (www.ons.gov.uk/ons/taxonomy/index.html?nscl=Life+Tables#tab-data-tables). It appears that 2010-12 figures were used, rather than the most recent estimates based on data from years 2012 to 2014. Rates were averaged for males and females, assuming a 50:50 gender mix at all ages. This is not realistic, but will not affect the cost-effectiveness results.

4.3.3.4 Resource use and costs

The company model includes costs for drug acquisition and administration, laboratory tests, and follow-up appointments with the patient's care team. A list of unit costs is provided in Table D3 (CS p 74). Table D4 (CS, p 75) summarises assumptions about staff time for implant injection visits. Other assumptions that govern resource use that were not reported in the CS were derived from the model. We present a summary of annual resource use and costs using the company base case assumption of [redacted] implants per year in Table 19. Note that the estimated costs of implants and administration do not accord with those reported in Table D6 of the company submission (CS, p 77).

Table 19 Summary of annual costs

Resource	Unit cost	Quantity per year			Cost per year	
		Afamelanotide	Standard care		Afamelanotide	Standard care
Medication						
Implant	£12,020	[redacted]	-		[redacted]	-
Vitamin D & calcium	£0.04	365	365		£15	£15
beta-carotene	£0.05	0	0		£0	£0
					[redacted]	£15
Administration						
Implant injection	£203.75	[redacted]	-		[redacted]	-
Final visit of year	£136.25	1	-		£136	-
					[redacted]	£0
Laboratory tests						
ETP	£2	2	1		£4	£2
Plasma porphyrin	£2	2	1		£4	£2
CBC	£2	2	1		£4	£2
Ferritin	£2	2	1		£4	£2
Liver functioning	£1	2	1		£2	£1
					£18	£9
Follow up						
Dermatology screen	£170	2	1		£340	£170
Photoprovocation	£135	1	1		£135	£135
					£475	£305
					[redacted]	**£329**

ETP - erythrocyte total protoporphyrin; CBC – complete blood count

Number of implants

The cost of treatment is largely governed by the mean number of implants per patient per year. The SmPC recommends three implants per year, up to a maximum of four. The costs in the company base case analysis are based on a mean of [redacted] implants per patient per year (CS Table D5, p 76), with the proportions of patients receiving zero to four implants cited as 'CLINUVEL data on file'. In response to a clarification question, the company explained that their estimates are based on 'real world' use: [redacted] (clarification response, 02/10/17). For cross validation, we checked the number of implants that the company model reported from the long-term follow-up study by Biolcati et al. (Table 20).[11] However, it should be noted that the Swiss centre allowed patients to have up to six implants per year, which is more than the capped value of four in the SmPC.

Table 20 Number of afamelanotide implants per year

Number of implants	N	Mean	SD
Company base case		[redacted]	
Swiss centre[a]	53	[redacted]	[redacted]
Italian centre[a]	120	[redacted]	[redacted]
Weighted average (Swiss & Italian)[a]	173	[redacted]	

[a] Biolcati et al. 2015

The mean number of implants can be changed in the company model. This changes the cost outputs but not the estimated effects. Thus the model generates the same number of DALYs avoided per patient treated, regardless of how many implants those patients were assumed to be using. In reality, treatment effectiveness is likely to be tied to the number of implants a patient receives. We note that if the mean number of implants costed in the company base case model ([redacted]) is not commensurate with average use underlying the effectiveness evidence, the results will be biased. CUV030 and CUV039 allowed up to three implants for patients in the intervention group, and CUV029 up to five implants. However, the mean number of implants per patient used in these three trials is not publicly available, and not reported in the CS. The CS includes a scenario analysis varying the mean number of implants per patient per year (CS Table D15, p 87). This analysis is helpful for understanding how the ICER might have been underestimated, if the company base case estimate of [redacted] implants is less than average use in the clinical trials. In additional ERG analysis, we also explicitly model how effectiveness (QALYs

gained), in addition to cost, is likely to change if the maximum number of implants per patient per year is varied (See section 4.4 below).

Drug acquisition costs

The cost per implant is reported as £12,020. Assuming a mean number of implants of [redacted] per year, this equates to [redacted] per year.

The model also assumed ongoing use of vitamin D and calcium for all patients, whether treated with afamelanotide or not. These costs cancel out of the incremental cost calculations. The company assumed that no patients received beta-carotene in either arm. We were advised by our clinical experts that routine beta-carotene use is uncommon, as it has questionable efficacy and causes orange pigmentation of the skin. Given this and the low cost of beta-carotene, its level of use is not an important issue for this appraisal.

Administration costs

In addition to drug acquisition costs, afamelanotide requires an appointment to inject each implant and a final visit after the last implant of the year. The company used estimates from Erasmus University to quantify the staff time required for each injection visit (CS Table D4, p 75). For each implant injection visit, this included: 15 minutes from the principal physician, 30 minutes from one consultant, 15 minutes from a second consultant and an hour from a nurse. The final visit of the year was assumed to require 15 minutes from the principal physician, 15 minutes each from two consultants, and one hour of nurse time. Based on PSSRU estimates of the cost per hour a medical consultant (£135) and band 5 nurse (£35)[35] and assuming [redacted] implants per year, the company estimates the administration cost of afamelanotide at [redacted] per patient per year. Experts consulted by the ERG believed that the resource use for injection visits may be higher than would be seen in UK practice. Thus the cost of administering the implants might be an overestimate.

Monitoring costs

The company included the cost of two full body skin examinations for patients on afamelanotide, as recommended in the SmPC. They assumed that patients on standard care would have one fully body skin examination per year. Each screening visit was assumed to take one hour of consultant and one hour of nurse time, costing £170 (at Personal Social Services Research Unit (PSSRU) estimates of the cost per hour). Experts consulted by the ERG thought that patients with EPP would not all be having an annual full dermatological scan under current NHS practice. They also suggested that the assumed staff time per visit was excessive.

The company also assumed that patients would have one photoprovocation test per year, whether or not they were using afamelanotide, at a cost of £135 (one hour of consultant time). Experts consulted by the ERG questioned whether this was necessary or acceptable.

Laboratory resource use and costs consisted of the following tests: erythrocyte total protoporphyrin (ETP), plasma porphyrin, complete blood count (CBC), ferritin, and liver functioning. The company assumed that under current practice patients have one of each test per year, and that with afamelanotide two tests per year would be needed. Costs for these tests were derived from NHS Reference Costs, in line with NICE guidance.

Costs of implementation

Conditions of marketing specify that the company should provide an educational training package for physicians, comprising face to face training material, educational video, SmPC and registry information sheet.

Productivity

The base case analysis only includes NHS costs. But the company highlights that EPP has an effect on employment, choice of profession, productivity and earnings: "a proportion of EPP patients is known to be unemployed, others are limited in their productivity, some have full employment, and others have taken up nocturnal employment" (CS, p 80)[17 31 32]. They explored the possible societal costs of EPP and assumptions about how they might be alleviated in a scenario analysis.

4.3.4 Cost effectiveness results

Results from the company model are presented section 12.5.1, of the CS (pp 81-82). For the base case analysis an incremental cost per DALY avoided of £278,471 is reported (see Table 21). The company notes that the incremental cost of [redacted] is largely driven by the cost of the afamelanotide implant, assuming a mean use of [redacted] implants per person per year and a cost per implant of £12,020. They note that the other costs included in the model have a small impact on total costs. The incremental benefit was [redacted] DALYs averted.

Table 21 Base case cost effectiveness results

	Discounted costs			Discounted DALYs		
Afamelanotide		[redacted]			[redacted]	
Standard care		[redacted]			[redacted]	

Incremental	[redacted]	[redacted]
ICER	**£278,471 per DALY averted**	

4.3.5 Assessment of uncertainty

The approach to sensitivity analysis is described in section 12.4 (pp 79-81) of the CS, and the results are reported in section 12.5.11 (pp 86-88). The company reported on four deterministic sensitivity analyses, which we discuss below, changing:

- the disability weights;
- the starting age and time horizon;
- the number of implants per patient per year that are costed; and
- the perspective, from NHS to societal.

The CS does not include a probabilistic sensitivity analysis (PSA).

This represents a very limited exploration of uncertainty. In particular, the CS does not present any sensitivity analysis over the parameters that reflect treatment effectiveness in the model or the methods and assumptions used to derive them. As discussed in section 4.3.2.2 (p 75) above, we believe that there is substantial uncertainty over the robustness of these parameters and assumptions.

4.3.5.1 Disability weights

The base case used a proxy of [redacted], with disability weights from the GBD 2010 survey (Table 22). The company tested two variations on this analysis, in which the disability weight for mild disease was changed from 0.03 in the base case, to 0.04 (Scenario 1) and 0.02 (Scenario 2). In each case, the ratios of the weights for moderate to mild disease (4.97) and severe to mild disease (3.51) were fixed at the base case values. A third scenario tested the effect of using the [redacted] estimates of disability weights for [redacted].[29]

Table 22 Scenario analysis: disability weights

Disability weight	[redacted]			[redacted]
	Base case	**Scenario 1**	**Scenario 2**	

Mild	0.030	[redacted]	[redacted]	[redacted]
Moderate	0.149	[redacted]	[redacted]	[redacted]
Severe	0.523	[redacted]	[redacted]	[redacted]
Incremental cost	[redacted]	[redacted]	[redacted]	[redacted]
Incremental DALYs	[redacted]	[redacted]	[redacted]	[redacted]
ICER (£ per DALY averted)	£278,471	£208,854	£417,707	£727,143

This analysis illustrates the sensitivity of the ICER to changes in the disability weights. Using weights for the proxy condition of [redacted] has the largest impact, because the gradient of the weights between mild, moderate and severe disease is less steep. Thus the benefit of reducing the proportion of patients with moderate or severe disease with the use of afamelanotide is lower. We note that the [redacted] estimates for [redacted] were derived from SF-6D scores, so are really utility decrements. This means that the ICER for the [redacted] scenario (£727,143) can be interpreted as an incremental cost per QALY gained.

4.3.5.2 Starting age and time horizon

The base case is for a cohort modelled from the age of [redacted]). The company presented a scenario analysis for a younger cohort, from the age of 18 to 78 years (60 year horizon).

Table 23 Scenario analysis: age and time horizon

	Base case (age 38 to 73)	Age 18 to 78
Incremental cost	[redacted]	[redacted]
Incremental DALYs	[redacted]	[redacted]
ICER (£ per DALY averted)	£278,471	£278,471

Although the incremental costs and DALYs are higher in this scenario than in the base case, reflecting the longer time horizon, the ICER is unchanged. As noted in the CS (p 69), the insensitivity of the ICER to the time horizon is a necessary result of model assumptions:

[redacted] These assumptions also mean that the ICER is insensitive to starting age. In reality there may be differences in the effects or costs of treatment at different ages, for example if younger patients are better able to make changes to their lifestyle. But there is insufficient evidence to model such possible effects.

4.3.5.3 Number of afamelanotide implants

The base case is costed, assuming a mean of [redacted] implants per person per year. The company tested the effect of changing this assumption: assuming a mean of 3 or 4 implants per year. The company notes that as the change is modelled, it only changes the cost of treatment, not the treatment effect. Thus with more implants per person, the incremental cost and ICER are higher (and conversely, with fewer implants the incremental cost and ICER are lower).

Table 24 Scenario analysis: number of implants

	Base case: [redacted] implants per year	3 implants per year	4 implants per year
Incremental cost	[redacted]	[redacted]	[redacted]
Incremental DALYs	[redacted]	[redacted]	[redacted]
ICER (£ per DALY averted)	£278,471	£378,561	£503,672

We interpret this analysis as demonstrating uncertainty over the ICER related to potential over or under-estimation of the mean number of implants that were associated with the clinical effectiveness results used to drive the model (from studies CUV029, CUV030 and CUV039).

4.3.5.4 Inclusion of societal costs

The base case analysis is conducted from an NHS perspective. In this scenario analysis, the company explored the possible effect of afamelanotide on earnings for people with EPP. This was based on the following assumptions:

- Mean weekly wage: £518 (source cited as a website that was not available)
- Retirement age of 62 (OECD)
- Proportion of mean wage without treatment: 50% (assumption)
- Proportion of mean wage with treatment: 67% year 1; 83% year 2 and 100% year 3+ (assumption).

Table 25 Scenario analysis: societal costs

	Base case: no societal costs	Scenario analysis: increase from 50% to 100% of mean wage over 3 years
Incremental cost	[redacted]	[redacted]
Incremental DALYs	[redacted]	[redacted]
ICER (£ per DALY averted)	£278,471	£172,302

Alternative assumptions about the gap in mean weekly earnings with and without treatment were also tested (see Table D15, CS p 87). However, these were not explained and we were unable to replicate them.

We acknowledge the occupational effects of EPP and their importance to patients and their families. Estimates of productivity costs are not usually taken into consideration in NICE appraisals, but they can be presented alongside a reference case analysis when appropriate. We note the high degree of uncertainty over the company's scenario analysis. Evidence of improving employment, productivity or earnings is not available from the clinical trial programme or long term follow up, up to eight years in the analysis presented by Biolcati et al.[11] Although they do present anecdotal evidence, citing cases where individuals reported being able to take up educational and occupational activities that they did not think they could do without treatment.[11] (see section 6 for the ERG's discussion of the impact of afamelanotide beyond direct health benefits).

4.3.6 Model validation

4.3.6.1 Internal consistency

The company states that internal validation of the model was conducted by a senior health economist not involved in the initial model build (CS section 12.7.1, p 89). No further information is given about how this validation was conducted.

The ERG conducted a series of checks on the model:

- We checked that all of the input parameters in the model were consistent with the numbers cited in the CS and also in the root source of evidence when possible.
- We visually checked the formulae throughout the model, to ensure that they were correctly connected to input parameters and the chain of calculations through the model.
- We replicated some aspects of the model, including the disability weight and cost calculations.
- We tested the reproducibility of the analyses reported in the CS, including the base case and sensitivity/scenario analyses.
- We ran a series of 'stress tests' on the model, checking that changes to input parameters had the expected results.

These tests did not identify any serious data entry or coding errors and we believe the model to be internally consistent. One small rounding error meant that the proportions of patients with

mild, moderate and severe disease in the company's model did not sum to 100% (at baseline for the afamelanotide arm, and 120 days for standard care, see Table C12 p59 CS). We corrected this in our additional analysis by rounding up the proportions of patients assumed to have mild disease. This led to a very small change to the company's base case ICER: from £278,471 per DALY averted to £278,386 per DALY averted.

4.3.6.2 External consistency

The company stated that "to our knowledge, this is the first economic evaluation of EPP attempted, therefore it was not possible to validate to external evidence sources." (CS page 89).

As stated earlier, the ERG identified a published abstract that reported some results from a model of afamelanotide for the treatment of EPP in adult patients (Thompson et al).[3]

[redacted] Other than this abstract we have not identified any other models or evidence sources that would provide a means of external validation.

4.3.7 Summary of ERG critique of company model

We consider that the structure of the submitted model is appropriate. It entails some strong simplifying assumptions:

[redacted] This is reasonable given current evidence.

However, we do have serious concerns about the way in which effectiveness was estimated and valued in the form of DALYs:

- There is insufficient information about the development and validation process of the EPP-QoL scale. It also appears that the items and scoring system may have been revised after initial analysis of trial results, which introduces risk of bias.

- The definition of mild, moderate and severe disease by division of the EPP-QoL scale into thirds is arbitrary and we cannot assess if it is consistent with the disability weights attached to these levels of severity in the DALY calculations.
- We do not know if [redacted] is an appropriate proxy condition for EPP. There are similarities in some of the psychological and functional impacts, but it is not clear if the magnitude and levels of severity are comparable. The same applies to the alternative proxy condition of [redacted]
- We have insufficient access to information about the EPP-QoL methods and results of studies CUV029 and CUV30 to be able to assess their quality or check the results.
- We also have insufficient information about how the results of the three trials, CUV029, CUV030 and CUV039 were analysed and pooled. There is a lack of clarity over whether ITT datasets were used, the number of patients included from each trial and whether the method of pooling accounted for clustering or randomisation.
- Results from a single time point (120 days) were used to estimate DALYs incurred over the whole year. The company stated that they chose 120 days as this was the longest time point available from all three trials, but the CS indicated that EPP-QoL data was also collected at 180 days. Results from two trials, to which we had access, suggest that the choice of 120 days rather than 180 days would have favoured afamelanotide as there was a larger difference between groups at that time. The use of a single time point also ignored information about how EPP-QoL changed during follow-up and failed to correct for baseline imbalance in EPP-QoL severity, which would have favoured afamelanotide.

We also have some questions about the cost estimates used. These were very largely driven by an assumption about the mean number of implants per person per year. This figure was estimated from ‘real world’ data, and it is not clear whether this was consistent with use in the clinical trials. If not, this would be a source of bias.

Finally, we note that the analysis of uncertainty presented in the CS was inadequate. In particular, there was no attempt to estimate the extent or consequences of uncertainty over the effectiveness parameters and assumptions. Given the discussion above, we think this could be considerable. There was also no PSA.

4.4 Additional work undertaken by the ERG

4.4.1 Overview and rationale for additional analyses

We developed two alternative versions of the company model as platforms to explore alternative assumptions and parameter uncertainty:

- A simple QALY version of the company model, applying utility estimates for mild moderate and severe disease for the company's proxy of [redacted] (section 4.4.2.1)
- An ERG base case analysis, in which we estimated QALYs from mean DLQI results at 0, 60, 120 and 180 days from CUV039 mapped to EQ-5D scores (section 4.4.2.2)

The key features of these analyses are summarised in Table 26, with further discussion below.

Table 26 Key features of company base case and ERG models

	Company base case	Simple QALY version	ERG base case
Value for money	Incremental cost per DALY averted	Incremental cost per QALY gained	
Source of clinical data	CUV029, CUV030 and CUV039 (method of pooling not specified)	No change	CUV039
Outcome measure	EPP-QoL 12 item	No change	DLQI
Effectiveness statistics	Proportion of sample by thirds of EPP-QoL scale at 120 days: intervention and control groups	No change	Between-group differences in mean change from baseline DLQI at 60, 120 and 180 days
Method of extrapolation	Assumed fixed within year and between years	No change	Standard care modelled assuming linear change between observations, with return to baseline at 12 months. For afamelanotide we assumed: linear onset of benefit over two months after the first implant of the year and linear loss of benefit over 2 months after last implant of year. Assumptions tested in scenario analysis.

Valuation	Disability weights from GBD 2010 for proxy of [redacted][28]	Utilities assumed as 1-GBD disability weights and scenario with utilities for proxy [redacted]	Utilities mapped from DLQI to EQ-5D from registry data for moderate to severe psoriasis [37]
Mean implant use	[redacted] per person per year (not related to effectiveness)	No change	No change for costing, but effectiveness data based on maximum of 3 implants per year (as in CUV039), and scenarios with up to 2 or 4 implants per year.
Uncertainty	Limited deterministic sensitivity analysis	Additional scenario analysis and deterministic sensitivity analysis, as well as probabilistic analysis	

4.4.1.1 Rationale for use of QALYs

Mathematically, DALYs and QALYs are similar. Both are calculated in relation to a weighted survival curve, with DALYs being the area above the curve (the healthy life that is lost) and QALYs the area below the curve (the imperfect quality life that remains). In economic evaluation, we are interested in the area between two weighted survival curves: one with the intervention of interest and one with an appropriate comparator. This area would be identical for DALYs avoided and QALYs gained, except that the meaning and method of estimation of the weights used to adjust survival differs. For DALYs the construct of interest is 'health loss', whereas for QALYs it is 'welfare loss'.[28] Welfare (or 'utility') is affected by health, but is also subject to other influences. This conceptual difference leads to different methods of eliciting weights. The GBD 2010 disability weights were based on a survey in which respondents were asked to make a series of judgements about which lay descriptions of states they considered to be 'healthier'. In contrast, the weights used to calculate QALYs are derived from trade-off questions designed to elicit preferences, in which the welfare is indirectly elicited by asking what sacrifices people would accept for defined improvements health. For example, the UK tariff for scoring the EQ-5D was based on survey in which respondents were asked how many years of life they would give up to avoid impairments (time trade-off). The scales of measurement do

also differ: DALY weights lie between 0 (no disability) and 1 (maximum disability); while QALY utility weights can be less than 0 if the state is considered worse than death.

The company believes that DALYs are more appropriate than QALYs for quantifying the effects of treatment for people with EPP. They argue that QALYs are conceptually inappropriate because of the way that people with EPP adapt to their condition:

> "Individuals with EPP are left to modify their natural behaviour by leading an indoors-based life deprived or starved of light sources (Lecha et al. 2009), while seeking ways to manage their anxiety of long-lasting burns. As a result, the ability to lead a 'normal' life in the community is severely impacted. Such impacts include choice of education at an early age, social development and interactions, access to further education and ultimately employment (Holme et al. 2006; Biolcati et al. 2015a)." CS, p 65.

Adaptation is common for people with lifelong or chronic conditions and has implications for evaluation of interventions as; patients may rate their pre-treatment health status or value it more highly than might be expected by people without the condition; the response to treatment may be lower or slower than expected, as learned behaviour can be difficult to change. The company suggested that such effects might explain poor results with the generic SF-36 quality of life questionnaire. In study CUV017, participants' SF-36 scores were higher at baseline than expected (higher than population norms) and showed no marked trends over time associated with treatment dose (CS, pp 29-30).

The phenomenon of adaptation and resulting 'disability paradox', have been cited as reasons for preferring an extra-welfarist or non-utilitarian approach (like DALYs) for public policy appraisal.[38] However, there are various possible explanations for adaptation that have different implications for the moral basis of using adapted patients' ratings of quality of life to inform allocation public resources.[39] For example, patients may make a higher than expected assessment of their health state or quality of life because of cognitive denial or lowered expectations, or because of more positive activity adjustment and altered conceptions of health. From an economic point of view, the use of a common metric to value health improvements across different conditions and patient groups is necessary to make judgements about the opportunity cost of new technologies. NICE has reached a considered position that QALYs should be the primary measure of effectiveness for use in economic evaluations.[40] This applies

in the HST programme, so in appraising value for money the committee is expected to consider incremental cost per QALY gained.[26]

In addition to conceptual arguments, the company make a more practical argument that QALYs could not be used because of the lack of robust utility data (CS, p 64). We disagree with this judgement and present two sources of utility estimates below:

- published utility values for the company's chosen proxy of [redacted]; and
- a published equation to map from the DLQI to EQ-5D utilities.

Although less robust than a generic utility instrument, such as the EQ-5D, or direct utility measurement by people with EPP, we believe that mapping from the DLQI is superior to the use of disability weights (or utilities) for proxy conditions. It is illogical to argue that QALYs cannot capture the unique and nuanced effects of EPP and then argue that DALYs from a common illness ([redacted]) can capture the effects of EPP, or that an illness that has very little in terms of symptoms in common ([redacted]) can also adequately capture the disease.

For comparison, we also present a simple QALY version model of the company model (Scenario 1.0), with utilities for mild, moderate and severe disease defined by subtracting the GBD disability weights from 1. This is a simplistic approach, but provides a baseline for comparison of our other analyses.

4.4.1.2 Rationale for use of DLQI

The appropriateness of the DLQI and EPP-QoL questionnaires for EPP is central to the interpretation of the clinical effectiveness and cost-effectiveness evidence. There was a difference in the results of the pivotal study CUV039 with EPP-QoL and DLQI: changes over time and between groups were mostly statistically significant with the former but not the latter (see section 3.3.5 of this report for HRQoL results). This might have been related to the different items included in the questionnaires and/or to their framing. We summarise arguments below.

Face validity of content and framing

See Table 27 for a summary comparison of the content of the DLQI and EPP-QoL (15-item version used in CUV039 and 12-item version used for scoring). For copyright reasons we cannot reproduce the full questionnaires, but they can be downloaded online.

- The DLQI questionnaire is available from the Cardiff University Department of Dermatology website, along with instructions for use and related references: see http://sites.cardiff.ac.uk/dermatology/quality-of-life/dermatology-quality-of-life-index-dlqi/.
- The 15-item version of the EPP-QoL is included in the protocol for study CUV039 published in the online material for the 2015 Langendonk et al. journal paper (http://www.nejm.org/doi/full/10.1056/NEJMoa1411481).[7] The 12-item version is also available in Table S1 in the supplementary appendix to this paper.

The DLQI contains 10 questions on the impact of skin problems over the last week on symptoms, feelings, daily activities, social and leisure activities, work and study, personal relationships and treatment, each measured on a four point scale from 'very much' to 'not at all'. The EPP-QoL has 15 (12) questions about the impact of EPP over the last two months on symptoms, daily activities, social and leisure activities, on a similar four point scale. The wording of several EPP-QoL questions relates specifically to effects on a sunny day and on outdoor activities. It includes additional questions on transport and the ability to be spontaneous, but excludes questions about feelings and personal relationships. Three items were removed in the 12-item version of the questionnaire used to score the study results: frequency of the need to seek out shade or to wear protective clothing; and impact on work or study. Unlike the DLQI, the EPP-QoL includes a direct question on well-being ('much better' to 'worse') and one on improvement in quality of life ('very much' to 'not at all').

The face validity of the two questionnaires and appropriateness for economic evaluation is unclear. The EPP-QoL asks about EPP-specific symptoms, which is important if people with this condition would not recognise the DLQI description of 'itchy, sore, painful or stinging' skin as applying to their symptoms. But the EPP-QoL (12-item version) excludes questions on feelings and ability to work or study, which are important aspects of life. The company argues that anxiety and depression are significant features of EPP, but then omit them from the questionnaire. The EPP-QoL also emphasises the ability to perform outdoor activities on sunny days, but does not measure the relative importance of these activities to the individual. The EPP-QoL does ask directly about well-being and quality of life. But we are concerned about the framing of the quality of life question (Q. 14), which does not allow for the possibility of deterioration. This is likely to have introduced bias. Another important difference between the two questionnaires is the recall period - one week in the DLQI and two months in the EPP-QoL. Again, it is unclear which is more appropriate, as a longer recall period reduces the risk of

missing periods of time when EPP may have had less of an effect on patients' lives, but it does also increase the risk of recall bias.

Table 27 Comparison of questions from DLQI and EPP-QoL

Concepts[a]	DLQI questions[b] **Over the last week, how much has skin affected...**	EPP-QoL questions[c] **Over the last two months, how much has EPP affected...**
Symptoms	Q1. Itchy, sore, painful or stinging	Q5. Frequency at risk of developing EPP symptoms Q13. Frequency of typical EPP skin complaints ~~Q3. Frequency of need to seek out shade~~ [d]
Feelings	Q2. Embarrassed or self conscious	
Daily activities	Q3. Going shopping, looking after home or garden Q4. Clothes you wear	Q10. Going shopping, looking after home or garden on sunny day Q4. Choice of clothes on sunny day ~~Q9. Frequency not wearing protective clothing on sunny day~~ [d] Q15. Transportation method or seating preference
Social and leisure activities	Q5. Social or leisure activitites Q6. Sport	Q6. Social or leisure activities on sunny day Q11. Outdoor social activities with family and friends Q12. Amount of outdoor activities Q7. Need to plan before leaving house Q8. Ability to undertake activities in spontaneous manner
Work and study	Q7. Prevented or problem with work or study	~~Q2. Capacity to go to work or school~~ [d]
Personal relationships	Q8. Problem with partner, close friends or relatives Q9. Sexual difficulties	
Treatment	Q10. Treatment problems, e.g. making home messy or taking time	
Overall		Q1. Well-being Q14. Quality of life

[a] Adapted from key concepts for DLQI from analysis by Ali et al. 2017.[41]

[b] DLQI: http://sites.cardiff.ac.uk/dermatology/quality-of-life/dermatology-quality-of-life-index-dlqi/

[c] EPP-QoL: 15-item version, CLINUVEL Protocol CUV039, Appendix 5 page 51-51. Available as online supplement to Langendonk et al. 2015.[7]

[d] Item deleted in 12-item version of EPP-QOL, Langendonk et al. 2015, Table S1 p5 online supplement.[7]

Responsiveness of DLQI in EPP

Whilst the DLQI has been criticised for inadequately measuring the effect of some skin conditions,[18] the study by Holme et al. (2006), a British study that is the single largest study to measure quality of life in EPP patients, does not support this claim in an EPP population.[17] This study found that DLQI showed marked gradations in the severity of HRQoL for patients with EPP in the UK. Rather than not capturing the severity of the disease, it was the first study to show the difficulties that EPP patients face using a HRQoL instrument, the DLQI. Table 28 shows the distribution of DLQI severity categories from Holme et al.[17] It can be seen that the range of quality of life in UK EPP population ranged from fairly normal quality of life, to severely impaired, with the majority of patients in the 'very large effect' category. Patients with severe DLQI scores have a very poor quality of life.[22]

Since 2007, there have been algorithms available to map DLQI to EQ-5D.[37 41-44] We applied two available algorithms to the Holme et al. results to estimate the distribution of EQ-5D utility scores in this UK EPP population (Table 28). It shows a wide range of utility: from values close to population norms for patients with no or small DLQI effects, to values between 0.3 or 0.4 for patients with severe effects. For context, Table 29 presents utility scores for a range of other disease areas using reputable UK sources.

Table 28 Mapping DLQI to EQ-5D in a UK EPP population

Severity	N[1]	Proportion [a]	Score (assume centre)	EQ-5D[b]	EQ-5D[c]
No effect (DLQI ≤ 1)	6	3.41%	0.5	0.8679	0.9433
Small (DLQI 2-5)	15	8.52%	3.5	0.8091	0.8668
Moderate (DLQI 6-10)	32	18.18%	8.0	0.7209	0.7522
Very large effect (DLQI 11-20)	92	52.27%	15.5	0.5739	0.5611
Severe (DLQI 21-30)	31	17.61%	25.5	0.3779	0.3063
Total	176	100.00%	14.4	0.5962	0.5900
Mean			14.0	0.6033	0.5993
Best possible			0	0.8777	0.9560
Worst possible			30	0.2897	0.1916

[a] N and proportions are derived from Holme et al. (2006), the assumed central points of each severity and the mapping are the work of the ERG[17]

[b] Norlin 2012 (whole population), EQ-5D = 0·8777 – 0·0196 DLQI[37]

[c] Currie & Conway 2006 EQ-5D = 0.956–0.0255 DLQI[43]

Table 29 Comparison of utility scores

Disease	EQ-5D score	Mapping (Yes/No)	Study
Metastatic breast cancer	0.685	Yes	Lidgren 2007[45], NICE TA424
Heart attack/angina	0.628	No	Ara & Brazier[46] [a]
Arthritis / rheumatism / fibrositis	0.597	No	Ara & Brazier[46] [a]
Fabry Disease with ESRD and heart complications	0.584	No	Rombach 2013[47], Migalastat (NICE HST4)

EPP Erythropoeitic protoporphyria; EQ-5D Euroqol five dimensions questionnaire; DLQI Dermatology quality of life index; KDQOL-36 Kidney disease quality of life 36; ESRD End-stage renal disease

[a] Patients with comorbidities

On balance, the ERG considers that the DLQI is a more robust choice for use in the economic evaluation than the EPP-QoL. This judgement is based on the lack of information about the development and validation process for the EPP-QoL. We are also seriously concerned that questions were removed from the EPP-QoL without adequate explanation, and the scoring system may have been revised after initial analysis of trial data, which poses a risk of bias. Despite some criticisms of the unidimensionality of the DLQI and the under-representation of emotional aspects of some skin conditions, it has been extensively studied and evidence for its validity, reliability and responsiveness is available.[18] Further, we consider that the Holme et al. study has shown that the DLQI is capable of detecting the severe impact that EPP has on patients' lives.[17] There are also mapping algorithms that allow estimation of EQ-5D utility values from DLQI scores. In particular, we note that the algorithm developed by Currie and Conway, has been validated in an independent dataset of 3542 people with a range of skin conditions.[41] [43] We use this algorithm in our base case model described in section 4.4.2.2 below.

4.4.1.3 Rationale for use of trial data

The third set of issues that we examine in additional ERG analysis relate to our criticisms of the company's use of trial data, as summarised in section 4.3.6 above.

First, we use our simple QALY version of the company model (Scenario 1.0) to test the impact of adjusting for baseline differences between the study arms and possible attenuation of treatment effects after the last implant of the year. These analyses use the estimated proportions of patients with mild, moderate and severe disease (as defined by thirds of the EPP-QoL scale) at baseline and 120 days pooled data from CUV029, CUV030 and CUV039 (Table C12 p59 CS). The company just used the 120 day results in their analysis, assuming that these values would remain unchanged within and between years. We tested two alternative

scenarios: adjusting the distribution of severity for baseline differences (Scenario 1.1); and assuming a linear loss of the treatment benefit between 180 days and the end of the year (Scenario 1.2).

Our second approach is more of a departure. The ERG base case model (Scenario 2.0) uses effectiveness data from CUV039 only – to address our concerns about the lack of information about the methods and results of trials CUV029 and CUV030 and about the company's methods of pooling data from the three trials. We also change the outcome measure used to drive the model to the mean DLQI mapped to EQ-5D utility values. We present three scenarios modelling alternative assumptions about how the estimated utilities from observed data might change over time: assuming immediate onset of treatment benefit after the first implant of the year (Scenario 2.1); assuming slower loss of treatment benefit after the last implant of the year (Scenario 2.2); and a combination of fast onset and slow loss of treatment effect (attenuation) (Scenario 2.3). In our base case model we assume a maximum of 3 implants per year, to match the effectiveness data from study CUV039. We also present two scenarios modelling changes to the maximum number of implants per year: two implants (Scenario 2.4) and four implants (Scenario 2.5).

We also introduce an exploration of uncertainty over the effectiveness data for both sets of analysis, with deterministic as well as probabilistic sensitivity analysis. Thus each scenario is accompanied by three sets of sensitivity analyses, investigating the effect of uncertainty over the key cost-effectiveness drivers: treatment effectiveness; weights used to adjust life years (disability weights and utilities); and mean utilisation of implants per year, which drives the costs.

4.4.2 ERG methods

4.4.2.1 Simple QALY model

We adapted the company model to calculate QALYs alongside DALYs. The simplest version of this model (Scenario 1.0) uses utilities for mild, moderate and severe EPP estimated from the GBD disability weights for the same proxy as in the company's base case model (utility = 1 – disability weight). This is intended to provide a platform to examine changes to the company's base case model and as a comparison for our preferred model. All parameters were the same as in the company base case (see Table 30).

Table 30 Simple QALY model: Input parameters

Parameters	Standard care	Afamelanotide	Source
Severity at baseline			
Proportion in mild category	[redacted]	[redacted]	CS Table C12 p 59
Proportion in moderate category	[redacted]	[redacted]	CS Table C12 p 59
Proportion in severe category	[redacted]	[redacted]	CS Table C12 p 59
Severity at 120 days			
Proportion in mild category	[redacted]	[redacted]	CS Table C12 p 59
Proportion in moderate category	[redacted]	[redacted]	CS Table C12 p 59
Proportion in severe category	[redacted]	[redacted]	CS Table C12 p 59
Sample size			
Total CUV029, CUV030 & CUV039	119[b]	125[b]	CS p33, p35, p37
Disability weights (GBD 2010)			
[redacted] proxy (mild)	[redacted]		Salomon 2012
[redacted] proxy (moderate)	[redacted]		Salomon 2012
[redacted] proxy (severe)	[redacted]		Salomon 2012
Utility estimates	*Mean (95% CI)*		
Mean EQ-5D mild (intercept)	[redacted]		[redacted]
Decrement for moderate	[redacted]		[redacted]
Decrement for severe	[redacted]		[redacted]
Implant utilisation		*Mean (SE)*	
% of maximum implants per year, used for costing (mean = [redacted])		[redacted]	CS Table C12 p 59

[a] Rounding changed to ensure total sums to 100%

[b] ERG assumption that all patients who received study drug (not ITT) were included in the company analysis of EPP-QoL by severity

We conducted three scenario analyses on this model:

Scenario 1.0: Company base case, adapted to calculate QALYs as well as DALYs

Scenario 1.1: Same as Scenario 1.0, except the mean disability weight per year with afamelanotide was adjusted for the difference in severity (vs. standard care) at baseline.

Scenario 1.2: Same as Scenario 1.1, except the benefit of treatment (mean difference in utility with afamelanotide vs. standard care) was assumed to attenuate after the last implant of the year. We assumed a linear decline between month six and eight.

Scenario 1.3: Same as Scenario 1.0, except utilities for the company proxy condition were taken from a published source. We used estimates for [redacted] The survey included EQ-5D questionnaires, with utilities calculated with the UK Tariff. Regression modelling was used to estimate mean utilities and additional decrements for moderate and severe symptoms.

Figure 2 to Figure 5 illustrate how estimates of mean utility over one year for these four scenarios. Note that for Scenario 1.0 (Figure 2) the mean utilities at month 4 are equal to 1 minus the disability weights in the company's base case model: [redacted] for standard care and [redacted] for afamelanotide (Table C13 p59 CS). The QALY gain per year is calculated as the area between the standard care and afamelanotide curves.

Figure 2 Simple QALY Scenario 1.0: company base case

Figure 3 Simple QALY Scenario 1.1: adjusted for baseline

Figure 4 Simple QALY Scenario 1.2: adjusted for baseline and attenuation

Figure 5 Simple QALY Scenario 1.3: utilities for proxy condition from literature

For sensitivity analysis we varied three sets of parameters:

- **Effects**: Note that sample sizes or measures of uncertainty were not reported around the severity distributions in Table C12 (CS p 59). For PSA, we used Dirichlet distributions for the four severity distributions, based on assumed sample sizes (all patients randomised and treated, as reported in section 9.4.1 of the CS). For deterministic sensitivity analysis we changed the proportion of patients treated with afamelanotide with mild disease at 120 days to between 60% and 90% (holding the ratio of patients with moderate to severe disease and other effectiveness parameters constant).
- **Weights**: For our scenarios with utilities calculated from proxy disability weights (1.0, 1.1 and 1.2), we used the same assumptions as the company (scenario 1 and 2 in Table D15 p87 CS). This entailed changing the disability weight for mild disease from [redacted] in the base case to [redacted] and [redacted], holding the ratios of mild to moderate and moderate to severe weights constant. We did not include the disability weights in the PSA. For Scenario 1.3, we fitted a beta distribution for the utility estimate for mild disease, and gamma distributions to the two decrement parameters. In deterministic sensitivity analysis, we varied the two decrement parameters between lower and upper 95% confidence limits.
- **Implants**: The mean number of implants costed per patient per year was based on the company's assumption ([redacted] per year). To include uncertainty over this parameter in the PSA, we assumed a beta distribution for the proportion of an assumed maximum number of implants that patients would actually receive ([redacted]). A standard error around this mean (0.049) was estimated from the implant utilisation reported for the Italian cohort (n=120) in the Biolcati et al. observational cohort (assumed maximum of three implants per year). For deterministic sensitivity analysis, we varied this parameter between 0.667 and 1.000, yielding a range of between two and four implants per year.

4.4.2.2 ERG preferred model

In our analysis:

- We used mean DLQI results from study CUV039 (at 0, 60, 120 and 180 days).[2]
- We first modelled mean DLQI through the year for the control group, starting from the observed baseline value and using change from baseline values to estimate mean DLQI

at 60, 120 and 180 days.[2] This approach enables correct propagation of uncertainty in the PSA, without treating repeated measures as independent variables.

- Then we modelled the DLQI curve for afamelanotide, using the between-group mean differences in DLQI at each time point. This retains patient randomisation in the trial, and builds in correlations between the control and intervention curves in the PSA.
- Utilities were estimated by mapping from the estimated mean DLQI values at each time point, using the mapping algorithm reported by Currie and Conway 2007.[43]
- We assumed a mean of three implants per person per year in our base case analysis (the maximum for the intervention group in study CUV039 and as recommended in the SmPC).
- We made the same assumptions about percentage utilisation as in the simple QALY model. Thus, we assumed that on average patients would use [redacted] of the maximum permitted number of implants per year, giving a mean of [redacted] implants per year for costing. This provides consistency with the company's assumptions based on 'real life' utilisation rates. We would have preferred to use the utilisation rate from CUV039, the same source as the effectiveness data, but data on the mean number of implants per patient was not available to us.

Input parameters for our preferred model are reported in Table 31 below.

Table 31 ERG preferred model: Input parameters

Parameter	Mean	SE	Source
DLQI standard care (placebo group)			
Baseline: day 0	10.4	0.87	EPAR (Table 18 and 20, pp 61-79)
Change: day 0 to day 60	-4.0	0.84	
Change: day 0 to day 120	-6.5	0.96	
Change: day 0 to day 180	-7.3	0.85	
Treatment effect (afamelanotide vs. placebo)			
Mean difference: day 0 to day 60	-2.0	1.20	EPAR (Table 18 and 20, pp 61-79)
Mean difference: day 0 to day 120	-1.3	1.30	
Mean difference: day 0 to day 180	-0.8	1.25	
EQ-5D mapping			
Maximum utility, DLQI=0 (Intercept)	0.878	0.039	Currie and Conway
Utility loss per unit increase in DLQI (slope)	0.020	0.004	
Implant utilisation			
% of maximum implants per year, used for costing (mean = [redacted])	[redacted]	[redacted]	CS Table C12 p 59

For the base case analysis (**Scenario 2.0**), we made the following assumptions about how utilities would be expected to change between modelled time points:

- **Baseline utility**: both groups were assumed to start with the same utility.
- **Onset of treatment effect**: there is a gradual increase in utility for the afamelanotide group over a two month period after the first implant of the year.
- **Effect with subsequent implants**: the treatment effect changes gradually between subsequent timepoints, with further increases in utility after the second and third implants.
- **Attenuation of treatment effect**: the relative treatment effect (mean difference between arms) gradually declines over a two month period after the last implant of the year (from day 180 to 240). Thus, the estimated utility for afamelanotide and standard care converge over two months.
- **End of year**: We assumed that both groups return to their baseline values at the end of the year, with no persistence of effect between years. This assumption is supported by EPP-QoL data at 360 days in study CUV039, which showed a mean change from baseline that was slightly lower in the afamelanotide group than in the placebo group

(not statistically significant) - see Figure 1 on page 80 above.[2] (We note that DLQI was not collected at 360 days in CUV039).

- This pattern is assumed to repeat in subsequent years, yielding the same mean QALY gain with treatment (vs. standard care) every year over the time horizon.

The resulting estimates of utility over a year are illustrated in Figure 6. Note that the observed datapoints (with adjustment for baseline) are shown with solid circles and squares, and assumed changes between these points by solid lines. The empty points and dotted lines represent ERG assumptions over extrapolation after the last DLQI observations at 6 months.

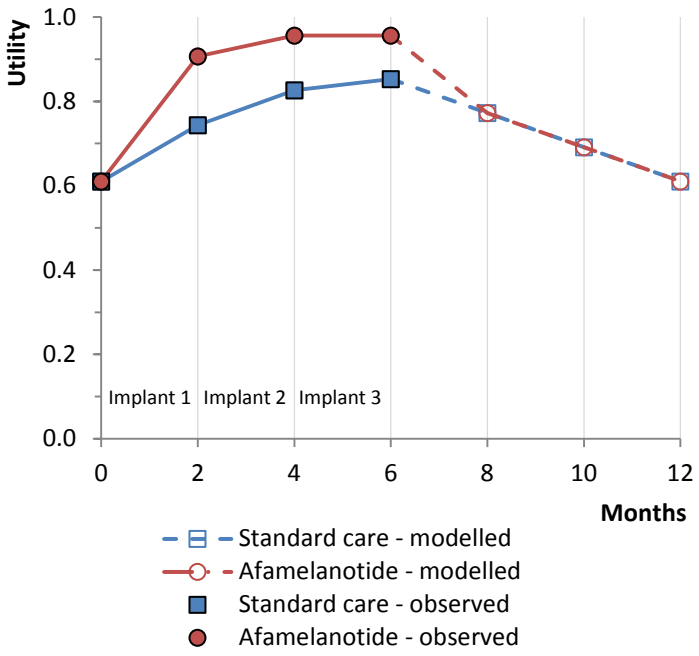

Figure 6 ERG Scenario 2.0: ERG base case

We conducted a set of three scenario analyses on our base case model to explore the effects of different assumptions about the speed of onset of treatment benefits after the first implant of the year and the speed of decline after the last implant of the year:

Scenario 2.1 Fast onset of effect (immediate) after the first implant of the year, with the observed mean difference in DLQI for afamelanotide vs. control at day 60 applied throughout the first two months. The rationale for this scenario is

bioavailability and pharmacodynamics information reported in the EPAR, which showed a peak of melanin density at day 15 (+0.68 from baseline) (CUV028 group 2, Table 4, p 44).[2] We note however that there is uncertainty about the plausibility of this scenario, because of uncertainty over how quickly a change in plasma levels translates to physical protection against light, how that translates to behaviour change (taking the risk of more sun exposure) and subsequently better utility.

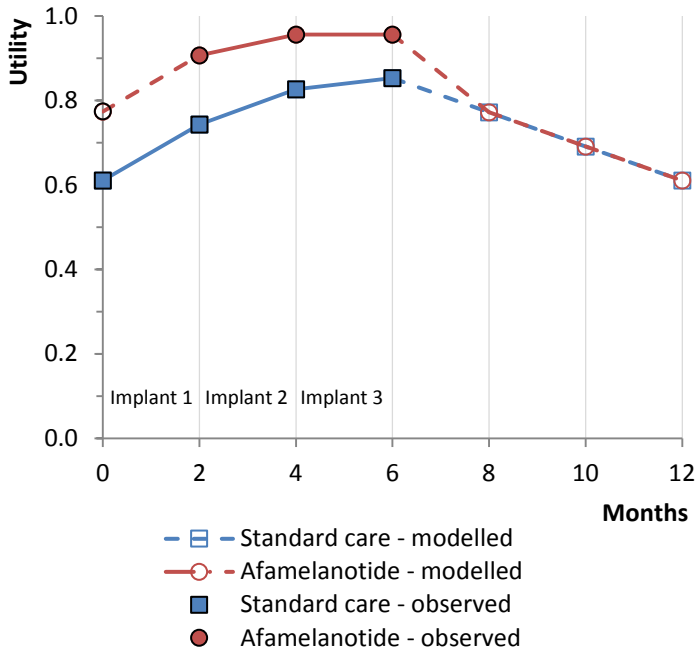

Figure 7 ERG Scenario 2.1: fast onset of effect

Scenario 2.2 Slow attenuation of effect (over six months). This illustrates a slower decline in treatment benefit after the last treatment of the year than in our base case, with a linear loss of the DLQI mean difference over six months (from day 180 to 360). Pharmacodynamic information from the EPAR (Table 4 p 44) shows that mean melanin density was starting to decline by day 60 (+0.38)).[2] Again, there is uncertainty over the plausibility of this scenario.

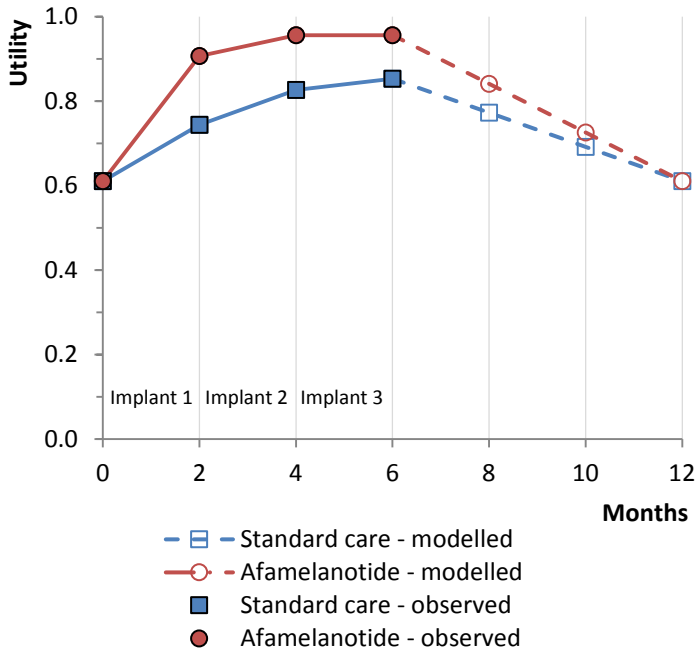

Figure 8 ERG Scenario 2.2: slow attenuation of effect

Scenario 2.3 Fast onset and slow attenuation, combining the assumptions in scenarios 2.1 and 2.2, with an immediate onset of benefit after the first implant of the year and gradual loss of benefit over six months after the last one. This is the most favourable variation on the ERG QALY model that we tested, producing the largest QALY gain (and lowest ICER).

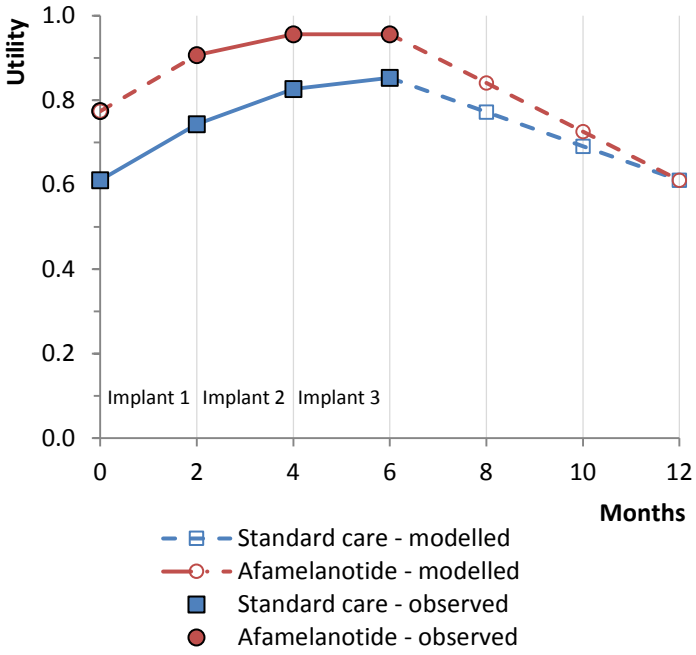

Figure 9 ERG Scenario 2.3: fast onset and slow attenuation

Our final pair of scenarios are designed investigate the impact of changing the maximum number of implants per patient per year:

Scenario 2.4 Assumes a maximum of two implants per year. Similar to our base case, there is a gradual loss of effect after the last implant, with utility in the afamelanotide arm declining to match that in the standard care arm over a two month period. These assumptions reduce the incremental effect, but also the incremental cost. Note that the same assumption about the mean proportion of implants that patients use is the same as in our base case ([redacted]), so a mean of only [redacted] implants is included in the cost calculations for this scenario.

Scenario 2.5 Assumes a maximum of four implants per year. Here we assume that the treatment effect at eight months is the same as that observed at six months, with attenuation of this effect over the next two months. And only

[redacted] ([redacted]) of the maximum four implants are included in the cost calculations.

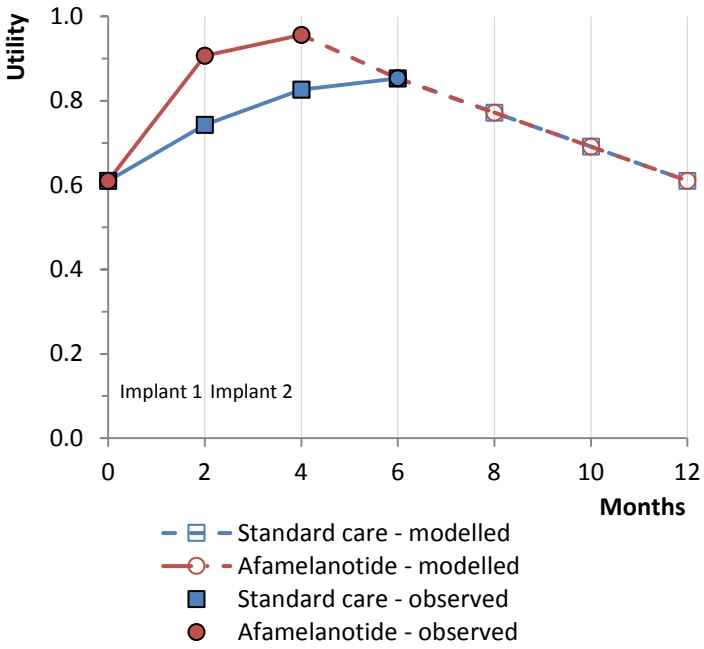

Figure 10 ERG scenario 4: fewer implants (up to 2 per year)

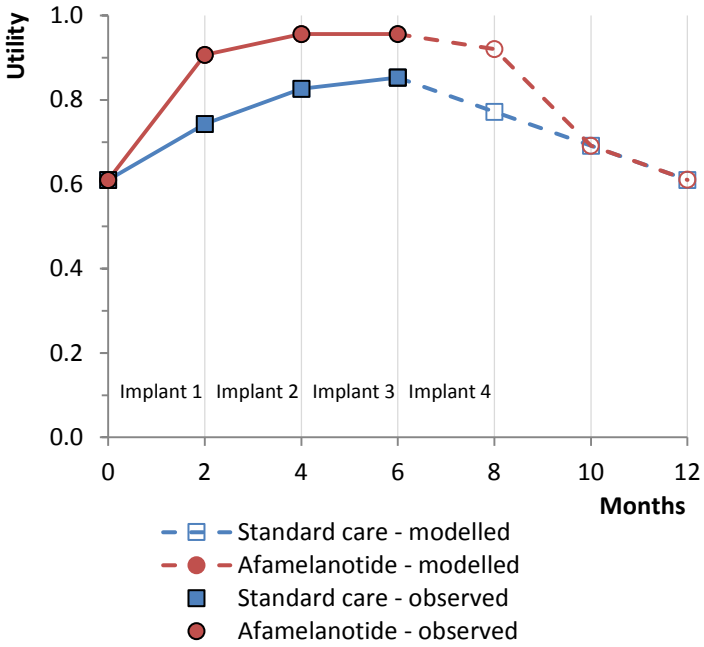

Figure 11 ERG scenario 5: more implants (up to 4 per year)

4.4.3 ERG results

4.4.3.1 Simple QALY model

Results from our simple QALY model are presented in Table 32. The ICER for the simplest QALY adaptation of the company's model is £278,386 per QALY gained. Note that the small difference between this and the company's base case ICER of £278,471 is purely due to the small rounding error in the effectiveness data which we corrected (see 4.3.5.1 above). Otherwise the models are identical. Scenarios 1.1 and 1.2 shows that the company's ICER would have been higher had they adjusted for baseline differences between study arms in EPP-QoL scores, and if they had made assumptions about attenuation of treatment benefit for the part of the year when the patients' did not have implants. The final scenario in this simple model shows that using estimates of utilities from the literature for the company's proxy condition yielded a much smaller QALY gain, and hence higher ICER.

Table 32 Simple QALY model results

Treatment	Cost (£)	QALYs	Incremental costs (£)	Incremental QALYs	ICER (£/QALY)
SCENARIO 1.0: company base case					
Standard care	[redacted]	[redacted]	-	-	-
Afamelanotide	[redacted]	[redacted]	[redacted]	[redacted]	*£278,386*
SCENARIO 1.1: adjustment for baseline					
Standard care	[redacted]	[redacted]	-	-	-
Afamelanotide	[redacted]	[redacted]	[redacted]	[redacted]	*£454,800*
SCENARIO 1.2: adjustment for baseline and attenuation of effect					
Standard care	[redacted]	[redacted]	-	-	-
Afamelanotide	[redacted]	[redacted]	[redacted]	[redacted]	£779,657
SCENARIO 1.3: utilities for proxy condition					
Standard care	[redacted]	[redacted]	-	-	-
Afamelanotide	[redacted]	[redacted]	[redacted]	[redacted]	£1,726,802

The ERG does not believe that any of these scenarios are plausible because they rely on an analysis of trial data that was post hoc and not transparent, the definitions of mild, moderate and severe disease were arbitrary and not related to the levels of severity in the disability weights/ utilities, which were also derived for a non-EPP population ([redacted] proxy).

ERG preferred model

Results for the ERG preferred version of the model are shown in Table 33. It can be seen that our base case was much higher than the company's base case, at £1.6 million per QALY gained. This result was similar to scenario 1.3, which used utility estimates from the literature rather than the simple estimates based on GBD disability weights. The ICERs were lower in scenario analyses exploring the impact of more favourable assumptions about the speed of onset after the first implant of the year and attenuation after the last implant of the year. However, our most favourable scenario (2.3) still yielded an ICER of over £1.1 million per QALY gained. Similarly, the ICER remained high when we modelled changes to the maximum number of implants per patient per year.

Table 33 ERG preferred model results

Treatment	Cost (£)	QALYs	Incremental costs (£)	Incremental QALYs	ICER (£/QALY)
SCENARIO 2.0: ERG base case					
Standard care	[redacted]	[redacted]	-	-	-
Afamelanotide	[redacted]	[redacted]	[redacted]	[redacted]	£1,605,478
SCENARIO 2.1: fast onset					
Standard care	[redacted]	[redacted]	-	-	-
Afamelanotide	[redacted]	[redacted]	[redacted]	[redacted]	£1,290,678
SCENARIO 2.2: slow attenuation					
Standard care	[redacted]	[redacted]	-	-	-
Afamelanotide	[redacted]	[redacted]	[redacted]	[redacted]	£1,343,359
SCENARIO 2.3: fast onset and slow attenuation					
Standard care	[redacted]	[redacted]	-	-	-
Afamelanotide	[redacted]	[redacted]	[redacted]	[redacted]	£1,115,671
SCENARIO 2.4: maximum 2 implants per year					
Standard care	[redacted]	[redacted]	-	-	-
Afamelanotide	[redacted]	[redacted]	[redacted]	[redacted]	£1,337,494
SCENARIO 2.5: maximum 4 implants per year					
Standard care	[redacted]	[redacted]	-	-	-
Afamelanotide	[redacted]	[redacted]	[redacted]	[redacted]	£1,785,957

The ERG believes that this model is preferable to our simple QALY adaptation of the company's DALY model. It relies on published data from the pivotal trial (CUV039) analysed in accordance with a pre-defined plan, and explicitly accounts for changes in quality of life across 12 months,

adjusting for baseline differences and changes under standard care. The utility estimates are derived from quality of life assessments by EPP patients, using a validated mapping algorithm from the DLQI to EQ-5D. There is uncertainty over which method of extrapolating between observed data points is more realistic. However, our scenario analysis demonstrates that the ICERs do not fall below £1,100,000 per QALY.

Deterministic sensitivity analysis

For each scenario, we used deterministic sensitivity analysis to examine the impact of changing three sets of input parameters: treatment effects; the disability or utility weights; and the mean number of implants per year that were costed. The results for the simple QALY model and ERG preferred model are shown in Table 34 and Table 35 respectively. It can be seen that in no case did the ICER fall below £150,000 per QALY.

The deterministic sensitivity analysis results are also shown in the Tornado graphs below Figure 12). These illustrate that the analyses based on GBD disability weights (Scenarios 1.0, 1.1 and 1.2) are much more favourable than those based on utility weights. They also illustrate the very wide range of uncertainty around the ERG preferred model ICERs. This is caused by the small magnitude of the mean differences in DLQI, which yielded very small estimates of incremental QALYs at the lower confidence limits.

Table 34 Simple QALY model: ICERs for lower and upper parameter ranges

Scenario	Effects[a]		GBD disability weight (mild)		Mean implants per year	
	Lower	Upper	Lower	Upper	Lower	Upper
	60.0%	90.0%	0.02	0.04	2	3
1.0	£221,520	£405,664	£208,790	£417,579	£253,371	£378,444
1.1	£320,421	£933,075	£341,100	£682,200	£413,934	£618,266
1.2	£549,293	£1,599,556	£584,743	£1,169,486	£709,600	£1,059,884
	Effects[a]		Disutilities (moderate; severe)[b]		Mean implants per year	
	60.0%	90.0%	(0.021;0.047)	(0.045;0.093)	2	3
1.3	£1,299,022	£2,889,993	£1,249,637	£2,542,183	£1,571,639	£2,347,455

[a] Proportion mild (120 days with treatment)
[b] Disutility vs. mild (moderate; severe)

Table 35 ERG preferred model: ICERs for lower and upper parameter ranges

Scenario	Effects[a]		Utility loss[b]		Mean implants per year	
	lower	Upper	lower	Upper	lower	Upper
	(-0.4;-0.0;-0.0)	(-4.9;-4.8;-4.5)	0.018	0.033	2	3
2.0	£552,284	£17,543,596	£1,198,119	£2,263,826	£1,461,217	£2,182,524
2.1	£457,817	£11,963,277	£963,194	£1,819,939	£1,174,704	£1,754,578
2.2	£438,286	£17,539,848	£1,002,508	£1,894,222	£1,222,651	£1,826,193
2.3	£376,615	£11,961,534	£832,591	£1,573,167	£1,015,422	£1,516,669
	Effects[a]		Utility loss[b]		Mean implants per year	
	(-0.4;-0.0;-0.0)	(-4.9;-4.8;-4.5)	0.018	0.033	1.3	2
2.4	£500,501	£11,766,004	£998,131	£1,885,952	£1,218,005	£1,815,451
	Effects[a]		Utility loss[b]		Mean implants per year	
	(-0.4;-0.0;-0.0)	(-4.9;-4.8;-4.5)	0.018	0.033	2.7	4
2.5	£534,044	£23,318,720	£1,332,805	£2,518,313	£1,625,012	£2,429,736

[a] Mean difference DLQI change (day 60;120;180)
[b] Utility loss per unit increase in DLQI

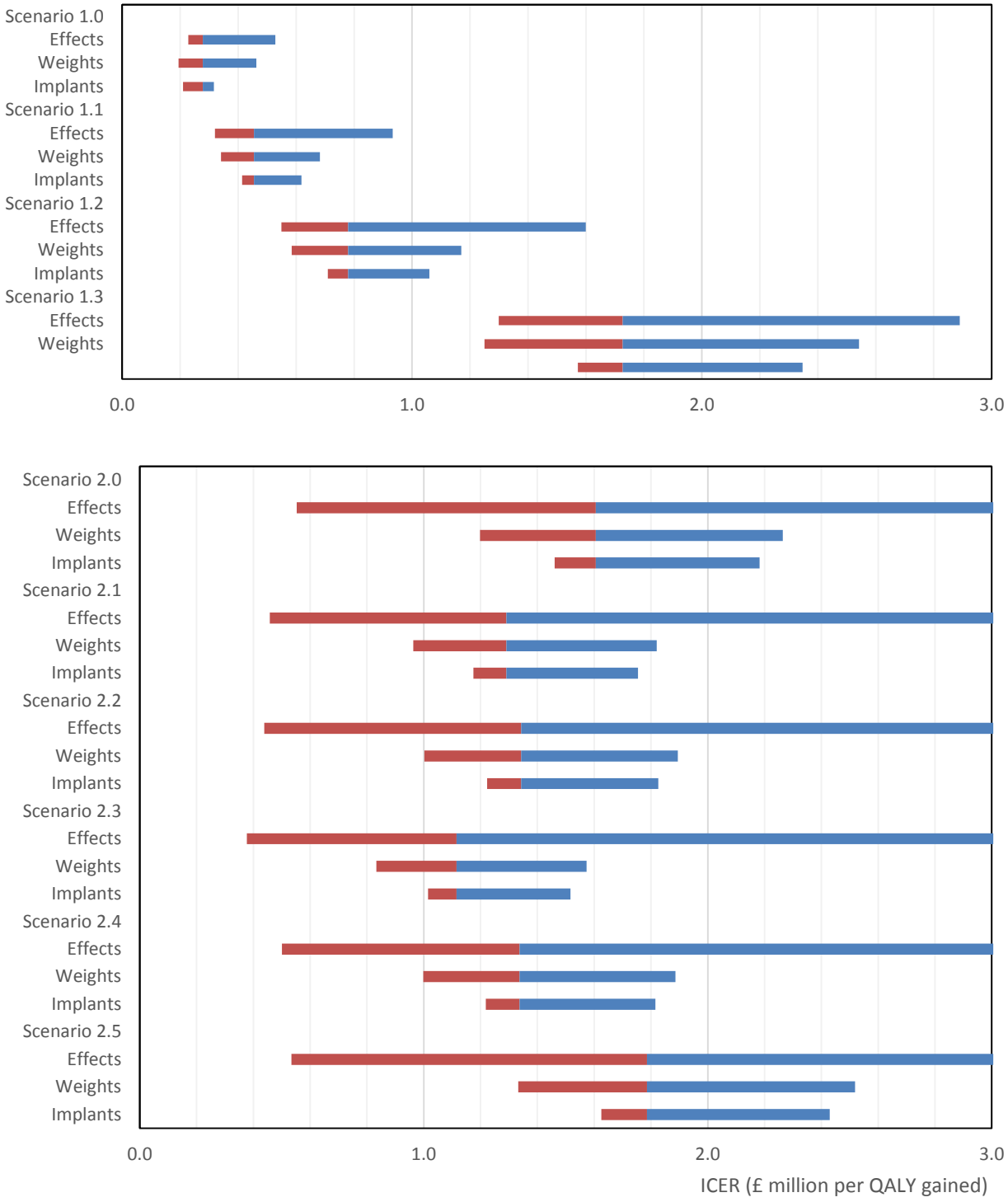

Figure 12 Tornado graphs for ERG scenarios

Probabilistic sensitivity analysis (PSA)

For all scenarios, the probability that afamelanotide was cost-effective at a threshold of £100,000 per QALY gained was 0%. When the threshold was increased to £150,000, the probability of cost-effectiveness remained negligible in all scenarios. We present three cost-effectiveness acceptability curves (CEACs) below. The company and ERG base cases are shown in Figure 13 and Figure 14 respectively, and reinforce the conclusion that these scenarios are unlikely to be cost-effective at a threshold of £150,000 per QALY gained.

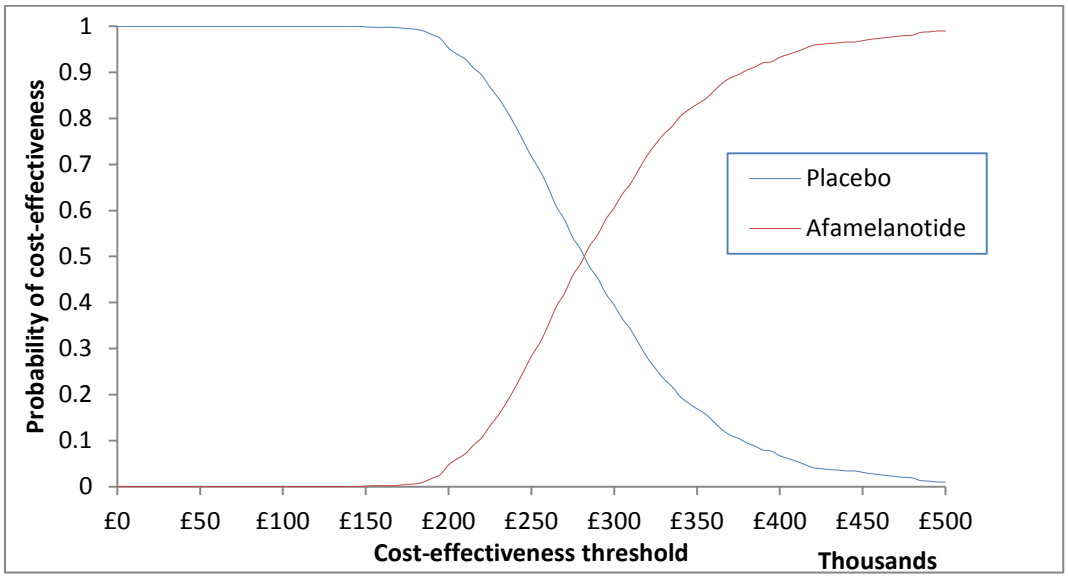

Figure 13 CEAC for Scenario 1.0 (company base case)

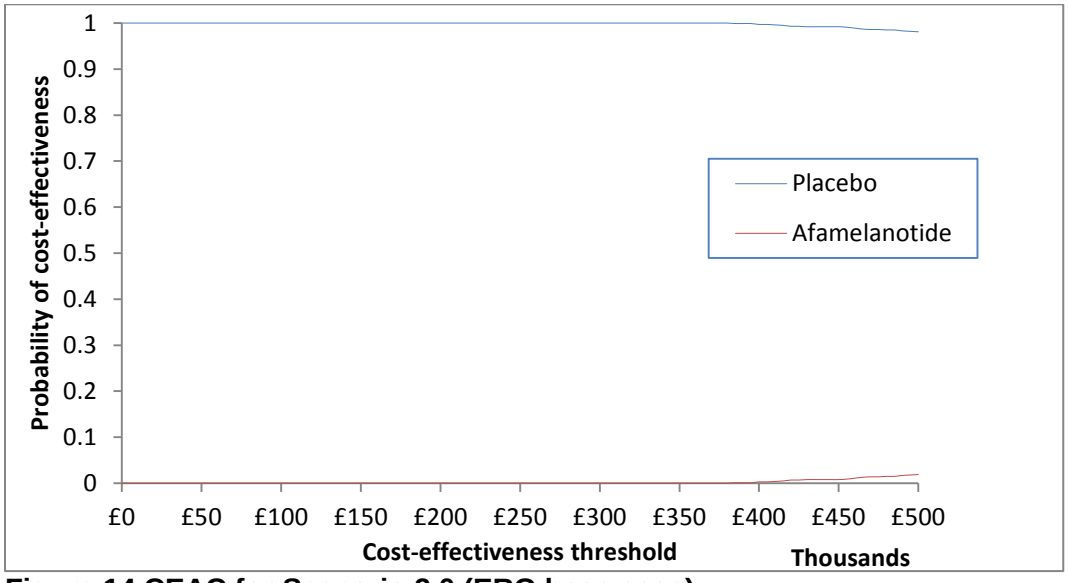

Figure 14 CEAC for Scenario 2.0 (ERG base case)

Best case analysis

Finally, we present the results of a best case PSA (Figure 15). This uses Scenario 1.0 (the QALY version of the company base case model), together with the most favourable limits for the three key sensitivity analyses: the upper limits for treatment effect and disability weights, and the lower limit for the mean number of implants per person per year used for costing. The best case scenario yielded an ICER of £151,212 per QALY gained. However, the ERG does not believe that this is a plausible scenario.

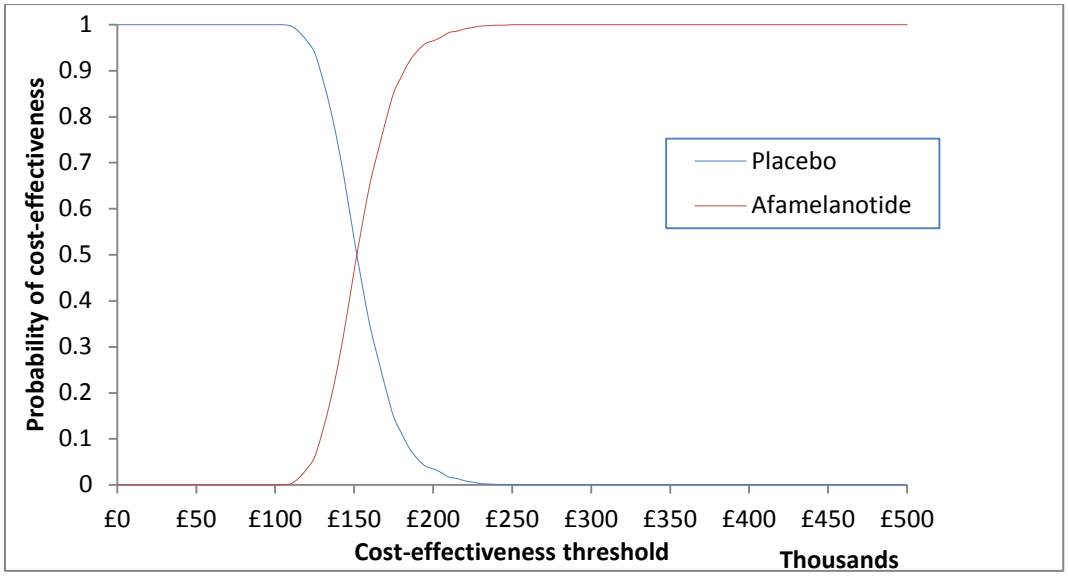

Figure 15 CEAC for best case scenario
(Company base case with upper limit for treatment effect and weights, and lower limit for number of implants)

5 Cost to the NHS and PSS

5.1 Base case budget impact

The company's model of budget impact is driven by three parameters (CS p 91):

- **EPP prevalence**. For the company's base case, 513 EPP patients were assumed to be eligible for treatment in England. This is higher than that previously cited, but clinical experts consulted by the ERG think that this figure is generally correct, or would not vary by more than 100 higher or lower (see section 2.3 above).
- **Uptake of afamelanotide**. The company assumed an uptake of [redacted] in the first year and [redacted] annually for the remaining four years.
- **Annual costs,** which are largely a function of the number of implants per patient per year. The company assumed that eligible patients would receive an average of [redacted] implants per year, in order to "[redacted]" (p91 CS).

The ERG notes that there are errors in the company's estimates of the budget impact over a five year period (section 13.7 of the CS). This seems to stems from them maintaining only [redacted] implant injection visits to administer [redacted] implants. We have corrected this to reflect that administering [redacted] implants will require [redacted] visits. Our corrected results are presented in Table 36.

Table 36 Corrected company budget impact estimated over the next 5 years

	Year 1	Year 2	Year 3	Year 4	Year 5	Total
Budget impact	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

5.2 Company and ERG sensitivity analyses

The CS does not report sensitivity analysis for the budget impact. We explored the budget impact, varying prevalence and the mean number of implants per year. The results of our sensitivity analysis are reported below in Table 37.

Table 37 Five year budget impact, varying prevalence and mean number of implants

	Year 1	Year 2	Year 3	Year 4	Year 5	Total
EPP prevalence						
300	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
400	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
600	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Mean implants per year						
2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
3	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

6 Impact of the technology beyond direct health benefits and on delivery of the specialised service

6.1 Impact on employment and income

The CS states that due to the lack of available EPP data, is it assumed that the majority of costs and savings of afamelanotide would be incurred within the NHS (CS section 14.1, p 93). No costs or savings to other government bodies or patients themselves are reported, though a sensitivity analysis which includes societal costs (which assumes an increase from 50% to 100% of mean wage over three years of afamelanotide treatment) is provided in the CS (see section 4.3.5.4 of this report). Although there is little empirical evidence on impacts of afamelanotide beyond direct health benefits, information from the NICE consultees and ERG clinical experts highlight the negative impact EPP has on patients' lives, including reduced study opportunities, job security and career development (see section 7 of this report for a discussion of consultee submissions). For example, as travel to a place of work or study can be difficult some patients can only engage in employment indoors, or undertake night work to avoid travelling during daylight hours. The CS mentions that a proportion of EPP patients are known to be unemployed, others are limited in their productivity, some however have full employment, whereas others have taken up nocturnal employment (CS section 12.4.2, p 80), though it should be noted that these proportions are not quantified in the CS.

The British Porphyria Association (BPA) submission to NICE suggests that patients with more severe EPP are unable to work under office lights and would therefore be restricted to working at home. The BPA also mentions a survey (reference not given) by an EPP patient organisation in the Netherlands which found that: 91% percent of patients changed careers because of EPP; 40% percent of patients reported losing a job because of EPP; 46% percent of patients took several (multiple consecutive) sick-days after an EPP-attack in the last five years and that 35% percent of patients can only work with adjustments (such adjustments are not defined). The BPA suggest that these figures are also applicable to the UK and the ERG agrees that this is a reasonable inference.

The BPA also states that patients tend to face economic dependence on the welfare state, along with the psychological burden that state dependence brings. They comment that "restricted options and preventative measures required to take part in other normal activities often adds hundreds, if not thousands of pounds sterling to the cost of living for both patients

and their families" (BPA submission, p 4-5). In summary, due to a reduced capacity to study and work, the socio-economic status of EPP patients and their families can be assumed to be lower overall, to the general population.

The afamelanotide RCTs included in the CS did not specifically measure the impact of treatment on ability to work or study, although the 18-item version of the EPP-QoL instrument does include an item assessing patient capacity to go to work or school (Question 12[11]) (this question appears to have been omitted from later revised versions of the EPP-QoL instrument). The long-term observational study of 115 EPP Swiss and Italian patients by Biolcati et al.[11], provided selected anecdotes from afamelanotide treated patients on their increased ability to study and to take employment and the financial benefits that this provided. It is reasonable to assume that the effects of treatment in terms of the ability to spend longer time in sunlight without pain, as described in section 3.3 of this report, will improve patients' education and employment opportunities and thus their income. Increased employment would also reduce demand on welfare benefits. However, there is no available data to quantify these impacts at present.

It is not clear what adjustments an employer would need to make (and therefore what the associated costs would be), to enable an EPP patient to attend the workplace. The ERG suggest that these could potentially include external/internal window screens, provision of suitable lighting, air conditioning facilities (e.g. to regulate the temperature without opening windows) and provision of car parking adjacent to building entrances/exits.

6.2 Impact on patient costs

The CS does not state any costs that patients would incur that would not be reimbursed by the NHS (CS section 14.3). However, to receive afamelanotide patients would need to travel to a specialist porphyria centre. Given the small number of centres in the UK that can potentially offer the treatment (the CS estimates that up to eight expert centres across the UK would provide treatment if recommended for the NHS), many patients would have to travel long distances to receive their implants and to be monitored (as required under the PASS protocol). These patients would therefore incur travel and potential accommodation costs, as well as potential loss of earnings from time away from work. The frequency of visits would depend on the number of implants required during the year. This frequency may vary between patients according to their specific needs, though it would not exceed four per year in line with the

marketing authorisation. These would be in addition to twice yearly monitoring appointments as required by the EMA (PASS protocol).

However, the British Association of Dermatologists (BAD) and BPA consultees in their submissions and expert clinical advice to the ERG, suggest that patients would not consider additional monitoring attendances as onerous or inconvenient, particularly compared to those associated with existing treatments such as UVB therapy which have a higher frequency of treatment appointments over a short time period.

6.3 Impact of the technology on delivery of the service

Administration of afamelanotide, if approved, in specialist centres in the UK was considered by both Royal College of Physician (RCP) and BAD consultees (in agreement with ERG clinical experts) to be feasible. However, it was noted by both BAD and RCP consultees that additional costs, in terms of time to train medical /nursing health professionals to administer the implants and provide additional follow up appointments would also need to be considered. The CS mentions that as part of the risk management plan agreed with the EMA, academic expert physicians will be trained and accredited by CLINUVEL to treat patients at the cost of the company. Only centres with existing, recognised expertise in EPP will be considered for training and accreditation (i.e., members of the European Porphyria Network (EPNET) and/or the British and Irish Porphyria Network (BIPNET)) (CS section 14.9, p 95). The company clarified that training should be conducted at least every two years and should apply to all staff involved in the care of adult patients with EPP (e.g. physicians, nurses, administrative staff, pharmacists) (clarification response question B7, 26/09/17). The duration of training (e.g. in terms of hours/days) and costs of training were not specified by the company.

[redacted]

7 Other submissions

Submissions were received from three consultee associations: the British Association of Dermatologists (BAD) (represented by four clinical experts); the British Porphyria Association (BPA) charity (represented by their vice-chairman who is also a helpline administrator) and the Royal College of Pathologists (RCP) (represented by a clinical expert from Salford Royal NHS Foundation trust). The ERG notes that the submissions from the BAD and the RCP both represent specialist porphyria services at Salford Royal NHS Trust which serves the greater Manchester area and other hospitals in north-west England. The BAD and RCP submissions represent the views of clinical EPP treatment specialists whilst the BPA submission represents the views of patients with EPP.

7.1 Number of patients with EPP

On referring to a 2006 academic paper (Holme et al. 2016[17]), the RCP quote the numbers of EPP patients in the UK to be 389 (which includes children under 18 years who are ineligible for treatment with afamelanotide). It should be noted that the BAD also quote this number but incorrectly state this number as those in England alone, where the reference quoted refers specifically to the number identified in the UK.

The BPA state that they currently have around 100 UK members who have EPP. They estimate that they have 25% of UK EPP patients on their database, which would agree with the number of around 400 in the UK previously quoted by Holme et al.[17] and Elder et al.[48]

7.2 Diagnosis and current treatment provision in the NHS

The BAD and RCP provide a general overview of the issues surrounding confirmation and average age of diagnosis as well as the lack of general practice and public knowledge of the condition.

The consultees acknowledge that there are no specific pharmacological treatments for EPP and the CS states that “The lack of available effective therapies for EPP means no formal treatment recommendations exist” (p 18). Current treatment options are limited to include effective sun protection, B carotene doses, correction of Vitamin D deficiency and narrow band UVB therapy.

In agreement with NICE and the CS, the use of currently available methods of managing the condition, high dose B carotene or Dundee cream were considered both ineffective and

impractical by the BAD, RCP and BPA consultees and ERG clinical experts. The BPA highlighted a systematic review of treatment options for dermal photosensitivity in EPP, stating that high dose beta-carotene is ineffective.[49] The use of narrow band UVB treatment by some patients (six treatments in quick succession every spring) was mentioned by the BAD and the RCP, as well as the ERG clinical experts. Although this was thought to be the best form of treatment, it was noted by these consultees that it can be problematic for patients who are working or live a long distance from a treatment centre given the frequency of administration necessary.

7.3 Impact on patients, families and carers

The BPA representative highlighted in detail, the patient's perspective of the effects of their condition on everyday life, stressing the distress experienced during a phototoxic reaction. Using quotes from EPP patients the BPA submission highlighted the effects of intense pain and extreme tiredness on not only the patients but families and carers. They discussed the impact on earning capacity for both the patients and families. The report quoted a study on the effect of EPP on work attendance, carried out by an EPP patient organisation in the Netherlands (as described earlier in Section 6.1 of this report), stating the negative impact on job retention and career choice. The BPA representative discussed the potential effects on mental health (anxiety) on patients. The CS in agreement with the BPA also stated that EPP severely impacts upon quality of life and ability to function normally, inhibiting social participation, education and employment (p16).

7.4 Advantages of the technology

It was noted by the BPA that despite the sub-optimal timing of trials for UK patients afamelanotide has a positive effect on symptoms (NB. They do not elaborate on the timing). As acknowledged by both the BPA representative and ERG clinical experts, afamelanotide may have a significant effect on the lifestyle of EPP patients. People who benefit most from the treatment are those who are willing and able to gradually recondition themselves to exposure to light. The BPA submission includes a number of emotive quotes to support the positive effect of afamelanotide on patients, stating their positive effect on family life, the parenting of young children and general lifestyle. This consultee stated that their information was obtained from consulting patient members who had participated in trials in the UK and comments correlated with consensus themes that emerged from presentations and discussions at a recent medical conference International Conference on Porphyrins and Porphyrias (ICPP2017) – Bordeaux

(June 2017; no data or reference provided). In addition, it is the BPA's opinion that, for those patients who are able to tolerate some degree of exposure to visible light (having less severe reactions), afamelanotide is a "complete life changer, effectively eliminating the impact of light exposure on working day life and opening up all but the most exposed of activities to EPP patients".

8 DISCUSSION

8.1 Summary of clinical effectiveness issues

The RCTs evaluating afamelanotide show statistically significant differences across outcomes in favour of the treatment. Compared to placebo patients were able to spend longer in sunlight without experiencing pain, or experiencing only mild pain. Statistically significant differences were observed in two of the RCTs (CUV029 and CUV039) demonstrating consistency in effects. The median increase in pain free sunlight exposure varied between approximately five hours to 24 hours depending on the study (taking into account its geographical location and overall study length, and the time period during each day in which outcomes were measured). The clinical significance of these results is unclear as these outcomes appear to have been devised specifically to evaluate this treatment and minimal important clinical differences have not yet been established. The effects could be interpreted as being modest. For example, in study CUV029 the median five hour increase in pain free direct sunlight exposure, measured between 10:00 to 15:00 hours per day, is only a small proportion of the total available daylight time over the nine study month period. However, there are a number of factors which influence an EPP patient's exposure to sunlight, including their long-standing fear of going outside, weather conditions, their daily activities (work, leisure, family commitments), and their physical mobility. Indeed, it has been commentated that the effects seen in the studies could be underestimated given patients' lifelong reluctance to expose themselves to light.[14]

The clinical significance of treatment effects is reinforced by patient testimonials, as reported in the consultee submissions to NICE (see section 7). Patients describe the positive impact that treatment has made on their lives, and say that even a relatively small increase in the time that light exposure can be tolerated can make a significant difference. The BPA in their submission states that they have not encountered a patient who has not received a significant benefit from afamelanotide. The BPA also suggest that people who would benefit most from treatment are those who are able to gradually recondition themselves to light exposure. Given that behaviour

takes time to change and maintain, the relatively short durations of the RCTs may be inadequate to demonstrate the optimum effectiveness of afamelanotide. Furthermore, it may be necessary for behavioural therapy to be provided to some patients receiving afamelanotide to enable them to overcome their fear of light exposure.

Another factor which may have influenced the results of the RCTs is a potential placebo effect. The journal publication for studies CUV029 and CUV039 mentions that a few patients who received placebo were convinced that they received afamelanotide and reportedly increased their sun exposure.[7] The ERG notes that placebo group EPP-QoL and DLQI scores improved during the study, indicating a potential unexplained placebo effect. However, it is also known that the tanning effect of afamelanotide unblinded some treated patients in the RCTs. This could have potentially encouraged treated patients to increase their sun exposure, thus mitigating the possible placebo effect in the studies. However, as stated above, the long-standing behavioural avoidance of sun exposure may have inhibited patients who had guessed that they were receiving afamelanotide from exposing themselves to light.

The generalisability of the size of the treatment effects from the studies to England and the UK is not straightforward. The CUV039 study was conducted in the USA and the trial journal publication suggests that the difference in the magnitude of sunlight exposure time gained between this trial and the European trial (CUV029) can be explained, in part, by differences in latitude. The European centres were at higher latitudes and it could be suggested that the amount of daylight available to patients in the European centres would, on average, be less than patients in the US centres. They would therefore have less opportunity to spend time outdoors and be exposed to light. Conversely it could be assumed that the strength of sunlight at lower latitudes would be greater and that this would limit the amount of time patients could spend in sunlight without experiencing pain. Furthermore, the results seen in the RCTs reflect a single period of months in time which may or may not have been typical in terms of weather patterns (and hence potential for sunlight exposure) in both continents. Thus, a number of factors need to be taken into consideration when generalising the results of the RCTs - particularly CUV039 - to the UK.

In summary, the ERG's interpretation of the evidence is that afamelanotide is associated with benefit to patients in the trials in terms of increased ability to be exposed to sunlight with little or no pain. In turn their HRQoL improved with an increased ability to take part in daily activities

outside of the home. However, there are a number of potential confounding factors which limit interpretation of the magnitude of the treatment effect and its generalisability to the UK. The ERG's interpretation is similar to that of the EMA assessment of the evidence as part of the marketing authorisation application.[2]

8.2 Summary of issues for costs and health effects

The ERG identified an abstract published in 2016 by Thompson et al.,[3] reporting an economic evaluation that estimated the cost-effectiveness of afamelanotide for EPP, using DALYs as the measure of effect. This abstract report a base case and sensitivity analysis range for DALYs averted that [redacted] Thompson et al. reported a base case ICER of £373,000 per DALY averted, which was higher than the company base case estimate of £278,471 per DALY averted. Thompson et al. also presented results for a sensitivity analysis using QALYs, reporting an ICER of £401,000 *per QALY gained* using the condition hereditary angioedema (swelling under the skin) as a proxy for EPP, and a range from £208,000 to £1.1 million per QALY with alternative sources for utility weights.

8.2.1 ERG critique of company model

The ERG critically appraised the company's submitted cost-effectiveness model. This estimates the value of health outcomes in the form of DALYs avoided, because the company did not consider QALYs to be an appropriate measure for EPP. We consider the basic structure of the company model – [redacted] - to be reasonable, although we note that it entails some strong assumptions: [redacted] These may be unlikely in practice, but we have not identified evidence to support alternative modelling assumptions. However, we do have serious concerns about the way in which the magnitude of treatment effect was estimated and valued in the company model.

The company's estimate of mean DALYs averted per year of treatment was based on EPP-QOL data from three randomised studies, CUV029, CUV030 and CUV039. The ERG is concerned that we have not had sufficient access to information about the methods and results of studies CUV029 and CUV30 to be able to assess their quality or check the results. We have also had insufficient information about how the results of the three trials were analysed and pooled. There is a lack of basic information about whether ITT datasets were used, the number of

patients included from each trial and whether the method of pooling accounted for clustering or randomisation. Furthermore, we are concerned about the lack of evidence over how the EPP-QOL scale was developed and validated. In particular, post hoc changes to the scoring system which were introduced after initial analysis of trial results, introduces a risk of bias.

With regard to the valuation of health effects, we do not have confidence that the disability weights for mild, moderate and severe disease in the company model are appropriate for EPP, or that they are consistent with the company's definitions of severity based on EPP-QOL scores. We do not know if [redacted] is an appropriate proxy condition for EPP – the clinical experts who we have consulted have suggested that it might not be. There may be some similarities in psychological and functional impacts, but it is not at all clear if the magnitude and severity of these conditions are comparable. The same applies to the alternative proxy condition of [redacted] The definition of mild, moderate and severe disease by division of the EPP-QoL scale into thirds is also arbitrary and we cannot assess if it is consistent with the definitions used to elicit the proxy disability weights.

Another set of problems with the company's approach, relate to how they have extrapolated treatment effects from a single time point to estimate mean DALY loss under standard treatment and with afamelanotide. The company's model only makes use of the 120 day results and assumes that these values apply for the whole year, including around half the year when patients would not have afamelanotide implants. We believe that this is simplistic and likely to have biased DALY estimates in favour of afamelanotide. The company stated that they chose 120 days as this was the longest time point available from all three trials, but the CS indicated that EPP-QoL data was also collected at 180 days in the three trials. The approach also fails to correct for baseline imbalance in EPP-QoL severity, which would have favoured afamelanotide. And we also question the assumption

We also have some questions about the cost estimates used. These were very largely driven by an assumption about the mean number of implants per person per year. This figure was estimated from 'real world' data, and it is not clear whether this was consistent with use in the clinical trials. If not, this would be a source of bias. We do also consider that the estimated administration and monitoring costs for afamelanotide, and usual care appear to be high for a UK context. However, these costs are small in relation to the drug acquisition costs, and so have little influence on the ICER.

Finally, we note that the analysis of uncertainty presented in the CS was inadequate. In particular, there was no attempt to estimate the extent or consequences of uncertainty over the effectiveness parameters and assumptions. Given the discussion above, we think this could be considerable. There was also no probabilistic analysis of uncertainty.

We conducted additional analysis based on the company's model. First, we developed a very simple QALY model as a platform to investigate alternative scenarios and sensitivity around the company's base case. This demonstrated that the company's incremental cost per DALY averted of £278,471 (£278,386 per QALY gained after a small correction by the ERG) is likely to be an underestimate. With correction for baseline differences in EPP-QoL, this rose to £454,800 per QALY gained. The ICER rose further, to £779,657 per QALY gained, when we assumed that treatment benefits would gradually decline over a 2 month period from month 6. Use of utility estimates from the literature for the same proxy condition as in the company base case, further increased the estimated ICER to over £1.7 million per QALY gained.

We conducted a 'best case' analysis, which combined the most favourable scenario that we had tested (our simple QALY conversion of the company's base case model), with the most favourable sensitivity analysis limits for treatment effects, disability weights and mean number of implants used for costing. This brought the ICER down to £151,212 per QALY gained. The ERG does not believe that this or any of the other ICER estimates based on our simple adaptation of the company model are plausible.

Our preferred set of analyses were based on mean DLQI data from the pivotal study (CUV039) mapped to EQ-5D utility values using a published algorithm. Results from this model were less favourable, and did not fall below £1.1 million per QALY gained in any of the scenarios that we tested. The ERG believes that this set of estimates is more plausible than the company's approach.

9 REFERENCES

1. Minder EI, Barman-Aksoezen J, Nydegger M, et al. Existing therapies and therapeutic targets for erythropoietic protoporphyria. *Expert Opinion on Orphan Drugs* 2016;4(6):577-89. doi: http://dx.doi.org/10.1517/21678707.2016.1171137
2. EMA (European Medicines Agency). Assessment Report. Scenesse. International non-proprietary name: afamelanotide. Procedure No. EMEA/H/C/002548/0000. London European Medicines Agency, 2014:1-107.
3. Thompson J, Mealing S, Vetrini D, et al. The cost effectiveness of afamelanotide for the treatment of Erythropoietic protoporhyria. *Value in Health* 2016;19(3):A248-A49.
4. Minder E I. Exploiting the photoprotective effect of afamelantoide. Erythropoietic Protoporphyria. 20th European Association of Dermatology and Venereology Congress, Lisbon, Portugal. 2011. *20th European Association of Dermatology and Venereology Congress, Lisbon, Portugal* 2011
5. Balwani M. A Phase II, multicentre, double-blind, randomised, placebo-controlled study to confirm the safety and efficacy of afamelanotide implants in patients with erythropoietic protoporphyria (EPP). *International Congress of Porphyrins and Porphyrias, EPP Satellite Symposium Luzern, Switzerland* 2013
6. Harms JH, Lautenschlager S, Minder CE, et al. Mitigating photosensitivity of erythropoietic protoporphyria patients by an agonistic analog of alpha-melanocyte stimulating hormone. *Photochemistry and Photobiology* 2009;85(6):1434-39. doi: http://dx.doi.org/10.1111/j.1751-1097.2009.00595.x
7. Langendonk J, Balwani M, Anderson K, et al. Afamelanotide for erthropoietic protoporphyria. *N Engl J Med* 2015;373:48-59.
8. Biolcati G, Marchesini E, Sorge F, et al. Long-term observational study of afamelanotide in 115 patients with erythropoietic protoporphyria. *British Journal of Dermatology* 2015;172(6):1601-12. doi: http://dx.doi.org/10.1111/bjd.13598
9. Langendonk JG. Afamelanotide (Scenesse, CLINUVEL) for the treatment of protoporphyria-induced phototoxicity, where are we? *International Congress on Porphyrins & Porphyrias, Bordeaux, France 27 June 2017* 2017
10. Harms J, Lautenschlager S, Minder C, et al. An α-melanocyte–stimulating hormone analogue in erythropoietic protoporphyria. *N Engl J Med* 2009;360:306-07.
11. Biolcati G, Marchesini E, Sorge F, et al. Long-term observational study of afamelanotide in 115 patients with erythropoietic protoporphyria. *Br J Dermatol* 2015;172(6):1601.

12. Harms J, Lautenschlager S, Minder CE, et al. An α-Melanocyte–Stimulating Hormone Analogue in Erythropoietic Protoporphyria. *New England Journal of Medicine* 2009;360(3):306-07. doi: 10.1056/NEJMc0805682
13. Clinuvel. Company announcement: Clinuvel demonstrates positive treatment effect of afamelanotide in US Phase II study. Melbourne, Australia: Clinuvel, 2011:1-3.
14. Mullard A. Clinuvel's tanning agent nears approval for porphyria. *Nat Biotechnol* 2015;33(1):7.
15. Minder E, Biolcati G, Marchesini E, et al. Update on long-term observation of afamelanotide (Scenesse®)-treatment in 146 Swiss and Italian patients with erythropoietic protoporphyria. International Congress on Porphyrins and Porphyrias. Düsseldorf, 2015.
16. Clinuvel. Company announcement: Clinuvel reports positive results in phase III porphyria (EPP) clinical trial. Melbourne, Australia: Clinuvel, 2010:1-5.
17. Holme S, Anstey A, Finlay A, et al. Erythropoietic protoporphyria in the U.K.: clinical features and effect on quality of life. *Br J Dermatol* 2006;155:574-81.
18. Basra MKA, Fenech R, Gatt RM, et al. The Dermatology Life Quality Index 1994–2007: a comprehensive review of validation data and clinical results. *British Journal of Dermatology* 2008;159(5):997-1035. doi: 10.1111/j.1365-2133.2008.08832.x
19. Jong CT, Finlay AY, Pearse AD, et al. The quality of life of 790 patients with photodermatoses. *British Journal of Dermatology* 2008;159(1):192-97. doi: 10.1111/j.1365-2133.2008.08581.x
20. Rosenkranz GK. A note on the Hodges-Lehmann estimator. *Pharmaceutical statistics* 2010;9(2):162-7. doi: 10.1002/pst.387 [published Online First: 2009/09/01]
21. Borenstein M, Hedges L, Higgins J, et al. Introduction to Meta-Analysis. Chichester: John Wiley & Sons, Ltd 2009.
22. Hongbo Y, Thomas CL, Harrison MA, et al. Translating the Science of Quality of Life into Practice: What Do Dermatology Life Quality Index Scores Mean? *Journal of Investigative Dermatology* 2005;125(4):659-64. doi: http://dx.doi.org/10.1111/j.0022-202X.2005.23621.x
23. Basra MK, Salek MS, Camilleri L, et al. Determining the minimal clinically important difference and responsiveness of the Dermatology Life Quality Index (DLQI): further data. *Dermatology (Basel, Switzerland)* 2015;230(1):27-33. doi: 10.1159/000365390 [published Online First: 2015/01/24]

24. Schuller Y, Hollak C, Biegstraaten M. The quality of economic evaluations of ultra-orphan drugs in Europe–a systematic review. *Orphanet Journal of Rare Diseases* 2015;10(1):92.
25. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal 2013. *NICE, London* 2013
26. National Institute for Health and Care Excellence. Interim Process and Methods of the Highly Specialised Technologies Programme Updated to reflect 2017 changes. *NICE, London* 2017
27. National Institute for Health and Care Excellence. Highly Specialised Technology Evaluation. Afamelanotide for treating erythropoietic protoporphyria. Final scope. *https://wwwniceorguk/guidance/gid-hst10009/documents/final-scope* 2017
28. Salomon JA, Vos T, Hogan DR, et al. Common values in assessing health outcomes from disease and injury: disability weights measurement study for the Global Burden of Disease Study 2010. *The Lancet* 2012;380(9859):2129-43.

[redacted]

30. Katugampola RP, Badminton MN, Finlay AY, et al. Congenital erythropoietic porphyria: a single-observer clinical study of 29 cases. *British Journal of Dermatology* 2012;167(4):901-13. doi: 10.1111/j.1365-2133.2012.11160.x
31. Rufener EA. Erythropoietic protoporphyria: a study of its psychosocial aspects. *British Journal of Dermatology* 1987;116(5):703-08. doi: 10.1111/j.1365-2133.1987.tb05904.x
32. Stafford R, Farrar MD, Kift R, et al. The impact of photosensitivity disorders on aspects of lifestyle. *British Journal of Dermatology* 2010;163(4):817-22. doi: 10.1111/j.1365-2133.2010.09905.x
33. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ* 2010;340 doi: 10.1136/bmj.c221
34. Whitehead A, Omar RZ, Higgins JPT, et al. Meta-analysis of ordinal outcomes using individual patient data. *Statistics in Medicine* 2001;20(15):2243-60. doi: 10.1002/sim.919
35. Curtis L, Burns A. Unit Costs of Health and Social Care 2016. *http://wwwpssruacuk/pdf/uc/uc2010/uc2010pdf* 2016

[redacted]

37. Norlin JM, Steen Carlsson K, Persson U, et al. Analysis of three outcome measures in moderate to severe psoriasis: a registry-based study of 2450 patients. *British Journal of Dermatology* 2012;166(4):797-802.
38. Wolff J, Edwards S, Richmond S, et al. Evaluating interventions in health: a reconciliatory approach. *Bioethics* 2012;26(9):455-63. doi: 10.1111/j.1467-8519.2011.01888.x
39. Menzel P, Dolan P, Richardson J, et al. The role of adaptation to disability and disease in health state valuation: a preliminary normative analysis. *Social science & medicine* 2002;55(12):2149-58.
40. Excellence NIfHaC. Social value judgements. Principles for the development of NICE guidance. London: NICE 2008.
41. Ali FM, Kay R, Finlay AY, et al. Mapping of the DLQI scores to EQ-5D utility values using ordinal logistic regression. *Quality of Life Research* 2017 doi: 10.1007/s11136-017-1607-4
42. Blome C, Beikert FC, Rustenbach SJ, et al. Mapping DLQI on EQ-5D in psoriasis: transformation of skin-specific health-related quality of life into utilities. *Archives of Dermatological Research* 2013;305(3):197-204.
43. Currie CJ, Conway P. PSK11 evaluation of the association between EQ5D utility and dermatology life quality index (DLQI) score in patients with psoriasis. *Value in Health* 2007;10(6):A470-A71.
44. Herédi E, Rencz F, Balogh O, et al. Exploring the relationship between EQ-5D, DLQI and PASI, and mapping EQ-5D utilities: a cross-sectional study in psoriasis from Hungary. *The European Journal of Health Economics* 2014;15(1):111-19.
45. Lidgren M, Wilking N, Jonsson B, et al. Health related quality of life in different states of breast cancer. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* 2007;16(6):1073-81. doi: 10.1007/s11136-007-9202-8 [published Online First: 2007/05/01]
46. Ara R, Brazier JE. Using health state utility values from the general population to approximate baselines in decision analytic models when condition-specific data are not available. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research* 2011;14(4):539-45. doi: 10.1016/j.jval.2010.10.029 [published Online First: 2011/06/15]

47. Rombach SM, Hollak CE, Linthorst GE, et al. Cost-effectiveness of enzyme replacement therapy for Fabry disease. *Orphanet Journal of Rare Diseases* 2013;8(1):29. doi: 10.1186/1750-1172-8-29
48. Elder G, Harper P, Badminton M, et al. The incidence of inherited porphyrias in Europe. *Journal of inherited metabolic disease* 2013;36(5):849-57. doi: 10.1007/s10545-012-9544-4 [published Online First: 2012/11/02]
49. Minder EI, Schneider-Yin X, Steurer J, et al. A systematic review of treatment options for dermal photosensitivity in erythropoietic protoporphyria. *Cellular and molecular biology (Noisy-le-Grand, France)* 2009;55(1):84-97. [published Online First: 2009/03/10]