NICE National Institute for Health and Care Excellence

Single Technology Appraisal

Crisaborole for treating mild to moderate atopic dermatitis in people aged 2 years and older [ID1195]

Committee Papers

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

SINGLE TECHNOLOGY APPRAISAL

Crisaborole for treating mild to moderate atopic dermatitis in people aged 2 years and older [ID1195]

Contents:

The following documents are made available to consultees and commentators:

The **final scope** **and** **final stakeholder list** are available on the NICE website.

1. **Company submission** from Pfizer
2. **Clarification questions and company responses**
3. **Patient group, professional group and NHS organisation submission** from:
   a. National Eczema Society
4. **Evidence Review Group report** prepared by the School of Health and Related Research (ScHARR)
   NB: The Evidence Review Group report was updated after the factual accuracy check
5. **Evidence Review Group report – factual accuracy check**
6. **Technical report**
7. **Technical engagement response from company**
8. **Technical engagement responses from consultees and commentators:**
   a. National Eczema society
   b. Centre of Evidence-Based Dermatology
9. **Technical engagement responses from experts:**
   a. Kymmene Dawson – patient expert, nominated by the National Eczema Society
10. **Evidence Review Group critique of company response to technical engagement** prepared by the School of Health and Related Research (ScHARR)

Any information supplied to NICE which has been marked as confidential, has been redacted. All personal information has also been redacted.

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Single technology appraisal

Crisaborole for Treating Mild to Moderate Atopic Dermatitis in People Aged 2 Years and Older [ID 1195]

Document B

Company evidence submission

File name	Version	Contains confidential information	Date
		Yes	

Contents

List of Tables

List of Figures

Abbreviations

AD	Atopic dermatitis
ADHD	Attention deficit hyperactivity disorder
ADSI	Atopic Dermatitis Severity Index
AE	Adverse event
AHRQ	Agency for Healthcare Research and Quality
BAD	British Association of Dermatologists
BID	Twice daily
BNF	British National Formulary
BSA	Body Surface Area
CADTH	Canadian Agency for Drugs and Technology in Health
cAMP	Cyclic adenosine monophosphate
CDLQI	Children's Dermatology Life Quality Index
CI	Confidence interval
CINAHL	Cumulative Index of Nursing and Allied Health Literature
CONSORT	Consolidated Standards of Reporting Trials
CRD	Centre for Reviews and Dissemination
CSR	Clinical Study Report
DALY	Disability-adjusted life year
DFI	Dermatitis Family Impact
DLQI	Dermatology Life Quality Index
EASI	Eczema Area Severity Index
EMA	European Medicines Agency
EPAR	European public assessment report
EQ-5D	EuroHRQoL 5 Dimensions
FAS	Full analysis set
FDA	Food and Drug Administration
GBD	Global Burden of Disease
GI	Gertner Institute
GP	General Practitioner
GREAT	Global Resource for Eczema Trials
HOME	Harmonising Outcomes Measures in Eczema
HPA	Hypothalamic-pituitary-adrenal
HRQoL	Health-Related Quality of Life
HS1	Health state 1
HS2	Health state 2
ICER	Institute for Clinical and Economic Review
ICER	Incremental cost-effectiveness ratio
ICF	Informed consent form
IGA	Investigators' Global Assessment
ISGA	Investigators' Static Global Assessment
IL	Interleukin
ISOLTE	International Study of Life with Atopic Eczema
ITT	Intent-to-Treat
LILACS	Literatura Latino-Americana e do Caribe em Ciências da Saúde (Literature in the Health Sciences in Latin America and the Caribbean)
LTE	Long-term extension
LYG	Life years gained
MA	Marketing Authorisation
MCID	Minimal clinically important difference
MCMC	Markov Chain Monte Carlo
mEASI	Modified Eczema Area and Severity Index
NHS	National Health Service
NICE	National Institute for Health and Clinical Excellence

NMA	Network Meta-Analysis
NRS	Numerical rating scale
PAS	Patient Access Scheme
PBAC	Pharmaceutical Benefits Advisory Committee
PbR	Payment-by-results
PDE4	Phosphodiesterase 4
PDE4i	Phosphodiesterase 4 inhibitor
PGA	Physician's Global Assessment
PI	Principal investigator
PICOS	Population, interventions, comparators, outcomes, study design
POEM	Patient-Oriented Eczema Measure
PP	Per protocol
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PSS	Personal Social Services
PSSRU	Personal Social Services Research Unit
QALY	Quality-adjusted life year
QoL	Quality of Life
RCT	Randomised controlled trial
SA	Sensitivity analysis
SAE	Serious adverse event
SAS	Safety analysis set
SCORAD	Scoring Atopic Dermatitis index
SD	Standard deviation
SLR	Systemic literature review
SMC	Scottish Medicines Consortium
SmPC	Summary of Product Characteristics
TCS	Topical corticosteroids
TCI	Topical calcineurin inhibitors
TCW	Totally controlled weeks
TEAE	Treatment-emergent adverse event
Th2	T-helper type 2
Th17	T-helper type 17
UK	United Kingdom
US	United States
UV	Ultraviolet
WCW	Well controlled weeks

B.1 Decision problem, description of the technology and clinical care pathway

B.1.1 Decision problem

This evaluation seeks to determine whether crisaborole (Staquis) 20 mg/g (2%) ointment is cost-effective in the treatment of mild to moderate atopic dermatitis in adults and children aged 2 years of age and older. Whilst crisaborole has been demonstrated to be safe and effective for the first line treatment of adults and children aged 2 years and older with mild to moderate AD, it is not anticipated that crisaborole will be recommended in the UK for first line treatment (TCS eligible patients), due to the very low cost of topical corticosteroids (TCS). Pfizer would consequently like to seek NICE recommendation for crisaborole in an optimised population for the treatment of adults and children aged 2 years and older with mild to moderate AD that has not been controlled by TCS or where there is a serious risk of important adverse effects from further TCS use, particularly irreversible skin atrophy (second-line use post-TCS treatment). Analyses for crisaborole versus first line (TCS eligible populations) have been presented in Appendices.

The full details of the decision problem, its alignment to the final scope issued by NICE (1), and how it has been addressed in this submission are presented in **Table 1.**

Table 1:The decision problem

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope
Population	• People aged 2 years and older with mild and moderate atopic dermatitis	• Adults and children aged 2 years and older with mild to moderate AD that has not been controlled by topical corticosteroids or where there is a serious risk of important adverse effects from further topical corticosteroid use, particularly irreversible skin atrophy (second-line use post-TCS treatment).	• Pfizer would like to seek NICE recommendation for crisaborole in an optimised population for the treatment of adults and children aged 2 years and older with mild to moderate AD that has not been controlled by topical corticosteroids or where there is a serious risk of important adverse effects from further topical corticosteroid use, particularly irreversible skin atrophy (second-line use post-TCS treatment). ICERs for crisaborole versus first line (TCS eligible patients) with mild to moderate AD have been presented in Appendices
Intervention	Crisaborole ointment, (20mg/g)	• Crisaborole ointment, (20mg/g)	• Consistent with the final scope

Comparator(s)	**For mild atopic dermatitis**: • Combination of emollients and mild to moderate potency topical corticosteroids **For moderate atopic dermatitis**: • High potency topical corticosteroids • Topical calcineurin inhibitors	**For mild atopic dermatitis:** Combination of emollients and mild to moderate potency topical corticosteroids • Topical calcineurin inhibitor: Adults and children: pimecrolimus For moderate atopic dermatitis: For moderate atopic dermatitis: • Combination of emollients and moderate to high potency topical corticosteroids • Topical calcineurin inhibitors: Adults: tacrolimus 0.1%, tacrolimus 0.03% Children: tacrolimus 0.03%, pimecrolimus 1%	Comparators are consistent with the NICE final scope. Pfizer has also presented additional analyses that evaluate crisaborole versus pimecrolimus for the treatment of adults and children with mild AD that has not been controlled by topical corticosteroids or where there is a serious risk of important adverse effects from further topical corticosteroid use (second-line use post-TCS treatment) Whilst we acknowledge that the TCIs are not currently recommended by NICE or included in the NICE scope for mild AD patients, data from the British Association of Dermatologists National Clinical Audit of atopic dermatitis in children, indicated that 8.26% mild AD patients currently receive TCIs in UK clinical practice.(2) This data suggests that TCIs appear to be a clinically relevant comparator for mild AD that has not shown a satisfactory clinical response to TCS or is otherwise contraindicated. Pimecrolimus is the only therapy licensed in adults and children in mild AD patients, hence, only this TCI has been considered for this scenario.
Outcomes	The outcome measures to be considered include: • Disease severity • Symptom control • Disease free period/maintenance of remission • Time to relapse/prevention of relapse • Adverse effects of treatment • Health-related quality of life	• Outcomes used in the economic model are consistent with the Final Scope: • Disease severity • Symptom control • Disease free period/maintenance of remission • Adverse effects of treatment • Health-related quality of life	• Comparative efficacy data were not adequately reported to evaluate time to relapse/prevention of relapse in the mild to moderate AD population.

Economic analysis	The reference case stipulates that the cost effectiveness of treatments should be expressed in terms of incremental cost per quality-adjusted life year. The reference case stipulates that the time horizon for estimating clinical and cost effectiveness should be sufficiently long to reflect any differences in costs or outcomes between the technologies being compared. Costs will be considered from an NHS and Personal Social Services perspective.	• The cost effectiveness model has been designed in line with the NICE reference case criteria. Outcomes are expressed in terms of incremental cost per quality-adjusted life year. Costs have been considered from an NHS and Personal Social Services	• Consistent with the final scope

Subgroups to be considered	If the evidence allows, the following subgroups will be considered: • adults and children • mild and moderate atopic dermatitis • people with different skin colour • people with atopic dermatitis affecting the hands • people with atopic dermatitis affecting sensitive areas (face, neck and flexures) • mild to moderate AD that has not been controlled by topical corticosteroids or where there is a serious risk of important adverse effects from further topical corticosteroid use, particularly irreversible skin atrophy (second-line use post-TCS treatment). Guidance will only be issued in accordance with the marketing authorisation. Where the wording of the therapeutic indication does not include specific treatment combinations, guidance will be issued only in the context of the evidence that has underpinned the marketing authorisation granted by the regulator.	Subgroups to be considered • Children mild AD • Children moderate AD • Adults mild AD • Adults moderate AD Adults and children aged 2 years and older with mild to moderate AD that has not been controlled by topical corticosteroids or where there is a serious risk of important adverse effects from further topical corticosteroid use, particularly irreversible skin atrophy (second-line use post-TCS treatment). Adults and children aged 2 years and older with mild to moderate TCS naive AD.	• Insufficient clinical evidence to present separate subgroup analyses for: • People with different skin colour* • People with atopic dermatitis affecting the hands* • People with atopic dermatitis affecting sensitive areas (face, neck and flexures)**

Abbreviations: NICE: National Institute for Health and Care Excellence; *Data for AD in patients with skin of colour were not consistently reported across comparator trials. Further, efficacy data for **ISGA clear and almost clear are not available by sensitive skin area for crisaborole.

B.1.2 Description of the technology being appraised

A summary and description of crisaborole ointment (20 mg/g) is provided in **Table 2** below.

Table 2:Technology being appraised

UK approved name and brand name	UK approved name: Crisaborole ointment (20 mg/g) Brand name: Staquis
Mechanism of action	Crisaborole topical ointment 20 mg/g is a novel, non-steroidal, topical therapy that inhibits phosphodiesterase 4 (PDE4). PDE4 is a key regulator of inflammatory cytokine production in AD through the degradation of cyclic adenosine monophosphate (cAMP).(3, 4) PDE4 activity is increased in circulating inflammatory cells of patients with AD, and the inhibition of PDE4 in monocytes in vitro has demonstrated reduction in the release of proinflammatory cytokines.(5-8) The novel boron chemistry of crisaborole enables synthesis of a low-molecular-weight compound (251 d) that facilitates effective penetration through human skin.(9) The specific mechanism(s) by which crisaborole exerts its therapeutic action for the treatment of atopic dermatitis is not well defined.(10) Crisaborole enhances cellular control of inflammation by inhibiting PDE4 and its ability to degrade intracellular cyclic adenosine monophosphate, thereby suppressing the release of cytokines by affecting downstream regulation of the nuclear factor-kB and nuclear factor of activated T-cell signalling pathways. (3, 11-14)
Marketing authorisation/CE mark status	Crisaborole is currently still under assessment by the EMA. Base case regulatory (EMA) assumptions for approval of crisaborole: • CHMP - 12th December 2019 • EC Decision (MA) - 17th February 2020
Indications and any restriction(s) as described in the summary of product characteristics (SmPC)	Crisaborole ointment (20 mg/g) (Staquis™) is indicated for treatment of mild to moderate atopic dermatitis in patients from 2 years of age with ≤40% body surface area (BSA) affected. The draft Summary of Product Characteristics (SmPC) includes the following: **Contraindications:** • Hypersensitivity to the active substance or to any of the excipients listed in section **Special warnings and precautions for use:** • Crisaborole is for cutaneous use only. Staquis is not for ophthalmic, oral, or intravaginal use. • Available data indicate that local skin reactions, such as burning or stinging, may be more likely to occur on sensitive skin areas such as the face and neck. • QT prolongation: Results from a thorough QT study of crisaborole applied to 60% BSA in healthy volunteers did not demonstrate QT prolongation. While there is uncertainty in extrapolating findings from healthy volunteers to patients with atopic dermatitis, clinical studies of crisaborole did not identify any cardiac effects including prolongation of QT interval. Pregnancy • There are no or limited amount of data from the use of crisaborole in pregnant women. Animal studies do not indicate direct or indirect harmful effects with respect to reproductive toxicity (see section 5.3). As a precautionary measure, it is preferable to avoid the use of Staquis during pregnancy.

	• Breast-feeding: Animal studies on milk excretion after topical application were not conducted, hence, Staquis should not be used in breast-feeding women. • Staquis is to be applied up to a maximum of 40% BSA. the small sample size of subjects with greater than 40% BSA didn't allow for a robust and conclusive determination of the benefit/risk balance across the age groups
Method of administration and dosage	Staquis is to be applied as a thin layer twice daily to affected skin areas up to a maximum of 40% BSA. Staquis can be used for up 4 weeks per treatment course. If any signs/and or symptoms persist, or new areas affected with atopic dermatitis appear, further treatment courses can be used. Staquis should be discontinued if signs and/or symptoms on treated areas persist after 3 consecutive treatment courses of 4 weeks each or if the signs and/or symptoms worsen during treatment. Staquis can be used on all skin areas except on the scalp. Use on the scalp has not been studied due to the ointment's thick formulation.
Additional tests or investigations	Crisaborole does not require special monitoring given that is a topical product and no systemic effects have been observed with its use. Only clinical evaluation is needed to prescribe and assess the treatment outcomes, which is included as an element of standard NHS Trust policies and should therefore, not be considered as additional to current clinical practice.
List price and average cost of a course of treatment	[redacted]

Abbreviations: UK: United Kingdom; mg: milligram; PDE4: phosphodiesterase 4; cAMP: cyclic adenosine monophosphate; SmPC: summary of product characteristics; w/w: by weight; NHS: National Health Service

B.1.2.1 Regulatory approval outside the UK

Crisaborole was approved in the US by the Food and Drug Administration (FDA) in December 2016, in Canada by Health Canada in November 2018, and in Israel by the Ministry of Health and Australia by the Department of Health in February 2019 for treatment of mild to moderate AD in patients two years of age and older.(10, 15, 16)

B.1.2.2 Ongoing HTAs in the rest of the UK

[redacted]

B.1.2.3 Changes in service provision and management

No additional monitoring requirements specific to crisaborole.

B.1.3 Health condition and position of the technology in the treatment pathway

- AD is a chronic inflammatory skin condition characterised by recurrent eczematous skin lesions, red patches with oozing and crusting that may lead to long-term scaling, thickening and cracking of the skin, discomfort, and persistent and severe itch.(17)

- AD can manifest at any age, however, onset peaks in infancy and approximately 50% of paediatric AD patients have recurrent symptoms into adolescence and adulthood.(18)
- An increase in the prevalence of AD in recent decades has been observed, especially in industrialised countries, with up to 90% of patients presenting with mild to moderate disease.(19-23)
- AD is most common in children between 5.3% and 23.1%. with the majority of cases of AD mild in severity.(24) Severity distribution in AD children (up to 10 years of age) was estimated up to 80% mild, up to 18% moderate with 2% severe.(25, 26)
- Many Children outgrow the disease, however, it remains common in adults, with a prevalence of 5.1% in 18-74 year olds and 8.7% in those aged 75-99.(27)
- Mild-to-moderate AD should be primarily managed at the primary care level. However, the 2015 BAD-BSPD-NICE. Paediatric National Audit on Atopic Eczema in Children showed that at the secondary care level 28.76% and 44.5% of AD patients present mild or moderate disease, respectively.(2)
- AD-associated pruritus results in frequent scratching, sleep disruption for both patients and caregivers/families, and contributes significant psychological, social and quality-of-life burdens to patients and their families.(28-30)
- NICE treatment 2007 guidelines recommend, besides emollients, the use of mild potency topical corticosteroids (TCSs) for mild AD and moderate potency TCSs for moderate AD. TCIs are recommended in moderate AD that has not been controlled by topical corticosteroids or where there is a serious risk of important adverse effects from further topical corticosteroid use, particularly irreversible skin atrophy (second-line use post-TCS treatment). TCSs and TCIs are associated with application site reactions and long-term safety concerns.
- Long-term TCS use is restricted to avoid local cutaneous atrophy (especially in sensitive and thin-skinned areas such as face and groin), striae formation, and systemic side effects.(31) Furthermore, labelling for TCSs carry Warnings and Precautions statements regarding the risk of hypothalamic-pituitary-adrenal (HPA) axis suppression and/or adrenal crisis due to systemic absorption of the corticosteroid. Cushing's syndrome, reversible HPA axis suppression, hyperglycaemia, and unmasking of latent diabetes mellitus can also result from systemic absorption of TCSs. Periocular use can lead to the development of glaucoma or cataracts. Some of the side effects described herein can be permanent. Although exact language varies across product labelling, the precautions are generally similar.
- TCIs are associated with local burning/stinging upon application and require enhanced patient education because of special warnings for increased risk of lymphoma.(32-34)
- Patients with skin of colour report a higher AD prevalence along with face specific challenges in the treatment of their AD. Treatment guidelines do not provide specific information for treating patients with skin of colour, but long-term TCS use may be associated with dyspigmentation, a concerning AE in patients with skin of colour.(31)
- Crisaborole provides a novel mode of action which directly targets pathophysiological mechanisms observed in patients with AD.(35)
- Crisaborole provides patients aged two and older with mild to moderate AD with an additional topical treatment option with a proven efficacy profile, including the rapid reduction of pruritus and multiple clinical signs.(12, 36-39)

- Crisaborole has a favourable safety profile for both short-term (28 days) and long-term use (48 weeks).(12, 36, 37, 40, 41) and is not associated with cutaneous adverse reactions, which have been reported with use of TCSs such as telangiectasia, application site atrophy, or hypopigmentation.(41)
- Crisaborole is associated with a low incidence of application site pain (defined as burning or stinging) (4.4%) (12) compared with rates of application site burning reported for tacrolimus (5-58%) (42-46) and pimecrolimus (8-26%).(42, 44, 45, 47) In contrast to TCIs, Crisaborole is not associated with special warning for a potential increased risk of cancer.(12)

B.1.3.1 Disease overview

Epidemiology

The prevalence of AD (also known as eczema) in young children has increased over the last 30 years.(21, 48, 49). It is thought this may be due to an increase in early-life exposure to environmental triggers associated with industrialisation and Western lifestyle in genetically susceptible individuals.(23). AD prevalence in children in UK, based on UK studies varies between 5.3% and 23.1 % with the majority of cases of AD mild in severity.(24) Severity distribution in AD children (up to 10 years of age) was estimated up to 80% mild, up to 18% moderate with 2% severe.(25, 26) Although mild-to-moderate AD should be managed primarily at the primary care level, the BAD-BSPD-NICE Paediatric National Audit on Atopic Eczema in Children showed that at the secondary care level 28.76% and 44.5% of AD patients present mild or moderate disease, respectively.(2) The prevalence of AD in adults is 5.1% in 18-74 year olds and 8.7% in those aged 75-99.(27)

AD onset can occur at any point in life, however, disease incidence peaks in infancy, with 60% of cases occurring within the first year of life with an additional 20% of cases occurring before 6 years of age.(23, 50) A part of adults with AD have child-onset AD, with disease persisting into adulthood while some AD adult patients have adult-onset disease. The results of a recent systematic review and meta-analysis of observational studies suggest that ~1 in 4 adults with AD report adult-onset disease.(51)

AD is also a risk factor for atopic march.(23) Atopic march was developed as a concept to describe the progression of atopic disorders from AD in infants to allergic rhinitis and asthma in children. AD is highlighted as the first step of the atopic march. The relationship between AD and respiratory allergy is influenced by AD severity. Atopic march is not present in all atopic individuals, and in particular individuals with adult-onset disease.

Disease Presentation

AD is a heterogeneous disorder associated with variable morphology, distribution, and disease course.(52) The lesions can affect any part of the body but typically show age-related morphology and distribution.(17) AD lesions show more localised distributions in children over the age of three and adults (> 18 years), presenting on the neck and flexor surfaces (body folds such as the knees and elbows).(53, 54) AD can be limited to the hands in some adults or the upper trunk, shoulders, and scalp.(17)

Infants and children tend to scratch lesions due to itch and discomfort, which may result in cutaneous infections. While acute signs of inflammation such as redness and swelling diminish with age, chronic lesions emerge and are characterised by dry, scaly, thickened, and cracked skin, which may be exacerbated by repetitive scratching. The clinical presentation may vary between ethnic groups, as Asian patients tend to have well-circumscribed, nummular lesions and lichenification, while involvement of the extensor sites and papules, a perifollicular distribution, or lichen planus-like lesions are seen mainly in dark-skinned patients.(55, 56)

The disease course of AD is unpredictable and cycles between disease flare and remission periods.(57) The seemingly unaffected skin in patients with AD show differences in skin barrier function, inflammatory cytokine production, and immune cell infiltration, suggesting ongoing subclinical disease activity.(58)

Diagnosis and Disease Severity

NICE guidelines recommend the diagnosis of AD in children when they present with an itchy skin condition plus three or more of the following criteria:(24)

- Visible flexural dermatitis involving the skin creases, such as the bends of the elbows or behind the knees (or visible dermatitis on the cheeks and/or extensor areas in children aged 18 months or under)
- Personal history of flexural dermatitis (or dermatitis on the cheeks and/or extensor areas in children aged 18 months or under)
- Personal history of dry skin in the last 12 months
- Personal history of asthma or allergic rhinitis (or history of atopic disease in a first degree relative of children aged under 4 years)
- Onset of signs and symptoms under the age of 2 years (this criterion should not be used in children aged under 4 years).

There is no consensus on which scoring instruments should be employed to measure AD severity in clinical trials. The instruments commonly used include the Investigator's Global Assessment (IGA)/Physician's Global Assessment (PGA), the Investigator's Static Global Assessment (ISGA) scale, and scoring systems such as the Eczema Area and Severity Index (EASI), and the Scoring Atopic Dermatitis index (SCORAD). These tools reflect the physician or health professional's perspective of the patient's disease during a clinical visit. All tools measure the core features of AD including the degree (on a 5 or 6-point scale) and extent of AD clinical signs (erythema, edema/swelling, excoriations, papulations, and lichenification). In the IGA/PGA/ISGA and the SCORAD oozing and crusting is also considered. The EASI and SCORAD incorporate the amount of body surface area affected by AD and, in addition, the SCORAD includes patient-reported subjective symptoms, such as pruritus, sleep disturbance and dryness.

In the pivotal trials for crisaborole, the ISGA scale (Table 3) was employed to determine baseline severity of disease and patient's subsequent response to treatment (see **Appendix D** for a complete description of the IGA and ISGA assessment tools). The ISGA provides a clinically meaningful snapshot of disease severity that is easily understood by physicians

and patients and is commonly used in clinical trials. Physicians or health professionals evaluate an ISGA score using the descriptors outlined in **Table 3** to best describe the overall appearance of lesions at the time of evaluation without reference to a previous time point (i.e. static). The ISGA scale evaluates the clinical characteristics of erythema, infiltration, papulation, oozing and crusting as guidelines for the overall severity assessment.

Table 3: Investigator's Static Global Assessment (ISGA) Scale(12)

Score	Grade	Definition
0	**Clear**	Minor residual hypo/hyperpigmentation; no erythema or induration/papulation; no oozing/crusting
1	**Almost Clear**	Trace faint pink erythema, with barely perceptible induration/papulation and no oozing/crusting
2	**Mild**	Faint pink erythema, with mild induration/papulation and no oozing/crusting
3	**Moderate**	Pink-red erythema with moderate induration/papulation with or without oozing/crusting
4	**Severe**	Deep or bright red erythema with severe induration/papulation and with oozing/crusting

Long-term assessments

The long term assessment of AD is challenging due to the chronic, relapsing nature of the disease.(59, 60) Patients with mild to moderate AD tend to experience periods of relative remission with intermittent periods of increased disease activity (i.e. flares). The rate of flare for each patient changes over time and occurs spontaneously and as such, during a long-term trial, patients may occupy several different disease states. The absence of a stable disease state precludes quantitative comparison between patients as the incidence rate of flare changes over time. Additionally, there is little consistency or consensus regarding the definition of disease flares and how to capture long-term control, across clinical studies.

A recent study undertaken by Pfizer, estimated flare rates in the UK population stratified by age and gender.(61)

[redacted]

Health-related quality of life

There are no well-validated disease-specific instruments for health-related quality of life (HRQoL) in AD. However, the most commonly used patient-reported dermatology-specific HRQoL tool that has been used in studies of AD is the Dermatology Life Quality Index (DLQI) along with a modified version validated in children and adolescents (Children's Dermatology Life Quality Index [CDLQI]).(62, 63) These HRQoL scales are used to assess the impact of AD on personal relationships, leisure, sleep, feelings, daily activities, work/school, along with evaluating the burden of treatment.

AD of all severities has been shown to be associated with important losses in HRQoL.(64) The median HRQoL loss in young adults with mild AD were similar to losses in young adults with chronic sinusitis, while those with moderate AD experienced losses similar to migraine, asthma, upper spinal problems and lower back disorders.(64)

It is noted that psychological and psychosocial comorbidities are common in AD due to the burden of pruritus and visible skin lesions, which can affect self-esteem and impede social interactions.(65, 66) Children with AD exhibit higher levels of irritability, mood changes, and sleep loss.(67) Sleep disruption, in particular, has a strong negative impact on HRQoL and can precipitate other AD-related co-morbidities.(68, 69) AD-related sleeping problems early in life are associated with an increased risk for subsequent diagnoses of attention deficit hyperactivity disorder (ADHD), and mental health disorders such as anxiety and depression.(66, 68, 70, 71) It is noted that psychological stress may also trigger AD flares, suggesting a bi-directional relationship, in which HRQoL is cyclically diminished.(72, 73) Psychological stress in early life may additionally lead to life-long-term dysfunction of the hypothalamic-pituitary-adrenal (HPA) axis, further exacerbating AD symptoms.(74)

An ongoing analysis of UK IQVIA-THIN database for the year 2017 provides an indication of the co-morbidity burden of AD in the UK.(61) Among all patients with AD,

Table 4: Demographic and Clinical Conditions of Mild-to-Moderate AD in the UK

Co-morbidity	Frequency (%)	
	All patients	>=18 years
Allergic co-morbidities e.g. asthma, allergic rhinitis)		
Neuropsychiatric		
Cardiovascular		
Metabolic and lifestyle e.g. obesity, smoking, hypertension, diabetes)		
Skin infections		
Eye disease		
Autoimmune disease		

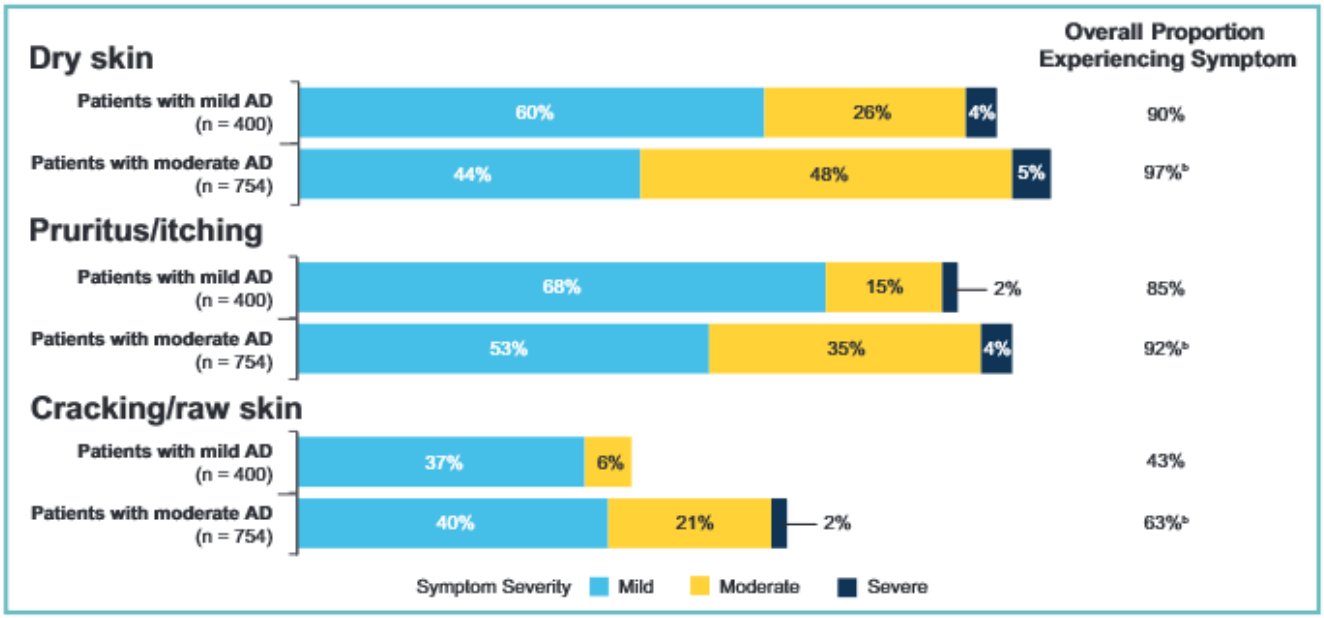
Social isolation is another commonly reported consequence of AD beginning at a very young age in children with AD.(75) In one study, more than half of interviewed parents stated that adults and children avoided interacting with their children with AD, most often for fear of a contagion.(76) Parents of children with AD also reported being worried that other children would not want to play with their child at school. Additionally, the International Study on Life with Atopic Eczema found that 27% of respondents experienced being teased or bullied due to their AD and 36% reported that AD negatively impacted their self-confidence.(77) AD has been found to negatively impact the school life of 30% of AD children (2-13 years) and 46% of AD adolescents (14-17 years).(77) In particular, AD flares caused children and adolescents an average of 2.0 and 3.5 days off school per year, respectively, and flares negatively impacted their school performance due to impaired concentration.(77)

Parents, caregivers, and families also report diminished HRQoL due to the burdens of treating and caring for, patients with AD.(36, 78) Families of children with AD report having impaired sleep, and higher levels of psychological and financial distress.(79, 80) In addition, mothers of children with AD report higher levels of depression, anxiety, and feelings of hopelessness when compared with their peers raising children unaffected by AD.(81, 82)

Adult patients with mild or moderate AD are severely impacted by disease flares despite daily treatment with currently available therapeutics. A recent real-world study drawn from the Adelphi AD Disease Specific Programme (survey results from France, Germany, and the UK), identified that more than one in five adult patients with mild AD were currently experiencing a flare.(83) Additionally, 22% of patients with mild AD experienced an increase in their AD severity to moderate or severe despite prophylactic treatment. It is estimated that current treatment options leave 88% of adults with mild and 78% of adults with moderate AD experiencing day-to-day symptoms. The most common day-to-day symptoms were dry skin, daily itch, and daily cracking/ raw skin (**Figure 1)**. Physician- and patient-reported data also indicates that a significant proportion of adult patients with mild-to-moderate AD experience flares despite ongoing treatment. Collectively, these data show a need for additional treatment options to manage symptoms during flares and quickly reduce exacerbations in pruritus.

AD flares negatively impact work performance and productivity.(77) Adults (18 years and older) reported missing an average of 2.5 days of work per year due to AD flares. It was estimated that 10% of their performance at work was negatively impacted by AD, whilst,14% of responding adults in this study believed that their career progression had been hindered by AD. An analysis of these productivity losses due to AD across the European Union (including data from the UK) estimated a cost of more than €2 billion per year.(77)

Figure 1: Proportion of adult patients with mild-to-moderate AD experiencing day-to-day symptoms by AD severity



Abbreviations: AD: atopic dermatitis
Note: Proportions among patients who experienced symptoms on a day-to-day basis
[b] *P* < 0.001 for overall proportion experiencing symptom for patients with mild AD versus patients with moderate AD

B.1.3.2 Burden of disease

Annual costs for AD in the UK exceed £800 million (including direct and indirect costs, adjusted for inflation), and Allergy UK reported that in 2015 General Practitioners (GPs) in England wrote 27 million prescriptions for the topical treatments used in the management of AD at a cost of approximately £169 million.(84-86) The 2013 Global Burden of Disease (GBD) Study reported that compared with other skin conditions dermatitis has the greatest effect on disability-adjusted life years (DALYs), a measurement that accounts for the quality and quantity of years lived.(87) According to updated GBD 2016 estimates, skin and subcutaneous diseases (primarily inflammatory in nature) represent the second leading cause of ill health in the UK.(88) This represents a significant economic burden due to high levels of healthcare resource utilisation, loss of productivity due to absenteeism (for patients and parents of children with AD), out-of-pocket costs for individual patients, and a high prevalence of chronic co-morbidities.(88)

Primary care visits in the UK for AD have been increasing annually, placing a significant strain on NHS resources.(89, 90) A recent study found that the annual GP consultation rate per person increased more than 10% from 2007 to 2014, from 4.67 to 5.16.(90, 91)

B.1.3.3 Clinical pathway of care

B.1.3.3.1 NICE and clinical guidance for mild-to-moderate AD

The primary aim of AD therapy is to reduce itch, clear inflamed and chronic lesions, and limit the frequency of disease flares. NICE recommends a holistic approach, whereby healthcare professionals assess severity as well as the impact on HRQoL including daily activities, sleep, social interactions, and psychological wellbeing (**Table 5**). It is noted that a global severity score used in clinical trials may underestimate the disease burden as experienced by the patient.(24, 36) The NICE and ISGA scale contain 4 and 5 levels of severity, respectively. In addition to clear, mild, moderate and severe, the ISGA severity scale includes 'almost clear' the definition of which aligns most closely with 'mild' on the NICE severity scale. The NICE severity definitions also include itch, a symptom not considered in the ISGA scale.

Table 5: NICE Holistic assessment of severity, quality of life, and psychosocial wellbeing

Skin/ physical severity	Definition	Impact on HRQoL and psychosocial wellbeing	Definition
Clear	Normal skin, no evidence of active atopic dermatitis	**None**	No impact on quality of life
Mild	Areas of dry skin, infrequent itching (with or without small areas of redness)	**Mild**	Little impact on everyday activities, sleep and psychosocial wellbeing
Moderate	Areas of dry skin, frequent itching, redness (with or without excoriation and localised skin thickening)	**Moderate**	Moderate impact on everyday activities and psychosocial wellbeing, frequently disturbed sleep
Severe	Widespread areas of dry skin, incessant itching, redness (with or without excoriation, extensive skin thickening, bleeding, oozing, cracking and alteration of pigmentation)	**Severe**	Severe limitation of everyday activities and psychosocial functioning, nightly loss of sleep

Abbreviations: HRQoL: health-related quality of life

Following diagnosis and severity assessments, healthcare professionals are advised to identify clinically relevant allergens and irritants that should be avoided, which may include food, inhalant and contact allergens, temperature extremes, and soaps and detergents. In addition, skin infections, such as, *staphylococcus aureus* (*S. aureus),* occur more frequently in patients with AD and are associated with lesional skin.(92) These bacterial infections exacerbate disease symptoms and require prompt treatment with antimicrobials.

The recommended treatment for AD utilises a stepwise approach according to severity. This approach aligns with NICE Clinical Guideline 57 and NICE Quality Standard 44, which focus

on children under the age of 12 years. There are currently no NICE guidelines or quality standards on the diagnosis, treatment, and management of mild-to-moderate AD in adults. Current European guidelines for the treatment and management of AD in adults are listed in **Appendix L.** The following outlines the stepwise approach noting where the NICE and European guidelines differ.

NICE's guideline currently recommends that healthcare professionals should offer children with AD a choice of unperfumed emollients to use every day for moisturising, washing and bathing. The daily use of emollients is the basis of AD management and is recommended even when skin is clear. Non-aqueous emollients soften the skin, create a physical barrier that maintains skin integrity, and reduces trans-epidermal water loss. Emollients are prescribed in large quantities (250-500g weekly) to cover all acute and chronically affected areas and application is recommended at least twice daily. Treatment is then stepped up or down in response to flares and changes in disease severity (**Table 6**.). Furthermore, the severity of AD may vary according to body site and a patient may have co-occurring mild and moderate lesions, requiring different treatment regimens.

Table 6: NICE Stepwise approach to AD management subsequent to treatment with TCS according to severity

Mild AD	Moderate AD	Severe AD
• Emollients	• Emollients	• Emollients
• Mild-potency topical corticosteroids	• Moderate potency topical corticosteroids	• Potent topical corticosteroids
	• Topical calcineurin inhibitors	• Topical calcineurin inhibitors
	• Bandages	• Bandages
		• Phototherapy
		• Systemic therapy

Abbreviations: AD: atopic dermatitis

Topical corticosteroids (TCSs) may be applied one to two times daily to areas of active AD defined as lesions that have been active in the past 48 hours. Only mild potency TCSs should be prescribed for mild AD lesions, or for areas at greater risk of skin atrophy such as the face, neck, and skin folds. In the case of severe flares, short-term (3-5 days) use of moderate potency TCSs may be warranted. Treatment for flares, as identified by increased dryness, itching, redness, swelling and general irritability, should be started as soon as possible and continued for 48 hours after symptoms subside. Moderate TCSs are appropriate for moderate AD, however, skin infections should be considered if treatment with TCSs has not controlled AD symptoms within 7-14 days. Furthermore, TCSs may be used prophylactically on problem areas for two consecutive days per week to reduce the frequency of flares in children experiencing two to three flares per month. This strategy is to be reviewed by physicians within three to six months to assess effectiveness. Further recommendations prior to stepping-up treatment, suggest that a different TCS of the same potency is considered for those patients who experience sudden drug intolerance.

An additional second-line treatment option for patients with AD are topical calcineurin inhibitors (TCIs), a class of non-steroidal, anti-inflammatory agents, which act to repair the

skin barrier and inhibit T-cell activation and proliferation.(93-95) TCIs may be prescribed where a serious risk of adverse effects from further TCS use exists, particularly irreversible skin atrophy. There are currently two NICE-recommended TCIs, tacrolimus and pimecrolimus (see **Table 7**) Neither are NICE recommended for the treatment of mild AD in the UK and they are not recommended as first-line therapy for AD of any severity. In the UK, tacrolimus is approved (label) for moderate to severe AD (the 0.03% concentration for children, adolescents and adults, and the 0.1% concentration for adolescents and adults only) and pimecrolimus is approved (label) for mild to moderate AD in children, adolescents, and adults. NICE recommendations cover the entire label of tacrolimus (i.e. tacrolimus 0.03% [children, adolescents, and adults] and tacrolimus 0.1% [adolescents (≥16-year old) and adults only] for the treatment of moderate-to-severe AD) but are more restrictive for pimecrolimus as NICE only recommends Pimecrolimus for individuals aged 2-16 with moderate AD of the face and neck. Treatment with TCIs should only be initiated by physicians specialising in dermatology, and only after discussing the risk/benefit profile of all second-line treatment options with patients and caregivers; as per the NICE guidelines. Furthermore, TCIs should only be applied to areas of active AD and should not be used under bandages or dressings without dermatological advice. In contrast to the NICE guideline, the European guideline recommends the use of TCIs as a second-line treatment for mild AD in children and adults.(54). Additionally, the European guideline recommends proactive therapy with topical tacrolimus as a first-line therapy for children and adults with moderate AD. Finally, dry bandages and medicated dressings can be used with emollients and TCSs for short-term (7-14 days) treatment of flares or areas of chronic lichenified AD.

Table 7: UK Licensed and Recommended Topical Calcineurin Inhibitors for Mild to Moderate AD

		Adult	Children
Licensed	**Mild**	Pimecrolimus 1%*	Pimecrolimus 1%*
	Moderate	Pimecrolimus 1% *, Tacrolimus 0.03%, Tacrolimus 0.1% **	Pimecrolimus 1% *, Tacrolimus 0.03%
NICE recommended	**Mild**	None	None
	Moderate	Tacrolimus 0.03%, Tacrolimus 0.1% **	Pimecrolimus 1%*, Tacrolimus 0.03%

*Note: Pimecrolimus 1% is not reimbursed for adults and adolescents >16-year old

**Note: Tacrolimus 0.1% is only licensed for the treatment of adults with moderate-to-severe AD.

B.1.3.3.2 Issues relating to current clinical practice

Prolonged or mis-use of TCSs can result in serious side-effects. Adverse events include spider-veins, skin thinning/atrophic changes, which can be permanent, increased hair, easy bruising, poor wound healing, secondary infection, acne and rosacea.(96, 97) TCSs have also been reported to produce systemic effects including immune and adrenal suppression when used in excess and over large body surface areas.(97). In addition, approximately 60-73% of patients and caregivers report “steroid phobia,” a term to describe all types of fear

about steroid use.(96, 98-102) A recent study showed that healthcare providers share TCS treatment concerns, which may adversely influence patient and caregiver behaviour.(103) In cases where steroid phobia persists despite education and counselling, other treatment options are required.

TCIs may be prescribed for patients who have failed to respond to TCSs, or for whom TCSs are inadvisable due to the affected area of the body, or concerns over low-adherence to TCS use.(104-107) However, topical tacrolimus and pimecrolimus have the special warning of potential risks of skin malignancies.(24) Patients and caregivers also consequently have significant concerns regarding the safety of TCIs, which may also lead to poor adherence. TCIs have also been associated with application-site burning, which has been reported to last up to 1 hour after application.(33, 104) In addition, TCIs may transiently worsen AD especially on acutely inflamed skin.(33, 104) and for some patients the side-effects are severe enough to justify treatment discontinuation.(33)

B.1.3.3.3 The need for additional treatment options in mild-to-moderate AD

It is noted that whilst TCS have been available to treat AD for over 50 years and TCIs for more than 15 years, recent literature still shows a significant burden of illness despite the availability of these products,(87, 88, 90, 91) indicating a high unmet need for a novel topical therapy that improves upon the perceived risk-benefit profile of TCSs and TCIs. A NICE study found that 70-80% of parents and carers of children with AD are concerned about the side effects of TCSs, and at least 25% do not use TCSs as prescribed due this concern.(24) Similar fears exist for TCIs, which carry warnings regarding increased risk of cancer. Treatment non-adherence or failure has costly reverberations given that uncontrolled AD may lead to severe HRQoL loss and caregiver burden and economic repercussions (see **Section B.1.3.1** and **B.1.3.2**).

AD-associated pruritus results in frequent scratching, sleep disruption, and contributes significant psychological, social and quality-of-life burdens to patients and their families.(28-30) NICE treatment guidelines recommend the use of mild potency TCSs for mild AD and moderate potency TCSs or TCIs for moderate AD; both TCSs and TCIs are associated with application site reactions and safety concerns. Long-term TCS use is restricted to avoid local cutaneous atrophy (especially in sensitive and thin-skinned areas such as face, neck, groin and skin folds), striae formation, and systemic side effects. TCIs are associated with local burning/stinging upon application and require enhanced patient education with special warning and precautions for use.(108, 109) Patients with skin of colour report a higher AD prevalence along with specific challenges in the treatment of their AD.(110, 111) Treatment guidelines do not provide specific information for treating patients with skin of colour, but long-term TCS use may be associated with dyspigmentation, a concerning AE in patients with skin of colour.(31)

B.1.3.3.4 Proposed positioning of crisaborole within the clinical pathway

The proposed reimbursement indication for crisaborole (**Figure 2)** is for adults and children aged 2 years and older with mild to moderate AD that has not been controlled by topical corticosteroids or where there is a serious risk of important adverse effects from further

topical corticosteroid use, particularly irreversible skin atrophy (second-line use post-TCS treatment).

Crisaborole offers a novel mode of action through its inhibition of PDE4, which modulates multiple cytokines specifically associated with the pathogenesis of AD.(35, 112) Crisaborole directly targets pathophysiological mechanisms observed in patients with AD.(40, 113)

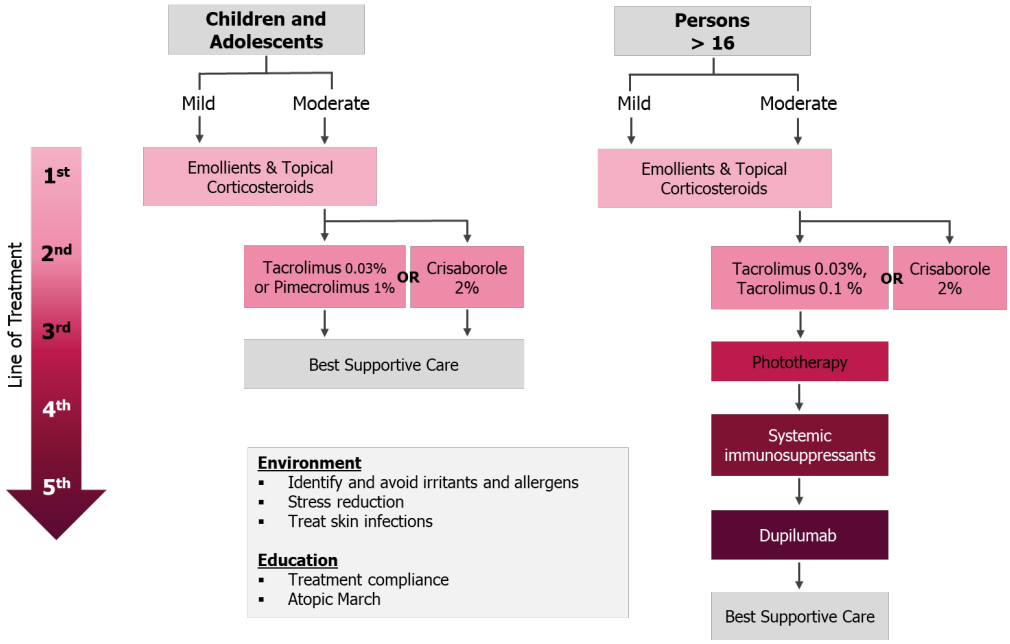
AD can have a detrimental impact on the lives of patients and their families with social, academic, and occupational impacts.(75-77) In particular, AD flares result in sleep deprivation, lack of concentration, impaired school-work productivity and absences.(77) Moreover, patients with AD report impact on self-esteem, mood, increased social isolation and decreased self-confidence. Crisaborole provides patients aged two and older with mild-to-moderate AD with an additional topical treatment option with a proven efficacy profile, including the rapid reduction of multiple AD clinical signs and symptoms.(12, 36-39) Crisaborole has a favourable safety profile for both short-term (28 days) and long-term use (48 weeks).(12, 36, 37, 40, 41) and is not associated with cutaneous adverse reactions, which have been reported with use of TCSs such as telangiectasia, application site atrophy, or hypopigmentation.(36)

In a long-term safety study crisaborole ointment demonstrated a low frequency of treatment-related AEs over 48 weeks of treatment.(41) Crisaborole shows an ability to quickly mitigate the signs and symptoms of AD with reduced safety concerns while providing long-term flare management. Additionally, crisaborole provides early (day 6) and sustained improvement in pruritus.(12, 37)

Crisaborole demonstrated a low incidence of application site pain (defined as burning or stinging) (4.4%) compared with rates of application site burning reported by tacrolimus (20-58%) and pimecrolimus (8-26%).(12, 108, 109) It is noted that a formal indirect comparison with TCS and TCI treatment was not feasible due inconsistent reporting of safety endpoints across comparator trials which resulted in a disconnected network of evidence.

Overall, crisaborole ointment represents a novel alternative to currently available pharmacologic topical therapies for the early relief and management of AD. Patients and physicians report the desire for safe treatments that reduce pruritus and manage symptom exacerbations during flares.

Figure 2: Proposed Positioning of Crisaborole in the Reimbursed Treatment Pathway



Note: Figure reflects reimbursed treatment options. Although pimecrolimus 1% is licensed for use in persons > 16, it is not reimbursed. Tacrolimus 0.1% is only licensed for the treatment of adults with moderate-to-severe disease. It is noted for simplicity bandages have not been included in the treatment pathway above.

B.1.4 Equality considerations

No equality issues are anticipated if crisaborole is recommended for use in England and Wales in accordance with its expected marketing authorisation.

B.2 Clinical Effectiveness

- The clinical effectiveness of crisaborole in mild-to-moderate AD was informed by two pivotal trials with identical design, AD-301 and AD-302 (4-week, randomised, vehicle-controlled clinical trials).
- The long-term safety of crisaborole was evaluated in AD-303, which was a long-term extension (LTE) study lasting 48 weeks and included patients who had completed either the AD-301 or AD-302 clinical trials.

AD-301 and AD-302

- **ISGA success at day 29** (defined as clear (0) or almost clear (1) with 2-grade or greater improvement from baseline): crisaborole was associated with a significant improvement in ISGA success compared to vehicle in both AD-301 and AD-302 (AD-301, 32.8% vs 25.4%, p=0.038; AD-302, 31.4% vs 18.0%, p<0.001).
- Patients treated with crisaborole achieved success in ISGA score earlier than those treated with vehicle with statistical significance at Day 8 (pooled data, p<0.001).
- **ISGA clear (0) or almost clear (1) at Day 29**: crisaborole was associated with a significant achievement of ISGA clear (0) or almost clear (1) at day 29 compared to vehicle (AD-301, 51.7% vs 40.6%, p=0.005; AD-302, 48.5% vs 29.7%, p<0.001)
- **ISGA clear (0) or almost clear (1) at Day 29 in children with moderate AD**, [redacted]
- **ISGA clear (0) or almost clear (1) at Day 29 in adults with moderate AD**, [redacted]
- **Signs of AD at Day 29,** a greater proportion of patients treated with crisaborole demonstrated improvement (defined as achieving none (0) or mild (1) with a 1-grade or more improvement from baseline) in the following signs of AD from a pooled analysis: erythema (59% vs 40%, p<0.001); exudation (40% vs 30%, p<0.001); excoriation (60% vs 48%, p<0.001); lichenification (52% vs 41%, p<0.001); and induration/papulation (55% vs 48%, p=0.008).
- **Pruritus at Day 29**, a greater proportion of crisaborole treated patients achieved improvement in pruritus Day 29 [p=0.002]) and, on average, achieved improvement earlier than patients treated with vehicle (pooled data, 1.37 vs 1.70 days, p=0.001).

Adverse Reactions

- The clinical trial programme (AD-301 and AD-302) demonstrated that treatment with crisaborole ointment 2% is well tolerated.
- No treatment-related serious adverse events (SAEs) were reported in the crisaborole treated patients including application-site atrophy.
- Adverse event rates were similar across crisaborole and vehicle study arms.
- The most frequent adverse events (AEs) reported throughout the Phase III trials were application-site pain (4.4%), which resolved within 1 day for 77.6% of patients.

- No treatment-related SAEs were reported during the long-term open-label 48-week safety study (AD-303).
- Analysis of TEAEs over time in the long-term safety study was broken into four 12-week treatment periods to monitor the accumulation of side effects. The frequency of reported AEs was similar across all four 12-week treatment periods showing that the safety profile of crisaborole is stable over time. Dermatitis atopic (i.e. worsening of disease) was the most frequently reported TEAE and occurred in 2.4-5.4% of patients over the 48 weeks.
- [redacted] In the pivotal (AD-301 and AD-302) and long-term extension (AD-303) studies, treatment-related AEs occurred in 10.3% of patients, and most (85.9%) were mild or moderate. The most frequently reported treatment-related AEs were worsening dermatitis atopic (3.1%), application-site pain (2.3%) and application-site infection (1.2%).

B.2.1 Identification and selection of relevant studies

B.2.1.1 Search Strategy

A systematic literature review (SLR) was undertaken to identify randomised controlled trials evaluating the clinical efficacy of treatments in mild to moderate Atopic Dermatitis. The SLR was performed in accordance with the methodological principles of conduct for systematic reviews as detailed in the University of York Centre for Reviews and Dissemination's (CRD) "Guidance for Undertaking Reviews in Health Care", described in **Appendix D**.

The SLR searched the Global Resource for Eczema Trials (GREAT) database in the first instance. The GREAT database contains a comprehensive list of systematic reviews and randomised controlled trials of eczema treatments identified until September 2017. An update of the GREAT database search was performed in the following electronic databases from 1 January 2017 to March 2019 as to the current time.:Medline, Embase, The Cochrane Library, CINAHL, Epistemonikos, LILACS, EconLIT and clinicaltrials.gov. Searches for publicly available information from NICE, the SMC, CADTH, PBAC, GI, AHRQ, ICER and HOME were also conducted by hand. A targeted hand search of the Tacrolimus and the Pimecrolimus FDA submissions were additionally conducted. Additional selected congresses were searched as described in **Appendix D**.

The SLR was consistent with the eligibility criteria outlined in **Appendix D, Table D10**. The inclusion and exclusion processes are summarised in a Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram in **Figure D1** and **Table 12** below. Overall, there were 2,965 citations found from the search, and of those, 2,297 citations were screened at the title/abstract phase. Further, 949 citations moved to full-text review and of those, 29 publications (17 trials represented by 16 primary publications and 14

companion publications) met the SLR inclusion criteria (**Table 12**). A large majority of studies were excluded due to not having an intervention of interest (n=718), followed by the incorrect study population (n=84). It is noted that only trials evaluating patients with mild-to-moderate AD (≥80% of the trial population) were included and therefore, most of the studies with predominantly severe populations, with no mild-to-moderate subgroup data reported, were excluded (including the tacrolimus pivotal trials, although the tacrolimus pivotal trial data were considered in the NMA sensitivity analysis). It is noted that only one TCS trial met the inclusion criteria for the SLR. However, this trial did not report the outcomes of interest and is therefore, excluded from the NMA. For further details of eligibility criteria, please refer to **Appendix D, Table D10.**

Table 8: Primary Studies and Companion Publications included in SLR

Trial number (Acronym)	Comparators	Primary publication	Companion publications
Abramovits, 2008	Tacrolimus 0.1% vs. pimecrolimus 1%	Abramovits, Journal of Drugs in Dermatology, 2008	
AD-301	Crisaborole 2% vs. vehicle	Paller, Journal of the American Academy of Dermatology, 2016	• Paller, Journal of the American Academy of Dermatology, 2016a • Paller, Pediatric Dermatology, 2016 • Hebert, Pediatric Dermatology, 2016 • Boguniewicz, Journal of Allergy and Clinical Immunology • Hebert, Pediatric Dermatology, 2016 • Zane, Journal of Immunology, 2016 • Paller, Journal of the American Academy of Dermatology, 2016b • Paller, Pediatric Dermatology, 2016 • Paller, Journal of Investigative Dermatology, 2016a • Paller, Journal of Investigative Dermatology, 2016b • Zane, Journal of Immunology, 2016 • Paller, Pediatric Dermatology, 2016c
AD-302	Crisaborole 2% vs. vehicle	Paller, Journal of the American Academy of Dermatology, 2016	• Paller, Journal of the American Academy of Dermatology, 2016a • Paller, Pediatric Dermatology, 2016 • Hebert, Pediatric Dermatology, 2016 • Boguniewicz, Journal of Allergy and Clinical Immunology • Hebert, Pediatric Dermatology, 2016 • Zane, Journal of Immunology, 2016 • Paller, Journal of the American Academy of Dermatology, 2016b • Paller, Pediatric Dermatology, 2016 • Paller, Journal of Investigative Dermatology, 2016a • Paller, Journal of Investigative

			Dermatology, 2016b • Zane, Journal of Immunology, 2016 • Paller, Pediatric Dermatology, 2016c
Boguniewicz, 1998	Tacrolimus 0.03% vs. tacrolimus 0.1% vs. vehicle	Boguniewics, The Journal of Allergy and Clinical Immunology, 1998	
Chapman, 2005	Tacrolimus 0.03% vs. vehicle	Chapman, Journal of the American Academy of Dermatology, 2005	
Eichenfield, 2002	Pimecrolimus 1% vs. vehicle	Eichenfield, Journal of the American Academy of Dermatology, 2002	
Fowler, 2007	Pimecrolimus 1% vs. vehicle	Fowler, Therapeutics for the Clinician, 2007	
Hanifin, 2016	OPA-15406 0.3% vs. OPA-15406 1% vs. vehicle	Hanifin, Journal of the American Academy of Dermatology,2016	
Hoeger, 2009	Pimecrolimus 1% vs. vehicle	Hoeger, British Journal of Dermatology, 2009	
Hordinsky, 2010	Pimecrolimus 1% vs. vehicle	Hordinsky, Dermatology, 2010	
Kaufmann, 2006	Pimecrolimus 1% vs. vehicle	Kaufmann, Allergy, 2006	
Kempers, 2004	Pimecrolimus 1% vs. tacrolimus 0.03%	Kempers, Journal of the American Academy of Dermatology, 2004	
Levy, 2005	Tacrolimus 0.03% vs. vehicle	Levy, The Journal of Allergy and Clinical Immunology, 2005	
Meurer, 2004	Pimecrolimus 1% vs. vehicle	Meurer, Pharmacology and Treatment, 2004	
Murrell, 2007	Pimecrolimus 1% vs. vehicle	Murrell, Clinical and Laboratory Investigations, 2007	
Paller, 2005	Tacrolimus 0.03% vs. pimecrolimus 1%	Paller, Journal of the American Academy of Dermatology, 2005	
Sears, 1997	hydrocortisone buteprate 0.01% vs. placebo	Sears, Clinical Therapeutics, 1997	
Schachner, 2005	Tacrolimus 0.03% vs. vehicle	Schachner, Pediatrics, 2005	
Wahn, 2002	Pimecrolimus 1% vs. control	Wahn, Pediatrics, 2002	• Papp, Pharmacology and Therapeutics, 2004

B.2.1.2 Study selection

The citations identified from the search strategy were evaluated against the pre-defined study selection criteria (PICOS: Population, Intervention, Comparators, Outcomes, Study design) listed in **Appendix D**. In the first stage of study selection, citations were evaluated based on the title, abstract and key words (where available). Those studies which appeared to meet the inclusion criteria or could not be excluded based on the exclusion criteria in this first stage were further evaluated in the second stage, based on an assessment of the full-text publications. These two stages were carried out independently by two blinded reviewers. The reviewers reconciled any discrepancies between included trials, as well as reasons for exclusion, once full-text screening was complete. If a consensus was not reached, a third reviewer provided arbitration. This resulted in the final list of included trials that proceeded to the data extraction phase.(115)

Data from included trials were extracted into a pre-specified extraction form in Microsoft Excel. Two independent, blinded reviewers extracted data on trial characteristics, intervention details, patient characteristics, and relevant outcomes for the final list of included trials. The reviewers reconciled any discrepancies between the trials, once data extraction was completed. Following reconciliation between the two reviewers, a third reviewer was included to reach consensus for any remaining discrepancies. See **Appendix D** for full details of the process and methods used to identify and select the clinical evidence relevant to the technology being appraised.

B.2.2 List of relevant clinical effectiveness evidence

The clinical SLR identified two Phase III randomised controlled trials (RCTs) of crisaborole and a Long-term extension (LTE) study in the populations relevant to the decision problem. The Phase III RCTS, AD-301 and AD-302, were identical in design and studied subjects 2 years of age and older with mild to moderate AD (**Figure 3**; **Table 9**). The LTE examined eligible patients at participating investigational sites who completed AD-301 or AD-302 (**Figure 6**; **Table 10**). All three trials contribute to the evidence base for crisaborole's AD indication.

Figure 3: Overview of the crisaborole clinical trial programme (Phase III to LTE study)

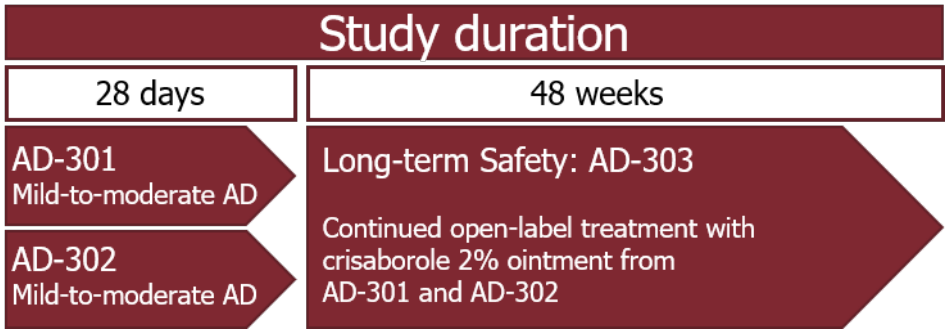


Table 9: Clinical effectiveness evidence: AD-301 and AD-302

Study	AD-301 AD-302
Study design	Identical Phase 3 randomised, multicentre, 4-week, double-blind, parallel-group, vehicle-controlled
Population	Subjects with mild-to-moderate AD, ≥ 2 years of age, who had a confirmed physician diagnosis of AD (Hanifin and Rajka criteria) affecting at least 5% of the body surface area (BSA) and a score of mild (2) or moderate (3) on a 5-grade investigator static global assessment scale (ISGA). Patients with active skin infections, use of biologics or systemic corticosteroids within 28 days, or topical corticosteroids or topical calcineurin inhibitors in the previous 14 days were excluded from the study.
Intervention(s)	Crisaborole ointment 2%, topical application BID (N=503, AD-301) (N=513, AD-302)
Comparator(s)	Vehicle ointment* topical application BID (N=256, AD-301) (N=250, AD-302) * **Use of vehicle as a comparator in clinical trials in Dermatology**: topical medications typically consist of two major components— the active ingredient and the delivery system or vehicle. Common vehicles are complex mixtures of diverse substances or excipients that serve a variety of functions, chiefly to liberate the active compound at the application site.(116-118) The vehicles also have essential properties that affect the permeation and penetration through the brick wall-like structure of the stratum corneum and percutaneous absorption of the active ingredient thereby influencing the drug efficacy. The vehicle in dermatological drug products are expected to exert an effect on its own, even without the active drug.(117, 118) Thus, the vehicle cannot be considered a true placebo.
Trial supports application for MA	Both trials support application for marketing authorisation
Trial used in the economic model	Both trials were used in the economic model
Rationale for use in the model	Phase III pivotal RCTs were included in the model because they include a population directly relevant to the decision problem

Reported outcomes specified in the decision problem[§]	• Disease severity and symptom control
	  ○ **The proportion of patients achieving an ISGA score of 0 (clear) or 1 (almost clear) at Day 29[a]**
	  ○ Time to success in ISGA
	  ○ The proportion of patients with improvement in signs of AD as measured on a 4-point scale of severity
	  ○ The proportion of patients achieving a pruritus severity score of 0 or 1 on a 4-point scale of severity on Day 2, 6, 8, 15, 22, and 29
	  ○ The proportion of patients with pruritus at baseline who became itch-free (a score of 0) on Day 2, 6, and 29
	  ○ The % change from baseline in the severity of signs of AD at Day 29
	• Health-related quality of life for patients, caregivers and families.
	  ○ **CDLQI[a], DLQI[a],** DFI
	• Adverse effects of treatment at Day 29

All other reported outcomes	• Disease activity

[§]Bolded outcomes were used to inform the economic model

Abbreviations: AD: Atopic Dermatitis; BID: twice daily; BSA: body surface area; CDLQI: Children's Life Quality Index; DLQI: Dermatology Life Quality Index; DFI: Dermatitis Family Impact; ISGA: Investigator's Static Global Assessment; MA: marketing authorisation

Table 10: Relevant non-RCT study: AD-303

Study	AD-303
Study design	Phase 3, long-term extension (LTE) study to assess the long-term safety of crisaborole in patients ≥ 2 years of age who previously participated in the controlled studies of crisaborole: AD-301 and AD-302
Population	Subjects with mild-to-moderate AD, ≥ 2 years of age, who participated in AD-301 or AD-302
Intervention(s)	Crisaborole ointment 2%, topical application BID (N=517, AD-303) Every 4 weeks, ISGA score was assessed, if the patient's ISGA score was mild or worse (≥2) at evaluation, an **on-treatment** period with crisaborole ointment was initiated; if the ISGA was clear (0) or almost clear (1), an **off-treatment** period was initiated. During on-treatment periods, **patients applied crisaborole ointment BID for 28 days** to all treatable AD-involved areas. Stopping rules included discontinuation of crisaborole ointment if there was no improvement in the subject's ISGA after three consecutive cycles of treatment **(3 months of continuous treatment)**.
Comparator(s)	N/A (None included in the study?)
Trial supports application for MA	Yes
Trial used in the economic model	Yes
Rationale for use in the model	The trial was included in the model because it includes a population directly relevant to the decision problem
Reported outcomes specified in the decision problem[§]	• Treatment exposure • Adverse effects of treatment stratified by treatment period: Week 1-12; Week 13-24; Week 25-36; and Week 37-48 • Adverse effects of treatment stratified by age: 2-11 years; 12-17 years; and ≥18 years

[a] Bolded outcomes were used to inform the economic model
Abbreviations: AD: Atopic Dermatitis; BID: twice daily; LTE: long-term extension

B.2.3 Summary of methodology of the relevant clinical effectiveness evidence

B.2.3.1 Comparative summary of RCT methodology

The methodology for the pivotal Phase III RCTs are summarised in **Table 11.**

Table 11: Comparative summary of trial methodology

Trial number (acronym)	**NCT02118766 (AD-301)**	**NCT02118792 (AD-302)**	**(AD-303)**
Location	United States (N=763)	United States (N=764)	United States (N=517)
Trial design	Phase 3 randomised, multicentre, 4-week, double-blind, vehicle-controlled, parallel treatment group		Phase 3, multicentre, long-term, open-label safety study
Eligibility criteriafo	Subjects with mild-to-moderate AD, ≥ 2 years of age, who had a confirmed physician diagnosis of AD (Hanifin and Rajka, 1980 criteria). Details of inclusion and exclusion criteria are provided in **Table 11**.		Subjects met the eligibility criteria for and successfully completed through Day 36 AD-301 or AD-302 at selected investigational sites. Details of inclusion and exclusion criteria are provided in **Table 12**.

r participants		
Settings and locations where the data were collected	The study was collected across 47 (AD-301) and 42 (AD-302) study sites across the United States	
Trial drugs Per	Crisaborole ointment, 2%, BID (N=503, AD-301) (N=513, AD-302) Vehicle ointment BID (N=256, AD-301) (N=250, AD-302) Subjects and/or parents/guardians were instructed to administer the study drug ointment BID to all treatable AD-involved areas (excluding the scalp) identified at Baseline/Day 1, regardless of whether they became clinically clear prior to Day 29. If needed, an additional amount of study drug could	[redacted]

mitted and disallowed concomitant medication	have been applied as needed to ensure coverage of any newly identified, treatable AD lesions that appeared after Baseline/Day 1. Doses were spaced approximately 8 to 16 hours apart. **Permitted and disallowed concomitant medication** After the Baseline/Day 1 Visit, the use of acceptable bland emollient(s) was permitted during the study to manage dry skin in areas surrounding but not on or overlapping the treatable AD-involved areas. [redacted] Prohibited medications during the study period included: [redacted]	[redacted] x

		[redacted]

Primary outcomes used in the economic model and/or specified in the scop	**Primary outcomes**
	• The proportion of subjects achieving success in ISGA at Day 29
	  ○ Success in ISGA was defined as an ISGA score of Clear (0) or Almost Clear (1) with at least a two-grade improvement from Baseline/Day 1.
	Supportive analysis of primary outcomes
	• Δ ISGA
	• Achievement of ISGA success
	• Proportion of subjects with pruritus scores of None (0) or Mild (1) for Days 8, 15, 22, and 29
	[redacted]Δ Treatable %BSA
	• Dermatology-related HRQoL questionnaires
	  ○ CDLQI for subjects 2-15 years; DLQI for subjects aged 16 years and older; DFI for parents/guardians of subjects aged 2-17
		Primary outcomes
		• Safety Analyses
		  ○ Safety tabulations were done by 12-week periods in order to provide a longitudinal analysis of the yearlong data

e¶		

Other outcomes used in the economic model and/or specified in the scope¶	**Secondary outcomes**
	• Proportion of subjects with an ISGA score of Clear (0) or Almost Clear (1) at Day 29
	• Time to success in ISGA
	  ○ Defined as a score of Clear (0) or Almost Clear (1) with at least a two-grade improvement from Baseline/Day 1
	
	Other outcomes
	• Time to Improvement in Pruritus
	  ○ Defined as a pruritus score of None (0) or Mild (1) with at least a one-grade improvement from Baseline/Day 1
	• Signs of AD
	  ○ Defined as erythema, induration/papulation, exudation (oozing or crusting), excoriation (evidence of scratching), and lichenification (epidermal thickening) evaluated globally on a four-point scale and not by body region
	
	Post-hoc analyses used in the economic model
	• Proportion of subjects with an ISGA score of Moderate (3) at baseline and Mild (2) at Day 29
	• Odds ratio for response vs the total population in the vehicle arm in the following subgroups:
	  ○ Mild disease
	  ○ Moderate disease
	  ○ Children
	  ○ Adults
	  ○ Patients without prior treatment
	Patients with prior treatment
		• Drug use per application (g)

Pre-pl	Subgroups of sex, age (ages 2-11 years, 12-17 years, and 18 years and older), race, and ethnicity were pre-planned.	

anned subgroups	

Abbreviations: Δ: change from baseline; AD: atopic dermatitis; ATC: anatomical therapeutic chemical classification system; BID: twice daily; BSA: body surface area; CDLQI: Children's Life Quality Index; DLQI: Dermatology Life Quality Index; DFI: Dermatitis Family Impact; ISGA: Investigator's Static Global Assessment; MA, marketing authorisation; PI, principle investigator; TCS: topical corticosteroid

B.2.3.2 Eligibility criteria

Key eligibility criteria for the pivotal Phase III RCTs are summarised in **Table 12 and Table 13** with additional eligibility criteria detailed in **Table D10** in **Appendix D**.

Table 12: Eligibility criteria for RCTs

Trial number (acronym)	NCT02118766 (AD-301) NCT02118792 (AD-302)
Inclusion criteria	• Male or female aged 2 years and older • Clinical diagnosis of AD according to the criteria of Hanifin and Rajka • AD involvement ≥ 5% treatable %BSA (excluding the scalp) • ISGA score of Mild (2) or Moderate (3) at Baseline/Day 1 • Adequate venous access to permit venepuncture for clinical safety laboratory sampling • Female subjects of childbearing potential must have agreed to use acceptable methods of contraception from the Screening Visit continuously until 30 days after stopping study drug • The ability to understand, agree to, and sign the ICF before initiation of any protocol-related procedures (adults or parent/guardian)
Exclusion criteria	• As determined by the study doctor, a medical history that may interfere with study objectives • Unstable AD or any consistent requirement for high potency topical corticosteroids • History of use of biologic therapy (including intravenous immunoglobulin) • Recent or anticipated concomitant use of systemic or topical therapies that might alter the course of AD • Recent or current participation in another research study • Females who are breastfeeding, pregnant, or with plans to get pregnant during the participation in the study • Participation in a previous crisaborole clinical trial

Abbreviations: AD: atopic dermatitis; BSA: body surface area; ICF: informed consent form; ISGA: Investigator's Static Global Assessment

Table 13: Eligibility criteria for non-RCT

Trial number (acronym)	(AD-303)
Inclusion criteria	• Male or female aged 2 years and older • Met eligibility criteria for AD-301 or AD-302, successfully completed AD-301 or AD-302 through Day 36, and enrolled into AD-303 within 8 days of Day 36 of the previous study at a participating investigator site. • Safety laboratory results from the Day 29 visit in AD-301 or AD-302 that were judged clinically acceptable in the opinion of the PI or designee • Female subjects of childbearing potential must have agreed to use acceptable methods of contraception from the Screening Visit continuously until 30 days after stopping study drug • The ability to understand, agree to, and sign the ICF before initiation of any protocol-related procedures (adults or parent/guardian)
Exclusion criteria	• Experienced a related or probably or possibly related AE or SAE during participation in AD-301 or AD-302 that precluded treatment with crisaborole ointment, 2%, in the judgement of the PI • Had a significant, active systemic or localised infection, including actively

	infected AD • Had an anticipated concomitant use of topical or systemic therapies that might alter the course of AD • Had enrolled in another drug or device research study within the 8 days between Day 36 of the previous pivotal study • Females who are breastfeeding, pregnant, or with plans to get pregnant during the participation in the study • Had a known sensitivity to any of the components of the study drug • Discontinued participation early from AD-301 or AD-302, for any reason • Had a history of noncompliance during AD-301 or AD-302 with study drug dosing, concomitant-medication restrictions, or study-required procedures, in the judgement of the PI

Abbreviations: AD: atopic dermatitis; AE: adverse events; ICF, informed consent form; pi, principal investigator; SAE: serious informed consent

B.2.3.3 Baseline characteristics and demographics

In AD-301 and AD-302, a total of 1,522 patients were randomised (2:1) to receive crisaborole (AD-301, n=503; AD-302, n=513) and vehicle (AD-301, n=256; AD-302, n=250). The enrolled patient population reflected the demographics of the overall AD population with at least 30% of patients between the ages of 2 and 6 years and no more than 15% adults (**Table 14**). The mean ages across treatment groups and the two studies were 11.8-12.6 years (range 2-79 years). The majority of the study population was white (57.6-63.3%), however, there was substantial representation across both studies of people of African family origin (23.8-31.2%). The majority of patients had moderate AD (ISGA grade 3) at baseline (60.0-63.7%) and the remainder had mild AD (ISGA grade 2). The severity of pruritus at baseline was similar across treatment arms and studies with the majority of patients reporting moderate (33.6-42.2%) or severe (29.8-35.0%) pruritus. Mean BSA was 17.7-18.8% across treatment groups in both studies. Similar baseline scores in CDLQI, DLQI, and DFI were observed across treatment groups with the majority of children, adolescents, and adults reporting a “moderate effect” of AD on HRQoL as assessed by severity bands. There were no significant demographic or baseline disease severity differences between those who received crisaborole or vehicle across both studies. The participants from AD-301 and AD-302 that were enrolled in AD-303 are further described in **Table 14**.

Table 14: Characteristics of participants across treatment groups in AD-301 and AD-302

	AD-301		AD-302	
Baseline characteristic	**Crisaborole (N=503)**	**Vehicle (N=256)**	**Crisaborole (N=513)**	**Vehicle (N=250)**
Age, (years)				
Mean	12.0	12.4	12.6	11.8
Range	2-65	2-63	2-79	2-79
Age groups, %				
2-6 y	32.3	30.5	33.7	37.2
7-11 y	30.8	28.5	26.7	28.4
12-17 y	24.1	26.2	24.6	22.8
≥18 y	12.9	14.8	15.0	11.6
Sex, %				

Male	43.5	44.1	45.0	44.8
Female	56.5	55.9	55.0	55.2
Ethnicity, %				
Hispanic or Latino	25.0	25.8	14.4	14.0
Not Hispanic or Latino	75.0	74.2	85.6	86.0
Race, %				
American Indian or Alaska Native	1.6	1.2	0.6	0.8
Asian	5.2	6.6	5.1	4.0
Black or African American	27.4	23.8	28.7	31.2
Native Hawaiian or Pacific Islander	0.0	1.6	1.4	1.6
White	61.2	63.3	60.2	57.6
Other	4.6	3.5	4.1	4.8
Baseline ISGA, %				
Mild (2)	39.0	36.3	38.4	40.0
Moderate (3)	61.0	63.7	61.6	60.0
% BSA				
Mean	18.8	18.6	17.9	17.7
Range	5-95	5-90	5-95	5-90
ISGA, %				
N	503	256	513	250
0 – Clear	0.0	0.0	0.0	0.0
1 – Almost Clear	0.0	0.0	0.0	0.0
2 – Mild	39.0	36.3	38.4	40.0
3 – Moderate	61.0	63.7	61.6	60.0
4 – Severe	0.0	0.0	0.0	0.0
Severity of Pruritus Scale, %				
N	446	223	457	218
0 – None	3.8	5.8	3.9	2.8
1 – Mild	25.8	28.7	24.9	25.2
2 – Moderate	35.4	33.6	37.9	42.2
3 – Severe	35.0	31.8	33.3	29.8
DLQI				
N	95	52	97	40
Mean (SD)	9.6 (6.37)	9.5 (6.52)	9.7 (6.24)	9.5 (6.52)
Median	9.0	8.0	8.0	8.0
Range	1-27	0-27	0-26	0-27
CDLQI				
N	393	199	404	204
Mean (SD)	9.7 (6.19)	9.1 (6.54)	9.0 (5.77)	8.9 (5.48)
Median	9.0	7.0	8.0	8.0
Range	0-28	0-30	0-28	0-27
DFI				
N	431	214	431	217
Mean (SD)	8.5 (6.63)	7.5 (6.66)	7.7 (6.57)	8.0 (5.65)
Median	7.0	6.0	6.0	7.0
Range	0-30	0-30	0-30	0-24

Abbreviations: AD: atopic dermatitis; BSA: body surface area; CDLQI: Children's Life Quality Index; DLQI: Dermatology Life Quality Index; DFI: Dermatitis Family Impact; ISGA: Investigator's Static Global Assessment score

Table 15: Characteristics of participants in AD-303

	AD-303
Baseline characteristic	**Crisaborole (N=517)**
Age, (years)	
Mean	11.7
Range	2-72

Age groups, %	
2-11 y	59.6
12-17 y	28.2
≥18 y	12.1
Sex, %	
Male	40.8
Female	59.2
Ethnicity, %	
Hispanic or Latino	15.9
Not Hispanic or Latino	84.1
Race, %	
American Indian or Alaska Native	0.2
Asian	5.4
Black or African American	29.4
Native Hawaiian or Pacific Islander	0.2
White	60.9
Other	3.9
Treatment received in AD-301 or AD-302, %	
Crisaborole ointment	69.1
Vehicle	30.9
Patients included in each 12-week period, n	
Week 1-12	482
Week 13-24	428
Week 25-36	368
Week 37-48	226

Abbreviations: AD, Atopic Dermatitis

B.2.4 Statistical analysis and definition of study groups in the relevant clinical effectiveness evidence

B.2.4.1 Analysis sets

The main analysis sets in the AD-301 and AD-302 are defined below.

The Full Analysis Set (FAS): All subjects who were randomised and dispensed study drug (Intention-to-treat [ITT] population) were included in the FAS. Subjects who prematurely discontinued from the study for any reason were included in the FAS. Efficacy analyses performed using the FAS were considered primary.

The Per-Protocol Set (PP): The PP included all subjects in the ITT who completed the Day 29 evaluation without any major protocol deviations. Subjects who prematurely discontinued from the study due to lack of efficacy, worsening of AD, or a treatment-related TEAE were included in the PP and the last ISGA value was carried forward. Efficacy analyses performed using the PP were considered supportive. The PP included subjects in the ITT who met all of the following criteria:

- Met all the Inclusion Criteria and none of the Exclusion Criteria (**Table 11**)
- Had not taken any interfering concomitant medications or therapies during the 29-day study period
- Completed the Day 29 Visit, including the Day 29 efficacy evaluation
- Had applied 80%-120% of the total number of expected doses during the Study Drug Application Period

- Had not missed six or more consecutive doses during the Study Drug Application Period
- Were in the visit window ($\pm$3 days) for the Day 29 Visit

The Safety Analysis Set (Safety): All subjects who were randomised, received at least one confirmed dose of study drug (crisaborole or vehicle), and had at least one post-baseline assessment were included in the Safety Set.

B.2.4.2 Statistical information

A summary of the statistical methods used in the AD-301 and AD-302 RCTs are presented in **Table 16.**

Table 16: Summary of statistical analyses

Trial number (acronym)	NCT02118766 (AD-301)	NCT02118792 (AD-302)
Hypothesis objective	To establish superiority of crisaborole ointment, 2%, to vehicle ointment for treatment of the severity, signs, and detrimental impact on the quality of life of mild-to-moderate AD in subjects 2 years and older.	
Multiple Comparisons/ Multiplicity	No adjustments of p-values for multiple comparisons were required for the primary efficacy endpoint.	
Sample Size, power calculation	For each study (AD-301 or AD-302), approximately 750 ITT subjects were randomised in a 2:1 ratio of approximately 500 subjects to active treatment (crisaborole) and approximately 250 subjects to vehicle treatment (vehicle). The sample size was selected for efficacy to yield at least 90% power to achieve a statistically significant difference (2-sided test at α = .05), assuming success rates of 20% (patients treated with crisaborole) and 10% (group of patients treated with vehicle). To achieve the target sample size, approximately 1000 subjects were screened, assuming a 33% screen failure rate.	
Statistical analysis of primary endpoints	Efficacy analyses were performed using the ITT population. **Primary Analysis:** Success in ISGA was defined as an ISGA score of Clear (0) or Almost Clear (1) with at least a two-grade improvement from Baseline/Day 1. The odds ratio of success in ISGA score at Day 29 were tested between treatment groups using logistical regression with factors for treatment group and analysis centre.	
Statistical analysis of secondary and other endpoints	Hypothesis testing for the secondary endpoints was conducted in a sequential (gatekeeping) manner, beginning with a test for secondary endpoint 1, followed by secondary endpoint 2. Secondary endpoint 2 was considered statistically significant only if secondary endpoint 1 was statistically significant. For example, time to success in ISGA was only tested if ISGA of Clear or Almost Clear at Day 29 was statistically significant. **Secondary Endpoint 1:** The proportion of subjects with an ISGA score of Clear (0) or Almost Clear (1) was tested between treatment groups using logistical regression with factors for treatment group and analysis centre **Secondary Endpoint 2:** Time to success in ISGA score was analysed by Kaplan-Meier methods and log-rank test. Subjects with missing values were censored at the time of the last observation.	

	Additional Efficacy Endpoint 1: Time to Improvement in Pruritus (defined as a pruritus score of None [0] or Mild [1] with at least one-grade improvement from Baseline/Day 1) was analysed using Kaplan-Meier methods and log-rank test. **Additional Efficacy Endpoint 2:** Signs of AD (erythema, induration/papulation, exudation, excoriation, and lichenification evaluated globally on a 4-point scale and not by body region. These results were summarised by treatment group using descriptive statistics. **Supportive Efficacy Analysis:** Other efficacy endpoints were summarised using descriptive statistics for each treatment group and each visit for the ITT and PP populations. Categorical variables were tabulated with frequencies and percentages. Continuous variables were tabulated with frequencies and percentages. The supportive efficacy analyses included ISGA scores; the proportion of subjects with pruritus scores of None (0) or Mild (1) for Days 8, 15, 22, and 29; dermatology-related quality of life (HRQoL) scores from the CDLQI, DLQI and DFI; and treatable percent body surface area. The established minimal clinically important difference (MCID) for each HRQoL instrument (CDLQI: ≥2.5-point change from baseline; DLQI: ≥3.3-point change from baseline) was used in the analysis assessment. No MCID has been established for the DFI. Differences between treatment groups in absolute change from baseline were analysed by Wilcoxon rank sum test, a nonparametric test. Differences between treatment groups regarding the percentage of patients who experienced MCID for the CDLQI and DLQI at day 29 were analysed using the Fisher exact test. A Wilcoxon rank sum test was used to analyse differences between treatment groups in severity bands. **Subgroup Analysis:** Subgroups of sex, age (ages 2-11 years, 12-17 years, and 18 years and older), race, and ethnicity were analysed for the ITT Population for the primary efficacy outcome and contained only descriptive statistics for Day 29. ***Post hoc* Analysis:** Populations from both studies were pooled. Early improvement of pruritus was defined as experiencing improvement at Day 6. Endpoints in the *post hoc* analysis included the proportion of patients who experienced early improvement in pruritus, the proportion who experienced pruritus symptom improvement at earliest assessment (Day 2), and the percentage reduction in pruritus severity over the first 6 days of treatment. The likelihood of early improvement in pruritus based on baseline demographics and disease characteristics was also evaluated. HRQoL scores for CDLQI and DLQI used in the correlation with early improvement of pruritus were based on the established MCID for each scale (CDLQI: ≥2.5-point change from baseline; DLQI: ≥3.3-point change from baseline). The correlation between early improvement in pruritus and treatment outcomes such as improvement in ISGA, HRQoL measures, other signs of AD and higher sleep scores (as a component of CDLQI) at Day 29 were examined in patients treated with crisaborole. This analysis also evaluated the proportion of patients who began the study with itch and became itch free (defined as achieving a pruritus severity score of 0 among patients with baseline pruritus severity of mild or worse [score ≥1]) on Days 2, 6 and 29, and the proportion of patients who experienced early improvement in pruritus and maintained improvement at day 29. Differences in proportions between treatment groups for improvement in pruritus were compared using the Fisher

	exact test. Differences in proportions between treatment groups for patients who became itch free were compared using normal approximation to binomial proportions. Differences in percentage change were compared using analysis of variance with a factor of treatment. Odds ratios with corresponding confidence intervals and p-values were found from a logistic regression with a factor of treatment. The statistical significance was set at a 0.05 level.
Data management, subject withdrawals	**Missing Data:** Missing Day 29 ISGA values were derived for the analysis using the method of Markov Chain Monte Carlo (MCMC) multiple imputation. The estimation was performed for each treatment group separately so that the pattern of missing data for one group did not influence the estimation of missing data for another group. Groups of complete data sets following the estimation were concatenated to form analysis data sets for the comparative analyses and subsequent imputation result inference. **Safety:** No imputation was made for missing safety data ***Post hoc* analysis:** No imputation was made for missing data **Withdrawals:** Subjects who withdrew from the trial were considered to have no response at any visit after discontinuation for *post hoc* analysis.

Abbreviations: AD: atopic dermatitis; CDLQI: Children's Life Quality Index; DLQI: Dermatology Life Quality Index; DFI: Dermatitis Family Impact; MCID: minimal clinically important difference; MCMC: Markov Chain Monte Carlo; PP: protocol population; ISGA: Investigator's Static Global Assessment score; ITT: intent-to-treat

B.2.4.3 Participant flow in the relevant randomised controlled trials

See **Figure 4**, **Figure 5** and **Figure 6** for details of the numbers of participants eligible to enter the trials and the participant flow throughout the trial duration.

Figure 4: Participant Flow for AD-301 and AD-302

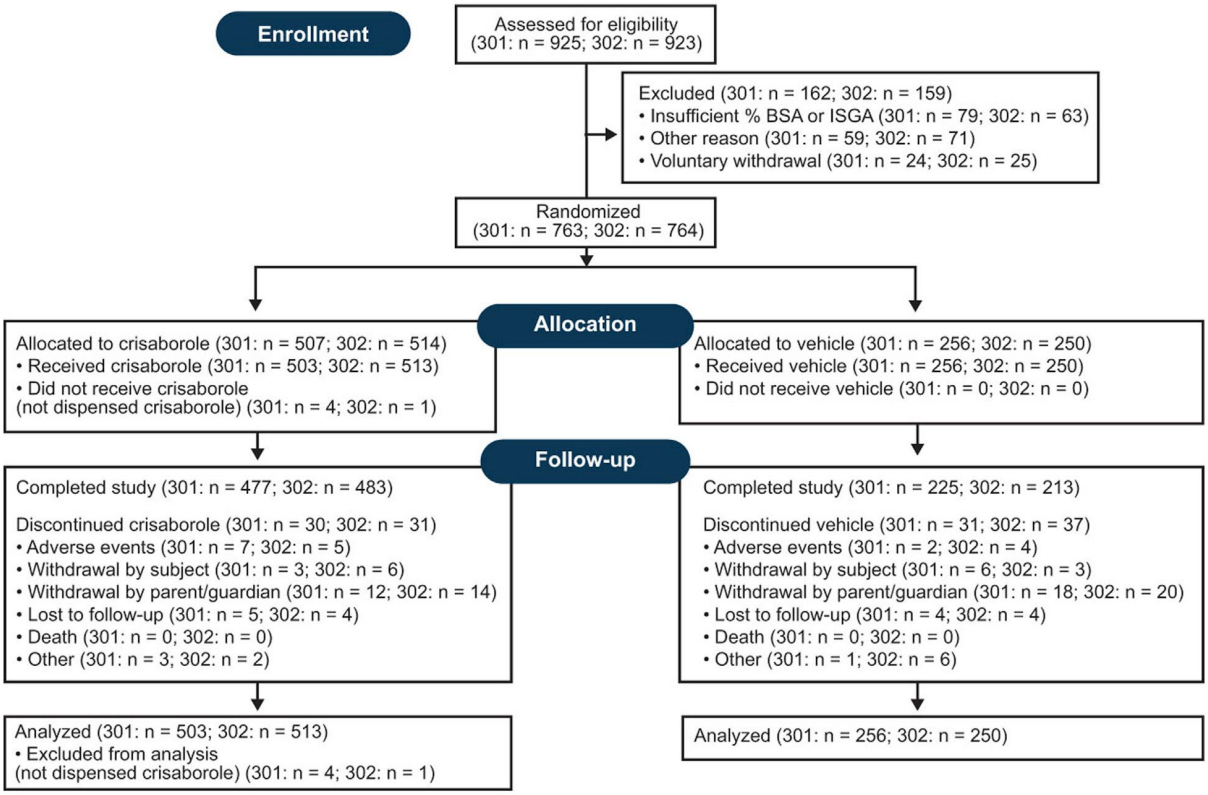
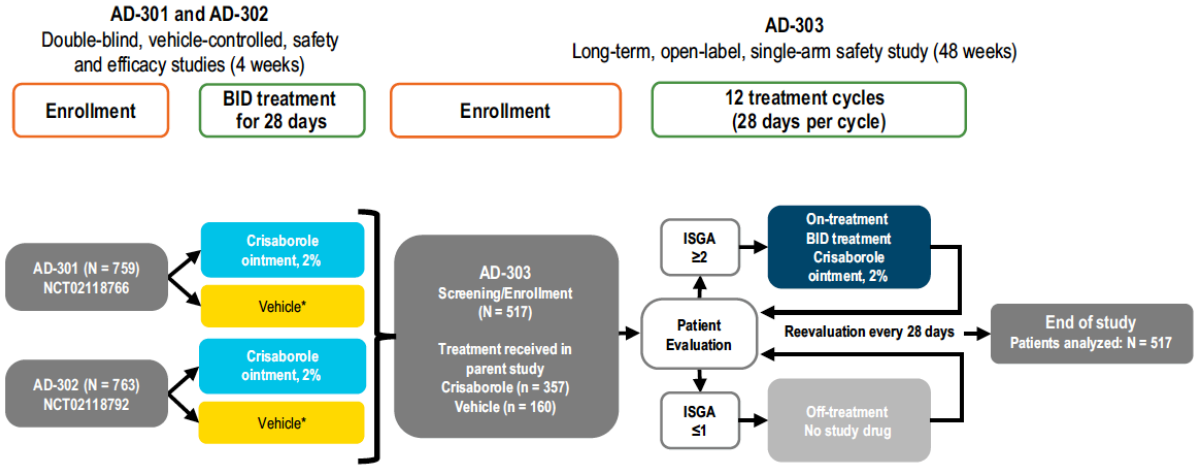
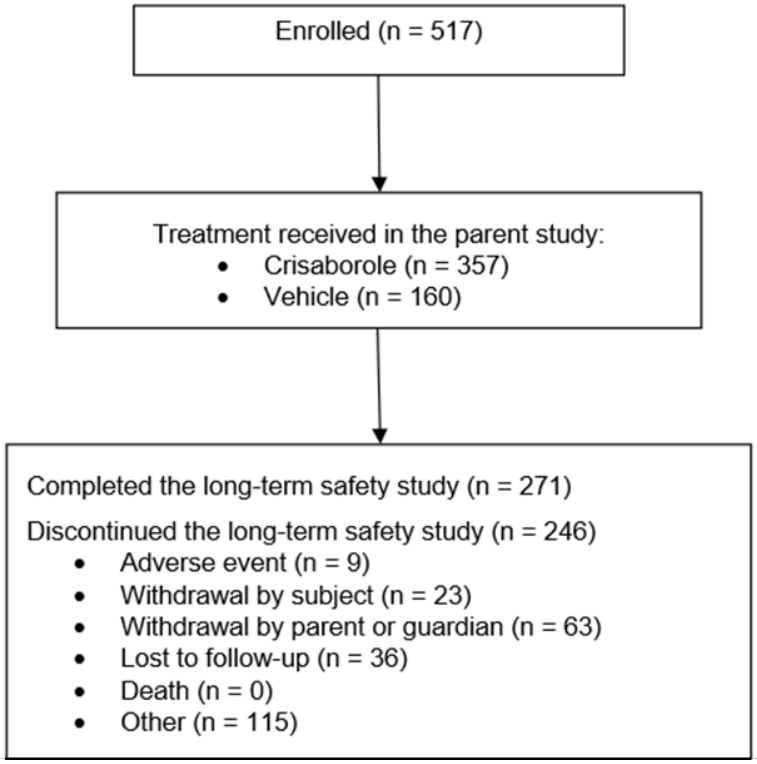


Figure 5: Trial design for AD-301, AD-302 and AD-303



Abbreviations: BID, twice daily; ISGA, Investigator's Static Global Assessment score

Figure 6: Participant Flow for AD-303



B.2.5 Quality assessment of the relevant clinical effectiveness evidence

See **Section D.1 (Appendix D – Tables D20 and D21)** for quality assessment of the relevant trials in the crisaborole trial programme.

B.2.6 Clinical effectiveness results of the relevant trials

B.2.6.1 Summary of outcome measures in AD-301 and AD-302

AD-301 and AD-302 were identical in design and in total enrolled 1522 mild-to-moderate AD subjects 2 years of age and older. All clinical trial sites were located in the United States and the primary efficacy end point was ISGA success.(12) The vehicle ointment was formulated with an emollient base, which is itself a treatment for AD as per NICE guidelines and clinical practice.(24) The vehicle in the crisaborole trials was the base ointment used for crisaborole ointment and it was a match in color, consistency, and packaging. Vehicles in different trials are formulated with different emollient properties.(119) It is known that some vehicle excipients, such as those used in crisaborole, have a more pronounced benefit than previously thought, and may improve barrier function and skin appearance.

[redacted]

[redacted].(120) Further discussion of vehicle properties and a comparison of vehicle response rates across clinical trials can be found in **Appendix D**.

B.2.6.1.1 Main efficacy outcomes

- ***ISGA Score: Proportion of subjects achieving success in ISGA***

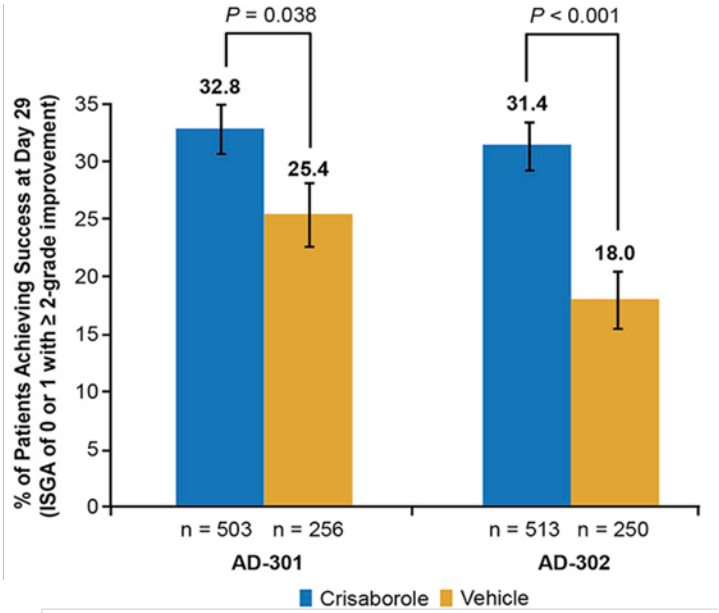
ISGA Success was defined as an ISGA score of Clear (0) or Almost Clear (1) with at least a 2-grade improvement from Baseline/Day 1. A significantly greater proportion of patients treated with crisaborole achieved success in ISGA at day 29 compared with vehicle treated patients in both trials, despite a strong vehicle effect (AD-301 32.8% vs 25.4%, p=0.038; AD-302 31.4% vs 18.0%, p<0.001; **Table 17** and **Figure 7**.

Table 17: Summary of primary efficacy results for AD-301 and AD-302 at Day 29

Outcome	AD-301		AD-302	
	Crisaborole	Vehicle	Crisaborole	Vehicle
Success in ISGA [a] at Day 29				
N	503	256	513	250
Success, %	32.8	25.4	31.4	18.0
p-value[b]	0.038	--	<0.001	--

[a]Success in ISGA was defined as ISGA of Clear or Almost Clear with at least a 2-grade improvement from Baseline/Day 1
[b]The p-value from a logistic regression (with Firth option) test with factors of treatment group and analysis centre
Abbreviations: AD, Atopic Dermatitis; ISGA, Investigator's Static Global Assessment

Figure 7: Proportion of patients achieving ISGA success at Day 29



Abbreviations: AD, Atopic Dermatitis; ISGA, Investigator's Static Global Assessment

B.2.6.1.2 Secondary and other efficacy outcomes

- ***ISGA Score: Proportion of subjects achieving an ISGA score of Clear (0) or Almost Clear (1)***

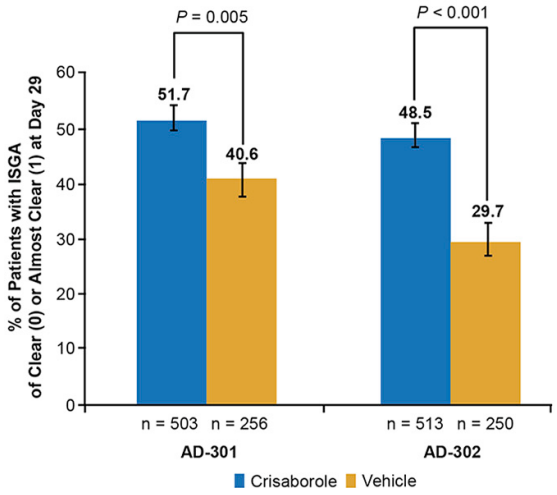
A significantly greater proportion of patients treated with crisaborole achieved ISGA scores of Clear (0) or Almost Clear (1) compared to vehicle treated patients at day 29 (AD-301 51.7% vs 40.6%, p=0.005; AD-302 48.5% vs 29.7%, p<0.001; **Table 18** and **Figure 8**). An additional 18.9% (AD-301) and 17.1% (AD-302) of patients treated with crisaborole, while not achieving the primary endpoint of ISGA success, were Almost Clear at day 29. It is noted that in order to achieve ISGA success, patients with baseline ISGA of Mild (2) had to be completely Clear at day 29 (score of 0).

Table 18: Proportion of patients with an ISGA score of Clear or Almost Clear for AD-301 and AD-302 at Day 29

Outcome	AD-301		AD-302	
	Crisaborole	Vehicle	Crisaborole	Vehicle
ISGA of Clear or Almost Clear at Day 29				
N	503	256	513	250
Success, %	51.7	40.6	48.5	29.7
p-value[a]	0.005	--	<0.001	--

[a]The p-value from a logistic regression (with Firth option) test with factors of treatment group and analysis centre.
Abbreviations: AD: atopic dermatitis; ISGA: Investigator's Static Global Assessment; N: population number

Figure 8: Proportion of patients achieving ISGA score of Clear (0) or Almost Clear (1) at Day 29



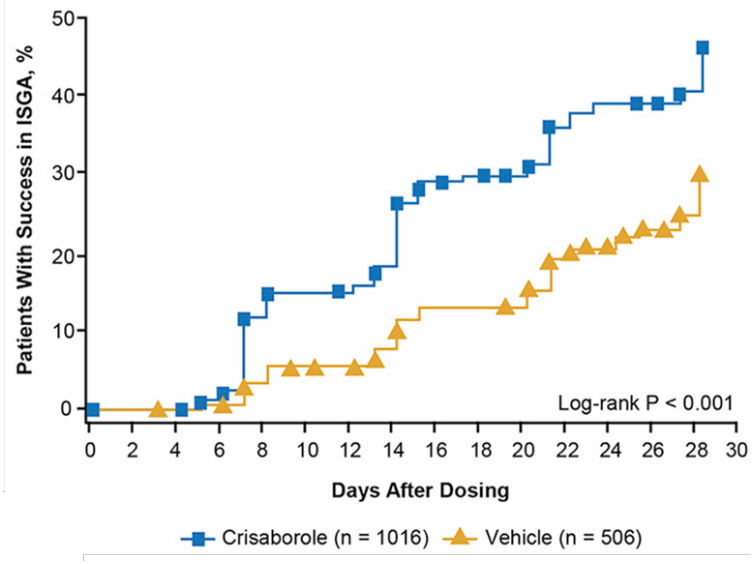
Abbreviations: AD, Atopic Dermatitis; ISGA, Investigator's Static Global Assessment

- ***ISGA Score: Time to Success in ISGA score***

Additionally, patients treated with crisaborole achieved success in ISGA score earlier than those treated with vehicle as demonstrated by Kaplan-Meier analysis **Figure 9.** The

separation in efficacy was observable by the first on-treatment visit (day 8) and continued throughout the treatment period (p<0.001).

Figure 9: Kaplan-Meier of pooled patients from AD-301 and AD-302 achieving success in ISGA from baseline through Day 29

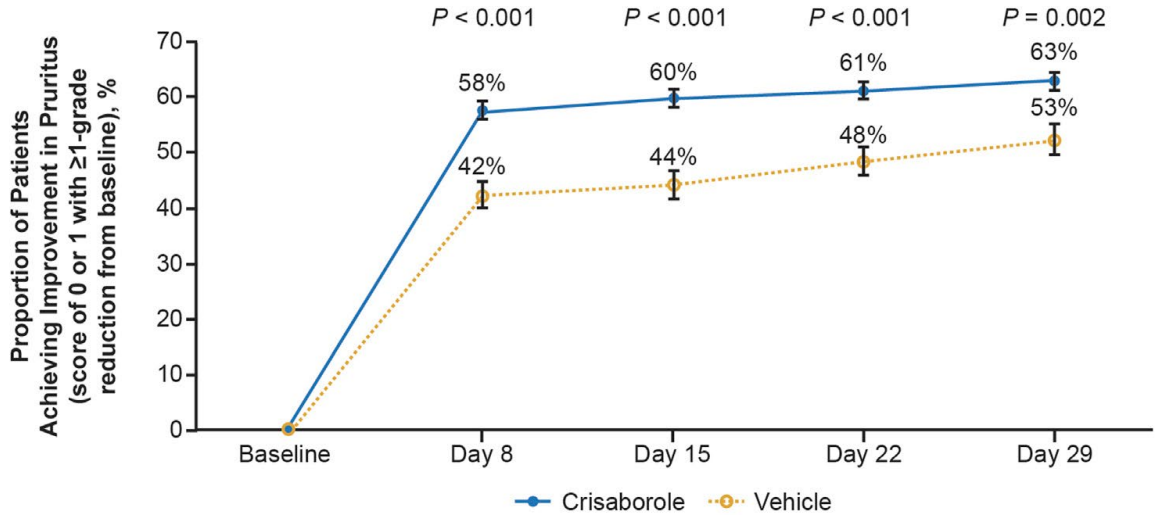


Abbreviations: AD, Atopic Dermatitis; ISGA, Investigator's Static Global Assessment
Success in ISGA was defined as ISGA of Clear or Almost Clear with at least a 2-grade improvement from Baseline/Day 1
The p-value was from a log-rank test. The median time to success in ISGA as a second secondary endpoint could not be calculated, as fewer than 50% of subjects reached success in ISGA

- ***Impact on Pruritus***

Patients rank pruritus as one of the most burdensome manifestations of AD see section **B.1.3.2**. Pruritus and pruritus-induced scratching precipitates disease-escalating side effects such as disrupted sleep, elevations in stress, increased inflammation, and increased risk of skin infection. Patients treated with crisaborole achieved improvement in pruritus earlier than patients treated with vehicle (pooled data, 1.37 vs 1.70 days, p = 0.001). Across all visits, a greater proportion of patients treated with crisaborole achieved improvement in pruritus (**Figure 10**).

Figure 10: Proportion of pooled patients from AD-301 and AD-302 achieving improvement in Pruritus at the earliest evaluation through Day 29



Abbreviations: AD, Atopic Dermatitis; ISGA, Investigator's Static Global Assessment
Improvement in Pruritus was defined as None or Mild with at least a 1-grade improvement from Baseline/Day 1

- ***Impact on disease severity***

A greater proportion of patients treated with crisaborole achieved improvement in multiple clinical signs of AD at Day 29 when compared with patients treated with vehicle (p<0.008-0.001, **Figure 11**). Additionally, patients treated with crisaborole showed greater mean reductions in severity of AD signs (p<0. 002-0.001,**Figure 12**).

Figure 11: Proportion of pooled patients from AD-301 and AD-302 with improvement in AD signs at Day 29

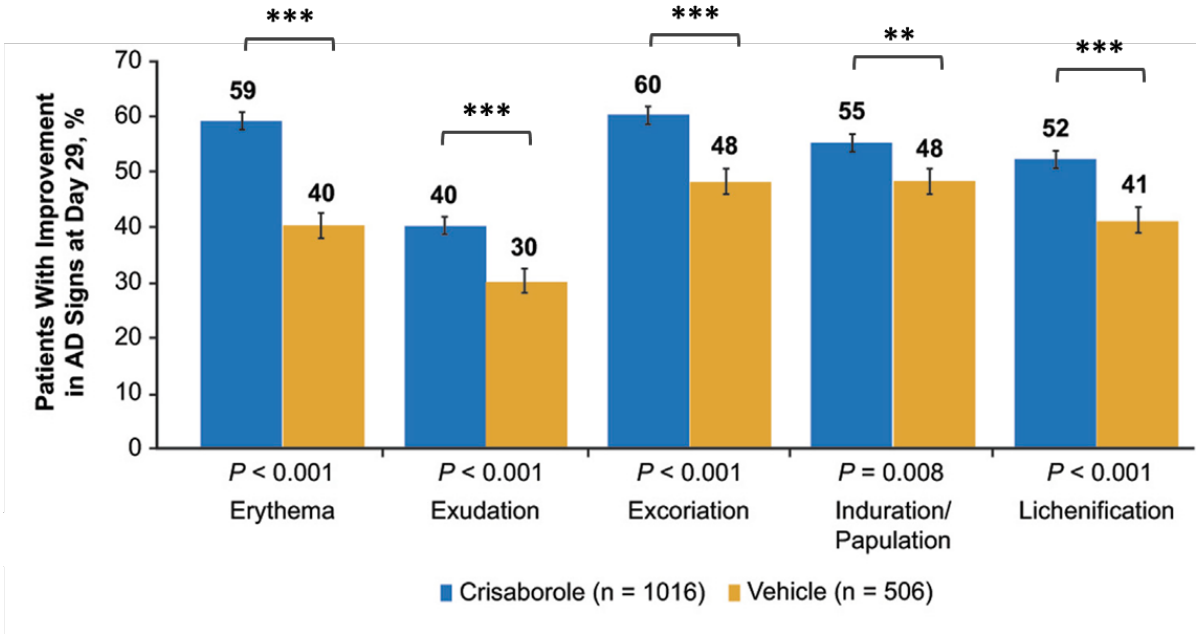
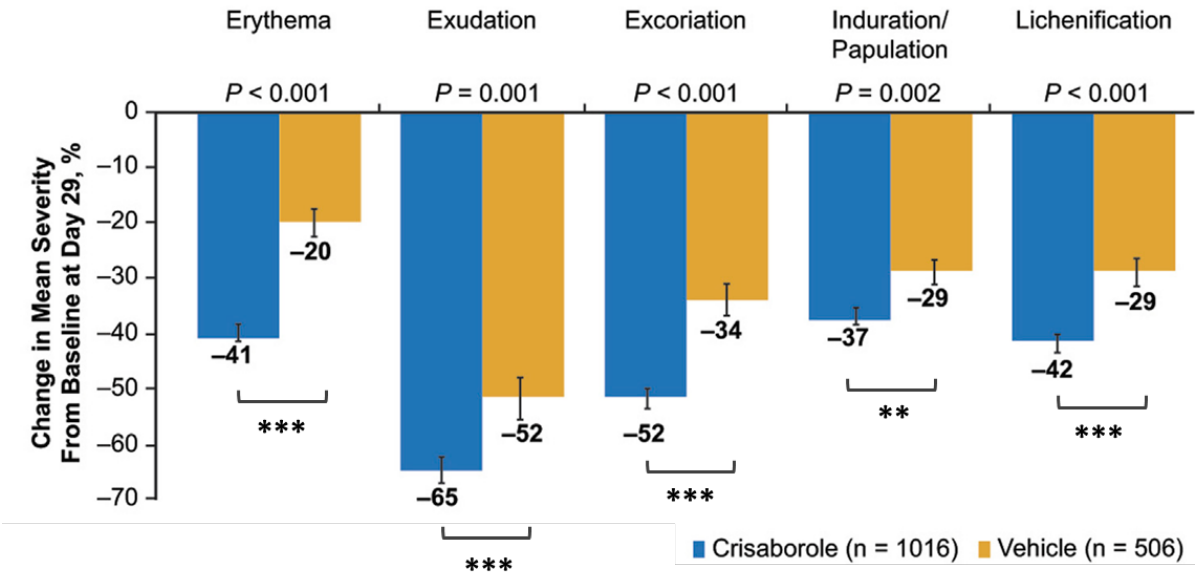


Figure 12: Change in mean severity of AD signs of pooled patients from AD-301 and AD-302 at Day 29



- ***Impact on Health Related Quality of Life for Patients, Caregivers, and Family***

Analyses for HRQoL are presented here based on a pooled analysis across both of the trials.(36) In these analyses, children and adolescents treated with crisaborole aged 2-15 years showed a significantly greater improvement in HRQoL at Day 29 as measured by the CDLQI assessment compared with patients with vehicle alone (mean change from baseline, crisaborole: -4.6; vehicle: -3.0; p<0.001; **Table 19**) The CDLQI is only validated for use in patients as young as 4 years old.(62) However, results were similar and retained significance among the subgroup of patients aged 4-15 years (mean change from baseline, crisaborole: - 4.5; vehicle: - 2.6; p<0.001; **Table 19**). The minimal clinically important difference (MCID) is the smallest amount of change in the assessment instrument score that would be identified as important and which would result in a change in the patient's management.(121-123) This may be related to reduction in symptoms or improvement in function the patient perspective, whilst for the physician this may be related to a change in treatment or in prognosis.(121, 122, 124, 125) The MCID for the CDLQI has been reported as ≥ 2.5-point change from baseline (93), whilst a longitudinal study in patients with inflammatory skin diseases showed evidence for an MCID of ≥3.3 point change from baseline for the DLQI.(124)

In children and adults aged 2-15 years, a significantly greater proportion of patients treated with crisaborole achieved MCID at Day 29, than patients treated with vehicle (crisaborole: 61.7%; vehicle: 52.1%; p=0.003; **Table 19**).

Table 19: CDLQI at Day 29

Outcome	2-15 years		4-15 years	
	Crisaborole	Vehicle	Crisaborole	Vehicle
Day 29				
N	750	355	614	297
Mean change from baseline in CDLQI	-4.6	-3.0	-4.5	-2.6
p-value	<0.001	--	<0.001	--
Proportion achieving MCID, %	61.7%	52.1%	--*	--*
p-value	0.003	--	--	--

Abbreviations: CDLQI: Children's Life Quality Index; MCID: minimal clinically important difference; N: population number
Note: Because the CDLQI is only validated for use in patients as young as 4 years old, change in CDLQI score was also assessed in the group of patients aged 4-15 years (62); the MCID for the CDLQI is ≥ 2.5-point change from baseline
p values are versus vehicle control, pooled data from AD-301 and AD-302
*MCID estimates are not available for patients of 4-15 years of age

Patients treated with crisaborole 16 years of age and older also showed significantly greater improvement in HRQoL as measured by DLQI despite a smaller sample size compared to patients under the age of 16 (mean change from baseline, crisaborole: - 5.2; vehicle: - 3.5; p = 0.015; **Table 20**). Additionally, a greater proportion of patients treated with crisaborole (53.9%) achieved MCID, than patients treated with vehicle (41.5%), with borderline significance (p=0.083; **Table 20**).

Table 20: DLQI at Day 29

Outcome	≥ 16 years	
Day 29	**Crisaborole**	**Vehicle**
N	182	82
Mean change from baseline in DLQI	-5.2	-3.5
p-value	0.015	--
Proportion achieving MCID, %	53.9%	41.5%
p-value	0.083	--

Abbreviations: DFQI: Dermatology Life Quality Index; MCID: minimal clinically important difference; N: population number
Note: the MCID for the DLQI is ≥ 3.3-point change from baseline
p values are versus vehicle control, pooled data from AD-301 and AD-302

AD is burdensome not only for the patient with AD, but for the family as a whole, as discussed in section **B.1.3.1**. The Dermatitis Family Impact (DFI) scale captures ten domains in which AD negatively impacts the HRQoL of parents, caregivers, and families of patients 2-17 years of age.(123) Assessment of improvement in HRQoL as measured by the DFI showed that the parents, caregivers, and families of patients treated with crisaborole treated patients had greater mean reduction overall at Day 29 in the impact of AD (mean change from baseline, crisaborole: -3.7; vehicle: -2.7; p=0.003; **Table 21**).

Table 21: DFI at Day 29

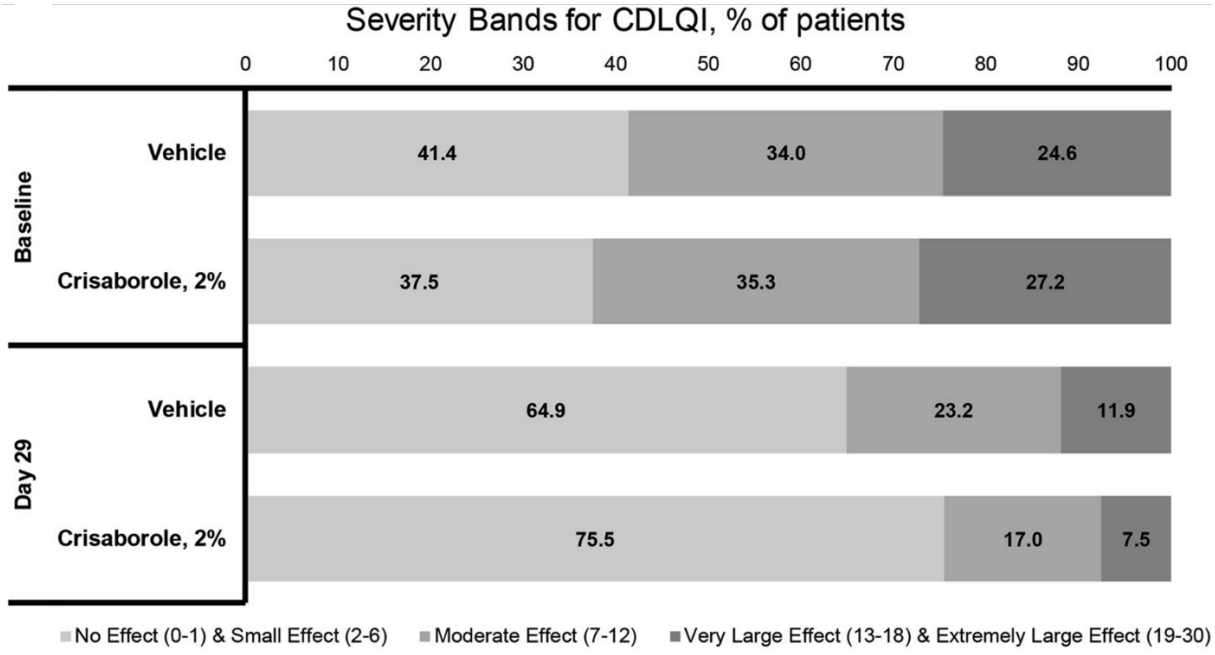
DFI at Day 29	Crisaborole	Vehicle
N	811	377
Mean change from baseline in DFI	-3.7	-2.7
p-value	0.003	--

Abbreviations: DFI: Dermatitis Family Impact, N: population number
p values are versus vehicle control, pooled data from AD-301 and AD-302

HRQoL score severity bands provide clinical interpretation of CDLQI and DLQI scores.(126, 127) Severity bands range from “no effect” to “extremely large effect.” At baseline, most children and adolescents had a “moderate effect” or higher of AD on HRQoL (crisaborole: 62.5%; vehicle 58.6%; p=0.1769). In total, 75.5% of patients treated with patients treated with crisaborole reported that their AD had “small effect” to “no effect” (**Figure 13**) by day 29, This improvement in HRQoL was significantly greater in patients treated with crisaborole when compared with patients treated with vehicle (p=0.0002; **Figure 13**). Similar trends were seen for patients 16 years of age and older: most reported that AD had a “moderate effect” or worse on HRQoL at baseline (crisaborole: 68.8%; vehicle: 64.1%; p=0.4030). Additionally, a greater proportion of patients treated with crisaborole reported that AD had a “small effect” to “no effect” than patients treated with vehicle at day 29 (crisaborole: 71.8%; vehicle 65.5%; p=0.1400; **Figure 15**). Crisaborole treatment reduced the impact of AD on the majority of treated patients and to a greater degree than vehicle treatment.

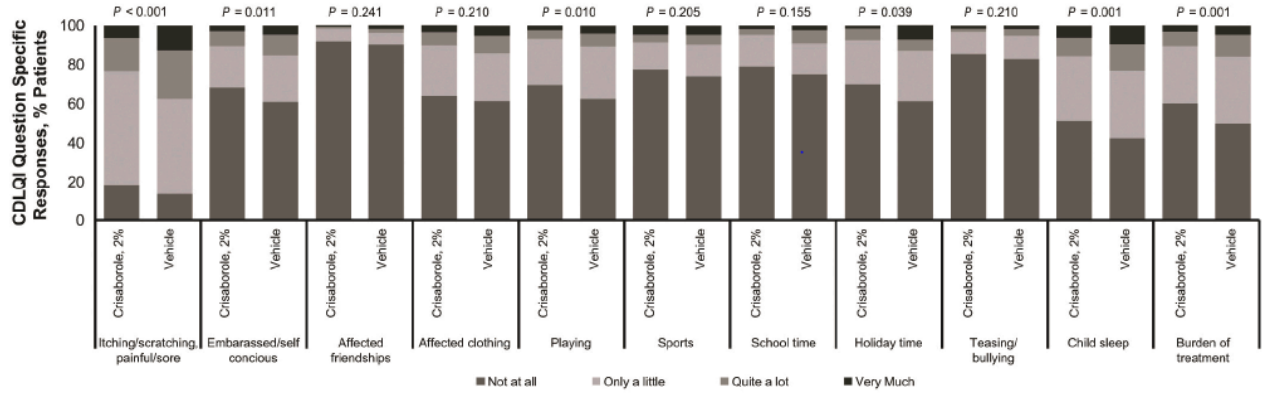
Examination of specific HRQoL domains revealed that children and adolescents treated with crisaborole experienced less impact of AD-related itching/scratching and painful/sore when compared with patients treated with vehicle (p<0.001). Additionally, children and adolescents treated with crisaborole experienced less impact on HRQoL related to feelings of embarrassment and self-consciousness (p=0.011), playing (p=0.010), holiday time (p=0.039), sleep (p=0.001), and the burden of treatment (p=0.001) when compared with patients treated with vehicle. Patients 16 years of age and older treated with crisaborole ointment reported similar improvements in HRQoL domains including less impact of itching/scratching and painful/sore (p=0.001), embarrassment and self-consciousness (p=0.024), and sexual difficulties between partners (p=0.017) when compared with patients treated with vehicle. Finally, parents/caregivers of children and adolescents treated with crisaborole experienced greater improvement in sleep (p=0.015), time shopping (p=0.026), and expenditure (p=0.042) than parents of patients treated with vehicle. Collectively, these data demonstrate the superior efficacy of crisaborole in reducing the impact of the most burdensome manifestations of AD on HRQoL when compared with vehicle.

Figure 13: Baseline and Day 29 severity bands for CDLQI



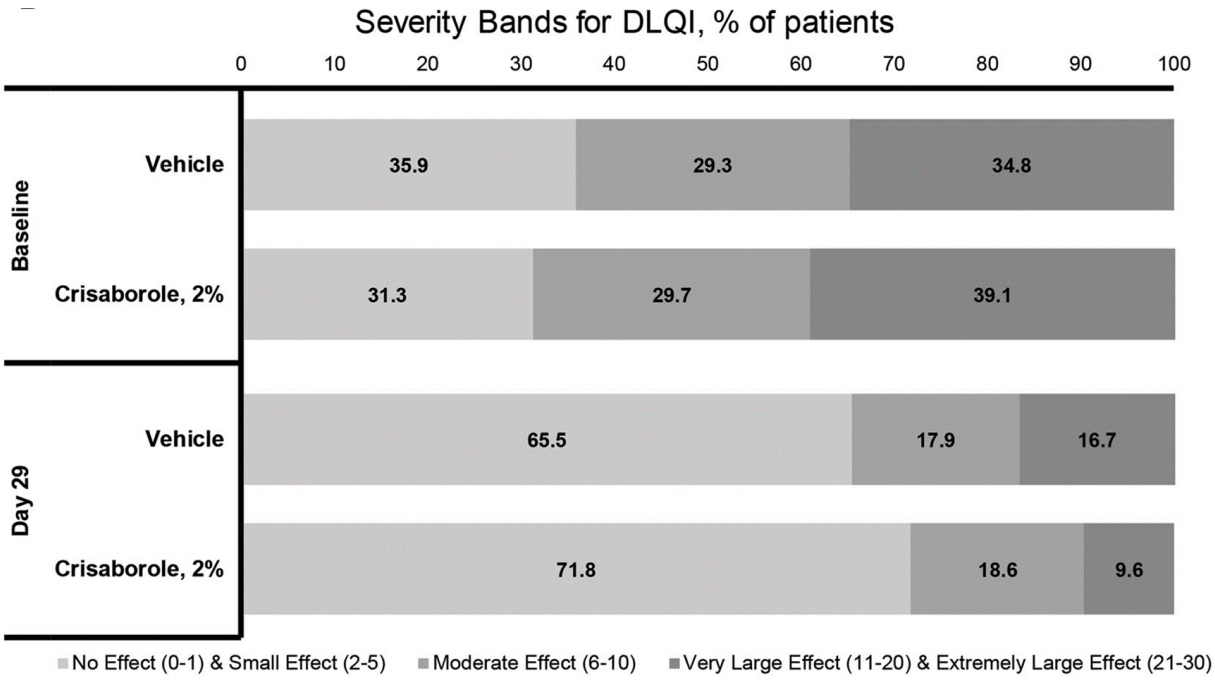
Abbreviations: CDLQI, Children's Dermatology Life Quality Index

Figure 14: CDLQI Item-Specific Responses



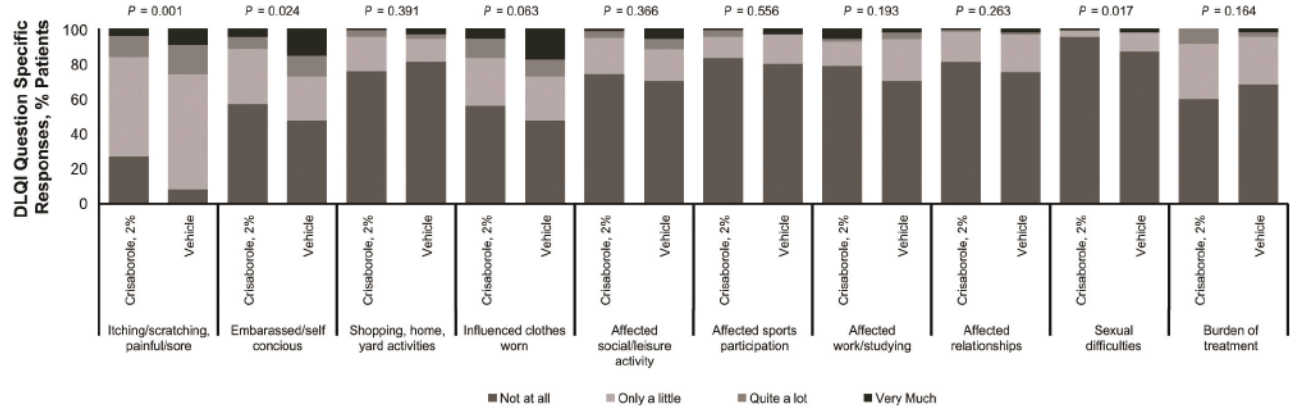
Abbreviations: CDLQI, Children's Dermatology Life Quality Index

Figure 15: Baseline and Day 29 severity bands for DLQI



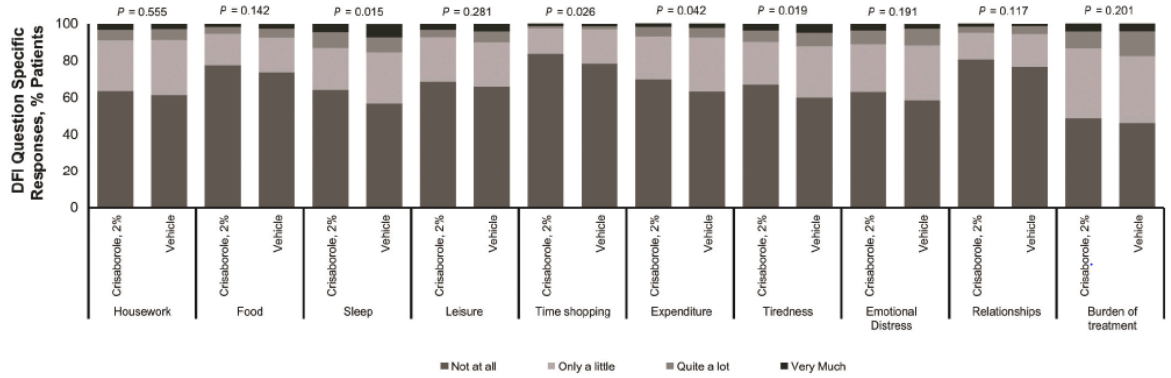
Abbreviations: DLQI, Dermatology Life Quality Index

Figure 16: DLQI Item-Specific Responses



Abbreviations: DLQI, Dermatology Life Quality Index

Figure 17: DFI Item-Specific Responses



Abbreviations: DFI, Dermatitis Family Impact Questionnaire

- ***EQ-5D Mapping from the CDLQI and the DLQI***

Generic HRQoL questionnaires assess health in the broadest sense and are applicable across a range of conditions. This generalisability enables comparisons within and between populations, health conditions and treatments. The EuroQol five-dimensional (EQ-5D) questionnaire is the preferred HRQoL questionnaire for UK NICE submissions, however, the EQ-5D was not assessed in the crisaborole trials. A post-hoc mapping of DLQI and CDLQI onto the EQ-5D-3L was performed to assess EQ-5D index utility scores at day 29, stratified by ISGA severity (**Table 22**).(128) Further details regarding the mapping algorithm and the EQ-5D index values used to inform the economic model are listed in **Section B.3.4.** It is noted that that the HRQoL benefit associated with treatment is assumed to be captured in the ISGA response and is not captured independently in EQ-5D estimates used to populate the economic analysis.

Table 22: EQ-5D Index values mapped from ISGA

EQ-5D Index Values			
	N	**Mean**	**95% CI of Mean**
ISGA at Day 29			
Age ≥ 16			
Clear/ Almost Clear (0/1)	[redacted]	[redacted]	[redacted]
Mild (2)	[redacted]	[redacted]	[redacted]
Moderate (3)	[redacted]	[redacted]	[redacted]
Severe (4)	[redacted]	[redacted]	[redacted]
Age 2 to < 16			
Clear/ Almost Clear (0/1)	[redacted]	[redacted]	[redacted]
Mild (2)	[redacted]	[redacted]	[redacted]
Moderate (3)	[redacted]	[redacted]	[redacted]
Severe (4)	[redacted]	[redacted]	[redacted]
Combined Ages			
Clear/ Almost Clear (0/1)	[redacted]	[redacted]	[redacted]
Mild (2)	[redacted]	[redacted]	[redacted]
Moderate (3)	[redacted]	[redacted]	[redacted]
Severe (4)	[redacted]	[redacted]	[redacted]

Note: the HRQoL benefit associated with treatment is assumed to be captured in the ISGA response and is not captured independently.

B.2.6.2 Summary of outcome measures in AD-303

Patients eligible for the long-term extension study had completed AD-301 or AD-302 without experiencing a crisaborole treatment-related AE or a serious AE. The LTE was a safety study and was not designed to assess efficacy. The study duration was 48 weeks broken into treatment periods of 28 days in duration. Patient's global disease severity, as measured by ISGA, was assessed at the beginning of every 28-day cycle. Patients with an ISGA score of 2 or greater then entered an on-treatment period and were instructed to apply crisaborole twice daily for the duration of the treatment period. Patients with an ISGA score of mild (0 or 1) entered an off-treatment period for the next 28 days. Investigators discontinued treatment with crisaborole for any subject who experienced no improvement in ISGA after three consecutive cycles of treatment. A total of 12 on-treatment periods were possible. Safety outcomes from the long-term extension study are detailed in **Section B.2.10.3.**

B.2.6.2.1 Main efficacy outcome

- ***Subsequent therapy use***

Subsequent therapy was defined as the need for concomitant nonconcurrent use of low- to mid-potency TCS or TCI. Most patients (77.8%) did not require subsequent therapy throughout the long-term study and an average of 150 treatment days preceded rescue therapy for those who needed it. The mean number of days on subsequent therapy was 21.4 for TCS (n = 155) and 24.2 for TCI (n = 6). Subsequent therapy was required in 22.4% of those 2-11 years of age; 26.0% of those 12-17 years of age, and 12.7% of those 18 years of age and older. The majority of patients (79.1%) resumed treatment with crisaborole and 75.7% of those receiving subsequent therapy remained in the study until week 48 or study closure.

B.2.6.2.2 Secondary outcome

- ***Treatment exposure***

The study population underwent an average of 6.2 on-treatment periods (a treatment period was defined as 28 days), including the pivotal trial. The mean number of drug applications was similar across age groups (347.7-348.9) as was the mean amount of drug applied per application (2.10-2.40 grams). Finally, the mean amount of crisaborole applied per month per patient was 133 grams (**Table 23**).

Table 23: Treatment exposure in AD-303

Cohort	2-11 years N = 308	12-17 years N = 146	≥ 18 years N = 63	Total patients N = 517
Applications, n				
Patients	304	146	60	510
Mean	349.0	349.4	347.7	348.9
SD	179.57	193.21	180.33	183.30

Total amount of drug used				
Patients, n	308	146	63	517
Mean, g	793.46	791.13	528.32	760.49
SD, g	1039.62	1052.15	722.22	1012.07
Amount of drug used per application				
Patients, n	304	146	60	510
Mean, g	2.40	2.29	2.10	2.34
SD, g	2.50	2.38	3.20	2.55
On-treatment periods, n				
Patients	308	146	63	517
Mean	6.2	6.3	5.9	6.2
SD	3.14	3.35	3.12	3.20
Duration of on-treatment periods				
On-treatment periods, n	1903	921	370	3194
Mean, days	28.4	28.3	28.6	28.4
SD, days	5.83	6.52	6.40	6.10

Abbreviations: n: population number; g: grams; SD: standard deviation

B.2.7 Subgroup analysis

A summary of the results for the subgroup analysis is provided in **Appendix E**. Post-hoc subgroup analyses were performed for age, baseline severity, and prior AD treatment. These subgroups align with clinical guidelines which recommend a stepped approach depending upon the patient's age, disease severity, and treatment history. A post-hoc subgroup analysis was also performed for partial response, defined as a one-point decrease in severity. Data on patient demographics in these subgroups is presented in **Section B.2.3.3**.

Data for the trials identified through the systematic literature review are presented in **Appendix E**.

B.2.7.1 Efficacy Outcomes

- ***Stratification by Age and Baseline Severity***

A statistically significant greater proportion of patients treated with crisaborole treated patients achieved success in ISGA across children with moderate AD compared to those treated with vehicle. Moreover, a greater proportion of patients treated with crisaborole children with mild or moderate AD achieved an ISGA score of clear (0) or almost clear (1) (**Table 24**).

Table 24: ISGA 0/1 Rates at 29 Days Per Treatment Arm Results for Adult and Child Subgroup RCT Data

Trial number (Acronym)	Arm	Total No. Patients/ % ISGA 0-1			
		Adults		Children (age groups in years)	
AD-301 and AD-302					
	Crisaborole 2%	142/47.8%	[redacted]	2-<7: 335/47.3%; 7-<12: 292/54.7%; 12-<18: 247/50.0%	2-<7: [redacted] 7-<12 [redacted]; 12-<18: [redacted]
	Vehicle	67/41.4%		2-<7: 171/31.6%; 7-<12: 144/37.8%; 12-<18: 124/34.7%	

Abbreviations: ISGA=Investigator's Static Global Assessment, NR=not reported, RCT=randomised controlled trial
Note: All values and percentages are rounded to the nearest whole number.

Table 25: ISGA Rates at 29 Days Per Treatment Arm Results for Mild and Moderate Subgroup RCT Data

Trial number (Acronym)	Arm	Total No. Patients/ % ISGA 0-1					
		Mild (All)	Mild Children (age groups in years)	Mild Adults	Moderate (All)	Moderate Children (age groups in years)	Moderate Adults
AD 301, 2016							
	Crisaborole 2%	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	Vehicle	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
AD 302, 2016							
	Crisaborole 2%	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Trial number (Acronym)	Arm	Total No. Patients/ % ISGA 0-1					
		Mild (All)	Mild Children (age groups in years)	Mild Adults	Moderate (All)	Moderate Children (age groups in years)	Moderate Adults
	Vehicle	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Abbreviations: ISGA=Investigator's Static Global Assessment, NR=not reported, RCT=randomised controlled trial
Note: All values and percentages are rounded to the nearest whole number

- ***Stratification by Prior AD Treatment Use***

In the crisaborole trials, nearly half of patients (42.5%) received prior treatment for their AD. [redacted]

Table 26: Proportion of patients with an ISGA score of Clear or Almost Clear for AD-301 and AD-302 at Day 29 Stratified by Prior Use of AD Treatment

Outcome	Achievement of Clear or Almost Clear ISGA score at Day 29			
	Prior Use of AD Treatment		Treatment Naïve	
	Crisaborole	Vehicle	Crisaborole	Vehicle
N	[redacted]			
Success, %				
p-value				

- ***Partial Response***

Post-hoc analyses were conducted to examine a partial response to treatment, defined as at least a one-point improvement in ISGA score. [redacted]

Table 27: Proportion of patients who achieved a partial response in ISGA score at Day 29

Outcome	At least One-Point Improvement in ISGA score at Day 29	
	Crisaborole	Vehicle
N	[redacted]	
Success, %		
Failure, %		
p-value		

Abbreviations: ISGA: Investigator's Static Global Assessment; N: population number
p values are versus vehicle control, pooled data from AD-301 and AD-302

- ***Stratification by Treatable % BSA***

Since ISGA only assesses the clinical signs of AD and does not account for %BSA, a post-hoc analysis was conducted to evaluate the change from baseline in %BSA across the same age and %BSA subgroup (**Table 28**).(129)

[redacted]

Table 28: Least-Square Mean Change from Baseline Treatable %BSA at Day 29 by Age Category and Baseline %BSA

Age	BSA 0.1-<16%		BSA 16-≤40%		BSA>40%	
	Crisaborole N (LS Mean)	Vehicle N (LS Mean)	Crisaborole N (LS Mean)	Vehicle N (LS Mean)	Crisaborole N (LS Mean)	Vehicle N (LS Mean)
2-6 years	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
7-11 years	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
≥12 years	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Abbreviations: BSA: Body Surface Area; N: population number; LS: Least-Square
Note: pooled data from AD-301 and AD-302

B.2.8 Indirect and mixed treatment comparisons

Full details of the methodology for the network meta-analysis (NMA) and unanchored matching adjusted indirect comparison (MAIC) are presented in **Appendix D**.

B.2.8.1 Objective of the indirect comparison

The objective of the indirect comparison was to compare Staquis™ (crisaborole 2%) with TCS, tacrolimus 0.1%, tacrolimus 0.03%, and pimecrolimus 1% in mild-to-moderate AD in patients aged 2 years of age or older using all relevant RCTs identified by the SLR described in **Section B.2.1**. No RCTs on TCS met the inclusion criteria of the SLR so only comparisons with TCIs could be conducted.

B.2.8.2 Selection of key outcome

The key outcome considered was ISGA/IGA 0/1. This outcome was selected as the key outcome for the NMA rather than ISGA success (the primary outcome in the crisaborole studies) because ISGA success was not consistently reported in comparator RCTs and could not be analysed appropriately (**Table 29**). The majority of RCTs on comparators also reported either ISGA 0/1 or IGA 0/1 and furthermore, this endpoint was consistent with the endpoint to be considered in the economic analysis. The pooling of ISGA and IGA outcomes is consistent with earlier NMA and meta-analysis in atopic dermatitis that merge physician global assessments and with clinical expert opinion, which judged that ISGA 0/1 and IGA 0/1 were sufficiently similar to analyse together.(34, 130) Furthermore, only one other RCT reported ISGA 0/1 (Levy 2005 comparing tacrolimus 0.03% to vehicle) so without merging IGA 0/1 with ISGA 0/1 indirect comparison with pimecrolimus 1% or tacrolimus 0.1% would not be possible.(131) We took the decision not to merge ISGA/IGA 0/1 with the PGE≥90% outcome as PGE is a dynamic assessment relative to baseline by the physician whereas the ISGA is static and the physician investigator does not compare the disease state with baseline but only evaluates patient condition at that moment in time.

B.2.8.3 Differences in vehicle across RCTs in mild to moderate AD

The reference treatment for an indirect comparison is typically an accepted standard of care therapy (or placebo), to which several other relevant treatments have been compared. In AD, most trials have compared the intervention of interest to vehicle. One obstacle to conducting any indirect comparison in mild to moderate AD is that vehicles used as controls in published trials must be assumed comparable to each other. Vehicles across trials are formulated with different excipients, with and without emollient or moisturizing properties. These differences across trials lead to substantial differences in response, which can bias an NMA that uses vehicle as an anchor to compare crisaborole, tacrolimus, and pimecrolimus. In particular, a relatively high rate of vehicle response was observed in the pivotal trials for crisaborole with regard to the primary [redacted] may reflect older trials where vehicle compounds could be irritating and induce dermatitis.(133) It is known that some vehicle excipients, such as those used in the crisaborole trials, have a more pronounced benefit than previously thought, and may improve barrier function and skin appearance; thus, vehicle cannot be considered a placebo.(120) Taken together, these differences in vehicle properties would make the relative benefits of the active therapies in older trials appear artificially high.(133) Differences in vehicle composition and response across RCTs are presented in **Appendix D**. Our base case NMA on ISGA/IGA 0/1 up to 6 weeks followed the recommended approach of the NICE Decision Support Unit (DSU).(134) Merging ISGA 0/1 and IGA 0/1 was consistent with earlier NMA and meta-analysis in atopic dermatitis that merged other physician global assessments.(34, 130) Although we included vehicle response meta-regression in the NMA to account for vehicle differences across trials, these aggregate data methods are not expected to fully adjust for potential bias. Following the recommendation of NICE DSU TSD 18, therefore, we used individual patient data (IPD) from the crisaborole trials to conduct an unanchored MAIC on ISGA/IGA 0/1 with details in **Section B.2.9.7**.(135)

B.2.8.4 Evidence networks for NMA

Head-to-head RCTs between all comparators with license/approval in all or some of mild to moderate AD in children and adults (namely TCS, pimecrolimus 1%, tacrolimus 0.03%, and tacrolimus 1%) have not been conducted; therefore, an NMA was undertaken to estimate the relative efficacy and safety between these treatments. NMA can provide comparative measures of effect (e.g. hazard ratios, odds ratios versus a reference treatment), for all relevant comparators in the absence of direct evidence and is especially suitable when there are multi-arm trials included within networks. Use of an NMA in preference to pairwise meta-analysis allowed for the inclusion of all available and relevant evidence and allowed for more precise treatment effects to be calculated. The results from the NMA feed into the economic

model described in **Section B.3**, evaluating the cost-effectiveness of crisaborole against relevant comparators. We have aligned our analysis with the methods described in the NICE DSU technical support documents.(134, 136-138). The key efficacy outcome we explored was ISGA 0/1, the proportion of patients achieved ISGA scores of clear (0) or almost clear (1). RCTs included in the efficacy analyses are summarised in **Table 29**. A cross-check of the studies identified by the SLR but excluded from the NMA is provided in **Table 31.** To enable an assumption of constant treatment effects over time (constant hazard ratios or proportional hazards), we included only studies reporting greater than 7 days and up to 8 weeks follow-up in the NMA. This restriction excluded the 24-week Meurer 2004 RCT (pimecrolimus 1% versus vehicle) and the 7-day Fowler 2007 and Kauffman 2006 RCTs (both pimecrolimus 1% versus vehicle) from all analyses.(139-141). All remaining RCTs had durations between 4 weeks and 43 days (**Table 29**). In addition to the pivotal AD-301 and AD-302 studies on crisaborole, RCTs on pimecrolimus 1%, tacrolimus 0.1%, tacrolimus 0.03%, were included in the network. No RCTs evaluating TCS were identified by the SLR so this treatment option was not included in the NMA. A full comparison of baseline characteristics is presented in **Table D13** while a comparison of ISGA/IGA success definitions is provided in **Table 30**. We note from the latter table that ISGA/IGA success was not consistently defined across the available studies. Studies on pimecrolimus 1%, tacrolimus 0.1% and tacrolimus 0.03% (Eichenfield 2002, Abramovits 2008, Paller 2005) used IGA or investigators global atopic dermatitis assessment (IGADA) 0/1 with an improvement of 1 or more points on IGA or IGADA, a weaker condition of success than the 2-grade improvement required in AD-301 and AD-302.(42, 45, 47)

Similarity and differences in baseline characteristics are discussed below. These differences can be accounted for by either random treatment effects models which allow treatment effects to vary across RCTs, covariate specific meta-regressions which adjust treatment effects for differences between RCTs, or baseline/vehicle response regressions which adjust for multiple characteristics at once through their impact on vehicle response; these approaches are discussed in the modelling methods section. The NMA models assumed constant treatment effects (proportional hazards assumption) but include an offset adjustment for follow-up, to account for differences in the follow-up period which ranged between 4 and 6 weeks. Subgroup analyses in mild, moderate, children, adult, mild-children, moderate-children, and moderate-adult are described in a later section, accounting for the most important differences identified between studies below.

Table 29: Summary of the trials used to carry out the efficacy NMA (primary outcome ISGA/IGA 0/1, secondary outcome ISGA/IGA Success)

References of trial	Vehicle	Crisaborole 2%	Pimecrolimus 1%	Tacrolimus 0.03%	Tacrolimus 0.1%	OPA-15406 1%	OPA-15406 0.3%	ISGA/IGA 0/1	ISGA/IGA Success	Available timepoint closest to 4 weeks
AD 301	Yes	Yes						Yes	Yes	29 days

AD 302	Yes	Yes						Yes	Yes	29 days
Chapman et al. 2005	Yes			Yes				Yes		46 weeks
Eichenfield et al. 2002	Yes		Yes					Yes	Yes	29 days
Kempers et al. 2004			Yes	Yes				Yes		29 days
Levy et al. 2005	Yes			Yes				Yes		4 weeks
Abramovits et al. 2008			Yes		Yes			Yes	Yes	6 weeks
Paller et al. 2005			Yes	Yes				Yes	Yes	6 weeks
Schachner et al. 2005	Yes			Yes				Yes		4 weeks

Table 30: Definitions of "IGA/ISGA success" used in RCTs for NMA*

Study	Global assessment	Definition of IGA/ISGA success
Abramovits et al. 2008	IGADA	IGADA 0/1 with 1 or more grades of improvement
Eichenfield et al. 2002	IGA	IGA 0/1 with improvement of 1 or more IGA scores
Paller et al. 2005	IGADA	IGADA 0/1 with 1 point or more improvement
AD301/302 2016	ISGA	At least 2 grade improvement

*IGA=Investigators global assessment, ISGA=Investigators static global assessment, IGADA=Investigator global atopic dermatitis assessment

Summary of Trials and Overview of Heterogeneity

Eichenfield 2002 compared pimecrolimus 1% to vehicle, reported the IGA 0/1 outcome and had similar baseline characteristics (age, proportion male, % BSA, and proportion Caucasian) to the AD-301 and AD-302 studies.(12, 47) Proportions mild and moderate AD in Eichenfield 2002 were similar and only ~10% of patients on both arms of this RCT were in the severe or very severe categories.(47) Kempers 2004 compared pimecrolimus 1% to tacrolimus 0.03%, reported the IGA 0/1 outcome, and had similar reported baseline characteristics (age, proportion male, proportion Caucasian) to AD-301 and AD-302.(12, 44) However, apart from one patient classified as severe, Kempers 2004 was exclusively in moderate AD; we excluded this study from subgroup analyses in mild AD.(44) Levy 2005

compared tacrolimus 0.03% to vehicle and reported the ISGA 0/1 outcome.(131) There was limited information on baseline characteristics but the study was restricted to adult (≥18 years old) patients and the overall (both arms) proportion mild or moderate AD was 88.7%; we therefore excluded Levy 2005 from subgroup analyses in children.(131)

The Chapman 2005 publication described two trials comparing tacrolimus 0.03% to vehicle, and reporting IGA 0/1; one was in adults (16+) and one was in children (2-15) (142) (46). Full results at 4-weeks for the children trial were provided in the Schachner 2005 publication (46). Results at 6-weeks were provided for the adult trial in the Chapman 2005 publication (142). Both trials had similar reported baseline characteristics (proportion male, and proportion Caucasian).(142) The % BSA at baseline across both studies in Chapman 2005 (11.0% on tacrolimus 0.03% and 11.2% on vehicle) was only somewhat lower than in AD-301 and AD-302 (means in range 17.7-18.8%).(12, 142). Conversely, the adult trial in Chapman 2005 had a median age of 38.5 years on tacrolimus 0.03% and 37.5 years on vehicle; we therefore excluded it in a sensitivity analysis.

Abramovits 2008 compared tacrolimus 0.1% to pimecrolimus 1%, reported the IGADA 0/1 outcome, and was similar on some reported baseline characteristics (proportion male, proportion Caucasian, % BSA) to AD-301 and AD-302.(12, 42) However, the mean age was 39.9 years on tacrolimus 0.1% and 38.3 years on pimecrolimus 1% and was exclusively on patients ≥16 years old; we therefore include Abramovits 2008 in an adult only subgroup analyses and exclude from child only analyses.(42) We also exclude from a sensitivity analysis removing the Chapman 2005 adult study due to higher aged populations. Abramovits 2008 was exclusively in moderate AD and thus contributes to the moderate AD subgroup analyses.(42) Paller 2005 compared tacrolimus 0.03% to pimecrolimus 1%, reported IGADA 0/1, and had comparable baseline

The three RCTs Hoeger 2009, Murrell 2007, and Hordinsky 2010 comparing pimecrolimus 1% to vehicle were excluded on the basis that they were restricted to facial AD; this restriction is expected to modify the treatment effect of pimecrolimus 1% and is consistent with the populations considered in other included studies and crisaborole trial data which were not in body specific sites (143-145) In addition, in Murrell 2007 over 80% of patients had been exposed to TCS on the face in the month prior to study commencement while in Hoeger 2009 80.5% of patients were considered TCS dependent.(144, 146) These are substantially higher than the average of 37.6% with prior TCS treatment in AD-301 and AD-302.(12) Although Hoeger 2009 was in patients aged 2-11, the mean age on pimecrolimus 1% in Murrell 2007 was 32 years and on vehicle was 28 years, while in Hordinsky 2010 the mean age was 43.9 years on pimecrolimus 1% and 44.1 on vehicle.(144, 146). These differences combined indicated that inclusion of these studies in the base case NMA would potentially bias efficacy estimates so they have been omitted.

Two RCTs (Fowler 2007, Kauffman 2006) comparing pimecrolimus 1% to vehicle and reporting IGA were excluded as their follow-up was only 7 days.(140, 141) Although our NMA model can adjust for different follow-up times using an offset, an assumption of constant treatment effects (hazard ratios) must be made for the treatment duration.(140, 141) It is difficult to justify the assumption that hazard ratios are the same after 7 days as after 4, 6, or 14 weeks. Similarly, Meurer 2004 comparing pimecrolimus 1% to vehicle was

excluded as its follow-up was 24 weeks; thus it is difficult to justify an assumption that treatment effects are constant going from 4, 6, or 14 weeks to 24 weeks.(139)

We excluded one RCT (Boguniewicz 1998 comparing vehicle, tacrolimus 0.1%, and tacrolimus 0.03%) that used the PGE≥90% scale as we decided against merging with the IGA/ISGA 0/1 outcome. This decision against merging was supported by a systematic review of investigator's global assessments in atopic dermatitis using all published randomised controlled trials of AD treatments in the Global Resource of Eczema Trials database (2000-2014) which excluded PGE (147). Their reasoning was that "a global assessment was defined as an overall evaluation of AD severity using an ordinal scale or an ordered categorical variable (e.g., clear, almost clear) that was scored using points by an investigator or a physician. We excluded scales that were extrapolated from calculated continuous scales such as the Eczema Area and Severity Index." Furthermore, PGE is a dynamic assessment relative to baseline by the physician whereas the ISGA is static and the physician investigator does not compare the disease state with baseline but only evaluates where he/she sees at that moment in time.

With the above inclusions and exclusions, the evidence network for ISGA 0/1 is presented in **Figure 15**. We see that crisaborole 2% can be indirectly compared with pimecrolimus 1%, tacrolimus 0.03%, and tacrolimus 0.1% on both outcomes.

Table 31: Summary of inclusion and exclusion of RCTs identified by the SLR in the NMA

Trial Name Clinical Trial No	Included in NMA, with reason for exclusion
Abramovits, 2008	Yes
AD 301 **NCT02118766**	Yes
AD 302 **NCT02118792**	Yes
Boguniewicz, 1998	Yes, but only in overall withdrawals and withdrawals due to AE. Not in ISGA/IGA 0/1 analysis as ISGA or IGA not repored
Chapman, 2005	Yes
Eichenfield, 2002	Yes
Fowler, 2007	No, only 7 day follow-up
Hanifin, 2016 **NCT02068352**	No, OPA-15406 not licensed
Hoeger, 2009 **NCT00130364**	No, population was facial AD only.
Hordinsky, 2010	No, population was facial AD only.
Kaufmann, 2006	No, only 7 day follow-up
Kempers, 2004	Yes
Levy, 2005	Yes
Meurer, 2004	No, follow-up of 24 weeks was beyond 8 week

Trial Name Clinical Trial No	Included in NMA, with reason for exclusion
	maximum for NMA.
Murrell, 2007	No, population was facial AD only.
Paller, 2005	Yes
Schachner, 2005	Yes
Sears, 1997	No, no useable outcomes.
Wahn, 2002	No, no useable outcomes.

Figure 18: Evidence network for ISGA/IGA 0/1 up to 6 weeks

B.2.8.5 Models and model selection for NMA

Full details of the modelling methods are provided in appendix D. The ISGA/IGA 0/1 primary outcome is a binary outcome, which has either been achieved or not achieved by end of RCT follow-up, but available RCTs vary in follow-up (**Table 29**). In line with the NICE TSD DSU 2, we therefore adopt a binomial likelihood and complementary log-log (cloglog) link function with offset for follow-up time for this outcome.(134) Pruritus 0/1 and all safety outcomes are similarly binary outcomes reported by RCTs varying in follow-up so the binomial likelihood cloglog with time offset has been adopted for all outcomes. Both fixed (common) and random (exchangeable) treatment effects across trials are considered.

Pimecrolimus and tacrolimus are both TCIs with the same method of action, hence, their treatment effects are potentially related to each other. We therefore considered treatment class effect models as described in Owens 2015; although these methods are recommended for covariate effects by the NICE DSU TSD 3 they have been used in previous technology appraisals in, for example, the TA445 AG on certolizumab pegol and secukinumab for psoriatic arthritis.(136, 148, 149) The expectation that TCIs have related effects gives prior preference for some form of class effect model. We have explored both fixed and random

class effect models, as described in Appendix D. Fixed class effects assume all treatments in the same class have the same effect while random class effects assume they are exchangeable. Fixed class effects would assume that pimecrolimus 1%, tacrolimus 0.03%, and tacrolimus 0.1% all have the same treatment effect, which is unlikely to be clinically plausible. Our prior preference was therefore random class effects, and we would only opt for fixed or no class effects only if they were strongly favoured by model assessment statistics (described below).

We have explored meta-regression on vehicle response due to the differences in vehicle response discussed in **Appendix D** and due to heterogeneity in the time, setting, and design of the included RCTs as recommended by NICE DSU TSD 3. This gives a total of 12 possible models based on assumptions on fixed versus random treatment effects, no class versus fixed class versus random class effects, and with and without vehicle response regression. We explored the total residual deviance in comparison to the number of data points in the network as well as the deviance information criterion (DIC) to compare models statistically; these were our primary model assessment statistics. In the base case and any networks where loops of trials were present, evidence of inconsistency was assessed using both independent means tests and node-splitting, in line with NICE DSU TSD 4.(137)

The comparison of deviance and DIC are presented in **Table 32**. The deviance and DIC evidence does not strongly favour fixed class effects or no class effects and due to their clinical implausibility, random class effects are selected. This gives a final selection of the fixed treatment effect, random class effect, vehicle response adjusted (FE-RCE-VR) model. The total residual deviance of this model (15.73) is below the number of datapoints (18), indicating good fit of the model to the data. The independent means inconsistency model for FE-RCE-VR did not indicate evidence of inconsistency as the DIC and residual deviance were higher than that the consistency FE-RCE-VR model (**Table 32**).To maintain uniformity of approach across outcomes the FE-RCE-VR model is preferred for all safety outcomes and in subgroup and sensitivity analyses. However, in cases where the FE-RCE-VR model did not converge a simple fixed or random treatment effect model with no class effect or vehicle response regression was used, with selection based on DIC or residual deviance.

As described in Appendix D, we use the Bayesian OpenBUGS Markov Chain Monte Carlo (MCMC) software to perform our NMAs. We used 50,000 MCMC iterations, with thin=10, for burn-in and 50,000 iterations for posterior sampling. When using binomial-cloglog we estimated hazard ratios of events (e.g. ISGA/IGA 0/1) comparing crisaborole to all treatments included in the evidence networks; when using binomial-logit in sensitivity analyses we estimated odds ratios of events. When interpreting hazard ratio estimates, we refer to treatments as having greater or lower hazards of events, whether the event in question is ‘good’ (e.g. ISGA/IGA 0/GA) or ‘bad’ (e.g. withdrawal due to adverse event). Due to skew of the hazard ratios and odds ratios, point estimates are the medians of the MCMC samples. We provide 95% credible intervals (CrI) which consist of the lower 2.5th and upper 97.5th percentiles of the MCMC samples. Bayesian probabilities that crisaborole is best (e.g. highest ISGA/IGA 0/1 or lowest adverse events) are the proportion of MCMC samples for which the hazard ratio or odds ratio was greater than 1 (if events are ‘good’) or less than 1 (if events are ‘bad’); these are labelled ‘p-best’. Probabilities that each treatment is 1st, 2nd,

3rd,… ranked are also generated and plotted for each treatment; such plots are called rankograms. Mean rank for each treatment are the final summary statistic.

Node splitting was used to test for inconsistency on specific loops.(137) These split the evidence networks into ‘direct’ and ‘indirect’ components and estimate hazard ratios restricting to these components. Differences between the hazard ratios suggest inconsistency. Direct evidence on each contrast would be limited to one or two RCTs, making it impossible to fit class effects or use vehicle response regression. Node splitting therefore used a simplified model with fixed treatment effects, a binomial likelihood, and a cloglog link function with an assumed offset of 4 weeks. To maintain a fair comparison between direct and indirect evidence, the same simplified model was employed. We used the mtc.nodesplit function of the GeMTC package in the R statistical software (150, 151) to conduct node splitting. The ‘network’ estimate is taken from the FE-RCE-VR model for ISGA/IGA 0/1 base case. This has the disadvantage that any identified inconsistency may be explained by differences in baseline risk between the direct and indirect evidence. The global independent means test for inconsistency, which does include baseline risk adjustment and results of which are presented in **Table 32**, is therefore a more reliable assessment for our networks.

Table 32: Comparison of model fit statistics for ISGA/IGA 0/1. Number of data points 18

Model			SD	Totresdev	DIC
Across studies fixed or random treatment effect	**Across treatment class (vehicle, TCI, PDE4 inhibitors) effect**	**Vehicle response regression**			
Random treatment effect	No class effect	No	0.27 (0.02 – 0.80)	17.87	129.2
Random treatment effect	Fixed class effect	No	0.24 (0.02 – 0.62)	18.3	128.6
Random treatment effect	Random class effect	No	0.24 (0.02 – 0.67)	17.88	128.8
Fixed treatment effect	No class effect	No		21.38	129.7
Fixed treatment effect	Fixed class effect	No		23.86	130.2
Fixed treatment effect	Random class effect	No		21.64	129.8
Random treatment effect	No class effect	Adjusted for vehicle response	0.12 (0.004 – 0.45)	15.94	126.4
Random treatment effect	Fixed class effect	Adjusted for vehicle response	0.15 (0.004 – 0.50)	18.58	128.5
Random treatment effect	Random class effect	Adjusted for vehicle response	0.11 (0.005 – 0.39)	16.05	126.4
Fixed treatment effect	No class effect	Adjusted for vehicle response		15.55	124.7
Fixed treatment effect	Fixed class effect	Adjusted for vehicle response		20.26	127.5
Fixed treatment effect	**Random class effect**	**Adjusted for vehicle response**		**15.73**	**124.8**
FE inconsistency – random class effects – adjusted for baseline risk				16.76	126.7

B.2.8.6 Baseline natural history model

The base case NMA on the key outcome of ISGA/IGA 0/1, as well as any secondary analyses that make use of vehicle response regression. This regression requires centering at some level of baseline risk (i.e. log hazard of response on vehicle) which will impact the hazard ratios of each treatment relative to vehicle but not to each other. We center baseline risk at the mean, on cloglog scale, on vehicle arms. This mean is estimated using a baseline natural history model to meta-analyse vehicle results across RCTs as described in NICE DSU TSD 5. (138) (DSU 5) The model uses a binomial likelihood with cloglog link and offset for log follow-up time, consistent with the model of the NMA.

[redacted]

Table 33:Estimated baseline log hazard ratio (i.e. on vehicle) for ISGA/IGA 0/1 base case and subgroups, and for safety outcomes

	Mean (95% CrI)	Mean age	Log hazard ratio age	Mean proportion moderate disease (%)	Log hazard ratio proportion moderate
ISGA/IGA 0/1 base case	1.21 (0.42, 1.97)	15.06	0.11 (-0.15, 0.38)	51.02	0.02 (-0.06, 0.09)
[redacted]					
[redacted]	[redacted]				
[redacted]	[redacted]				
[redacted]	[redacted]				
[redacted]	[redacted]				
[redacted]					
[redacted]	[redacted]				
[redacted]	[redacted]				

B.2.8.7 [redacted]Results of the NMA

The estimated hazard ratios of ISGA/IGA 0/1, with 95% CrI, on crisaborole compared to all included treatments is presented in Figure 16. Bayesian probabilities that crisaborole are superior (P-best) are also provided. Recall that these are generated using the FE-RCE-VR model which adjusts for vehicle response in trials and assumes random class effects on TCIs (pimecrolimus 1%, tacrolimus 0.03%, and tacrolimus 0.1%). Cross tables of pairwise hazard ratios comparing all treatments are provided in Table 34. Rankograms for all treatments are provided in Figure 17 and Table 35, with mean rank in Table 36.[redacted]

[redacted]

[redacted]Figure 19: Forest plot of NMA estimated hazard ratios (95% CrI) with P-best (probability crisaborole superior) of ISGA/IGA 0/1 of comparators versus crisaborole.

[redacted]

Table 34: Pairwise hazard ratios (95% CrI) on ISGA/IGA 0/1 estimated by NMA.

Vehicle	[redacted]		[redacted]	[redacted]
[redacted]	**Tacrolimus 0.03%**		[redacted]	
[redacted]	[redacted]	**Tacrolimus 0.1%**	[redacted]	
[redacted]	[redacted]	[redacted]	**Pimecrolimus 1%**	
[redacted]	[redacted]	[redacted]	[redacted]	**Crisaborole**

Figure 20: Rankograms: Probability of each treatment occupying each rank on ISGA/IGA 0/1 estimated by NMA

Table 35: Ranking probabilities on ISGA/IGA 0/1 estimated by NMA.

Rank	Vehicle	tacrolimus 0.03%	tacrolimus 0.1%	pimecrolimus 1%	Crisaborole 2%
1	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
3	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
4	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
5	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Table 36: Mean rank on ISGA/IGA 0/1 estimated by NMA.

Treatment	Mean rank
Vehicle	[redacted]
tacrolimus 0.03%	[redacted]
tacrolimus 0.1%	[redacted]
pimecrolimus 1%	[redacted]
Crisaborole 2%	[redacted]

Figure 21: Node splitting test for inconsistency in the ISGA/IGA 0/1 base case analysis

B.2.8.7.1 Exploring heterogeneity through network meta-regression

Effect modifiers are variables that change treatment response; imbalance in effect modifiers across RCTs in an NMA can bias treatment effect. Our base case model (FE-RCE-VR) included vehicle response regression to account for heterogeneity in effect modifiers, represented by their impact as prognostic variables on vehicle response, across RCTs (152). To explore the impact of heterogeneity in specific potential effect modifiers, network meta-regressions on individual characteristics were conducted. Results are presented in **Table 37**. These are aligned with the FE-RCE-VR model adopted for ISGA/IGA 0/1 but replace the vehicle response regression with regression on each of the covariates. The limited evidence necessitated an assumption of a common regression coefficient for each treatment.(136) The regression coefficient is the hazard ratio of ISGA/IGA 0/1 response in patients with 1 unit higher of the regression coefficient (i.e. 1 year higher mean age, 100% moderate versus 0% mild severity, 100% BSA versus 0% BSA, 100% Caucasian versus 0% Caucasian, and 100% male versus 0% male); if the 95% CrI crosses 1 this suggests no evidence of effect modification. Models with lower total residual deviance and DIC are preferred; if these statistics are lower in a model with a regression coefficient compared to that without (labelled 'None' in **Table 37**) this suggests evidence of effect modification.

Based on the total residual deviance, DIC, and regression coefficients reported in **Table 37** there is strongest evidence that proportion male impacts treatment effects. However, the effect is low (hazard ratio 1.04; 95% CrI (1.01, 1.08)) and the DIC (126.0) is only marginally lower than a model not adjusted for age (129.8). There is weaker evidence that age (hazard ratio 0.99; 95% CrI (0.97, 1.00); DIC=128.2) and % BSA (1.04; 95% CrI (1.00, 1.09); DIC=128.3) modify treatment effect while there is no evidence that severity or proportion Caucasian modify treatment effect. The model adjusting for vehicle response (FE-RCE-VR) has lower DIC (124.8) than all other regression models and has the smallest regression coefficient (hazard ratio 0.44; 95% CrI (0.29, 0.66)) with a 95% CrI clearly excluding 1. We

are therefore reassured that our selected model, that with vehicle response regression, is the optimal method to adjust for heterogeneity in reported potential effect modifiers.

We would note that the network meta-regressions are underpowered as only 2 at most RCTs are available on each treatment contrast in the ISGA/IGA 0/1 network (**Table 29**). These multiple RCTs on contrasts are Kempers 2004 and Paller 2005 comparing pimecrolimus 1% to tacrolimus 0.03% and Chapman 2005 and Levy 2005 comparing tacrolimus 0.03% to vehicle. The evidence is somewhat boosted by use of class effects on TCI.

Table 37: Results of network meta-regression on ISGA/IGA 0/1 for all possible covariates

Covariate All FE – random class effect	Total residual deviance	DIC	Regression coefficient on log hazard ratio scale (95%CrI)	Regression coefficient on hazard ratio scale (95% CrI)
None	21.64	129.8	[redacted]	[redacted]
Adjusted for vehicle response **(FE-RCE-VR)**	15.73	124.8	[redacted]	[redacted]
Age	19.04	128.2	[redacted]	[redacted]
Severity (proportion moderate)	21.41	129.6	[redacted]	[redacted]
Mean % BSA	19.09	128.3	[redacted]	[redacted]
Proportion Caucasian	23.01	131.7	[redacted]	[redacted]
Proportion male	16.82	126.0	[redacted]	[redacted]

B.2.8.7.2 [redacted]Uncertainties in the indirect and mixed treatment comparisons

Our NMA has taken evidence from all RCTs in both adults and children and in both mild and moderate disease. The estimates of treatment effect reflect outcomes for such a mixed population. However, this does not reflect the reimbursement and marketing authorisation for these comparators. For example, pimecrolimus 1% is not reimbursed in adults, and tacrolimus 0.1% and 0.03% are not licensed or reimbursed in mild disease. Differences in reimbursement and licensing reflect variation in treatment effect by age and severity; this variation is a very important source of uncertainty in our NMA. The meta-regressions on age and severity reported above found only limited evidence of effect modification but were limited by the availability of only aggregate data on comparators. Subgroup analyses stratifying by age and severity are reported in **Section B.2.9.8.**

Many assumptions were necessary to form connected and reasonably homogeneous evidence networks (e.g. exclusion of facial AD RCTs on pimecrolimus, inclusion of studies

with much greater mean age, inclusion of RCTs whose follow-up was greater than 4 weeks). Sensitivity to these model and data assumptions are explored in **Section B.2.9.9**.

A key uncertainty, as discussed in **Section B.2.9.3**, is the variation in vehicle response across RCTs. This is caused by variation in vehicle ingredients and differences in the characteristics of patients enrolled in RCTs. We used vehicle response regression to adjust for these differences, but this was limited by the use of aggregate data. As a sensitivity analysis, we employed unanchored matching adjusted indirect comparison (MAIC) to remove vehicle response and attempt to balance effect modifiers and prognostic variables across RCTs. The results and general methods are described in **Section B.2.9.10**. A full table of baseline characteristics across RCTs, many of which are important and need to be matched, is reported in **Table D18**.

All studies included in the NMA were on patients with active disease at baseline, indicating homogeneity on at least this issue.

Inconsistency is a potential difficulty in our NMAs as only indirect evidence was available to compare crisaborole to pimecrolimus 1%, tacrolimus 0.03%, and tacrolimus 0.1%. Node-splitting did indicate inconsistency in the loop of vehicle, tacrolimus 0.03%, and pimecrolimus 1%. Our global independent means test did not detect evidence of inconsistency, such tests are not highly powered. Given the differences in baseline characteristics discussed above, there remains potential for inconsistency to affect our results. We attempted to mitigate this impact through targeted subgroup analyses in **Section B.2.9.5**. Differences in reported and unreported baseline characteristics remain in these subgroups but there was no evidence of inconsistency in these networks. Potential inconsistency should be borne in mind when drawing conclusions from any of our analyses.

B.2.8.8 Subgroup analyses by AD severity and age

The AD-301 and AD-302 RCTs included both adult and child patients and also those with both mild and moderate AD. The RCTs on pimecrolimus 1%, tacrolimus 0.03%, and tacrolimus 0.1% were also mixed over these patient characteristics. Although our network meta-regression found limited evidence of effect modification by age and no evidence of effect modification by severity, we noted that these are underpowered as only 2 at most RCTs are available on each treatment contrast. Additionally, licenses and NICE approvals vary by age and AD severity, a situation summarised in **Table 38**. We therefore conducted targeted subgroup NMAs on the ISGA/IGA 0/1 outcome in mild AD, moderate AD, children with mild/moderate AD, adults with mild/moderate AD, children with mild AD, and children with moderate AD. There was no evidence available on comparators in adults with mild AD.

As the model from the base case (FE-RCE-VR) combines evidence on class effects and vehicle risk across trials, we included all available evidence, regardless of license or reimbursement situation outlined in **Table 38**. However, we focus interpretation only on treatments that are licensed or reimbursed in the specific subgroups.

Table 38: Licenses and reimbursements (NICE approval) across subgroups in AD (as specified in TA82)

		Adult	Children
Licensed	**Mild**	Pimecrolimus 1%*	Pimecrolimus 1%*
	Moderate	Pimecrolimus 1% *, Tacrolimus 0.03%, Tacrolimus 0.1% **	Pimecrolimus 1% *, Tacrolimus 0.03%
Reimbursed (relevant NICE TA)	**Mild**	None TCI	None TCI
	Moderate	Tacrolimus 0.03%, Tacrolimus 0.1% **	Pimecrolimus 1%*, Tacrolimus 0.03%

*Note: Pimecrolimus 1% is not reimbursed for adults and adolescents >16-year old

**Note: Tacrolimus 0.1% is only licensed for the treatment of adults with moderate-to-severe AD.

The available evidence on subgroup analyses is presented in **Table 39.** This includes FDA subgroup analyses (PIM FDA B307 and PIM FDA B305) of the Eichenfield 2002 pimecrolimus 1% versus vehicle RCT (47, 153). An important limitation of all these subgroup data is that patients were not randomised within these age and severity subgroups and are therefore observational in nature and subject to selection bias and confounding. These analyses are therefore ad-hoc and the NMA can be considered to be based on observational evidence (154). Also, a network limited to moderate adults was not connected crisaborole to vehicle and tacrolimus 0.03% as, in addition to data for AD-301/AD-302 (comparing crisaborole versus vehicle), only Abramovits 2008 (comparing pimecrolimus 1% to tacrolimus 0.1%) reported this subgroup (42, 142). The definition of adult and children age groups varied across the trials with the cut-off age varying across 16, 17, and 18. No comparative evidence was reported on mild adults so this subgroup was also not analysed. The network plots for subgroups with connected networks are presented in **Figure 22** and **Figure 23**.

Table 39: Summary of the trials used to carry subgroup NMA for the ISGA/IGA 0/1 primary outcome.

References of trial	Vehicle	Crisaborole 2%	Pimecrolimus 1%	Tacrolimus 0.03%	Tacrolimus 0.1%	OPA-15406 1%	OPA-15406 0.3%	Mild	Moderate	Adult	Children	Mild Children	Moderate Children	Moderate Adult	Available timepoint closest to 4 weeks
AD 301	Yes	Yes						Yes	Yes	Yes. >=18	Yes. <18	Yes. <18	Yes. <18	Yes. >=18	29 days
AD 302	Yes	Yes						Yes	Yes	Yes. >=18	Yes. <18	Yes. <18	Yes. <18	Yes. >=18	29 days
Chapman et al. 2005	Yes			Yes				Yes	Yes	Yes. >=16.					6 weeks
Eichenfield et al. 2002 (PIM FDA B307 and PIM FDA B305)	Yes		Yes					Yes	Yes		Yes. Only in 1-17.				29 days
Kempers et al. 2004			Yes	Yes					Yes		Yes. Only in 2-17.		Yes. 2-17 moderate.		29 days
Levy et al. 2005	Yes			Yes											4 weeks

Abramovits et al. 2008			Yes		Yes				Yes. Only moderate	Yes. Only >=16				Yes. Only >=16 moderate.	6 weeks
Paller et al. 2005			Yes	Yes				Yes, but 2-15 only	Yes (child and adult		Yes	Yes, 2-15			6 weeks
Schachner et al. 2005	Yes			Yes				Yes			Yes, 2-15.	Yes	Yes.		4 weeks

Figure 22: Subgroup ISGA/IGA 0/1 up to 6 weeks evidence networks by severity and in children

Figure 23: Subgroup ISGA/IGA 0/1 up to 6 weeks evidence networks for moderate AD children and mild AD children

Model assessment and selection in the ISGA/IGA 0/1 age and severity subgroups is summarised in Table 39. Our default was to use the same model as in the base case ISGA/IGA 0/1 networks, which was FE-RCE-VR. However, in children with AD and adults with AD

Random effects

the total residual deviance of all selected models

Global independent means inconsistency tests

M
ean ranks across the networks are presented in Table 40, and discussed in combination with the forest plots below. Table 40: Mean ranks of treatments across age and severity ISGA/IGA 0/1 subgroup analyses. Blank cells indicate treatment not in network.

	Base case all patients	Moderate AD	Mild AD	Children	Adults	Moderate children	Mild children
Vehicle							
Tacrolimus 0.03%							
Tacrolimus 0.1%							
Pimecrolimus 1%							
Crisaborole 2%							

Results for moderate AD using a FE-RCE-VR model are presented in **Figure 24**.

Results for mild AD using the FE-RCE-VR model are presented in **Figure 25**.

Results for children (mixed across mild and moderate AD) using a simple fixed treatment effects model are presented in

Figure 26.

Results for adults (mixed across mild and moderate AD) using a simple fixed treatment effects model are presented in Figure 27.

In children with moderate AD using the FE-RCE-VR model (Figure 28

In children with mild AD using the FE-RCE-VR model (Figure 29),

These NMAs stratifying by age and severity may give more clinically relevant comparisons as they are specific to the actual licensing and reimbursement landscape in the UK. However, the data available to inform these analyses was limited and the networks of evidence sparsely populated. In the children and adult networks there was insufficient evidence to perform vehicle response regression, potentially biasing the results. The data are also ad-hoc subgroup analyses and should be considered observational in nature. Furthermore, meta-regressions indicated limited evidence of effect modification by age or severity and that the more important difference to model was vehicle response.

Table 41: Assessment of model fit in subgroups of ISGA/IGA 0/1 evidence networks.

Subgroup	Model	Total residual deviance	Number of datapoints	DIC
Moderate AD	FE-RCE-VR	18.11	18	116.7
	FE-RCE-VR inconsistency	18.51	18	117.5
Mild AD	FE-RCE-VR	12.65	14	90.65
	FE-RCE-VR inconsistency	13.28	14	92.06
Children with AD	FE-RCE-VR	Doesn't converge		
	Random effects	Doesn't converge		
	Fixed effects	8.676	10	72.69

	FE inconsistency	9.531	10	74.53
Adults with AD	FE-RCE-VR	Doesn't converge		
	Random effects	Doesn't converge		
	Fixed effects	4.026	4	28.46
	FE inconsistency	4.027	4	28.46
Moderate AD in children	FE-RCE-VR	11.36	12	76.05
	FE-RCE-VR inconsistency	11.89	12	77.13
Mild AD in children	FE-RCE-VR	11.8	12	78.37
	FE-RCE-VR inconsistency	12.46	12	79.85

Figure 24: Subgroup of moderate AD. Forest plot of NMA estimated hazard ratios (95% CrI) with P-best (probability crisaborole superior) of ISGA/IGA 0/1 of comparators versus crisaborole using a FE-RCE-VR model.

Figure 25: Subgroup of mild AD. Forest plot of NMA estimated hazard ratios (95% CrI) with P-best (probability crisaborole superior) of ISGA/IGA 0/1 of comparators versus crisaborole.

Figure 26: Subgroup of children with AD. Forest plot of NMA estimated hazard ratios (95% CrI) with P-best (probability crisaborole superior) of ISGA/IGA 0/1 of comparators versus crisaborole. Using simple fixed treatment effects model.

Figure 27: Subgroup of adults with AD. Forest plot of NMA estimated hazard ratios (95% CrI) with P-best (probability crisaborole superior) of ISGA/IGA 0/1 of

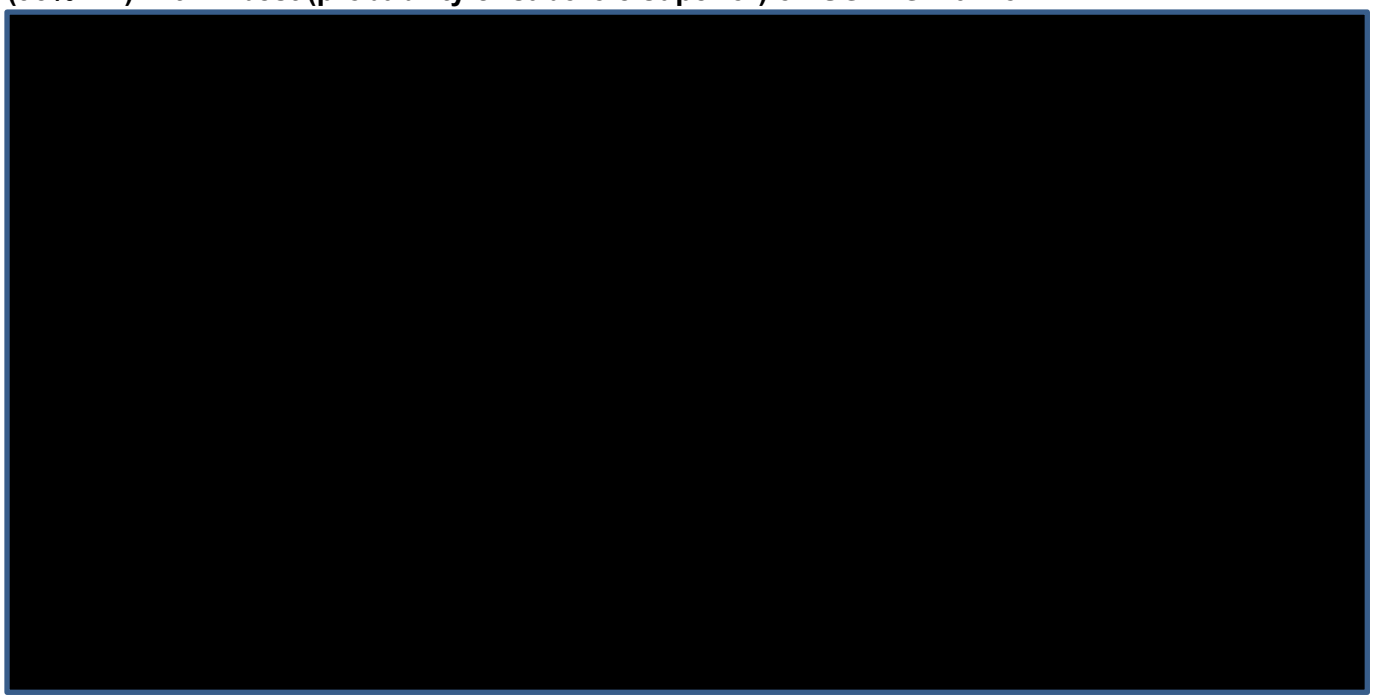


Figure 28: Subgroup of moderate AD in children. Forest plot of NMA estimated hazard ratios (95% CrI) with P-best (probability crisaborole superior) of ISGA/IGA 0/1 of comparators versus crisaborole.

Figure 29: Subgroup of mild AD in children. Forest plot of NMA estimated hazard ratios (95% CrI) with P-best (probability crisaborole superior) of ISGA/IGA 0/1 of comparators versus crisaborole.

B.2.8.9 Sensitivity analyses

Sensitivity analyses on the ISGA/IGA 0/1 outcome are provided in Appendix D. These explored sensitivity to the data and to the model. The results of our base case were robust to these sensitivities as the conclusions didn't change

Pruritus 0/1 was not connected as it was not reported in any studies on pimecrolimus 1%, tacrolimus 0.03%, or tacrolimus 0.1% except for the pimecrolimus 1% versus vehicle studies in facial AD (144-146). These latter studies were excluded from the NMA due to greater sensitivity in the face and neck area and thus greater treatment response. Inclusion would substantially bias the analysis. A further reason to avoid NMA on pruritus is the absence of a standard pruritus scale, complicating NMA using available data.

B.2.8.10 The definitions of ISGA success are provided in **Table 30**. This outcome was only reported heterogeneously by comparator studies and no connected network could be formed. No NMA was therefore conducted. **Unanchored MAIC on ISGA 0-1 conduced due to differences in vehicle**

As discussed in **Section B.2.9.4.1** and **Appendix D**, there are substantial differences in the vehicle composition and thus vehicle response across trials, the evidence network for

ISGA/IGA 0/1 presented in **Figure 22** may not truly be connected as vehicle is not the same treatment across RCTs. Although our NMA model FE-RCE-VR included vehicle response regression this is inevitably underpowered as only aggregate data are used.(135, 136) We therefore conducted a sensitivity analysis using the individual patient data (IPD) from crisaborole arms of AD-301 and AD-302 to conduct an unanchored MAIC. Note that anchored MAIC was considered but there is limited evidence of effect modification so limited justification for using this method over standard (135); in addition, anchored MAIC can only make use of RCTs linked to vehicle so loses indirect evidence between tacrolimus doses and pimecrolimus 1% (e.g. Kempers 2004, Abramovits 2008, Paller 2005).(42, 44, 45)

Full details of the unanchored MAIC methodology are presented in appendix D. In brief, we use a binomial logistic regression model for ISGA 0/1 for both the AD-301/AD-302 crisaborole trials and the comparator trials on pimecrolimus 1%, tacrolimus 0.03%, and tacrolimus 0.1%. Based on regression analyses of the AD-301 and AD-302 RCTs, a targeted literature review, and medical expert opinion, we determined a list of potential effect modifiers or prognostic variables to match:

[redacted]If more than one trial arm was available on a comparator (e.g. 4 arms on pimecrolimus 1%: Kempers 2004, Paller 2005, Abramovits 2008, Eichenfield 2002; 5 arms on tacrolimus 0.03%: Kempers 2004, Chapman 2005, Paller 2005, Levy 2005, Schachner 2005; while only 1 arm available on tacrolmus 0.1%: Abramovits 2008), fixed effects meta-analysis was applied to the odds of achieving ISGA/IGA 0/1.The matching characteristics, listed above, were then combined in a weighted average using weights from the fixed effects meta-analysis. If only a subset of RCTs reported the characteristic only they contributed to the average; this made a limiting assumption that the characteristic is similar in studies that do not report to those that do report. The reweighed ISGA/IGA 0/1 response was estimated from pooled AD-301 and AD-302 IPD, with a covariate for study effect, using a robust ‘sandwich’ estimator as recommended by the NICE DSU TSD 18.(135) We furthermore note that as the comparator RCTs varied on their patient baseline characteristics there were different implicit target populations for each comparison. In each case, the target population is that of the weighted average of the comparator RCTs; however, the importance of the unanchored MAIC is as an assessment of the impact of differences vehicle response on indirect comparisons, not differences in effect modifiers or prognostic variables.

RCTs included for each comparator are summarised in Table 42. We report also the effective sample size (ESS) of the crisaborole arms of AD-301 and AD-302 after matching. The ESS is the number of independent non-weighted individuals that would be required to give an estimate with the same precision as the weighted sample estimate; weighting always reduces the effective sample size. The ESS for crisaborole arms of AD-301 and AD-302 is reduced from a possible maximum of 1021. The extent of reduction gives an assessment of similarity/overlap of the crisaborole and comparator populations on the reported matching variables. The extent of overlap with the tacrolimus 0.1% RCTs is seen to be very low, meaning unanchored MAIC with this treatment have limited validity. The overlap with pimecrolimus 1% and tacrolimus 0.03% is reasonable with the ESS over 50% and 30% of the unadjusted sample size of 1,021. This robustness is reinforced by the patient weights histograms in Appendix D. A comparison of patient characteristics before and after matching

to pimecrolimus 1% and tacrolimus 0.03% is presented in Appendix D. These indicate [redacted]

Table 42: Evidence on comparators and assessment of difference in populations using ESS

Comparator	ESS of crisaborole arms from AD-301 and AD-302 after matching (max=1021)	Number of patients in AD-301 and AD-302 assigned zero weight	Studies with arms on the comparator treatment
Pimecrolimus 1%	[redacted]	[redacted]	Kempers 2004, Paller 2005, Abramovits 2008, Eichenfield 2002
Tacrolimus 0.03%	[redacted]	[redacted]	Kempers 2004, Chapman 2005, Schachner 2005, Paller 2005, Levy 2005
Tacrolimus 0.1%	[redacted]	[redacted]	Abramovits 2008

The results of the unanchored MAIC are presented in **Figure 30** and these can be compared with the base case NMA results presented in **Figure 19**.

[redacted] We align with a checklist of recommendations for reporting MAICs provided in Appendix B of Phillippo 2018 (135), as detailed in **Table 43**.

The unanchored MAIC indicates [redacted]

Despite the important differences, the difference in results between NMA and MAIC is most likely driven by the removal of vehicle controls from the indirect comparison. The unanchored MAIC indicates that vehicle response regression may not fully account for

differences in vehicle response across RCTs, possibly because there are too few RCTs on each treatment contrast. Although the unanchored MAIC should not be viewed as a base case analysis, it suggests sensitivity of the NMA to unaccountable differences in vehicle response.

Figure 30: Forest plot of odds ratios (with 95% Cs, two-sided p-values and effective sample sizes) comparing crisaborole to comparators using unanchored MAIC and naïve (no reweighting) comparison.

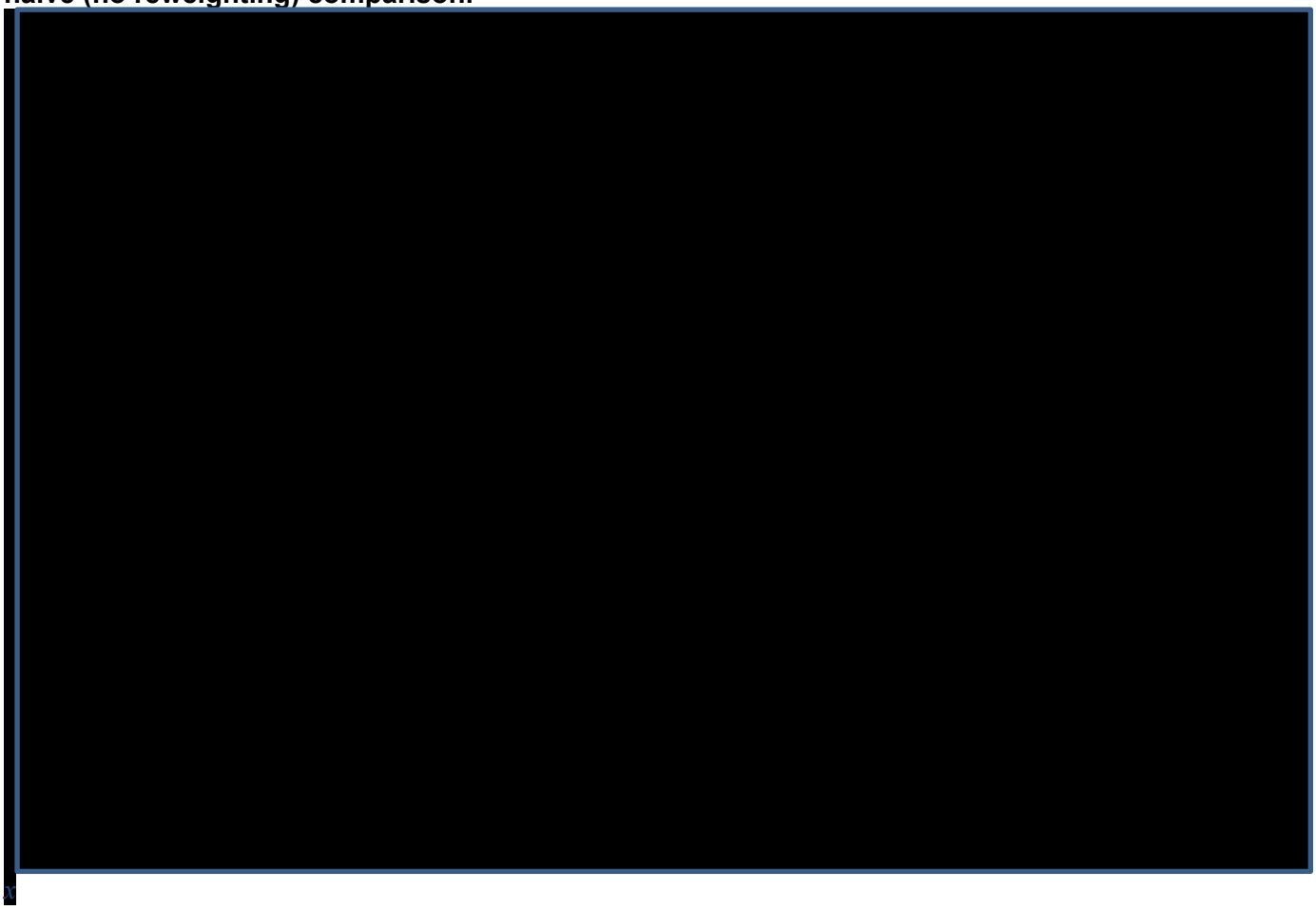


Table 43 Checklist for population adjusted indirect comparison from Phillippo 2018

Recommendation of Appendix B of Phillippo 2018	Details of recommendation	How we followed the recommendation
Recommendation 1.	The variables available in each study should be listed, along with their distributions Sufficient covariate overlap between the populations should be assessed: for population reweighting methods (such as MAIC), the number of individuals assigned zero weight should be reported;	Variables are listed in **Table D40** and **Table D41** of **Appendix D.** The number of patients with zero weight are reported in **Table 42**. There are zero such patients for all three comparisons.

	For unanchored comparisons all variables relevant to outcome should be presented.	All potential prognostic variables and effect modifiers are presented.
Recommendation 2.	Evidence for effect modifier status should be given, along with the proposed size of the interaction effect and the imbalance between the study populations.	Evidence for effect modifier or prognostic variable status are provided in **Appendix D**. This includes literature reviews, regression analyses using AD-301 and AD-302 data, and expert opinion.
Recommendation 3.	The distribution of weights should be presented for population weighting analyses, and used to highlight any issues with extreme or highly variable weights. Presentation of the effective sample size may also be useful.	The distribution of patient weights for comparisons with tacrolimus 0.03% and pimecrolimus 1% are presented in **Figure D33** of **Appendix D**.
Recommendation 4.	Measures of uncertainty, such as confidence intervals, should always be presented alongside any estimates. Care should be taken that uncertainty is appropriately propagated through to the final estimates. For population reweighting methods, a robust sandwich estimator (as typical for MAIC) provides estimates of standard error which account for all sources of uncertainty.	Confidence intervals (95% CI) are presented in the results forest plot (**Figure 30**) A robust sandwich estimator was used for the unanchored MAIC using the code provided in NICE DSU TSD 18.(135)
Recommendation 5.	For an unanchored comparison, estimates of systematic error before and after population adjustment should be presented.	Methods for estimating this systematic error have not been published. We include naïve (unweighted estimates) alongside unanchored MAIC estimates. We note the potential for systematic error when presenting results.
Recommendation 6.	Present estimates for the appropriate target population using the shared effect modifier assumption if appropriate, or comment on the representativeness of the aggregate population to the true target population.	The target population was that of the comparator RCTs. The importance of unanchored MAIC was to assess impact of differences in vehicle response, not difference in populations.
Recommendation 7.	In order to convey some clarity about the impact of any population adjustment, for an unanchored comparison, a crude	A naïve indirect comparison (no reweighting) was included alongside the

	unadjusted difference should be presented alongside the MAIC estimate.	MAIC estimate in **Figure 30**.

B.2.8.11 [redacted]Conclusions of the indirect comparison

Our NMA was limited by the need to assume that vehicle arms were comparable to placebo and thus could be used as an anchor to indirectly compare crisaborole with pimecrolimus and tacrolimus. However, differences in vehicle ingredients across RCTs lead to substantial differences in response. This indicated that the network is potentially disconnected and assuming vehicles are comparable, even with vehicle response regression, is likely to have biased the NMA (**Section B.2.9.2** and **Appendix D**). Although our base case analysis followed the NMA approach recommended by the NICE DSU, we supplemented this with a sensitivity employing unanchored MAIC, recommended by the NICE DSU in cases where evidence networks are disconnected. Our approach is fully aligned with the NICE DSU recommendations, with a full checklist provided in **Appendix D (Table D42)**(155).

[redacted]

[redacted]These findings were robust to sensitivity analyses to both the model and underlying dataset (**Section B.2.9.6**).

As explained in **Section B.2.9.5**, pimecrolimus 1% is only licensed and reimbursed in the UK in children with moderate AD (and only for the face and neck). Furthermore, tacrolimus 0.03% is licensed and reimbursed in children or adults with moderate AD while tacrolimus 0.1% is only licensed and reimbursed in adults with moderate AD. In light of this landscape we conducted subgroup NMAs in children with mild AD, children with moderate AD, and adults with moderate AD. However, the data available to inform these analyses was limited and the networks of evidence sparsely populated. The data are also ad-hoc subgroup analyses and should be considered observational in nature. There was no comparative evidence on adults with mild AD. In children with mild AD using the FE-RCE-VR model (**Figure 29**) results indicate

[redacted]Meta-regression was additionally undertaken, the meta-regressions indicated limited evidence of effect modification by age or

severity and that the more important difference to model was vehicle response, which we included in the base case NMA.

Results of the MAIC are presented in **Figure 30** of **Section B.2.9.7**. This found [redacted]). Notably, unweighted naïve comparisons agree with these conclusions. [redacted] The unanchored MAIC with pimecrolimus 1% and tacrolimus 0.03% suggest heavy sensitivity of the NMA to unaccountable differences in vehicle response. They indicate that crisaborole increases the odds of an ISGA 0/1 response compared to those TCIs licensed in children, directionally consistent with the base case NMA. Although unanchored MAIC should not be considered the base case due to lack of randomization, its use was supported by the strong impact, and potential bias, that vehicle response had on treatment effects.

B.2.9 Adverse reactions

Overall, the crisaborole safety database supporting this submission includes safety data from a total of 1,511 subjects. Crisaborole was found to have a favourable safety profile for both short-term (28 days) and long-term use (48 weeks).(12, 41) For both short- and long-term use, the majority of treatment-emergent adverse events (TEAEs) were mild to moderate in nature and deemed to be unrelated to the crisaborole treatment.

In AD-301, five treatment-emergent serious adverse events (TESAEs) were reported in 5/502 (1.0%) patients treated with crisaborole versus one TESAE in 1/252 (0.4%) subjects treated with vehicle. No treatment-related TESAEs were reported, except for severe cellulitis reported as possibly related to the vehicle treatment. The rates of patients reporting at least one TEAE on study were similar between treatment groups (147/502 [29.3%] and 50/252[19.8%] patients in the crisaborole versus vehicle treated group, respectively). The rates of patients who discontinued from study participation due to TEAE were 7/502 [1.4%] and 2/252 [0.8%] patients treated with crisaborole and vehicle, respectively. There were no clinically important differences by age group in the frequencies or types of TEAEs reported during the study, in the severity of events reported, the relationship to study drug, or premature discontinuations due to TEAEs.

In AD-302, three TAEs were reported in 3/510 (0.6%) patients treated with crisaborole versus none reported in patients treated with vehicle. None of the TESAEs were considered treatment related. The rates of patients reporting at least one TEAE through the day 29 visit (regardless of relationship) were similar between treatment groups (150/510 [29.4%] and 79/247[32.0%] in the crisaborole and vehicle groups, respectively). Similar to AD-301, there were no clinically important differences by age group in the frequencies or types of TEAEs reported during the study, in the severity of events reported, the relationship to study drug, or premature discontinuations due to TEAEs.

In the long-term safety (LTS) study (AD-303), 396 patients were exposed to crisaborole treatment for 6 months and 271 patients were exposed to crisaborole treatment for 12 months. Seven TESAEs were reported in the long-term safety trial, none of which were considered treatment-related. Nine subjects (1.7%) discontinued the LTS due to TEAEs. Eight patients (1.5%) had study drug usage interrupted due to one or more AEs and did not restart treatment. Eighteen patients (3.5%) had an AE for which study usage was interrupted and later resumed.

During the pivotal studies and AD-303, 65% of patients reported ≥1 TEAE, most of which were mild (51.2%) or moderate (44.6%) and considered unrelated to treatment (93.1%).(41) The frequency and severity of TEAEs were consistent.(12, 41) The most commonly reported treatment-related adverse events (AEs) were worsening of atopic dermatitis (exacerbation, flare, or flare-up) (48 weeks: 3.1%), application-site pain (reported as burning and/or stinging) (48 weeks: 2.3%), and application-site infection (48 weeks: 1.2%).(12, 41) No serious AEs were reported.

In these phase III studies of crisaborole, there was no evidence of application-site cutaneous adverse reactions such as atrophy or telangiectasia, side effects that have been reported

with long-term TCS treatment. Additionally, mild treatment-related application site discoloration was reported in a single Caucasian patient who experienced grey skin discoloration of hands and feet at the application site. Thus, the safety profile of crisaborole suggest that it may be an appropriate treatment option for patients with skin of colour and those at risk for application-site cutaneous adverse reactions (e.g. AD on sensitive skin areas such as the face, neck, and skin folds).

B.2.9.1 AD-301 and AD-302

The frequency of TEAEs were collected for all randomised study participants receiving at least one dose of the study drug. Short-term treatment (28 days) with crisaborole had similar rates of TEAEs as vehicle treatment (**Table 44**). The majority of TEAEs were mild or moderate for patients treated with crisaborole treated patients (94.3%) compared to patients treated with vehicle (96.9%), and the rates of discontinuation were equal for both treatment groups at 1.2%(12) A summary of TEAEs that occurred in at least 1% of subjects overall is presented by treatment group in **Table 44** for the pivotal trials, AD-301 and AD-302.

Table 44: Summary of treatment-related adverse events and treatment-emergent adverse events (≥1% of patients) for AD-301 and AD-302

Cohort	Crisaborole ointment, N = 1012	Vehicle, N = 499
Treatment-related adverse event, n (%)		
Application-site pain	45 (4.4)	6 (1.2)
Treatment-emergent adverse event, n (%)		
Gastrointestinal disorders Vomiting	**27 (2.7)** 15 (1.5)	**12 (2.4)** 5 (1.0)
General disorders and administration site conditions Application-site pain Application-site pruritus Pyrexia	**75 (7.4)** 45 (4.4) 5 (0.5) 19 (1.9)	**25 (5.0)** 6 (1.2) 6 (1.2) 7 (1.4)
Infections and infestations Nasopharyngitis Staphylococcal skin infection Upper respiratory tract infection	**118 (11.7)** 18 (1.8) 1 (0.1) 30 (3.0)	**59 (11.8)** 6 (1.2) 5 (1.0) 15 (3.0)
Injury, poisoning, and procedural complications	**20 (2.0)**	**9 (1.8)**
Investigations	**10 (1.0)**	**6 (1.2)**
Nervous system disorders Headache	**14 (1.4)** 11 (1.1)	**2 (0.4)** 1 (0.2)
Respiratory, thoracic, and mediastinal disorders Cough Oropharyngeal pain	**47 (4.6)** 12 (1.2) 11 (1.1)	**15 (3.0)** 8 (1.6) 2 (0.4)
Skin and subcutaneous tissue disorders Dermatitis atopic	**37 (3.7)** 7 (0.7)	**21 (4.2)** 8 (1.6)

Abbreviations: AD: atopic dermatitis; n: population number

A post-hoc analysis of TEAEs stratified by age and %BSA was conducted (**Table 45**). Across both treatment groups, [redacted] Among subjects treated with crisaborole, [redacted]

Table 45: Number and Percent of Subjects with TEAEs by Age and %BSA

%BSA	2-6 years				7-11 years				≥12 years			
	Crisaborole		Vehicle		Crisaborole		Vehicle		Crisaborole		Vehicle	
	n/N	%	n/N	%	n/N	%	n/N	%	n/N	%	n/N	%
0.1-<16	[redacted]											
16-≤40	[redacted]											
>40	[redacted]											

Abbreviations: BSA: Body Surface Area; N: population number; n: number of subjects with TEAEs by each subgroup; TEAEs: Treatment-emergent adverse events
Note: pooled data from AD-301 and AD-302

B.2.9.2 Safety overview

The majority of treatment-related AEs were mild to moderate over a combined 52-weeks of safety analysis of crisaborole for the treatment of AD. Application-site pain (described as burning and/or stinging) was one of the most commonly reported treatment-related adverse events that typically resolved within one day and diminished in frequency overtime; half (n=6; 1.2%) of these reactions were reported in the short-term (29 day) trials, and the majority of the remaining cases (n=5; 1%) were reported in the first three months of the long-term (48 week) open-label study.(12, 41) Incidence of AD (i.e. exacerbation, flare) and application-site infection related to crisaborole use, was also low at 3.1% (n=16) and 1.2% (n=6), respectively; most of which were resolved within one treatment cycle (approximately one month). Most patients (78%?) did not require rescue therapy throughout the duration of the long-term study. For those receiving rescue therapy, the mean duration of use was 21.4 days for TCS and 24.2 days for TCI with the majority of these patients (79.1%) returning to crisaborole treatment.(41) It was noted that patients experience difficulty adhering to treatment recommendations, which are time-consuming and challenging to perform.(156, 157) Therefore, low-treatment adherence leads to persistence, or worsening of AD symptomology, further affecting HRQoL for distressed patients and their caregivers. The low rate of treatment-related AEs and withdrawals from the LTE study suggest good adherence to crisaborole and demonstrates a favourable risk/benefit profile. In summary, crisaborole demonstrated a favourable safety profile as a monotherapy treatment for AD across a combined 52 weeks and did not result in cutaneous side effects commonly reported with long-term use of TCS.(41)

B.2.9.3 Short- and Long-term Safety Studies

Treatment exposure was similar across age groups (mean number of applications: 2-11 years: 349.0; 12-17 years: 349.4; ≥ 18 years: 347.7) with an average of 6.2 on-treatment periods (which includes treatment during the pivotal phase III AD-301 and AD-302 trials).(41) The most commonly reported TEAEs from the pivotal and long-term studies were atopic dermatitis (11.2%) and upper respiratory tract infection (10.3%). The majority of TEAEs reported were mild (51.2%) or moderate (44.6%), and 93.1% determined to be unrelated to the study drug.(41) The most frequently reported treatment-related AEs were atopic dermatitis (defined as worsening, exacerbation, flare or flare-up; 3.1%), application-site pain (reported as burning and/or stinging) (2.3%), and application-site infection (1.2%). A summary of TEAEs that occurred in at least 5% of subjects in AD-301, AD-302 AND AD-303 are presented by age group in **Table 46.**

Table 46: Summary of treatment-emergent adverse events (≥5% of patients) for the Short-term and Long-term Safety Studies of Patients exposed to Crisaborole

Cohort	2-11 years (N = 308)	12-17 years (N = 146)	≥ 18 years (N = 63)	Total patients (N = 517)
Treatment-emergent adverse event, n (%)				
General disorders and administration site conditions Pyrexia	**41 (13.3)** 27 (8.8)	**12 (8.0)** 2(1.4)	**12 (8.0)** 0 (0.0)	**58 (11.2)** 29 (5.6)
Infections and infestations Nasopharyngitis Upper respiratory tract infection	**157 (51.0)** 21 (6.8) 38 (12.3)	**56 (38.4)** 15 (10.3) 12 (8.2)	**14 (22.2)** 4 (6.3) 3 (4.8)	**227 (43.9)** 40 (7.7) 53 (10.3)
Respiratory, thoracic, and mediastinal disorders Cough	**55 (17.9)** 27 (8.8)	**26 (17.8)** 6 (4.1)	**5 (7.9)** 2 (3.2)	**86 (16.6)** 35 (6.8)
Skin and subcutaneous tissue disorders Dermatitis atopic	**65 (21.1)** 37 (12.0)	**35 (24.0)** 16 (11.0)	**9 (14.3)** 5 (7.9)	**109 (21.1)** 58 (11.2)

Abbreviations: n/N: population number

The evaluation of TEAEs in the long-term safety study (48 weeks) was carried out over four 12-week treatment periods. Across the four 12-week treatment periods there were no clinically relevant differences in the types of adverse events. Safety signals from vital signs or laboratory assessments were not identified during the treatment period for the pivotal trials or long-term safety study. Additionally, infections or the occurrence of neoplasms did not increase with long-term use of the study drug.(41) Treatment-related AEs of application-site infection had a mean duration of 12 days with 75% of study participants reporting resolution within one treatment cycle.(41) Similarly, 90% of participants experiencing application-site burning (treatment-related AE) had resolution within one treatment cycle.(41). The majority of participants (77.8%) did not require the use of rescue treatment during the 48 week treatment phase.(41) A summary of TEAEs that occurred in at least 1% of subjects for any 12-week period is presented in **Table 47.**

Table 47: Summary of treatment-emergent adverse events (≥1% of patients) in any 12-Week Period

Cohort	Days 1-85 N = 482	Days 86-169 N = 428	Days 170-253 N = 368	≥Days 254 N = 226
Treatment-emergent adverse event, n (%)				
Gastrointestinal disorders Diarrhoea Vomiting	**8 (1.7)** 2 (0.4) 2 (0.4)	**14 (3.3)** 5 (1.2) 4 (0.9)	**10 (2.7)** 1 (0.3) 4 (1.1)	**3 (1.3)** 0 (0.0) 0 (0.0)
General disorders and administration site conditions Application-site pain Pyrexia	**23 (4.8)** 5 (1.0) 13 (2.7)	**12 (2.8)** 1 (0.2) 5 (1.2)	**9 (2.4)** 0 (0.0) 7 (1.9)	**5 (2.2)** 0 (0.0) 3 (1.3)

Immune system disorders	**3 (0.6)**	**3 (0.7)**	**8 (2.2)**	**2 (0.9)**
Seasonal allergy	2 (0.4)	2 (0.5)	6 (1.6)	2 (0.9)
Infections and infestations	**93 (19.3)**	**89 (20.8)**	**65 (17.7)**	**36 (15.9)**
Application-site infection	8 (1.7)	6 (1.4)	4 (1.1)	2 (0.9)
Croup infections	2 (0.4)	0 (0.0)	0 (0.0)	3 (1.3)
Gastroenteritis viral	2 (0.4)	1 (0.2)	4 (1.1)	1 (0.4)
Influenza	4 (0.8)	6 (1.4)	1 (0.3)	1 (0.4)
Nasopharyngitis	15 (3.1)	12 (2.8)	6 (1.6)	3 (1.3)
Otitis media	5 (1.0)	2 (0.5)	1 (0.3)	3 (1.3)
Pharyngitis	3 (0.6)	7 (1.6)	2 (0.5)	3 (1.3)
Pharyngitis streptococcal	9 (1.9)	8 (1.9)	2 (0.5)	2 (0.9)
Sinusitis	8 (1.7)	9 (2.1)	8 (2.2)	1 (0.4)
Upper respiratory infection	19 (3.9)	16 (3.7)	18 (4.9)	6 (2.7)
Viral infection	1 (0.2)	6 (1.4)	3 (0.8)	1 (0.4)
Nervous system disorders	**6 (1.2)**	**6 (1.4)**	**6 (1.6)**	**2 (0.9)**
Headache	3 (0.6)	6 (1.4)	2 (0.5)	2 (0.9)
Respiratory, thoracic, and mediastinal disorders	**24 (5.0)**	**38 (8.4)**	**17 (5.0)**	**8 (3.5)**
Asthma	2 (0.4)	5 (1.2)	4 (1.1)	4 (1.8)
Cough	12 (2.5)	17 (4.0)	6 (1.6)	2 (0.9)
Nasal congestion	5 (1.0)	1 (0.2)	2 (0.5)	1 (0.4)
Oropharyngeal pain	3 (0.6)	8 (1.9)	2 (0.5)	1 (0.4)
Skin and subcutaneous tissue disorders	**42 (8.7)**	**36 (8.4)**	**13 (3.5)**	**28 (12.4)**
Dermatitis atopic	25 (5.2)	23 (5.4)	9 (2.4)	10 (4.4)
Dermatitis contact	5 (1.0)	1 (0.2)	3 (0.8)	6 (2.7)
Eczema	3 (0.6)	7 (1.6)	0 (0.0)	5 (2.2)

Abbreviations: n/N: population number

B.2.9.3.1 Severe treatment-emergent adverse events

A total of 33 subjects experienced at least one severe TEAE. The only severe TEAEs reported for more than one subject were atopic dermatitis (worsening, exacerbation, flare, or flare-up), application-site pain (burning or stinging), application-site infection, sinusitis, depression, and asthma. Overall, seven subjects (7/517 [1.4%]) had two types of severe treatment-related AEs. The most common severe treatment-related AE was atopic dermatitis, which occurred in five subjects who experienced a total of six severe events. Two subjects experienced application-site pain, which was considered to be related to the study drug.

Table 48: Summary of treatment-emergent severe adverse events during the Short- and Long-term Safety Studies

Cohort	2-11 years N = 308	12-17 years N = 146	≥ 18 years N = 63	Total patients N = 517
Treatment-emergent adverse event, n (%)				
Blood and lymphatic system disorders	**1 (0.3)**	**0 (0.0)**	**0 (0.0)**	**1 (0.2)**
Lymphadenopathy	1 (0.3)	0 (0.0)	0 (0.0)	1 (0.2)
Gastrointestinal disorders	**0 (0.0)**	**0 (0.0)**	**2 (3.2)**	**2 (0.4)**
Abdominal pain upper	0 (0.0)	0 (0.0)	1 (1.6)	1 (0.2)
Gastritis	0 (0.0)	0 (0.0)	1 (1.6)	1 (0.2)
Nausea	0 (0.0)	0 (0.0)	1 (1.6)	1 (0.2)
Vomiting	0 (0.0)	0 (0.0)	1 (1.6)	3 (0.6)
General disorders and administration site conditions	**1 (0.3)**	**1 (0.7)**	**1 (1.6)**	**3 (0.6)**
Application-site dermatitis	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.2)
Application-site pain	1 (0.3)	1 (0.7)	1 (1.6)	2 (0.4)
Immune system disorders	**0 (0.0)**	**1 (0.7)**	**0 (0.0)**	**1 (0.2)**
Anaphylactic reaction	0 (0.0)	1 (0.7)	0 (0.0)	1 (0.2)
Infections and infestations	**5 (1.6)**	**3 (2.1)**	**1 (1.6)**	**9 (1.7)**
Appendicitis	1 (0.3)	0 (0.0)	0 (0.0)	1 (0.2)
Application-site infection	1 (0.3)	1 (0.7)	0 (0.0)	2 (0.4)
Mycoplasma infection	0 (0.0)	1 (0.7)	0 (0.0)	1 (0.2)
Pharyngitis	0 (0.0)	1 (0.7)	0 (0.0)	1 (0.2)
Pharyngitis streptococcal	1 (0.3)	0 (0.0)	0 (0.0)	1 (0.2)
Pneumonia bacterial	1 (0.3)	0 (0.0)	0 (0.0)	1 (0.2)
Sinusitis	0 (0.0)	1 (0.7)	1 (1.6)	2 (0.4)
Upper respiratory tract infection	1 (0.3)	0 (0.0)	0 (0.0)	1 (0.2)
Injury, poisoning and procedural complications	**1 (0.3)**	**0 (0.0)**	**0 (0.0)**	**1 (0.2)**
Ligament sprain	1 (0.3)	0 (0.0)	0 (0.0)	1 (0.2)
Musculoskeletal and connective tissue disorder	**0 (0.0)**	**1 (0.7)**	**0 (0.0)**	**1 (0.2)**
Back pain	0 (0.0)	1 (0.7)	0 (0.0)	1 (0.2)
Nervous system disorders	**1 (0.3)**	**1 (0.7)**	**0 (0.0)**	**2 (0.4)**
CNS ventriculitis	1 (0.3)	0 (0.0)	0 (0.0)	1 (0.2)
Migraine without aura	0 (0.0)	1 (0.7)	0 (0.0)	1 (0.2)
Psychiatric disorders	**0 (0.0)**	**2 (1.4)**	**0 (0.0)**	**2 (0.4)**
Depression	0 (0.0)	2 (1.4)	0 (0.0)	2 (0.4)
Suicide attempt	0 (0.0)	1 (0.7)	0 (0.0)	1 (0.2)
Respiratory, thoracic and mediastinal disorders	**2 (0.6)**	**0 (0.0)**	**0 (0.0)**	**2 (0.4)**
Asthma	2 (0.6)	0 (0.0)	0 (0.0)	2 (0.4)
Skin and subcutaneous tissue disorders	**10 (3.2)**	**2 (1.4)**	**1 (1.6)**	**13 (2.5)**
Dermatitis atopic	10 (3.2)	1 (0.7)	1 (1.6)	12 (2.3)
Urticaria	0 (0.0)	1 (0.7)	0 (0.0)	1 (0.2)

Abbreviations: n/N: population number

B.2.9.4 Safety outcomes NMAs

NMA of reported safety outcomes was conducted using the methods described briefly in **Section B.2.8** and in detail in **Appendix D**.(158) The available data from RCTs on pimecrolimus 1%, tacrolimus 0.03%, and tacrolimus 0.1% are summarised in **Table 49**. The outcomes for which comparisons of crisaborole 2% with pimecrolimus and tacrolimus doses were feasible were: withdrawal due to AE, overall AE, and overall withdrawals. No comparator data on serious AEs was available. Application site pain was not reported by key studies linking pimecrolimus 1% to vehicle (47) or either dose of tacrolimus to vehicle (43). The evidence network for application site pain hence could not connect crisaborole to pimecrolimus 1%, tacrolimus 0.03%, or tacrolimus 0.1%.

Table 49: Summary of the trials used to carry safety NMAs

References of trial	Vehicle	Crisaborole 2%	Pimecrolimus 1%	Tacrolimus 0.03%	Tacrolimus 0.1%	OPA-15406 1%	OPA-15406 0.3%	Withdrawals due to AE	Overall AE	Serious AE	Overall withdrawals	Available timepoint closest to 4 weeks
AD 301	Y	Y						Y	Y	Y	Y	29 days
AD 302	Y	Y						Y	Y	Y	Y	29 days
Chapman et al. 2005	Y			Y				Y	Y		Y	6 weeks
Eichenfield et al. 2002	Y		Y					Y	Y		Y	6 weeks
Hanifin et al. 2016	Y					Y	Y	Y	Y	Y	Y	8 weeks
Kempers et al. 2004			Y	Y				Y	Y		Y	43 days
Levy et al. 2005	Y			Y					Y			4 weeks
Abramovits et al. 2008			Y		Y			Y	Y		Y	6 weeks
Paller et al. 2005			Y	Y				Y	Y		Y	6 weeks

Boguniewicz et al. 1998	Y			Y	Y			Y			Y	36 days

Prior to discussing results, it should be noted that the overall AEs analysis is subject to the serious limitation of high variability in adverse events reported by RCTs included in the networks.(158) Reported AEs are summarised in **Table 50**. For the reason of this variability, the withdrawals due to AE and overall withdrawals likely provide a less heterogeneous comparison of safety.

Table 50: Adverse events reported by RCTs included in the safety NMAs

Trial Name	Intervention	List of adverse events reported and included in NMA
Boguniewicz, 1998	Tacrolimus 0.03	not reported
Boguniewicz, 1998	Tacrolimus 0.1	not reported
Boguniewicz, 1998	Vehicle	not reported
Kempers 2004	Pimecrolimus	herpes simplex, staphylococcal infection NOS, atopic dermatitis NOS
Kempers 2004	Tacrolimus 0.03	herpes simplex, impetigo NOS, atopic dermatitis NOS, pruritus NOS, rash NOS, erythema
Chapman 2005(43)(114)(112)(112)(94)(94)	Tacrolimus 0.03	skin burning or sensation, increased itching, skin erythema, folliculitis, skin infection, acne, herpes simplex, eczema herception
Chapman 2005	Vehicle	skin burning or sensation, increased itching, skin erythema, folliculitis, skin infection, acne, herpes simplex, eczema herception
Paller 2005	Tacrolimus 0.03	application site burning, application site pruritus, application site pain, application site erythema, acne, herpes simplex
Paller 2005	Pimecrolimus	application site burning, application site pruritus, application site pain, application site erythema
Hanifin 2016	OPA-15406 0.3%	worsening atopic dermatitis, nasopharyngitis, headache, upper respiratory tract infection, vomiting, toothache, impetigo, diarrhea
Hanifin 2016	OPA-15406 1%	worsening atopic dermatitis, nasopharyngitis, headache, worsening pruritus, upper respiratory tract infection
Hanifin 2016	Vehicle	worsening atopic dermatitis, nasopharyngitis, worsening pruritus, vomiting, toothache, excoriation
Levy 2005	Tacrolimus 0.03%	not reported
Levy 2005	Vehicle	not reported
Abramovits 2008	Tacrolimus 0.1%	application site burning, application site pruritus, application site pain, application site warmth,

		application site erythema, application site temperature intolerance, alcohol intolerance, folliculitis, skin infection, infected dermatitis
Abramovits 2008	Pimecrolimus 1%	application site burning, application site pruritus, application site warmth, application site erythema
Eichenfield 2002	Pimecrolimus 1%	AEs reported in >10% of sample: upper respiratory trac infection NOS, headache NOS, cough, nasopharyngitis, application site burning
Eichenfield 2002	Vehicle	AEs reported in >10% of sample: upper respiratory trac infection NOS, headache NOS, cough, nasopharyngitis, application site burning
AD-301	Crisaborole	Treatment emergent AEs reported in >=1% of sample: gastrointestinal disorders (vomiting), general disorders and administration site conditions (application site pain, application site pruritus, pyrexia), infections and infestations (nasopharyngitis, upper respiratory tract infection), respiratory, thoracic and mediastinal disorders (nasal congestion)
AD-301	Vehicle	Treatment emergent AEs reported in >=1% of sample: gastrointestinal disorders (vomiting), general disorders and administration site conditions (application site pain, application site pruritus, pyrexia), infections and infestations (upper respiratory tract infection), respiratory
AD-302	Crisaborole	Treatment emergent AEs reported in >=1% of sample: gastrointestinal disorders (diarrhea, vomiting), general disorders and administration site conditions (application site pain, application site pruritus, pyrexia), infections and infestations (nasopharyngitis, staphylococcal skin infection, upper respiratory tract infection), injury, poisoning, and procedural complications, investigations, nervous system disorders (headache), respiratory, thoracic and mediastinal disorders (cough, oropharyngeal pain), skin and subcutaneous tissue disorders (dermatitis atopic, eczema, pruritus)
AD-302	Vehicle	Treatment emergent AEs reported in >=1% of sample: gastrointestinal disorders (diarrhea, vomiting), general disorders and administration site conditions (application site pain, application site pruritus, application site urticaria, pyrexia), infections and infestations (nasopharyngitis, staphylococcal skin infection, upper respiratory tract infection), injury, poisoning, and procedural complications, investigations, nervous system disorders (headache), respiratory, thoracic and mediastinal disorders (cough, oropharyngeal pain), skin and subcutaneous tissue disorders (dermatitis atopic, eczema, pruritus)

The network diagrams for safety comparisons are presented in **Figure 31**. These networks were analysed using the FE-RCE-VR model to maintain uniformity of approach with the ISGA/IGA 0/1 NMA. Model assessment for each of the outcomes is presented in **Table 51;** the residual deviance of FE-RCE-VR for each of the three safety outcomes is close to the number of data points and the DIC is only marginally different from that of simple fixed and random treatment effects models, with the possible exception of overall AEs. The class effect aspect of this model assumes treatment effect on hazard of each safety outcome is exchangeable within TCIs (pimecrolimus and tacrolimus). The vehicle response regression adjusts for differences in "baseline risk" across the trials and their influence on the hazard of safety outcomes.

Figure 31: Evidence networks of the safety NMAs on outcomes up to 8 weeks

Table 51: Assessment of model fit in safety outcomes evidence networks.

Outcome	Model	Total residual deviance	Number of datapoints	DIC
Overall AEs	FE – random class effect – adjusted for baseline risk	23.2	18	131.8
Withdrawals due to AE	Random effects	20.04	20	93.86
	Fixed effects	21.77	20	93.1
	FE – random class effect – adjusted for baseline risk	20.0	20	94.69
Overall	Random effects	20.15	20	128.5

withdrawals	Fixed effects	21.39	20	127.3
	FE – random class effect – adjusted for baseline risk	22.22	20	128.1

The results on overall AEs (Figure 32) indicate

The results on withdrawals due to AE (Figure 33)

The overall withdrawals analysis (Figure 34)

Overall, the safety NMAs on overall AEs, withdrawals due to AE, and overall withdrawals

Figure 32: Overall AEs up to 8 weeks. Forest plot of results (hazard ratios, 95% CrI, p-best) of NMA

Figure 33: Withdrawals due to AE up to 8 weeks. Forest plot of results (hazard ratios, 95% CrI, p-best) of NMA using simple random treatment effects model with no class effect or vehicle response adjustment

Figure 34: Overall withdrawals up to 8 weeks. Forest plot of results (hazard ratios, 95% CrI, p-best) of NMA

B.2.10 Ongoing studies

A Phase 3B/4, multicentre, 4-week, randomised, assessor blinded, vehicle and active controlled, parallel group study is ongoing to evaluate the safety and efficacy of crisaborole ointment, 2%, crisaborole vehicle, hydrocortisone butyrate cream, 0.1%, and pimecrolimus cream, 1% in subjects 2 years of age and older with mild-to-moderate AD. This study (NCT03539601) is expected to enroll 600 participants and [redacted]

A single-center two arm, open label observational prospective study, that will evaluate the safety and efficacy of crisaborole ointment, 2% alone compared to a combination therapy of crisaborole and a topical corticosteroid (Triamcinolone Acetonide Ointment, 0.1%) over an 8 week period for the treatment of mild-to-moderate AD is currently recruiting patients (NCT04008784).

B.2.11 Innovation

Crisaborole employs a unique mechanism of action for a topically applied drug in the treatment of mild-to-moderate AD. Crisaborole is a non-steroidal compound and first-in-class topical PDE4 inhibitor. Increasing intracellular cAMP levels by inhibiting PDE4 has been proven to be an effective therapeutic strategy to decrease inflammatory cytokines involved in AD. Crisaborole is a topical ointment, hence, has limited systemic effects and is not associated with the serious adverse events reported with oral PDE4 inhibitors, such as nausea, vomiting, emesis, and headache.(41, 159) Results from the pivotal trials show that crisaborole provides rapid and sustained improvement in pruritus, a symptom of AD that is responsible for a significant proportion of disease burden but is not adequately captured by the EQ-5D. Crisaborole reversed biomarker profiles of skin inflammation and barrier function, with associated improvements in clinical efficacy measures, highlighting the therapeutic utility of targeting PDE4 in AD.(35)

B.2.12 Interpretation of clinical effectiveness and safety evidence

B.2.12.1 Principal findings from the clinical evidence highlighting the clinical benefits and harms of the technology

Overall, crisaborole 20 mg/g ointment has been shown to be an effective treatment for patients 2 years of age and older with mild-to-moderate AD. Crisaborole was statistically superior to vehicle ointment on the primary and secondary efficacy endpoints in both pivotal trials. Moreover, a greater proportion of patients treated with crisaborole experienced improvement in all signs and symptoms of AD at day 29 than those treated with vehicle. It is noted that moderate to severe baseline pruritus was reported in approximately 70% of patients in the pivotal trials, and a significantly greater proportion of patients treated with crisaborole treated patients became itch-free by day 6. Furthermore, early improvement in pruritus was associated with improvements in HRQoL including achievement of MCID in CDLQI and DLQI. Patients treated with crisaborole treated patients reported an overall reduction in the impact of AD on their HRQoL. Additionally, crisaborole treatment lessened

the burden of AD on parents and caregivers. Relevant subgroup analyses for age, baseline severity, prior treatment use, and race and ethnicity revealed numerically greater proportions of patients treated with crisaborole treated patients than patients treated with vehicle experienced improvement in ISGA. Additionally, crisaborole improved AD signs and symptoms common in patients with skin of colour, a patient population at higher risk for AD yet underrepresented in AD clinical trials.

NMA was used to compare crisaborole to pimecrolimus 1%, tacrolimus 0.1%, and tacrolimus 0.03% in a mixture of adults and children and of mild and moderate RCTs. The key outcome was merged ISGA 0/1 with IGA 0/1, following prior meta-analyses and expert opinion; there was no comparator evidence on ISGA success, the primary outcome of the AD301/AD302 RCTs, to conduct NMA. Vehicle response regression was used to adjust for the substantial variation in vehicle response across RCTs. A random class effect was used across TCIs (pimecrolimus 1% and tacrolimus doses). The NMA found [redacted] These findings were robust to sensitivity analyses to both the model and underlying dataset. The safety database included 1012 patients who were exposed to crisaborole and 499 patients who were exposed to vehicle for 28 days in the pivotal trials and 517 patients who continued treatment with crisaborole in an open-label long-term study for up to 48 weeks. In the pivotal trials, the safety profile of crisaborole treatment was similar to vehicle treatment and no serious AEs were reported. The safety profile of crisaborole was favourable overall for long-term treatment of patients, and crisaborole demonstrated a low frequency of TEAEs and treatment-related AEs, most of which were mild to moderate. Additionally, most patients did not require rescue therapy throughout the long-term study, and the majority of those who did receive rescue therapy resumed crisaborole monotherapy within one month.

Safety NMAs were conducted on overall AEs, withdrawals due to AE, and overall withdrawals. These [redacted] The only exception was [redacted] The addition of crisaborole as a topical therapeutic option with a favourable safety profile, clinical efficacy, and limited systemic exposure for the treatment of mild-to-moderate AD is important to patients, caregivers, and physicians. While TCSs and TCIs are effective therapies, they are associated with significant

potential adverse events with chronic use. A theoretical benefit of crisaborole is the potential to reduce the use of TCSs and TCIs.

B.2.12.2 Strengths and limitations of the clinical evidence base for the technology

B.2.12.2.1 Strengths of the evidence base

The two Phase III clinical trials (AD-301 and AD-302) of crisaborole 2%, were multi-centre, double-blind, randomised, placebo-controlled studies, representing the highest standard of clinical evidence.

The crisaborole trials addressed the decision problem and included patient populations and clinical outcomes relevant to the final NICE scope. The trials included patients 2 years of age and older with mild-to-moderate AD, which represents patients who may receive crisaborole in clinical practice. Baseline demographics and disease-specific characteristics were similar across the trials and were well-balanced between the treatment and vehicle arms in each trial.

The endpoints measured across the Phase III trials were well-recognised, clinically-relevant outcomes, and the assessed AD signs and symptoms are consistent with clinical practice in the UK.(24) Moreover, HRQoL assessments align with recommendations by NICE as part of a holistic approach to AD management.(24) Age, baseline severity, and prior treatment use subgroups are relevant to the final NICE scope. Furthermore, the LTE study, AD-303, examined the safety profile of crisaborole over a 48-week treatment period and confirmed the low frequency of treatment-related AEs seen in the pivotal trials.

The statistical analyses employed across the crisaborole clinical trials were robust and conservative in nature. The primary endpoint of achievement of an ISGA score of clear (0) or almost clear (1) *and* an improvement in ISGA score by at least 2, meant that a patient with mild AD must be completely clear to achieve the primary endpoint and a patient with moderate AD must become at least ‘almost clear.’

The data supporting the safety of crisaborole are comprehensive and include the LTE study reporting safety data up to 48 weeks. Overall, crisaborole is well tolerated with a stable AE profile over time. In contrast to TCSs, long-term use of crisaborole is not associated with systemic and cutaneous side-effects; nor is it associated with special warnings as is the case for TCIs (160-162). Further, crisaborole does not pose a risk of application site discolouration, a relevant concern for patients of colour.

The crisaborole clinical trials did not include active comparators. To address this an SLR and NMA comparing crisaborole to key comparators has been undertaken (pimecrolimus 1%, tacrolimus 0.1%, tacrolimus 0.03%). This merged across adults and children and across mild and moderate AD, representing a comparison in this merged population. The key outcome was ISGA 0/1 merged IGA 0/1, ensuring a connected evidence network. Extensive exploration of potential models was conducted, covering fixed and random treatment effects, vehicle response regression, and fixed and random class effects. The selected fixed

treatment effect, random class effect, vehicle response regression (FE-RCE-VR) model adjusts for many potential effect modifiers through the regression on vehicle response. To further explore heterogeneity, meta-regressions on age, gender, % BSA, proportion Caucasian, and severity were conducted; none were superior in fit to our selected FE-RCE-VR model indicating heterogeneity was being accounted for in an optimal manner. Global and local inconsistency tests were conducted, with only mixed evidence on inconsistency being identified. Sensitivity analyses to both the model assumptions and underlying data did not indicate changes in our base case conclusions.

B.2.12.2.2 Potential limitations of the evidence base

The base case NMA merged across adults and children and across mild and moderate AD but this does not reflect the licensing and reimbursement landscape (e.g. tacrolimus 0.1% is not licensed or reimbursed in mild AD). In order to better address this landscape, analyses in children with mild AD, children with moderate AD, and adults with moderate AD were explored. Nevertheless, it should be noted that the data available to inform these analyses was limited and the networks of evidence sparsely populated. The data are also ad-hoc subgroup analyses and should be considered observational in nature. Furthermore, there was no comparative evidence on adults with mild AD. In context of these limitations, in children with mild AD, results indicate [redacted]A comparison to TCSs was not possible due to trial population differences and the lack of subgroup data reporting with which to compare to the mild to moderate population treated with crisaborole.

[redacted]The NMA assumed that vehicle arms were comparable across trials. However, differences in vehicle ingredients across RCTs lead to substantial differences in response. [redacted]

Taken together, these differences in vehicle properties would make the relative benefits of the active therapies in older trials appear artificially high.(133) Differences in vehicle

response and properties suggest the evidence networks used in NMA are actually disconnected.

To account for these differences in vehicle properties across RCTs, we employed vehicle response regression in our base case NMA. However, this analysis was limited by only using aggregate data. The strength of the regression also suggested that comparison of absolute outcomes may be optimal, supporting the view that our networks are disconnected due to differences in 'vehicle' across RCTs. We therefore conducted an unanchored matching adjusted indirect comparison (MAIC) as a sensitivity analysis as described in **Section B.2.9.10**. This found strong evidence that crisaborole has higher odds of achieving ISGA/IGA 0/1 than tacrolimus 0.03 and pimecrolimus 1%. The unanchored MAIC suggests heavy sensitivity of the NMA to unaccountable differences in vehicle response. They indicate that crisaborole increases the odds of an ISGA 0/1 response compared to those TCIs licensed in children. Although unanchored MAIC should not be considered the base case due to lack of randomisation, its use was supported by the strong impact, and potential bias, that vehicle response had on treatment effects.

B.3 Cost effectiveness

B.3.1 Published cost-effectiveness studies

An SLR was conducted to identify relevant cost-effectiveness evidence for use in the cost-effectiveness analysis. This search was conducted in March 2019. A detailed description of the search, its methods, and results is provided in **Appendix G.**

B.3.2 Economic analysis

Crisaborole was not identified as a comparator in any economic evaluation identified in the systematic review, it was consequently necessary to develop a de novo cost-effectiveness model.

One of three modelling approaches have been used in previous CE models for other therapies in this indication:

1. Treatment state models (e.g. Garside/TA82 model) (163)
2. Treatment response models (e.g. ICER dupilumab model) (133)
3. Markov model using health states (e.g IGA Ellis et al) (164)

[redacted] Whilst, the second approach undertaken in the dupilumab submission does not appear to capture the relapsing nature of AD.

The third approach defines disease severity (based on IGA score), which in turn is associated with resource use and HRQoL patterns. This approach allows the model to

capture disease remission and acute flares and a similar approach is taken in the de novo model.

B.3.2.1 Patient population

The economic evaluation considers children and adults aged two years and over with mild and moderate AD, in line with the anticipated marketing authorisation, and the clinical trial populations from AD-301 and 302.(12) Sub-populations considered have been based on age and disease severity, in alignment to those subgroups in which topical calcineurin inhibitors (TCIs) are approved for use and recommended by NICE. Comparators are varied according to their reimbursement and marketing authorisation conditions as follows:

1. Children (aged 2–17) with mild AD
2. Children (aged 2–17) with moderate AD
3. Adults (aged ≥18) with mild AD
4. Adults (aged ≥18) with moderate AD

Whilst crisaborole has been demonstrated to be safe and effective for treatment of adults and children aged 2 years and older with mild to moderate AD it is not anticipated that crisaborole will be recommended in the UK for first line use in TCS eligible patients, due to the very low cost of TCS. Pfizer is consequently seeking reimbursement for crisaborole in an optimised population for the treatment of adults and children aged 2 years and older with mild to moderate AD that has not been adequately controlled by topical corticosteroids or where there is a serious risk of important adverse effects from further topical corticosteroid use, particularly irreversible skin atrophy (second-line use post-TCS treatment). Cost-effectiveness analyses for AD TCS eligible (first line) populations have been presented in **Appendix N**.

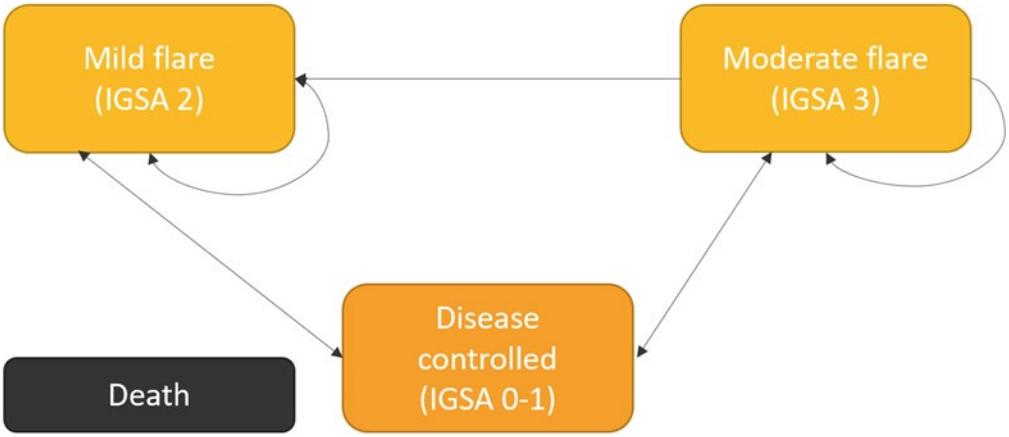
B.3.2.2 Model structure

A Markov cohort model was developed in Microsoft Excel to evaluate the cost-effectiveness of crisaborole and comparators, from the perspective of the National Health Service (NHS) and Personal Social Services (PSS). A half cycle correction has been applied. Markov models have previously been used in NICE appraisals for AD, including TA82.(163)

The modelled health states are defined by disease severity using ISGA criteria (**Figure 35**):

- Moderate flare – ISGA 3
- Mild flare – ISGA 2
- Controlled disease – ISGA 0–1
- Death

Figure 35: Model schematic



People may enter the death state at any point in the model
Abbreviations: ISGA, Investigator's Static Global Assessment

The ISGA health state definitions were considered more appropriate than the treatment states utilised in TA82 (163), which encompassed groups of patients with different levels of disease severity. The health states differ slightly from the Ellis et al (164) modelled health states as they use ISGA in place of IGA. It is noted that clinically a clear/almost clear score on the IGA and ISGA were considered clinically equivalent and hence, these outcomes were considered comparable and have been pooled in the NMA. The model does not capture severe or very severe disease (ISGA4/5) as crisaborole is not licensed in this population.

All patients enter the model in either the moderate or mild state and may either persist in this state or improve. Patients who improve are at risk of a disease flare, returning them to their original health state. This allows the model to capture the relapsing nature of AD, where patients have disease-controlled periods punctuated by acute flares.

The following treatments are assumed to be used as standard of care for acute flares, based on current marketing authorisation and NICE recommendations.

Mild AD

In mild disease, for both children and adults, emollients are used as standard care management and mild TCS are used during acute flares. It is noted that whilst TCIs are not currently recommended by NICE for mild AD patients, data from the British Association of Dermatologists National Clinical Audit of atopic dermatitis in children, indicated that 8.26% mild AD patients currently receive TCIs in UK clinical practice.(2) It is noted that pimecrolimus is the only licensed TCI in adults and children with mild AD patients. This suggests that pimecrolimus may be a clinically relevant comparator for mild AD that has not been adequately controlled by topical corticosteroids or where there is a serious risk of important adverse effects from further topical corticosteroid use, particularly irreversible skin

atrophy. hence, this treatment has been considered as a comparator for mild disease, second-line post-TCS scenario analyses.

Moderate Disease:

In moderate disease, standard care emollients are also used as standard care management, whilst patients with acute flares will receive initial treatment with moderate TCS. Patient whose disease is not controlled using TCS may be eligible for TCIs.

Children aged 2 to 16 with moderate AD that has not been controlled by topical corticosteroids, where there is a serious risk of important adverse effects from further topical corticosteroid use, particularly irreversible skin atrophy may receive tacrolimus 0.03% twice daily for up to 3 weeks, then once daily until the lesions have cleared.(109) Alternatively, children with AD of the face and neck may receive pimecrolimus 1% twice daily until clearance of lesions.(108, 162)

Patients aged 16 years and older with moderate AD that has not been controlled by topical corticosteroids, where there is a serious risk of important adverse effects from further topical corticosteroid use, particularly irreversible skin atrophy may receive treatment with tacrolimus 0.1% twice daily until clearance of lesions.(162) Tacrolimus 0.03% is also licensed for use in adults, however the SPC states that patients should start treatment with tacrolimus 0.1%.(109) Pimecrolimus is not recommended by NICE for use in adults.(162)

It is noted whilst clinical guidelines state that mild TCS should be used in mild disease and moderate TCS should be used in moderate disease, in clinical practice patients may be receive a more potent TCS prior to receiving TCIs. BAD national audit data indicates that approximately 24% of patients receive a moderate potency TCS and 1% receive a potent TCS. In order to capture this in the model a weighted average of mild, moderate and potent TCS is assumed for TCS eligible patients (Section B.3.3.1).

Treatment Application:

TCS may be used once or twice daily for a maximum of two weeks. Tacrolimus and pimecrolimus can be used twice daily for up to 6 weeks, but patients should stop applying treatment when lesions have cleared, or after 2 weeks if there are no signs of improvement. Children receiving tacrolimus 0.03% should reduce use to once daily after 3 weeks.

In the economic analysis, responders and partial responders to TCIs or crisaborole are assumed to receive 4 weeks of treatment (56 applications in total), whilst non-responders discontinue after 2 weeks. The total number of applications per flare are reported in **Table 52**. In a sensitivity analysis, patients treated with TCIs are modelled to receive 6 weeks of therapy (84 applications in total) whilst crisaborole treated patients receive only 4 weeks of therapy (56 applications in total). These assumptions tested due to differences in maximum usage for TCIs (i.e. that tacrolimus and pimecrolimus can be used twice daily for up to a maximum 6 weeks) and clinical efficacy data considered in the NMA (which captures outcomes for up to 6 weeks in TCIs). Crisaborole may be used for up to 4 weeks per treatment cycle and efficacy estimates from crisaborole trials is based on day 29 data.

Table 52: Treatment applications per treatment cycle

	TCS	Tacrolimus 0.03%	Tacrolimus 0.1%	Pimecrolimus	Crisaborole
Responders	28[a]	Children: 49[b] Adults:56[c]	56[c]	56[c]	56[c]
Partial responders	28[a]	Children: 49[b] Adults:56[c]	56[c]	56[c]	56[c]
Non-responders	28[a]	28[a]	28[a]	28[a]	28[a]

a. Twice daily for 14 days
b. Twice daily for 21 days then once daily for 7 days
c. Twice daily for 28 days
Abbreviations: TCS, topical corticosteroids.

Patients with moderate disease that respond to tacrolimus may continue to use it twice weekly as prophylactic therapy to reduce the incidence of flares.(109) Prophylactic use of tacrolimus has been considered in a separate sensitivity analysis.

[redacted]

In the economic model if a patient fails to respond to treatment, or discontinues due to an adverse event, they enter a composite state with costs and outcomes based on a weighted average costs and outcome associated with a combination of subsequent therapies (Section B.3.3.3). It is assumed that patients who have been treated with TCIs due to a serious risk of important adverse effects from further topical corticosteroid use, particularly irreversible skin atrophy, will not be eligible for further TCS treatment post TCIs or crisaborole (and other therapies are considered in lieu of TCS). Data from the British Association of Dermatologists National Clinical Audit of atopic dermatitis in children has been used to inform subsequent therapy use. (2)

Table 53: Features of the economic analysis

	Previous appraisals	Current appraisal	
Factor	**TA82**	**Chosen values**	**Justification**
Time horizon	• 1 year for adults • Up to age 18 for children	• Lifetime for adults • Up to age 18 for children	A lifetime time horizon is in line with the NICE reference case
Treatment waning effect?	No	No	
Source of utilities	• Disease controlled/remission states were informed by assumptions • Active disease states in children were informed by the Novartis submission • Active disease states in adults were informed by a utility panel	Mapped CDLQI/DLQI to EQ-5D data from the crisaborole clinical trials. Mean EQ-5D estimates derived by ISGA state	• Utilities in TA82 come from a range of sources, not all of which are patient-reported. Using mapped data from the clinical trials was deemed to be more in line with the reference case. • Values from TA82 are used in scenario analysis.
Source of costs	• Drug costs: BNF • Medical resource use: PSSRU with resource use estimates from Su et al	• Drug costs: BNF & eMIT • Medical resource use: PSSRU and NHS reference costs with resource use estimates from Pfizer epidemiology report	Resource use estimates have been updated with data from the Pfizer UK epidemiology study. These are deemed more appropriate as they are more recent and UK specific.

Abbreviations: BNF, British National Formulary; CDLQI, Children's Dermatitis Life Quality Index; DLQI, Dermatitis Life Quality Index; eMIT, electronic market information tool; PSSRU, Personal Social Services Research Unit.

B.3.2.3 Intervention technology and comparators

The intervention technology is crisaborole 2%, which is anticipated to receive a marketing authorisation for the treatment of mild to moderate AD in adults and children from 2 years of age.

[redacted]. The cost-effectiveness analysis consequently compares crisaborole with other topical therapies used to treat flares. **Table 54** summarises the modelled populations and their relevant comparators. A scenario analysis considers the prophylactic use of tacrolimus. The model assumes that patients use emollients at all times.

Table 54: Summary of modelled populations and the relevant comparators

	Children	Adults
Mild disease – First line (TCS eligible)	TCS	TCS
Mild disease Second line (not adequately controlled by TCS)	Pimecrolimus	Pimecrolimus
Moderate disease – First line (TCS eligible)	TCS	TCS
Moderate disease Second line (not adequately controlled by TCS)	Tacrolimus 0.03% Pimecrolimus (for patients with AD of the face and neck)	Tacrolimus 0.03% Tacrolimus 0.1%

B.3.3 Clinical parameters and variables

B.3.3.1 Response to treatment

Patients enter the model in either the mild or moderate disease state and begin treatment. Treatment efficacy is measured by the proportion of patients achieving an ISGA score of 0 or 1. The probability of responding to treatment is taken from the network meta-analysis (NMA). The base-case analysis uses the results of the analysis assessing ISGA 0–1 response rates up to week six.

A natural history model (NHM) (165) is used to predict vehicle response rates based on age and disease severity. The NHM provides an event rate λ for vehicle, allowing the response rate at time t to be calculated as:

$$p = 1 - \exp(-\lambda t),$$

where

$$\lambda = \exp\left(\alpha + \beta * (age - \overline{age}) + \gamma * \left(\%moderate - \overline{\%moderate}\right)\right).$$

Here $\overline{age}$ is the mean age of patients included in the meta-analysis for the natural history model and $\overline{\%moderate}$ is the average proportion of patients with moderate disease.

The NMA provided a hazard ratio for response for each comparator versus vehicle. The hazard ratios are applied to the NHM provides response rates for crisaborole, pimecrolimus, tacrolimus 0.1% and 0.03% and vehicle in the four main subgroups. The response rates are calculated at 4 weeks.

The NHM can be used to predict age- and severity-dependent response probabilities, however this is based on relatively few data points with estimates of the effects of age and severity having very wide credible intervals. The estimated response rates generated by the NHM become implausible as the target age and severity move away from the mean values observed in the included trials.

The base-case analysis uses estimated response probabilities from the NHM with the age and severity parameters set to their average population values. The effect of varying response rates with age and disease severity is then tested in scenario analysis, by adjusting response rates using odds ratios calculated from the vehicle arm of the AD-301 and AD-302 trials.

AD-301 and AD-302 included patients who were TCS naïve and TCS experienced, and additional subgroup analyses for these populations have been performed and used to inform two scenario analyses exploring the difference in response rates by adjusting the baseline response rates in the NHM.

Table 55 presents the odds ratios used in these scenarios. Odds ratios are calculated compared to the vehicle population as a whole.

Table 55: Odds ratios used to adjust response probabilities

Parameter	Subgroup	Odds ratio for vehicle response vs whole population	Source
Disease severity	Mild disease	[redacted]	Pfizer data on file. AD-301 & 302 ad hoc analysis. Table 11.1.1.3 (166)
	Moderate disease		
Age group	Children		**Appendix E Table E1**
	Adults		
Prior treatment	Treatment naïve		Pfizer data on file. AD-301 & 302 ad hoc analysis. Table 44.2.2.5 (167)
	Treatment exposed		

It was not possible to incorporate mild or moderate TCS into the NMA due to the lack of evidence identified in the systematic review. In order to facilitate a comparison of crisaborole with TCS, the transition probabilities from TA82 were applied.(165) When comparing to TCS, it is assumed that patients may be treated with mild, moderate and high potency TCS. The proportion of patients using mild, moderate and high potency TCS has been estimated using a weighted average based on the proportion of patients receiving each form of TCS in the BAD audit data collected in for NG52. These weights have been calculated separately for mild and moderate disease, see **Table 57**.

Table 56: Proportion of patients receiving mild, moderate and potent TCS in mild and moderate disease

	Mild TCS	Moderate TCS	Potent TCS
% of patients - Mild disease	66%	33%	2%
% of patients - Moderate disease	9%	77%	14%

Response rates used in the model are summarised in **Table 57 and Table 58** below.

Two sets of scenario analyses are performed exploring the effect of using alternative response data:

1. Response rates for crisaborole and TCIs are generated from the NMAs adjusted using the odds ratios from the vehicle arms of AD-301 & -302 (**Section B.2.8.5**)

2. Response rates for crisaborole and TCIs are generated from the unadjusted MAIC (**Section B.2.8.10**)

Table 57: Response rates applied in the base-case analyses and using subgroup adjustments

Therapy	Base-case (Credible interval)	Adjusted for severity - mild	Adjusted for severity - moderate	Adjusted for age - children	Adjusted for age - adults	Adjusted for prior therapy – treatment naive	Adjusted for prior therapy – treatment exposed
Mild TCS							
Moderate TCS							
Tacrolimus 0.03%							
Tacrolimus 0.1%							
Pimecrolimus							
Crisaborole							

Abbreviations: TCS, Topical corticosteroids.

Table 58: Response rates using the MAIC analyses

	Vs tacrolimus 0.03%	Vs tacrolimus 0.1%	Vs pimecrolimus
Comparator			
Crisaborole			

B.3.3.1.1 Partial response rates

It is possible that people with moderate disease will have a partial response to treatment after 28 days and will have moved to having mild disease (ISGA 2). Partial response rates are not well reported in the literature and no evidence synthesis has been performed comparing partial response rates for crisaborole with those for TCIs. Partial response rates are available from AD-301 and AD-302, where [redacted] of people with moderate disease had a partial response to crisaborole and [redacted] had a partial response to vehicle.(166) This represents [redacted] of non-responders for crisaborole and vehicle respectively.

In the model base-case, it was conservatively assumed that [redacted] of non-responders to crisaborole, tacrolimus and pimecrolimus achieve a partial response and receive a further cycle of treatment.

B.3.3.1.2 Second-line treatment response rates

Response rates for second-line treatments have been taken from TA82.(163)

Table 59: Response rates for second-line treatments

	Response rate - Mild disease	Response rate - Moderate disease
Tacrolimus 0.03%	0.4	0.4
Tacrolimus 0.1%	0.4	0.4
Pimecrolimus	0.2	0.2
Systemic treatment	0.7	0.7
UV therapy	0.7	0.7

Footnotes: AD, atopic dermatitis; UV, ultraviolet

B.3.3.2 Flare rates

Mild to moderate AD is characterised by periods of low disease activity, punctuated by acute flares where more intensive treatment is required. In the model, patients in the disease-controlled state are assumed to be at risk of a flare and returning to the mild or moderate disease state, dependent on their baseline severity. The Pfizer epidemiology (**Appendix M**) report provides flare rates for patients by age and disease severity [redacted] The survey assessed symptoms experienced, visits to healthcare professionals, flare rates, treatments used, concerns about therapies and out of pocket expenses. **Table 60** presents the average number of flares (assumed to be annual rates), and four-weekly probabilities derived from these rates.(61)

Table 60: Number and probability of flares in mild and moderate disease

Age range	Mild disease		Moderate disease	
	Annual number of flares	4-weekly probability	Annual number of flares	4-weekly probability
2-11	[redacted]	[redacted]	[redacted]	[redacted]
12-17	[redacted]	[redacted]	[redacted]	[redacted]
18+	[redacted]	[redacted]	[redacted]	[redacted]

In a scenario analysis, it is assumed that tacrolimus is used as a maintenance treatment to reduce the rate of flares. In order to inform the reduction in flares, a rate ratio for tacrolimus compared to vehicle has been derived from a long-term study of tacrolimus vs vehicle (168). This study compares 3-times-weekly tacrolimus to vehicle and reports the number of patients with 0, 1, 2 or the or more flares at 40 weeks (**Table 61**). Assuming all patients with more than 3 disease flares had 4 flares, the average number of flares at 40 weeks was 0.854 in the tacrolimus arm and 1.183 in the vehicle arm, indicating a rate ratio of 0.72. Pimecrolimus is not licenced as a prophylactic treatment and has not been considered for prophylactic use in the model.

Table 61: Number of disease flares by week 40 (168)

Flares	Tacrolimus	Vehicle
0	37.9%	33.8%
1	44.4%	38.0%
2	12.1%	11.3%
3	5.6%	9.9%
>3	0.0%	7.0%

This study assessed 3-times-weekly tacrolimus, however the SPC states that tacrolimus should be used twice-weekly as a maintenance therapy. Given the lack of evidence for the reduction in flares rates with twice weekly tacrolimus, this rate ratio is assumed to be valid for twice-weekly tacrolimus and the model assumes 50% of patients receive twice-weekly tacrolimus and 50% receive three-times-weekly tacrolimus.

B.3.3.3 Treatment withdrawal and subsequent therapies

Patients who have not responded to therapy after four weeks (both TCIs and crisaborole) are assumed to withdraw from treatment and start a new therapy. Thus, any patient not achieving ISGA0-1 after 4 weeks is assumed to discontinue treatment. This is consistent with previous models.(163)

Patients who do not respond adequately to treatment are assumed to start a new therapy. BAD UK national audit data presents a breakdown of treatment regimen by disease severity in paediatric patients and this data has been used to inform the proportion of patients receiving each subsequent therapy. It has been assumed that patients that discontinue will have their treatment escalated. **Table 62** presents the modelled probabilities of starting each therapy in mild disease.

Table 62: Probability of starting different treatments having failed the primary therapy in mild disease

	Probability - Children	Probability - Adults
Tacrolimus 0.03%	0.44	0.00
Tacrolimus 0.1%	0.00	0.44
Pimecrolimus	0.44	0.44
Systemic therapy	0.13	0.13
Phototherapy	0.00	0.00

Abbreviations: TCI, topical calcineurin inhibitor; TCS, topical corticosteroids.

Table 63 presents the modelled probabilities of starting each therapy having discontinued first-line treatment (moderate TCS or crisaborole) in moderate disease.

Table 63: Probability of starting different treatments having failed the primary therapy in moderate disease

	Probability - Children	Probability - Adults
Tacrolimus 0.03%	0.41	0.00
Tacrolimus 0.1%	0.00	0.41
Pimecrolimus	0.41	0.41
Systemic therapy	0.13	0.13
Phototherapy	0.05	0.05

Abbreviations: TCI, topical calcineurin inhibitor; TCS, topical corticosteroids.

According to the BAD audit data, phototherapy is not used in mild disease. Thus all patients that have failed on TCS and TCIs move to systemic treatment. This applies to both adults and children.

Table 64 presents the modelled probabilities of starting each therapy having discontinued second-line treatment (TCIs or crisaborole) in moderate disease.

Table 64: Probability of starting different treatments having failed the second-line therapy in moderate disease

	Probability - Children	Probability - Adults
Systemic therapy	71.48%	71.48%
Phototherapy	28.52%	28.52%

Abbreviations: TCI, topical calcineurin inhibitor; TCS, topical corticosteroids.

B.3.3.4 Resolution of atopic dermatitis

It is thought around 75% of children outgrow their AD in childhood or early adolescence.(169) In order to capture this in the model it has been assumed that 75% of patients outgrow their AD by age 16. This gives a 4-week probability of non-recurrence of 0.708%.

B.3.3.5 Mortality

Age-specific mortality is taken from lifetables for England and Wales (170). Atopic dermatitis is not expected impact mortality and no standardised mortality ratio is applied.

B.3.4 Measurement and valuation of health effects

B.3.4.1 Health-related quality-of-life data from clinical trials

The AD-301, -302 and -303 trials collected HRQoL data using the DLQI, CDLQI and DFI questionnaires.(12, 41)

- In AD-301 and -302 these were collected both at baseline and Day 29, or when patients discontinued
- In AD-303 these were collected at the start and end of each on-treatment period, and at the end of the study

The questionnaires were completed by:

- Subjects aged 2–15 at the baseline study visit (CDLQI)
- All subjects aged 16 or older at baseline (DLQI)
- Parents/guardians of patients aged 2–17 at baseline (DFI)

These measures do not provide utility values suitable for use in an economic evaluation and are not in line with the reference case.

B.3.4.2 Mapping

A single algorithm for mapping the DLQI to the EQ-5D-3L was identified.(128) An ordinal regression model was used to predict the probability of responding 1, 2 or 3 on each item of the EQ-5D questionnaire based on a patient's age, sex and DLQI responses. Once a patient's responses were predicted the UK EQ-5D-3L tariff was applied to obtain utility values.

Table 65: Results of the ordinal logistic regression algorithm for mapping the DLQI to the EQ-5D-3L

	Mobility	Self-care	Usual activities	Pain/discomfort	Anxiety/depression
Threshold a_1	4.500 (0.190)	4.854 (0.251)	3.574 (0.171)	2.204 (0.133)	1.469 (0.128)
Threshold a_2	9.506 (0.368)	9.074 (0.438)	7.231 (0.237)	6.052 (0.178)	4.775 (0.162)
Age	0.051 (0.003)	0.033 (0.004)	0.027 (0.003)	0.025 (0.002)	0.003 (0.002)
Sex[a]	0.046 (0.089)	−0.213 (0.120)	0.133 (0.087)	0.177 (0.073)	0.465 (0.073)
DLQI 1	0.087 (0.055)	0.176 (0.074)	0.270 (0.052)	0.685 (0.047)	0.035 (0.044)
DLQI 2	0.013 (0.061)	0.052 (0.079)	−0.114 (0.059)	0.014 (0.049)	0.378 (0.048)
DLQI 3	0.209 (0.068)	0.278 (0.085)	0.351 (0.063)	0.199 (0.060)	0.107 (0.057)
DLQI 4	0.071 (0.058)	0.053 (0.072)	0.051 (0.055)	0.097 (0.050)	−0.099 (0.048)
DLQI 5	0.113 (0.075)	0.064 (0.095)	0.209 (0.070)	−0.122 (0.064)	0.205 (0.062)
DLQI 6	0.116 (0.060)	0.014 (0.071)	0.215 (0.055)	0.310 (0.054)	−0.075 (0.052)
DLQI 7	0.251 (0.053)	0.236 (0.063)	0.283 (0.049)	−0.048 (0.046)	0.186 (0.044)
DLQI 8	−0.008 (0.076)	−0.013 (0.091)	−0.081 (0.071)	0.163 (0.066)	0.121 (0.064)
DLQI 9	−0.094 (0.065)	0.002 (0.075)	0.068 (0.060)	0.132 (0.057)	0.194 (0.054)
DLQI 10	0.233 (0.061)	0.478 (0.071)	0.210 (0.057)	0.245 (0.054)	0.155 (0.052)

The 10 DLQI questions are represented in order by DLQI 1, DLQI 2, etc

[a] Sex was coded male = 0, female = 1

The study used split-half cross-validation whereby the dataset was split into five separate estimation and validation sets. The estimation set was used to derive the mapping models, whilst the out-of-sample validation set was utilised for validating the fitted models. This process was repeated with each of the five estimation/ validation sets after which the sets were reversed, resulting in a total of 10 complete models. The mean overall difference of the difference between predicted mean health utility estimates and observed mean health utility estimates was -0.0120. This is an underestimate of 1.59% which the authors state represents a clinically unimportant effect.

The algorithm has been specifically [redacted]

B.3.4.3 Health-related quality-of-life studies

An SLR was conducted to identify relevant HRQoL evidence for use in the cost-effectiveness analysis. This search was conducted in March 2019. A detailed description of the search, its methods, and results is provided in **Appendix H.**

B.3.4.4 Adverse reactions

No adverse events have been incorporated into the economic analysis. The key safety outcome of interest is application site pain; however, this is a very short-term outcome and the associated disutility is difficult to accurately assess for incorporation into the economic model. Other disease related adverse events, such as pruritus, will be correlated with ISGA and consequently effect of treatment on such events is assumed to be implicitly captured within the modelled health states. Other adverse events associated with TCIs are longer-

term outcomes such as lymphomas, however there is not enough long-term data on crisaborole for these to be incorporated into the economic model.

B.3.4.5 Health-related quality-of-life data used in the cost-effectiveness analysis

HRQoL data in the cost-effectiveness model were generated by mapping the DLQI and CDLQI to the EQ-5D-3L using the Ali et al algorithm (128) described in **Section B.3.4.2**. Descriptive statistics exploring the average EQ-5D score have been produced for each age and disease severity subgroup. These scores have also been used to generate utility multipliers relative to ISGA 0-1.

In the base-case, the multipliers generated using ISGA score for adults has been applied in both the adult and child populations, as the mapping algorithm has been developed using an adult population and the DLQI. Patients in the ISGA 0-1 health state are assumed to revert to population utility norms for their age group in the UK.(172) Analysis using children's scores in the child populations and using mean EQ-5D scores directly in patient with ISGA 2 or 3 are applied in a scenario analysis, as are the values from TA82. The utility scores from TA82, the average scores from the crisaborole clinical trials, and utility multipliers are compared in **Table 66**. Population norms are presented in **Table 67**.

Table 66: Comparison of utility values used in the model

Health state/disutility	TA82 utility values (163)	TA82 – Multiplier vs ISGA 0-1	Mean EQ-5D index Value AD 301 and AD 302	Multiplier vs ISGA 0-1 based on EQ-5D index Value AD 301 and AD 302	Mean EQ-5D index Value AD 301 and AD 302 – combined adult and child scores	Multiplier vs ISGA 0-1 based on EQ-5D index Value AD 301 and AD 302– combined adult and child scores
Mild AD (ISGA 2) - Adults	0.985	0.995	[redacted]	[redacted]	[redacted]	[redacted]
Moderate AD (ISGA 3) - Adults	0.875	0.884	[redacted]	[redacted]	[redacted]	[redacted]
Controlled disease (ISGA 0-1) - Adults	0.99	1	[redacted]	[redacted]	[redacted]	[redacted]
Mild AD (ISGA 2) - Children	0.8625	0.880	[redacted]	[redacted]	[redacted]	[redacted]
Moderate AD (ISGA 3) - Children	0.69	0.704	[redacted]	[redacted]	[redacted]	[redacted]
Controlled disease (ISGA 0-1) - Children	0.98	1	[redacted]	[redacted]	[redacted]	[redacted]
Non-recurrence	1	1	1	1	1	1

(Children)						

Abbreviations: AD, atopic dermatitis; ISGA, Investigator's Static Global Assessment.

Table 67: EQ-5D index value population norms by age group (172)

Age group	Utility
Under 18	1.00
18-24	0.94
25-34	0.93
35-44	0.91
45-54	0.85
55-64	0.80
65-74	0.78
75+	0.73

Note: The value for the under 18 population is based on an assumption of perfect health.

B.3.5 Cost and healthcare resource use identification, measurement and valuation

An SLR was conducted to identify relevant cost and healthcare resource use evidence for use in the cost-effectiveness analysis. This search was conducted in March 2019. A detailed description of the search, its methods, and results is provided in **Appendix I.**

B.3.5.1 Intervention and comparators' costs and resource use

Intervention and comparator costs are assumed to include drug costs. Administration and monitoring costs are not included.

B.3.5.1.1 Drug dosing

The amount of drug for topical therapies varies by age and it is assumed that all topical therapies are used at the same rate per application. The amount of drug used per application is taken from the long-term safety follow-up study AD-303 and is presented in **Table 68** below.(41)

Table 68: Drug use per application by age group

Age range	Drug use per application (g)
2-11 year	2.40
12-17 years	2.29
Adults	2.10

Crisaborole, tacrolimus 0.1% and pimecrolimus are applied twice daily to manage flares. Tacrolimus 0.03% is applied twice daily for 3 weeks, then once daily thereafter. TCS may be applied once or twice a day. In the base-case it assumed that crisaborole and TCIs applied for 4 weeks by responders and partial responders. Non-responders are assumed to cease therapy after 2 weeks. TCS are used for 2 weeks in all patients. **Table 71** summarises the number of applications per flare in the base-case analysis.

Patients using tacrolimus as maintenance therapy in the sensitivity analysis have been assumed to use the same volume of drug per application. It is assumed that 50% of patients apply maintenance therapy twice-weekly and 50% apply it three-times-weekly.

Patients progressing to systemic treatment are assumed to receive 1.25 mg/kg of ciclosporin twice daily for 4 weeks.

B.3.5.1.2 Drug acquisition costs

Drug acquisition costs are taken from the British National Formulary (BNF) (173) and for eMIT.(174)

Table 69: Treatment costs and pack sizes

Drug	Pack size (g/capsules)	Cost per pack	Source
Crisaborole	60	[redacted]	
Tacrolimus 0.03%	60	£42.55	BNF
Tacrolimus 0.1%	60	£37.82	BNF
Pimecrolimus	100	£59.07	BNF
Hydrocortisone 1% cream	30	£0.66	eMIT
Betamethasone valerate 0.025% ointment	100	£2.99	BNF
Betamethasone valerate 0.1% cream	100	£3.24	BNF
Diprobase cream	500	£5.32	BNF
Ciclosporin 50 mg	30	£21.80	BNF
Phototherapy	N/A	£93	NHS reference costs – Currency code JC47Z Phototherapy or photochemotherapy – Dermatology (175)

Abbreviations: BNF, British National Formulary.

B.3.5.2 Health-state unit costs and resource use

People with mild to moderate AD are assumed to be managed in primary care. Data from the Pfizer epidemiology study presented the average number of visits to healthcare professionals by disease severity.(61) People with mild and moderate disease had on [redacted], respectively. These visits are assumed to be to GPs.

It is assumed that adults and children post-TCS who additionally fail second-line crisaborole or TCI treatment in the model will require increased visits to healthcare professionals. It has been assumed that 50% of patients that progress to subsequent therapy will see a dermatologist. Additionally, the number of annual GP visits increases to [redacted] for children and [redacted] for adults reflective of data form a retrospective cross-sectional analysis of the THIN (The Health Improvement Network) database, designed to describe demographic and clinical characteristics of AD patients and to estimate their healthcare resource use.

The increase in GP and dermatology visits is to reflect escalation of treatment in difficult to treat mild and moderate AD, that would not be initiated in primary care and would require additional monitoring.

Table 70: Unit costs and resource use in mild to moderate disease

Resource	Annual use in mild disease	Annual use in moderate disease	Cost per unit	Source
GP visit	2	3	£37	PSSRU 2018 (176)
Dermatologist visit - adults	On treatment failure and in 50% of flares thereafter	On treatment failure and in 50% of flares thereafter	£111	NHS reference costs – Service code 330 (175)
Dermatologist visit - children	On treatment failure and in 50% of flares thereafter	On treatment failure and in 50% of flares thereafter	£148	NHS reference costs – Service code 257 (175)

Abbreviations: GP, general practitioner; NHS, National Health Service; PSSRU, Personal Social Services Research Unit.

B.3.5.3 Adverse reaction unit costs and resource use

No costs associated with adverse events have been incorporated into the model.

B.3.5.4 Miscellaneous unit costs and resource use

No other costs are considered.

B.3.6 Summary of base-case analysis inputs and assumptions

B.3.6.1 Summary of base-case analysis inputs

Table 71: Summary of variables applied in the economic model

Variable	Value (reference to appropriate table or figure in submission)	Measurement of uncertainty and distribution: CI (distribution)	Reference to section in submission
Discount rate (costs)	0.035	Not varied	
Discount rate (outcomes)	0.035	Not varied	
Baseline age (children)	2	Not varied	
Baseline age (adults)	18	Not varied	
% female	55.7%	Not varied	
Cost of crisaborole	[redacted]	Not varied	
Response probability for vehicle	[redacted]	[redacted]	Section B.3.3.1
Response probability tacrolimus 0.03%	[redacted]	[redacted]	Section B.3.3.1
Response probability tacrolimus 0.1%	[redacted]	[redacted]	Section B.3.3.1
Response probability pimecrolimus 1%	[redacted]	[redacted]	Section B.3.3.1
Response probability crisaborole	[redacted]	[redacted]	Section B.3.3.1
MAIC response crisaborole vs tac 0.03%	[redacted]	[redacted]	Section B.3.3.1
MAIC response crisaborole vs tac 0.1%	[redacted]	[redacted]	Section B.3.3.1
MAIC response crisaborole vs pimecrolimus	[redacted]	[redacted]	Section B.3.3.1
MAIC response tac 0.03%	[redacted]	[redacted]	Section B.3.3.1
MAIC response tac 0.1%	[redacted]	[redacted]	Section B.3.3.1
MAIC response pimecrolimus	[redacted]	[redacted]	Section B.3.3.1
Response rate among non-responders - Crisaborole	[redacted]	[redacted]	Section B.3.3.1
Response rate: Mild TCS - mild disease	0.520	0.39 - 0.65 (Beta)	Section B.3.3.1
Response rate: Mild TCS - moderate disease	0.520	0.39 - 0.65 (Beta)	Section B.3.3.1
Response rate: Moderate TCS - mild disease	0.600	0.45 - 0.75 (Beta)	Section B.3.3.1
Response rate: Moderate TCS - moderate disease	0.600	0.45 - 0.75 (Beta)	Section B.3.3.1
Response rate: Potent TCS - mild disease	0.700	0.525 - 0.875 (Beta)	Section B.3.3.1
Response rate: Potent TCS - moderate disease	0.700	0.525 - 0.875 (Beta)	Section B.3.3.1
% of patients - Mild disease Mild TCS	[redacted]	[redacted]	Section B.3.3.1
% of patients - Mild disease Moderate TCS	[redacted]	[redacted]	Section B.3.3.1
% of patients - Mild disease	[redacted]	[redacted]	Section B.3.3.1

Potent TCS		[redacted]	
% of patients - Moderate disease Mild TCS	[redacted]	[redacted]	Section B.3.3.1
% of patients - Moderate disease Moderate TCS	[redacted]	[redacted]	Section B.3.3.1
% of patients - Moderate disease Potent TCS	[redacted]	[redacted]	Section B.3.3.1
Partial response rate: Mild TCS - moderate disease	[redacted]	[redacted]	Section B.3.3.1
Partial response rate: Moderate TCS - moderate disease	[redacted]	[redacted]	Section B.3.3.1
Mild disease flares per year: Age 2 to 11	[redacted]	[redacted]	Section B.3.3.2
Mild disease flares per year: Age 12 to 17	[redacted]	[redacted]	Section B.3.3.2
Mild disease flares per year: Age Adults	[redacted]	[redacted]	Section B.3.3.2
Moderate disease flares per year: Age 2 to 11	[redacted]	[redacted]	Section B.3.3.2
Moderate disease flares per year: Age 12 to 17	[redacted]	[redacted]	Section B.3.3.2
Moderate disease flares per year: Age Adults	[redacted]	[redacted]	Section B.3.3.2
Mild disease Probability - Children Tacrolimus 0.03%	43.6%	0.436 - 0.436 (Dirichlet)	Section B.3.3.3
Mild disease Probability - Children Tacrolimus 0.1%	0.0%	0 - 0 (Dirichlet)	Section B.3.3.3
Mild disease Probability - Children Pimecrolimus	43.6%	0.436 - 0.436 (Dirichlet)	Section B.3.3.3
Mild disease Probability - Children Systemic therapy	12.9%	0.129 - 0.129 (Dirichlet)	Section B.3.3.3
Mild disease Probability - Children Phototherapy	0.0%	0 - 0 (Dirichlet)	Section B.3.3.3
Mild disease Probability - Adults Tacrolimus 0.03%	0.0%	0 - 0 (Dirichlet)	Section B.3.3.3
Mild disease Probability - Adults Tacrolimus 0.1%	43.6%	0.436 - 0.436 (Dirichlet)	Section B.3.3.3
Mild disease Probability - Adults Pimecrolimus	43.6%	0.436 - 0.436 (Dirichlet)	Section B.3.3.3
Mild disease Probability - Adults Systemic therapy	12.9%	0.129 - 0.129 (Dirichlet)	Section B.3.3.3
Mild disease Probability - Adults Phototherapy	0.0%	0 - 0 (Dirichlet)	Section B.3.3.3
Moderate disease - post moderate TCS/crisaborole Probability - Children Tacrolimus 0.03%	41.0%	0.41 - 0.41 (Dirichlet)	Section B.3.3.3
Moderate disease - post moderate TCS/crisaborole Probability - Children Tacrolimus 0.1%	0.0%	0 - 0 (Dirichlet)	Section B.3.3.3
Moderate disease - post moderate TCS/crisaborole Probability - Children Pimecrolimus	41.0%	0.41 - 0.41 (Dirichlet)	Section B.3.3.3
Moderate disease - post moderate TCS/crisaborole	12.8%	0.128 - 0.128 (Dirichlet)	Section B.3.3.3

Probability - Children Systemic therapy			
Moderate disease - post moderate TCS/crisaborole Probability - Children Phototherapy	5.1%	0 - 0 (Dirichlet)	Section B.3.3.3
Moderate disease - post moderate TCS/crisaborole Probability - Adults Tacrolimus 0.03%	0.0%	0 - 0 (Dirichlet)	Section B.3.3.3
Moderate disease - post moderate TCS/crisaborole Probability - Adults Tacrolimus 0.1%	41.0%	0.41 - 0.41 (Dirichlet)	Section B.3.3.3
Moderate disease - post moderate TCS/crisaborole Probability - Adults Pimecrolimus	41.0%	0.41 - 0.41 (Dirichlet)	Section B.3.3.3
Moderate disease - post moderate TCS/crisaborole Probability - Adults Systemic therapy	12.8%	0.128 - 0.128 (Dirichlet)	Section B.3.3.3
Moderate disease - post moderate TCS/crisaborole Probability - Adults Phototherapy	5.1%	0 - 0 (Dirichlet)	Section B.3.3.3
Moderate disease - post TCI/crisaborole Probability - Children Tacrolimus 0.03%	0.0%	0 - 0 (Dirichlet)	Section B.3.3.3
Moderate disease - post TCI/crisaborole Probability - Children Tacrolimus 0.1%	0.0%	0 - 0 (Dirichlet)	Section B.3.3.3
Moderate disease - post TCI/crisaborole Probability - Children Pimecrolimus	0.0%	0 - 0 (Dirichlet)	Section B.3.3.3
Moderate disease - post TCI/crisaborole Probability - Children Systemic therapy	71.5%	0.467 - 0.912 (Dirichlet)	Section B.3.3.3
Moderate disease - post TCI/crisaborole Probability - Children Phototherapy	28.5%	0 - 0 (Dirichlet)	Section B.3.3.3
Moderate disease - post TCI/crisaborole Probability - Adults Tacrolimus 0.03%	0.0%	0 - 0 (Dirichlet)	Section B.3.3.3
Moderate disease - post TCI/crisaborole Probability - Adults Tacrolimus 0.1%	0.0%	0 - 0 (Dirichlet)	Section B.3.3.3
Moderate disease - post TCI/crisaborole Probability - Adults Pimecrolimus	0.0%	0 - 0 (Dirichlet)	Section B.3.3.3
Moderate disease - post TCI/crisaborole Probability - Adults Systemic therapy	71.5%	0.467 - 0.912 (Dirichlet)	Section B.3.3.3
Moderate disease - post TCI/crisaborole Probability - Adults Phototherapy	28.5%	0.088 - 0.533 (Dirichlet)	Section B.3.3.3
Odds ratio for response vs all	[redacted]	[redacted]	Section B.3.3.1

patients - No prior treatment			
Odds ratio for response vs all patients - Prior treatment	[redacted]	[redacted]	Section B.3.3.1
Odds ratio for response vs all patients - Mild disease	[redacted]	[redacted]	Section B.3.3.1
Odds ratio for response vs all patients - Moderate disease	[redacted]	[redacted]	Section B.3.3.1
Odds ratio for response vs all patients - Child	[redacted]	[redacted]	Section B.3.3.1
Odds ratio for response vs all patients - Adult	[redacted]	[redacted]	Section B.3.3.1
% patients who outgrow AD	0.75	Not varied	Section B.3.3.4
Mortality multiplier	1	Not varied	Section B.3.3.5
Utility AD-301 & -302 Children - non-recurrence	1.000	Not varied	Section B.3.4.5
Utility AD-301 & -302 Adults - mild disease	[redacted]	[redacted]	Section B.3.4.5
Utility AD-301 & -302 Adults - moderate disease	[redacted]	[redacted]	Section B.3.4.5
Utility AD-301 & -302 Adults - controlled disease	[redacted]	[redacted]	Section B.3.4.5
% returning to population norm	100%	Not varied	Section B.3.4.5
Population utility norms Under 18	1	Not varied	Section B.3.4.5
Population utility norms 18-24	0.94	Not varied	Section B.3.4.5
Population utility norms 25-34	0.927	Not varied	Section B.3.4.5
Population utility norms 35-44	0.911	Not varied	Section B.3.4.5
Population utility norms 45-54	0.847	Not varied	Section B.3.4.5
Population utility norms 55-64	0.799	Not varied	Section B.3.4.5
Population utility norms 65-74	0.779	Not varied	Section B.3.4.5
Population utility norms 75+	0.726	Not varied	Section B.3.4.5
Drug use per application – age 2-11 years	2.40	2.17-2.63 (Normal)	Section B.3.5.1
Drug use per application – age 12-17 years	2.29	1.97-2.61 (Normal)	Section B.3.5.1
Drug use per application – adults	2.10	1.44-2.76 (Normal)	Section B.3.5.1
Hydrocortisone 1% Grams per tube (g)	30	Not varied	Section B.3.5.1
Hydrocortisone 1% Cost per tube (£)	0.66	Not varied	Section B.3.5.1
Hydrocortisone 1% Applications per flare - responders	28	Not varied	Section B.3.5.1
Hydrocortisone 1% Applications per flare - responders	28	Not varied	Section B.3.5.1
Hydrocortisone 1% Applications per flare - responders	28	Not varied	Section B.3.5.1
Betamethasone valerate 0.025% ointment Grams per tube (g)	100	Not varied	Section B.3.5.1
Betamethasone valerate 0.025% ointment Cost per tube (£)	2.99	Not varied	Section B.3.5.1
Betamethasone valerate	28	Not varied	Section B.3.5.1

0.025% ointment Applications per flare - responders			
Betamethasone valerate 0.025% ointment Applications per flare - partial responders	28	Not varied	Section B.3.5.1
Betamethasone valerate 0.025% ointment Applications per flare - non responders	28	Not varied	Section B.3.5.1
Tacrolimus 0.03%Grams per tube (g)	60	Not varied	Section B.3.5.1
Tacrolimus 0.03%Cost per tube (£)	42.55	Not varied	Section B.3.5.1
Tacrolimus 0.03%Applications per flare - responders	56	Not varied	Section B.3.5.1
Tacrolimus 0.03%Applications per flare - partial responders	56	Not varied	Section B.3.5.1
Tacrolimus 0.03%Applications per flare - non responders	28	Not varied	Section B.3.5.1
Tacrolimus 0.1%Grams per tube (g)	60	Not varied	Section B.3.5.1
Tacrolimus 0.1%Cost per tube (£)	37.82	Not varied	Section B.3.5.1
Tacrolimus 0.1%Applications per flare - responders	56	Not varied	Section B.3.5.1
Tacrolimus 0.1%Applications per flare - partial responders	56	Not varied	Section B.3.5.1
Tacrolimus 0.1%Applications per flare - non responders	28	Not varied	Section B.3.5.1
Pimecrolimus Grams per tube (g)	100	Not varied	Section B.3.5.1
Pimecrolimus Cost per tube (£)	59.07	Not varied	Section B.3.5.1
Pimecrolimus Applications per flare - responders	56	Not varied	Section B.3.5.1
Pimecrolimus Applications per flare - partial responders	56	Not varied	Section B.3.5.1
Pimecrolimus Applications per flare - non responders	28	Not varied	Section B.3.5.1
Crisaborole Grams per tube (£)	60	Not varied	Section B.3.5.1
Crisaborole Applications per flare - responders	56	Not varied	Section B.3.5.1
Crisaborole Applications per flare - partial responders	56	Not varied	Section B.3.5.1
Crisaborole Applications per flare - non responders	28	Not varied	Section B.3.5.1
Betamethasone valerate 0.15Grams per tube (g)	100	Not varied	Section B.3.5.1
Betamethasone valerate 0.15Cost per tube (£)	3.24	Not varied	Section B.3.5.1
Betamethasone valerate 0.1% Applications per flare - responders	28	Not varied	Section B.3.5.1
Betamethasone valerate 0.1% Applications per flare -partial responders	28	Not varied	Section B.3.5.1
Betamethasone valerate 0.1% Applications per flare - non	28	Not varied	Section B.3.5.1

responders			
Diprobase Units per pack (g)	500	Not varied	Section B.3.5.1
Diprobase Cost per pack (£)	6.32	Not varied	Section B.3.5.1
Diprobase Units per flare - responders	500	Not varied	Section B.3.5.1
Ciclosporin 25 mg Units per pack (g)	1500	Not varied	Section B.3.5.1
Ciclosporin 25 mg Cost per pack (£)	21.8	Not varied	Section B.3.5.1
Ciclosporin 25 mg Units per flare - responders	70	Not varied	Section B.3.5.1
Phototherapy Cost per pack (£)	93	Not varied	Section B.3.5.1
Phototherapy Units per flare - responders	1	Not varied	Section B.3.5.1
Diprobase maintenance use	250	Not varied	Section B.3.5.2
Annual GP visits Mild	2	Not varied	Section B.3.5.2
Annual GP visits Moderate	3	Not varied	Section B.3.5.2
Cost per unit - GP visit	37	Not varied	Section B.3.5.2
Cost per unit Dermatologist visit - adult	111	Not varied	Section B.3.5.2
Cost per unit Dermatologist visit - children	148	Not varied	Section B.3.5.2
Subsequent therapy Response rate - mild disease Tacrolimus 0.03%	0.374	0.281 - 0.468 (Beta)	Section B.3.3.3
Subsequent therapy Response rate - mild disease Tacrolimus 0.1%	0.385	0.289 - 0.481 (Beta)	Section B.3.3.3
Subsequent therapy Response rate - mild disease Pimecrolimus	0.249	0.187 - 0.311 (Beta)	Section B.3.3.3
Subsequent therapy Response rate - mild disease Systemic treatment	0.7	0.525 - 0.875 (Beta)	Section B.3.3.3
Subsequent therapy Response rate - mild disease UV therapy	0.7	0.525 - 0.875 (Beta)	Section B.3.3.3
Subsequent therapy Response rate - Moderate disease Tacrolimus 0.03%	0.374	0.281 - 0.468 (Beta)	Section B.3.3.3
Subsequent therapy Response rate - Moderate disease Tacrolimus 0.1%	0.385	0.289 - 0.481 (Beta)	Section B.3.3.3
Subsequent therapy Response rate - Moderate disease Pimecrolimus	0.249	0.187 - 0.311 (Beta)	Section B.3.3.3
Subsequent therapy Response rate - Moderate disease Systemic treatment	0.7	0.525 - 0.875 (Beta)	Section B.3.3.3
Subsequent therapy Response rate - Moderate disease UV therapy	0.7	0.525 - 0.875 (Beta)	Section B.3.3.3
Proportion of flares incurring a dermatology visit in subsequent therapy Mild disease	50%	37.5%-42.5% (Beta)	Section B.3.5.2

Proportion of flares incurring a dermatology visit in subsequent therapy Moderate disease	50%	37.5%-42.5% (Beta)	Section B.3.5.2
GP visits while on subsequent therapy - children	[redacted]	[redacted]	Section B.3.5.2
GP visits while on subsequent therapy - adults	[redacted]	[redacted]	Section B.3.5.2

B.3.6.2 Assumptions

Table 72: Model assumptions and justification

Model input	Assumption	Justification
Assessment point for response	Response rates are assessed at 4 weeks.	This assumption has been made to simplify the model structure. The response rates included in the model base-case for each comparator reflect the assessment time point in the clinical trials used to inform the NMA.
Partial response	The proportion of non-responders with a partial response is constant across treatments.	This assumption is made due to a lack of data on partial response rates from the clinical trials.
Flare rates	Flare rates are independent of treatment.	While some patients will receive maintenance therapy to reduce the rate of disease flares, there is little comparative evidence on efficacy. Both tacrolimus and pimecrolimus have been shown to reduce flare rates when compared to vehicle, but there is not head-to-head data comparing TCIs to TCS. Additionally, flare rates in model have been taken from the Pfizer epidemiology report. The population included patients being treated with TCIs and represents an overall flare rate for the population, including those using TCIs. This assumption is tested in scenario analysis.
HRQoL mapping	EQ-5D values generated by mapping the DLQI to the EQ-5D can be used in the model for children.	No algorithm for mapping the CDLQI to the EQ-5D was identified. The DLQI and CDLQI are very similar in construct and the author has indicated that the same methods are applicable. A scenario using EQ-5D data mapped from the CDLQI is included.
Emollient use	Patients use emollients at all times and patients experiencing a flare use twice as much emollient as those with controlled disease.	Emollients form the basis of management of AD and should always be used even when the AD is clear.
Drug use	The amount of drug used per application is constant across all topical treatments.	All treatments are topical and little data is available comparing application rates. There were no significant differences in application rates for crisaborole and vehicle in the clinical trial and the same assumption was made in TA82.

Subsequent therapies	Patients that discontinue treatment within the model will receive a more intensive subsequent treatment.	The primary reason for patients withdrawing from treatment in the model is lack of response. It is not clinically plausible that an AD patient that is uncontrolled by TCS or TCIs will be managed using emollient only or lower potency TCS given the multiple treatment options currently available (i.e. phototherapy, systemic therapies) for patients unresponsive to first and second line of therapy.
Monitoring of subsequent therapies	Patients progressing to subsequent therapy receive increased dermatologist and GP visits.	Higher potency treatments such as systemic therapies are subject to a high risk of adverse events and require additional monitoring. Additionally, off-label therapies are unlikely to be initiated by a GP.

Abbreviations: CDLQI, Children's' Dermatitis Life Quality Index; DLQI, Dermatitis Life Quality Index; NMA, network meta-analysis; TCI, topical calcineurin inhibitor; TCS, topical corticosteroids.

B.3.7 Base-case results

B.3.7.1 Base-case incremental cost-effectiveness analysis results

In adults and children aged 2 years and older with mild to moderate AD that has not been controlled by topical corticosteroids or where there is a serious risk of important adverse effects from further topical corticosteroid use, particularly irreversible skin atrophy (second-line use post-TCS treatment) crisaborole is the dominant treatment option in all populations.

Table 73: Children with mild AD post-TCS: Base-case results

Technologies	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER versus baseline (£/QALY)	ICER incremental (£/QALY)
Crisaborole	[redacted]	[redacted]	[redacted]	[redacted]	-	Dominant
Pimecrolimus	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominated

Abbreviations: ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years

Table 74: Adults with mild AD post-TCS: Base-case results

Technologies	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER versus baseline (£/QALY)	ICER incremental (£/QALY)
Crisaborole	[redacted]	[redacted]	[redacted]	[redacted]	**-**	**Dominant**
Pimecrolimus	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominated

Abbreviations: ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years

Table 75: Children with moderate AD post-TCS: Base-case results

Technologies	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER versus baseline (£/QALY)	ICER incremental (£/QALY)
Crisaborole	[redacted]	[redacted]	[redacted]	[redacted]	**-**	**Dominant**
Tacrolimus 0.03%	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominated
Pimecrolimus	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominated

Abbreviations: ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years

Table 76: Adults with moderate AD post-TCS: Base-case results

Technologies	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER versus baseline (£/QALY)	ICER incremental (£/QALY)
Crisaborole	[redacted]	[redacted]	[redacted]	[redacted]	**-**	**Dominant**
Tacrolimus 0.1%	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominated
Tacrolimus 0.03%	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominated

Abbreviations: ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years

B.3.8 Sensitivity analyses

Sensitivity analysis results are presented here for the post-TCS populations. The remaining analyses are presented in **Appendix N.**

B.3.8.1 Probabilistic sensitivity analysis

Joint parameter uncertainty was explored through probabilistic sensitivity analysis (PSA), in which all parameters are assigned distributions and varied jointly. 1000 Monte Carlo simulations were recorded. Where parameters have been taken from an NMA they have been varied using the CODA output. Results were plotted on the cost-effectiveness plane (CEP) and cost-effectiveness acceptability curves (CEAC) were generated.

B.3.8.1.1 Children with mild AD post-TCS

Table 77 presents the mean costs and QALYs from the PSA, alongside the incremental costs, QALYS and ICERs. **Figure 36** and **Figure 37** present the cost-effectiveness plane and cost-effectiveness acceptability curve (CEAC) respectively. The results of the PSA in children with post-TCS mild disease are congruent with the deterministic ICERs. There are marginal differences in costs and QALYs between treatments, nevertheless, crisaborole remains dominant for the PSA mean estimates and is cost-effective in 79% of simulations at a willingness-to-pay threshold of £20,000.

Table 77: Children with mild AD post-TCS - PSA results

Technologies	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER versus baseline (£/QALY)	ICER incremental (£/QALY)
Crisaborole	[redacted]	[redacted]	[redacted]	[redacted]	-	-
Pimecrolimus	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominated

Figure 36 Children with mild AD post-TCS - cost-effectiveness plane

Figure 37 Children with mild AD post-TCS - CEAC

B.3.8.1.2 Adults with mild AD post-TCS

Table 78 presents the mean costs and QALYs from the PSA, alongside the incremental costs, QALYS and ICERs. **Figure 38** and **Figure 39** present the cost-effectiveness plane and cost-effectiveness acceptability curve (CEAC) respectively. The results of the PSA in adults with post-TCS mild disease are congruent with the deterministic ICERs. There are

marginal differences in costs and QALYs between treatments, nevertheless, crisaborole remains dominant for the PSA mean estimates and is cost-effective in 94% of simulations at a willingness-to-pay threshold of £20,000.

Table 78: Adults with mild AD post-TCS - PSA results

Technologies	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER versus baseline (£/QALY)	ICER incremental (£/QALY)
Crisaborole	[redacted]	[redacted]	[redacted]	[redacted]	-	-
Pimecrolimus	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominated

Figure 38 Adults with mild AD post-TCS - cost-effectiveness plane

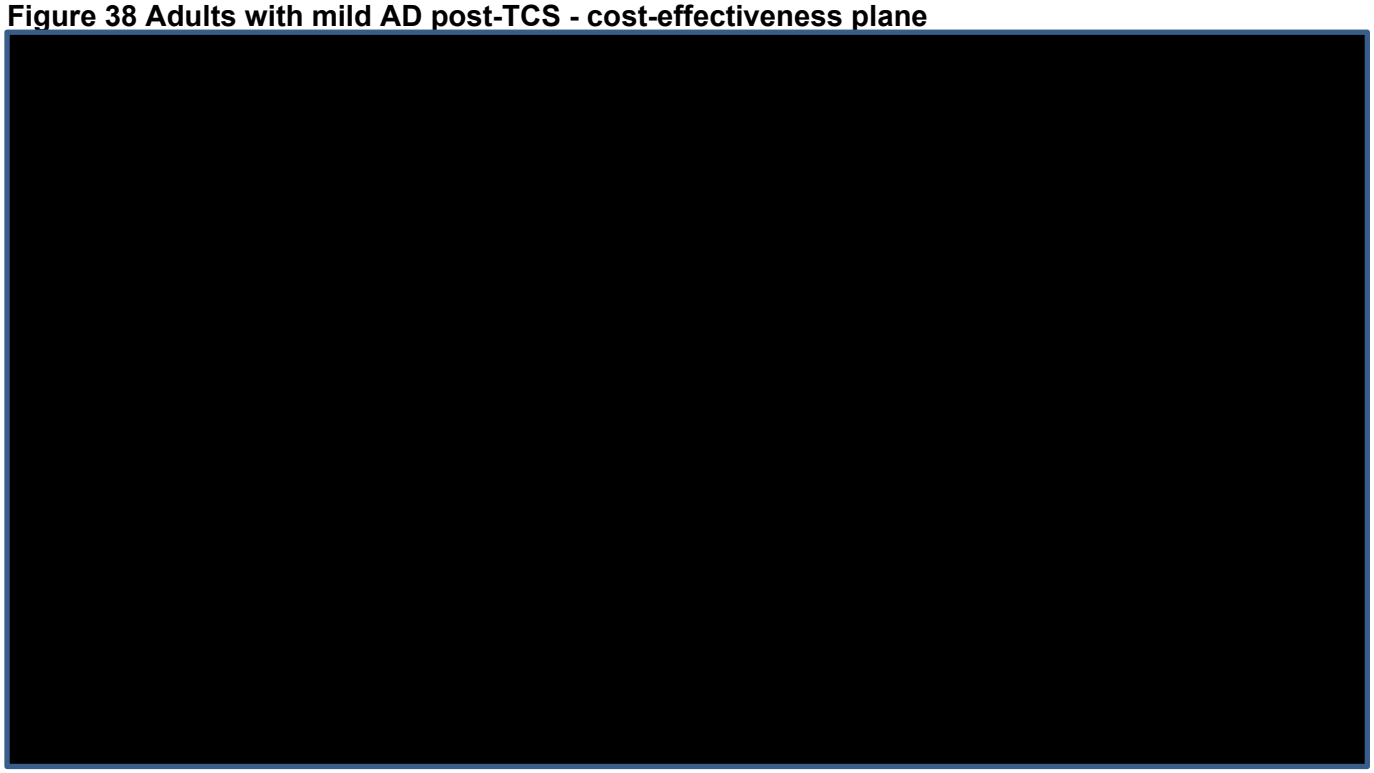


Figure 39 Adults with mild AD post-TCS - CEAC

B.3.8.1.3 Children with moderate AD post-TCS

Table 79 presents the mean costs and QALYs from the PSA, alongside the incremental costs, QALYS and ICERs. **Figure 40** and **Figure 41** present the cost-effectiveness plane and cost-effectiveness acceptability curve (CEAC) respectively. The results of the PSA in children with post-TCS moderate disease are congruent with the deterministic ICERs. There are marginal differences in costs and QALYs between treatments, nevertheless, crisaborole remains dominant for the PSA mean estimates and is cost-effective in 76% of simulations at a willingness-to-pay threshold of £20,000.

Table 79: Children with moderate AD post-TCS - PSA results

Technologies	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER versus baseline (£/QALY)	ICER incremental (£/QALY)
Crisaborole	[redacted]	[redacted]	[redacted]	[redacted]	-	-
Tacrolimus 0.03%	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominated
Pimecrolimus	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominated

Figure 40 Children with moderate AD post-TCS - cost-effectiveness plane

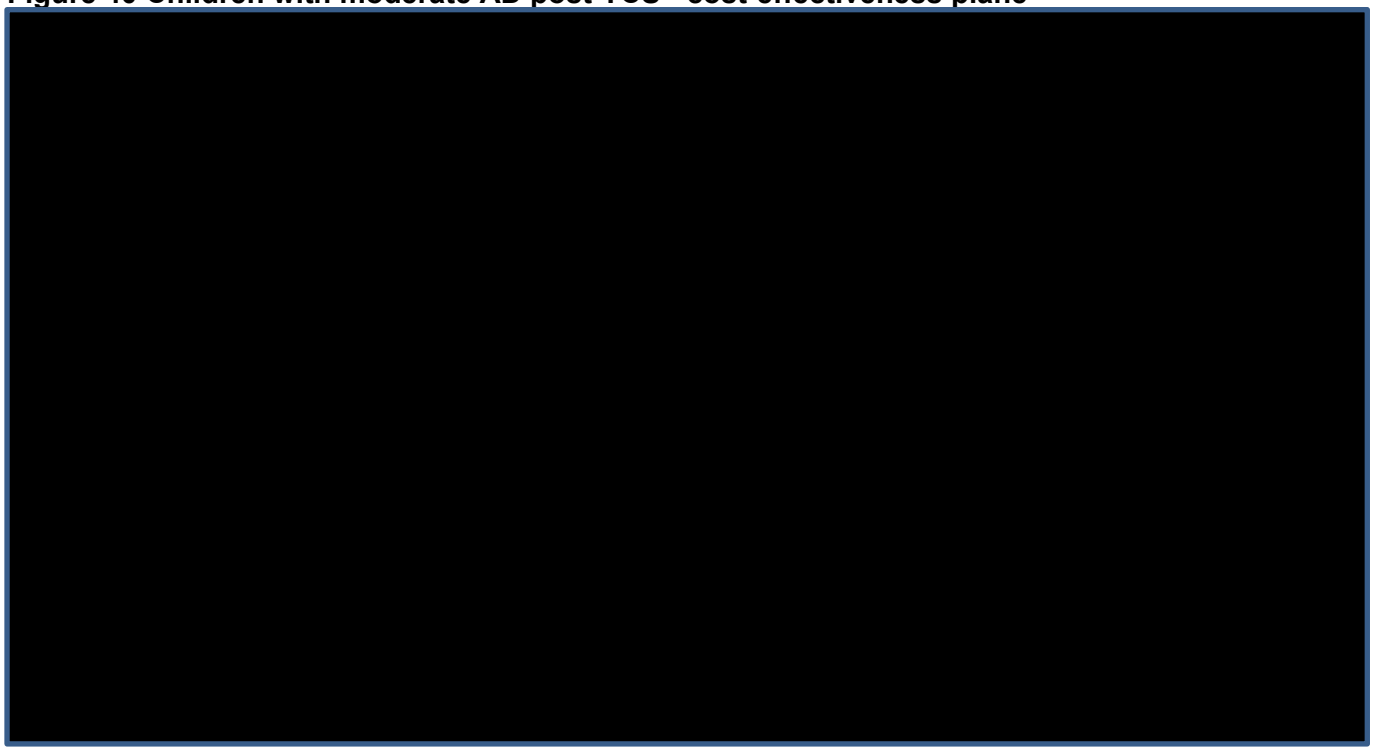


Figure 41 Children with moderate AD post-TCS - CEAC

B.3.8.1.4 Adults with moderate AD post-TCS

Table 80 presents the mean costs and QALYs from the PSA, alongside the incremental costs, QALYS and ICERs. Figure 42 and Figure 43 present the cost-effectiveness plane and cost-effectiveness acceptability curve (CEAC) respectively. The results of the PSA in adults with post-TCS moderate disease show there are marginal differences in costs and QALYs

between treatments, hence, the PSA in this instance is generating a higher ICER than the deterministic analysis, due to the wide credible intervals for crisaborole versus tacrolimus and due to Jensen's inequality (economic model is a nonlinear function).This leads to a higher ICER for crisaborole vs tacrolimus 0.1%. Nevertheless, crisaborole is cost-effective in 54% of simulations at a willingness-to-pay threshold of £20,000 and cost saving in 51% of scenarios.

Table 80: Adults with moderate AD post-TCS - PSA results

Technologies	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER versus baseline (£/QALY)	ICER incremental (£/QALY)
Tacrolimus 0.1%	[redacted]	[redacted]	[redacted]	[redacted]	-	-
Crisaborole	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Tacrolimus 0.03%	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominated

Figure 42 Adults with moderate AD post-TCS - cost-effectiveness plane

Figure 43 Adults with moderate AD post-TCS - CEAC

B.3.8.2 Deterministic sensitivity analysis

Deterministic sensitivity analysis is presented for the post-TCS populations. This sensitivity analysis has been run on the net monetary benefit (NMB) at a willingness to pay threshold of £20,000. In populations with more than one comparison each comparator has been considered separately.

B.3.8.2.1 Children with mild AD post-TCS

Figure 44 present the results of the one-way sensitivity analysis (OWSA) for children with mild AD vs pimecrolimus. The results of the OWSA are sensitive to the absolute response probabilities for crisaborole and the comparators, reflective of the relatively wide credible intervals in the NMA. The results also show some sensitivity to the assumptions about subsequent therapy and the utility values associated with mild and controlled disease. Nevertheless, overall the NMB remains positive when these are varied and indicate that crisaborole remains cost-effective at a WTP threshold of £20,000, indicating results are robust to plausible variations in assumptions.

Figure 44 Children with mild AD post-TCS - tornado diagram vs pimecrolimus

B.3.8.2.2 Adults with mild AD post-TCS

Figure 45 presents the results of the one-way sensitivity analysis (OWSA) for adults with mild AD vs pimecrolimus. The results of the OWSA for adults also show sensitivity to changes in the response probabilities for crisaborole or comparators and similar to results in children with mild disease, results show some sensitivity to the assumptions around subsequent therapy and the utility values for mild and disease-controlled states. The results are also sensitive in adults to the assumed drug usage since, there were a smaller number of adults in the clinical trial, the confidence interval around drug use is larger in adults. Overall, however, the NMB remains positive in most scenarios indicating that crisaborole remains cost-effective at a WTP threshold of £20,000 and results are quite robust to plausible variations in assumptions.

Figure 45 Adults with mild AD post-TCS - tornado diagram vs pimecrolimus

B.3.8.2.3 Children with moderate AD post-TCS

Figure 46 presents the results of the one-way sensitivity analysis (OWSA) for children with moderate AD vs tacrolimus 0.03%. **Figure 47** presents the same data for the comparison with pimecrolimus. The OWSA results are also sensitive to the assumptions around subsequent therapy and monitoring, and the utility values associated with moderate disease, however crisaborole remains cost-effective when these are varied. The results show sensitivity to changes in the response probabilities for crisaborole or comparators and similar to results in children, mild disease, results show some sensitivity to the assumptions around subsequent therapy and the utility values for mild and disease-controlled states. Overall, however, the NMB remains positive in most scenarios indicating that crisaborole remains cost-effective at a WTP threshold of £20,000 and results are robust to plausible variations in assumptions.

Figure 46 Children with moderate AD post-TCS - tornado diagram vs tacrolimus 0.03%

Figure 47 Children with moderate AD post-TCS - tornado diagram vs pimecrolimus

B.3.8.2.4 Adults with moderate AD post-TCS

Figure 48 presents the results of the OWSA for adults with moderate AD vs tacrolimus 0.03%. **Figure 49** presents the same data for the comparison with tacrolimus 0.1%. ICERs are sensitive to the treatment efficacy, utility values, assumptions about subsequent treatments, drug use assumptions and dermatology visits. Overall, however, the NMB

remains positive in most scenarios indicating that crisaborole remains cost-effective at a WTP threshold of £20,000 and results are robust to plausible variations in assumptions.

Figure 48 Adults with moderate AD post-TCS - tornado diagram vs tacrolimus 0.03%

Figure 49 Adults with moderate AD post-TCS - tornado diagram vs tacrolimus 0.1%

B.3.8.3 Scenario analysis

Table 81 present the results of the scenario analyses for children and adults with post-TCS mild and moderate AD respectively. Crisaborole remains dominant in the majority of scenarios. It is noted that the base case economic analysis presents ICERs using NMA evidence. An unanchored MAIC was performed due to a potential disconnect in the network resulting from heterogeneity in vehicle response. Conclusions from this analysis show that the NMA results are robust and reinforced by additional evidence from the MAIC and [redacted].

.

Table 81: Children with moderate AD post-TCS – Scenario analysis

Scenario name	Details	Children with mild AD post TCS – incremental ICER	Adults with mild AD post TCS – incremental ICER	Children with moderate AD post TCS – incremental ICER	Adults with moderate AD post TCS – incremental ICER
Base-case		Dominant	Dominant	Dominant	Dominant
MAIC - vs tacrolimus 0.03%	Compare to tacrolimus 0.03% using MAIC response probabilities	N/A	N/A	Dominant	Dominant
MAIC - vs pimecrolimus	Compare to pimecrolimus using MAIC response probabilities	Dominant	Dominant	Dominant	N/A
Adjust for line of therapy	Adjust response probabilities using the odds ratio for having received prior therapy	Dominant	Dominant	Dominant	[redacted]
Adjust for severity of disease	Adjust response probabilities using the odds ratio for moderate disease	Dominant	Dominant	Dominant	[redacted]
Adjust for age	Adjust response probabilities using the odds ratio for children	Dominant	Dominant	Dominant	Dominant
Adjust for line, severity and age	Apply adjustments for line of therapy, age and disease severity	Dominant	Dominant	Dominant	[redacted]
MAIC - vs tacrolimus 0.03% Adjust for line, severity and age	Compare to tacrolimus 0.03% using MAIC response probabilities and including all adjustments	N/A	N/A	Dominant	Dominant
MAIC - vs pimecrolimus Adjust for line, severity and age	Compare to pimecrolimus using MAIC response probabilities and including all adjustments	Dominant	Dominant	Dominant	N/A

Use tacrolimus 0.03% response rate	Assume equal efficacy between tacrolimus 0.03% and crisaborole	N/A	N/A	Dominated	Dominated
Use tacrolimus 0.1% response rate	Assume equal efficacy between tacrolimus 0.1% and crisaborole	N/A	N/A	N/A	Dominated
Use pimecrolimus response rate	Assume equal efficacy between pimecrolimus and crisaborole	Dominated	Dominated	Dominated	N/A
Increase response rates by 20%	Increase all response probabilities by 20%	Dominant	Dominant	Dominant	Dominant
Decrease response rates by 20%	Decrease all response probabilities by 20%	Dominant	Dominant	Dominant	[redacted]
Remove partial response	Set partial response rates to 0	N/A	N/A	Dominant	Dominant
Tacrolimus Prophlaxis use	Assume tacrolimus is used while in the controlled disease state and apply a rate ratio to flare rates	N/A	N/A	Cheapest therapy Pimecrolimus: Dominated Tacrolimus 0.03%: [redacted]*	Dominated
50% outgrow AD	Assume only 50% of patients outgrow AD by adulthood	Dominant	N/A	Dominant	N/A
Use child-specific utilities	Use utility values derived from children in AD-301 & -302	Dominant	N/A	Dominant	N/A
Garside utilities	Use TA82 utility values	Dominant	Dominant	Dominant	Dominant
6 weeks TCI	Assume responders to tacrolimus 0.03% and pimecrolimus are treated	Dominant	Dominant	Dominant	Dominant

	for 6 weeks				
Time horizon 16 years	Set the time horizon to 16 years old	Dominant	N/A	Dominant	N/A

Footnotes:

[redacted]

[redacted]

B.3.8.4 Summary of sensitivity analyses results

One-way sensitivity analyses show that ICERs show some sensitivity to assumptions around efficacy, drug use, utilities and dermatology visits. Overall, however, the NMB remains positive in most scenarios indicating that crisaborole remains cost-effective at a WTP threshold of £20,000 in all populations and results are robust to plausible variations in most assumptions.

Results of the PSA are congruent with the deterministic ICERs in three of the four populations. In children with mild disease post-TCS, adults with mild disease post-TCS and children with moderate disease post-TCS crisaborole is cost-effective in 79%, 94% and 76% of simulations at a willingness-to-pay threshold of £20,000 respectively. In adults with moderate disease crisaborole was cost-effective in 54% of simulations at a WTP threshold of £20,000.

B.3.9 Subgroup analysis

See base case results for analyses by disease severity and adult/child subgroups.

B.3.10 Validation

B.3.10.1 Validation of cost-effectiveness analysis

The cost-effectiveness model has been validated internally by the model developers and by health economists not involved in the construction of the model. The model was validated using standard procedures:

- Cell-by-cell checks of logic and consistency
- Logical check of model outputs
- Comparison of outputs to those from previous economic analyses

Results of model verification led to modifications to the electronic model; model outputs were subsequently compared to that of TA82 and were considered consistent.

B.3.11 Interpretation and conclusions of economic evidence

Crisaborole has been shown to be a highly cost-effective treatment option versus current standard care for adults and children aged 2 years with mild to moderate AD that has not been controlled by topical corticosteroids or where there is a serious risk of important adverse effects from further topical corticosteroid use.

Whilst ICERs show some sensitivity to assumptions around efficacy, drug use, utilities and dermatology visits, one-way sensitivity analyses indicate the results are robust to plausible variations in parameter estimates and key structural assumptions.

This analysis is relevant to all patient groups that may receive crisaborole and population-specific data from the crisaborole clinical trial programme is used where available. The model is designed to imitate the treatment pathway used in England and Wales. Data used in the model are generalisable to the UK. The main strengths of this analysis are:

- The effectiveness of crisaborole has been demonstrated in two large randomised clinical trials

- The model is flexible in that it considers adult and child subgroups by disease severity and hence covers the range of populations and treatment scenarios in which crisaborole may be used. The model consequently presents ICERs for crisaborole versus comparators in relevant patient subgroups according to their UK licensing and NICE reimbursement recommendations.

- The base case NMA was a comprehensive analysis that used advanced evidence synthesis methods to evaluate crisaborole versus existing standard care therapies. This NMA adjusted both for heterogeneity in the time-points for treatment outcomes assessment and class effects for TCIs. Furthermore, this model adjusted for heterogeneity in vehicle response using meta-regression. A substantial number of sensitivity analyses were explored.

- An unanchored MAIC was performed consistent with NICE guidance, due to a potential disconnect in the network resulting from heterogeneity in vehicle response. Whilst the base case economic analysis presents ICERs using NMA evidence, conclusions from this analysis are robust and reinforced by additional evidence from the MAIC.

- HRQoL weights (utility data) in the cost-effectiveness model were generated by mapping DLQI and CDLQI data collected in AD 301 and AD 302 to the EQ-5D-3L using a validated for mapping DLQI to EQ-5D-3L published algorithm by Ali et al (96). Mean EQ-5D utility scores have been produced for each age and disease severity subgroup in order to appropriately inform the economic model.

The main limitations of this analysis are:

- Limitations of the underlying clinical evidence. NMA analyses stratified by age and/or severity (e.g. children with mild AD, children with moderate AD, adults with mild AD, adults with moderate AD) were sparsely populated or networks were disconnected. These analyses consequently are not considered sufficiently reliable to inform the cost-effectiveness analysis. In the base case analysis response estimates for all populations have been based on the overall NMA (all populations combined) with estimates from the MAIC available in sensitivity analyses. A comparison to TCSs was also not possible due to trial population differences and the lack of subgroup data reporting with which to compare to the mild to moderate population treated with crisaborole.

B.4 References

1. National Institute for Health and Care Excellence. Crisaborole for treating mild to moderate atopic dermatitis in people aged 2 years and older. Final scope. 2019.
2. BAD-BSPD-NICE. Paediatric National Audit on Atopic Eczema in Children. March 2015.
3. Jimenez JL, Punzon C, Navarro J, Munoz-Fernandez MA, Fresno M. Phosphodiesterase 4 inhibitors prevent cytokine secretion by T lymphocytes by inhibiting nuclear factor-kappaB and nuclear factor of activated T cells activation. J Pharmacol Exp Ther. 2001;299(2):753-9.
4. Baumer W, Hoppmann J, Rundfeldt C, Kietzmann M. Highly selective phosphodiesterase 4 inhibitors for the treatment of allergic skin diseases and psoriasis. Inflamm Allergy Drug Targets. 2007;6(1):17-26.
5. Grewe SR, Chan SC, Hanifin JM. Elevated leukocyte cyclic AMP-phosphodiesterase in atopic disease: a possible mechanism for cyclic AMP-agonist hyporesponsiveness. J Allergy Clin Immunol. 1982;70(6):452-7.
6. Heskel NS, Chan SC, Thiel ML, Stevens SR, Casperson LS, Hanifin JM. Elevated umbilical cord blood leukocyte cyclic adenosine monophosphate-phosphodiesterase activity in children with atopic parents. J Am Acad Dermatol. 1984;11(3):422-6.
7. Butler JM, Chan SC, Stevens S, Hanifin JM. Increased leukocyte histamine release with elevated cyclic AMP-phosphodiesterase activity in atopic dermatitis. J Allergy Clin Immunol. 1983;71(5):490-7.
8. Hanifin JM, Chan SC, Cheng JB, Tofte SJ, Henderson WR, Jr., Kirby DS, et al. Type 4 phosphodiesterase inhibitors have clinical and in vitro anti-inflammatory effects in atopic dermatitis. J Invest Dermatol. 1996;107(1):51-6.
9. Jarnagin K, Chanda S, Coronado D, Ciaravino V, Zane LT, Guttman-Yassky E, et al. Crisaborole Topical Ointment, 2%: A Nonsteroidal, Topical, Anti-Inflammatory Phosphodiesterase 4 Inhibitor in Clinical Development for the Treatment of Atopic Dermatitis. J Drugs Dermatol. 2016;15(4):390-6.
10. Pfizer Inc. Eucrisa (Crisaborole 2%). Summary of product characteristics. Last updated 2019. Available from: https://www.pfizermedicalinformation.com/en-us/eucrisa/description
11. Jimenez JL, Iniguez MA, Munoz-Fernandez MA, Fresno M. Effect of phosphodiesterase 4 inhibitors on NFAT-dependent cyclooxygenase-2 expression in human T lymphocytes. Cell Signal. 2004;16(12):1363-73.
12. Paller AS, Tom WL, Lebwohl MG, Blumenthal RL, Boguniewicz M, Call RS, et al. Efficacy and safety of crisaborole ointment, a novel, nonsteroidal phosphodiesterase 4 (PDE4) inhibitor for the topical treatment of atopic dermatitis (AD) in children and adults. J Am Acad Dermatol. 2016;75(3):494-503 e6.
13. Akama T, Baker SJ, Zhang YK, Hernandez V, Zhou H, Sanders V, et al. Discovery and structure-activity study of a novel benzoxaborole anti-inflammatory agent (AN2728) for the potential topical treatment of psoriasis and atopic dermatitis. Bioorg Med Chem Lett. 2009;19(8):2129-32.
14. Freund YR, Akama T, Alley MR, Antunes J, Dong C, Jarnagin K, et al. Boron-based phosphodiesterase inhibitors show novel binding of boron to PDE4 bimetal center. FEBS Lett. 2012;586(19):3410-4.

15. Pfizer Inc. Eucrisa (Crisaborole 2%). Summary of product characteristics. Last updated 2019. Available from: https://www.pfizermedicalinformation.ca/en-ca/eucrisa#.
16. Australian Department of Health. Final scheduling decisions and reasons: NCEs and Appendix H: Australia Department of Health; 2019 [Available from: https://www.tga.gov.au/book-page/21-crisaborole.
17. Weidinger S, Novak N. Atopic dermatitis. Lancet. 2016;387(10023):1109-22.
18. Leung DY. New insights into atopic dermatitis: role of skin barrier and immune dysregulation. Allergol Int. 2013;62(2):151-61.
19. Barbarot S, Auziere S, Gadkari A, Girolomoni G, Puig L, Simpson EL, et al. Epidemiology of atopic dermatitis in adults: Results from an international survey. Allergy. 2018;73(6):1284-93.
20. Odhiambo JA, Williams HC, Clayton TO, Robertson CF, Asher MI, Group IPTS. Global variations in prevalence of eczema symptoms in children from ISAAC Phase Three. J Allergy Clin Immunol. 2009;124(6):1251-8 e23.
21. Finch J, Munhutu MN, Whitaker-Worth DL. Atopic dermatitis and nutrition. Clin Dermatol. 2010;28(6):605-14.
22. Arkwright PD, Motala C, Subramanian H, Spergel J, Schneider LC, Wollenberg A, et al. Management of difficult-to-treat atopic dermatitis. J Allergy Clin Immunol Pract. 2013;1(2):142-51.
23. Weidinger S, Beck LA, Bieber T, Kabashima K, Irvine AD. Atopic dermatitis. Nat Rev Dis Primers. 2018;4(1):1.
24. National Institute for Health and Care Excellence. Atopic eczema in under 12s: diagnosis and management. 2007.
25. Ben-Gashir MA, Seed PT, Hay RJ. Predictors of atopic dermatitis severity over time. J Am Acad Dermatol. 2004;50(3):349-56.
26. Emerson RM, Williams HC, Allen BR. Severity distribution of atopic dermatitis in the community and its relationship to secondary referral. Br J Dermatol. 1998;139(1):73-6.
27. Abuabara K, Magyari A, McCulloch CE, Linos E, Margolis DJ, Langan SM. Prevalence of Atopic Eczema Among Patients Seen in Primary Care: Data From The Health Improvement Network. Ann Intern Med. 2019;170(5):354-6.
28. Schneider L, Tilles S, Lio P, Boguniewicz M, Beck L, LeBovidge J, et al. Atopic dermatitis: a practice parameter update 2012. J Allergy Clin Immunol. 2013;131(2):295-9 e1-27.
29. Blume-Peytavi U, Metz M. Atopic dermatitis in children: management of pruritus. J Eur Acad Dermatol Venereol. 2012;26 Suppl 6:2-8.
30. Yosipovitch G, Papoiu AD. What causes itch in atopic dermatitis? Curr Allergy Asthma Rep. 2008;8(4):306-11.
31. Hengge UR, Ruzicka T, Schwartz RA, Cork MJ. Adverse effects of topical glucocorticosteroids. J Am Acad Dermatol. 2006;54(1):1-15; quiz 6-8.
32. Siegfried EC, Jaworski JC, Hebert AA. Topical calcineurin inhibitors and lymphoma risk: evidence update with implications for daily practice. Am J Clin Dermatol. 2013;14(3):163-78.
33. Carr WW. Topical calcineurin inhibitors for atopic dermatitis: review and treatment recommendations. Paediatr Drugs. 2013;15(4):303-10.

34. Chen SL, Yan J, Wang FS. Two topical calcineurin inhibitors for the treatment of atopic dermatitis in pediatric patients: a meta-analysis of randomized clinical trials. J Dermatolog Treat. 2010;21(3):144-56.
35. Bissonnette R, Pavel AB, Diaz A, Werth JL, Zang C, Vranic I, et al. Crisaborole and atopic dermatitis skin biomarkers: an intrapatient randomized trial. J Allergy Clin Immunol. 2019.
36. Simpson EL, Paller AS, Boguniewicz M, Eichenfield LF, Feldman SR, Silverberg JI, et al. Crisaborole Ointment Improves Quality of Life of Patients with Mild to Moderate Atopic Dermatitis and Their Families. Dermatol Ther (Heidelb). 2018;8(4):605-19.
37. Yosipovitch G, Gold LF, Lebwohl MG, Silverberg JI, Tallman AM, Zane LT. Early Relief of Pruritus in Atopic Dermatitis with Crisaborole Ointment, A Non-steroidal, Phosphodiesterase 4 Inhibitor. Acta Derm Venereol. 2018;98(5):484-9.
38. Clinical Study Report: AD-301. A multicenter, randomized, double-blind vehicle -controlled study of the safety and efficacy of AN2728 topical ointment, 2% in children, adolescents, and adults (ages 2 years and older) with atopic dermatitis. Palo Alto (CA); 2015 Nov 23.
39. Clinical Study Report: AD-302. A multicenter, randomized, double-blind vehicle -controlled study of the safety and efficacy of AN2728 topical ointment, 2% in children, adolescents, and adults (ages 2 years and older) with atopic dermatitis Palo Alto (CA); 2015 Nov 23.
40. Ciaravino V, Coronado D, Lanphear C, Chanda S. 2-Year animal carcinogenicity results for crisaborole, a novel phosphodiesterase 4 inhibitor for atopic dermatitis. J Dermatol Sci. 2017;87(2):116-22.
41. Eichenfield LF, Call RS, Forsha DW, Fowler J, Jr., Hebert AA, Spellman M, et al. Long-term safety of crisaborole ointment 2% in children and adults with mild to moderate atopic dermatitis. J Am Acad Dermatol. 2017;77(4):641-9 e5.
42. Abramovits W, Fleischer AB, Jr., Jaracz E, Breneman D. Adult patients with moderate atopic dermatitis: tacrolimus ointment versus pimecrolimus cream. J Drugs Dermatol. 2008;7(12):1153-8.
43. Chapman MS, Schachner LA, Breneman D, Boguniewicz M, Gold MH, Shull T, et al. Tacrolimus ointment 0.03% shows efficacy and safety in pediatric and adult patients with mild to moderate atopic dermatitis. J Am Acad Dermatol. 2005;53(2 Suppl 2):S177-85.
44. Kempers S, Boguniewicz M, Carter E, Jarratt M, Pariser D, Stewart D, et al. A randomized investigator-blinded study comparing pimecrolimus cream 1% with tacrolimus ointment 0.03% in the treatment of pediatric patients with moderate atopic dermatitis. J Am Acad Dermatol. 2004;51(4):515-25.
45. Paller AS, Lebwohl M, Fleischer AB, Jr., Antaya R, Langley RG, Kirsner RS, et al. Tacrolimus ointment is more effective than pimecrolimus cream with a similar safety profile in the treatment of atopic dermatitis: results from 3 randomized, comparative studies. J Am Acad Dermatol. 2005;52(5):810-22.
46. Schachner LA, Lamerson C, Sheehan MP, Boguniewicz M, Mosser J, Raimer S, et al. Tacrolimus ointment 0.03% is safe and effective for the treatment of mild to moderate atopic dermatitis in pediatric patients: results from a randomized, double-blind, vehicle-controlled study. Pediatrics. 2005;116(3):e334-42.
47. Eichenfield LF, Lucky AW, Boguniewicz M, Langley RG, Cherill R, Marshall K, et al. Safety and efficacy of pimecrolimus (ASM 981) cream 1% in the treatment of

mild and moderate atopic dermatitis in children and adolescents. J Am Acad Dermatol. 2002;46(4):495-504.
48. Asher MI, Montefort S, Bjorksten B, Lai CK, Strachan DP, Weiland SK, et al. Worldwide time trends in the prevalence of symptoms of asthma, allergic rhinoconjunctivitis, and eczema in childhood: ISAAC Phases One and Three repeat multicountry cross-sectional surveys. Lancet. 2006;368(9537):733-43.
49. Shaw TE, Currie GP, Koudelka CW, Simpson EL. Eczema prevalence in the United States: data from the 2003 National Survey of Children's Health. J Invest Dermatol. 2011;131(1):67-73.
50. Perkin MR, Strachan DP, Williams HC, Kennedy CT, Golding J, Team AS. Natural history of atopic dermatitis and its relationship to serum total immunoglobulin E in a population-based birth cohort study. Pediatr Allergy Immunol. 2004;15(3):221-9.
51. Lee HH, Patel KR, Singam V, Rastogi S, Silverberg JI. A systematic review and meta-analysis of the prevalence and phenotype of adult-onset atopic dermatitis. J Am Acad Dermatol. 2019;80(6):1526-32 e7.
52. Silverberg JI. Adult-Onset Atopic Dermatitis. J Allergy Clin Immunol Pract. 2019;7(1):28-33.
53. Wollenberg A, Barbarot S, Bieber T, Christen-Zaech S, Deleuran M, Fink-Wagner A, et al. Consensus-based European guidelines for treatment of atopic eczema (atopic dermatitis) in adults and children: part I. J Eur Acad Dermatol Venereol. 2018;32(5):657-82.
54. Wollenberg A, Barbarot S, Bieber T, Christen-Zaech S, Deleuran M, Fink-Wagner A, et al. Consensus-based European guidelines for treatment of atopic eczema (atopic dermatitis) in adults and children: part II. J Eur Acad Dermatol Venereol. 2018;32(6):850-78.
55. Vachiramon V, Tey HL, Thompson AE, Yosipovitch G. Atopic dermatitis in African American children: addressing unmet needs of a common disease. Pediatr Dermatol. 2012;29(4):395-402.
56. Noda S, Suarez-Farinas M, Ungar B, Kim SJ, de Guzman Strong C, Xu H, et al. The Asian atopic dermatitis phenotype combines features of atopic dermatitis and psoriasis with increased TH17 polarization. J Allergy Clin Immunol. 2015;136(5):1254-64.
57. Margolis JS, Abuabara K, Bilker W, Hoffstad O, Margolis DJ. Persistence of mild to moderate atopic dermatitis. JAMA Dermatol. 2014;150(6):593-600.
58. Tang TS, Bieber T, Williams HC. Are the concepts of induction of remission and treatment of subclinical inflammation in atopic dermatitis clinically useful? J Allergy Clin Immunol. 2014;133(6):1615-25 e1.
59. Langan SM, Schmitt J, Williams HC, Smith S, Thomas KS. How are eczema 'flares' defined? A systematic review and recommendation for future studies. Br J Dermatol. 2014;170(3):548-56.
60. Thomas KS, Stuart B, O'Leary CJ, Schmitt J, Paul C, Williams HC, et al. Validation of treatment escalation as a definition of atopic eczema flares. PLoS One. 2015;10(4):e0124770.
61. Pfizer Inc. Data on file. Atopic Dermatitis Prevalence - UK.
62. Lewis-Jones MS, Finlay AY. The Children's Dermatology Life Quality Index (CDLQI): initial validation and practical use. Br J Dermatol. 1995;132(6):942-9.

63. Finlay AY, Khan GK. Dermatology Life Quality Index (DLQI)--a simple practical measure for routine clinical use. Clin Exp Dermatol. 1994;19(3):210-6.
64. Poole CD, Chambers C, Sidhu MK, Currie CJ. Health-related utility among adults with atopic dermatitis treated with 0.1% tacrolimus ointment as maintenance therapy over the long term: findings from the Protopic CONTROL study. Br J Dermatol. 2009;161(6):1335-40.
65. Chamlin SL, Chren MM. Quality-of-life outcomes and measurement in childhood atopic dermatitis. Immunol Allergy Clin North Am. 2010;30(3):281-8.
66. Yaghmaie P, Koudelka CW, Simpson EL. Mental health comorbidity in patients with atopic dermatitis. J Allergy Clin Immunol. 2013;131(2):428-33.
67. Alanne S, Nermes M, Soderlund R, Laitinen K. Quality of life in infants with atopic dermatitis and healthy infants: a follow-up from birth to 24 months. Acta Paediatr. 2011;100(8):e65-70.
68. Simpson EL. Comorbidity in Atopic Dermatitis. Curr Dermatol Rep. 2012;1(1):29-38.
69. Dogan DG, Canaloglu SK, Kivilcim M, Kum YE, Topal E, Catal F. Sleep patterns of young children with newly diagnosed atopic dermatitis. Postepy Dermatol Alergol. 2017;34(2):143-7.
70. Riis JL, Vestergaard C, Deleuran MS, Olsen M. Childhood atopic dermatitis and risk of attention deficit/hyperactivity disorder: A cohort study. J Allergy Clin Immunol. 2016;138(2):608-10.
71. Paller AS LT, Herbert AA, Zaenglein AL, Silverberg JI, Tallman AM, Tan H, Ports WC, Zielinski MA. . Crisaborole Ointment Improves Global Atopic Dermatitis Severity Across Patients with Varying Baseline Characteristics: Pooled Results from Two Phase 3 Trials. International Symposium on Atopic Dermatitis; Utrecht, The Netherlands2018.
72. Suarez AL, Feramisco JD, Koo J, Steinhoff M. Psychoneuroimmunology of psychological stress and atopic dermatitis: pathophysiologic and therapeutic updates. Acta Derm Venereol. 2012;92(1):7-15.
73. Oh SH, Bae BG, Park CO, Noh JY, Park IH, Wu WH, et al. Association of stress with symptoms of atopic dermatitis. Acta Derm Venereol. 2010;90(6):582-8.
74. Lin TK, Zhong L, Santiago JL. Association between Stress and the HPA Axis in the Atopic Dermatitis. Int J Mol Sci. 2017;18(10).
75. Drucker AM, Wang AR, Li WQ, Sevetson E, Block JK, Qureshi AA. The Burden of Atopic Dermatitis: Summary of a Report for the National Eczema Association. J Invest Dermatol. 2017;137(1):26-30.
76. Chamlin SL, Frieden IJ, Williams ML, Chren MM. Effects of atopic dermatitis on young American children and their families. Pediatrics. 2004;114(3):607-11.
77. Zuberbier T, Orlow SJ, Paller AS, Taieb A, Allen R, Hernanz-Hermosa JM, et al. Patient perspectives on the management of atopic dermatitis. J Allergy Clin Immunol. 2006;118(1):226-32.
78. Jang HJ, Hwang S, Ahn Y, Lim DH, Sohn M, Kim JH. Family quality of life among families of children with atopic dermatitis. Asia Pac Allergy. 2016;6(4):213-9.
79. Moore K, David TJ, Murray CS, Child F, Arkwright PD. Effect of childhood eczema and asthma on parental sleep and well-being: a prospective comparative study. Br J Dermatol. 2006;154(3):514-8.

80. Filanovsky MG, Pootongkam S, Tamburro JE, Smith MC, Ganocy SJ, Nedorost ST. The Financial and Emotional Impact of Atopic Dermatitis on Children and Their Families. J Pediatr. 2016;169:284-90 e5.
81. Pauli-Pott U, Darui A, Beckmann D. Infants with atopic dermatitis: maternal hopelessness, child-rearing attitudes and perceived infant temperament. Psychother Psychosom. 1999;68(1):39-45.
82. Al Robaee AA, Shahzad M. Impairment quality of life in families of children with atopic dermatitis. Acta Dermatovenerol Croat. 2010;18(4):243-7.
83. Egeberg A AP, Cappelleri J, Piercy J, Moon R, Tallman A, et al, editor Impact of mild or moderate atopic dermatitis on adult patients: Results from a real-world study in 3 European countries. European Academy of Dermatology and Venerology Congress 2018 Annual Meeting; 2018 Sept 12-16; Paris, France.
84. Allergy UK. ECZEMA: Are We Just Scratching The Surface? [Available from: https://www.allergyuk.org/about/latest-news/310-eczema-are-we-just-scratching-the-surface#_edn5.
85. Herd RM, Tidman MJ, Prescott RJ, Hunter JA. The cost of atopic eczema. Br J Dermatol. 1996;135(1):20-3.
86. Tsakok T, Woolf R, Smith CH, Weidinger S, Flohr C. Atopic dermatitis: the skin barrier and beyond. Br J Dermatol. 2019;180(3):464-74.
87. Karimkhani C, Dellavalle RP, Coffeng LE, Flohr C, Hay RJ, Langan SM, et al. Global Skin Disease Morbidity and Mortality: An Update From the Global Burden of Disease Study 2013. JAMA Dermatol. 2017;153(5):406-12.
88. Steel N, Ford JA, Newton JN, Davis ACJ, Vos T, Naghavi M, et al. Changes in health in the countries of the UK and 150 English Local Authority areas 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet. 2018;392(10158):1647-61.
89. Simpson A, Maeda S, Marsella R. Temporal dynamic changes of phenotypic expression of peripheral CD4 cells during environmental allergen challenge in an experimental model of canine atopic dermatitis: a pilot study. J Vet Med Sci. 2009;71(9):1177-81.
90. Hobbs FDR, Bankhead C, Mukhtar T, Stevens S, Perera-Salazar R, Holt T, et al. Clinical workload in UK primary care: a retrospective analysis of 100 million consultations in England, 2007-14. Lancet. 2016;387(10035):2323-30.
91. Owen K, Hopkins T, Shortland T, Dale J. GP retention in the UK: a worsening crisis. Findings from a cross-sectional survey. BMJ Open. 2019;9(2):e026048.
92. Tauber M, Balica S, Hsu CY, Jean-Decoster C, Lauze C, Redoules D, et al. Staphylococcus aureus density on lesional and nonlesional skin is strongly associated with disease severity in atopic dermatitis. J Allergy Clin Immunol. 2016;137(4):1272-4 e3.
93. Gisondi P, Ellis CN, Girolomoni G. Pimecrolimus in dermatology: atopic dermatitis and beyond. Int J Clin Pract. 2005;59(8):969-74.
94. Jensen JM, Weppner M, Dahnhardt-Pfeiffer S, Neumann C, Brautigam M, Schwarz T, et al. Effects of pimecrolimus compared with triamcinolone acetonide cream on skin barrier structure in atopic dermatitis: a randomized, double-blind, right-left arm trial. Acta Derm Venereol. 2013;93(5):515-9.
95. Ohtsuki M, Morimoto H, Nakagawa H. Tacrolimus ointment for the treatment of adult and pediatric atopic dermatitis: Review on safety and benefits. J Dermatol. 2018;45(8):936-42.

96. Furue M, Terao H, Rikihisa W, Urabe K, Kinukawa N, Nose Y, et al. Clinical dose and adverse effects of topical steroids in daily management of atopic dermatitis. Br J Dermatol. 2003;148(1):128-33.
97. Coondoo A. Topical corticosteroid misuse: the Indian scenario. Indian J Dermatol. 2014;59(5):451-5.
98. Charman CR, Morris AD, Williams HC. Topical corticosteroid phobia in patients with atopic eczema. Br J Dermatol. 2000;142(5):931-6.
99. Hoare C, Li Wan Po A, Williams H. Systematic review of treatments for atopic eczema. Health Technol Assess. 2000;4(37):1-191.
100. Hon KL, Kam WY, Leung TF, Lam MC, Wong KY, Lee KC, et al. Steroid fears in children with eczema. Acta Paediatr. 2006;95(11):1451-5.
101. Krejci-Manwaring J, Tusa MG, Carroll C, Camacho F, Kaur M, Carr D, et al. Stealth monitoring of adherence to topical medication: adherence is very poor in children with atopic dermatitis. J Am Acad Dermatol. 2007;56(2):211-6.
102. Aubert-Wastiaux H, Moret L, Le Rhun A, Fontenoy AM, Nguyen JM, Leux C, et al. Topical corticosteroid phobia in atopic dermatitis: a study of its nature, origins and frequency. Br J Dermatol. 2011;165(4):808-14.
103. Bos B, Antonescu I, Osinga H, Veenje S, de Jong K, de Vries TW. Corticosteroid phobia (corticophobia) in parents of young children with atopic dermatitis and their health care providers. Pediatr Dermatol. 2019;36(1):100-4.
104. Ruzicka T, Bieber T, Schopf E, Rubins A, Dobozy A, Bos JD, et al. A short-term trial of tacrolimus ointment for atopic dermatitis. European Tacrolimus Multicenter Atopic Dermatitis Study Group. N Engl J Med. 1997;337(12):816-21.
105. Van Leent EJ, Graber M, Thurston M, Wagenaar A, Spuls PI, Bos JD. Effectiveness of the ascomycin macrolactam SDZ ASM 981 in the topical treatment of atopic dermatitis. Arch Dermatol. 1998;134(7):805-9.
106. Reitamo S, Wollenberg A, Schopf E, Perrot JL, Marks R, Ruzicka T, et al. Safety and efficacy of 1 year of tacrolimus ointment monotherapy in adults with atopic dermatitis. The European Tacrolimus Ointment Study Group. Arch Dermatol. 2000;136(8):999-1006.
107. Meurer M, Folster-Holst R, Wozel G, Weidinger G, Junger M, Brautigam M, et al. Pimecrolimus cream in the long-term management of atopic dermatitis in adults: a six-month study. Dermatology. 2002;205(3):271-7.
108. Mylan Ltd. Elidel (Pimecrolimus 1%). Summary of product characteristics. Last updated 2019. Available from https://www.medicines.org.uk/emc/product/4966/smpc#companyDetails.
109. Leo Laboratories Limited. Summary of product characteristics. Protopic ointment (tacrolimus monohydrate) Last updated 2018. Available from: https://www.ema.europa.eu/en/documents/product-information/protopic-epar-product-information_en.pdf.
110. Callender VD AA, Stein Gold LF, Lebwohl M, Desai SR, Paller AS, Tan H, Vranic I, Ports WC, Zielinski MA, Tallman AM. . Long-Term Safety of Crisaborole Ointment, 2%, Across Racial and Ethnic Groups. American Academy of Dermatology Annual Meeting; San Diego, CA2018.
111. Horii KA, Simon SD, Liu DY, Sharma V. Atopic dermatitis in children in the United States, 1997-2004: visit trends, patient and provider characteristics, and prescribing patterns. Pediatrics. 2007;120(3):e527-34.

112. Zane LT, Kircik L, Call R, Tschen E, Draelos ZD, Chanda S, et al. Crisaborole Topical Ointment, 2% in Patients Ages 2 to 17 Years with Atopic Dermatitis: A Phase 1b, Open-Label, Maximal-Use Systemic Exposure Study. Pediatr Dermatol. 2016;33(4):380-7.
113. Dong C, Virtucio C, Zemska O, Baltazar G, Zhou Y, Baia D, et al. Treatment of Skin Inflammation with Benzoxaborole Phosphodiesterase Inhibitors: Selectivity, Cellular Activity, and Effect on Cytokines Associated with Skin Inflammation and Skin Architecture Changes. J Pharmacol Exp Ther. 2016;358(3):413-22.
114. Pfizer Inc. Data on file. Analysis of clinical trials AD-301 and AD-302. MO-48 Table 2.
115. Tacconelli E. Systematic reviews: CRD's guidance for undertaking reviews in health care. The Lancet Infectious Diseases. 2010;10(4):226.
116. Huang X, Tanojo H, Lenn J, Deng CH, Krochmal L. A novel foam vehicle for delivery of topical corticosteroids. J Am Acad Dermatol. 2005;53(1 Suppl 1):S26-38.
117. Daniels R, Knie U. Galenics of dermal products--vehicles, properties and drug release. J Dtsch Dermatol Ges. 2007;5(5):367-83.
118. Shamsudin N, Fleischer AB, Jr. Vehicle or placebo? Investigators use incorrect terminology in randomized controlled trials half of the time: a systematic review of randomized controlled trials published in three major dermatology journals. J Drugs Dermatol. 2010;9(10):1221-6.
119. Franz TJ, Parsell DA, Halualani RM, Hannigan JF, Kalbach JP, Harkonen WS. Betamethasone valerate foam 0.12%: a novel vehicle with enhanced delivery and efficacy. Int J Dermatol. 1999;38(8):628-32.
120. Pfizer Inc. Efficacy and Safety of Crisaborole and other Topical Agents for Mild-Moderate Atopic Dermatitis: Systematic Literature Review and Network Meta-analysis 2018.
121. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials. 1989;10(4):407-15.
122. Salek MS, Jung S, Brincat-Ruffini LA, MacFarlane L, Lewis-Jones MS, Basra MK, et al. Clinical experience and psychometric properties of the Children's Dermatology Life Quality Index (CDLQI), 1995-2012. Br J Dermatol. 2013;169(4):734-59.
123. Dodington SR, Basra MK, Finlay AY, Salek MS. The Dermatitis Family Impact questionnaire: a review of its measurement properties and clinical application. Br J Dermatol. 2013;169(1):31-46.
124. Basra MK, Gada V, Ungaro S, Finlay AY, Salek SM. Infants' Dermatitis Quality of Life Index: a decade of experience of validation and clinical application. Br J Dermatol. 2013;169(4):760-8.
125. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. J Clin Epidemiol. 2003;56(5):395-407.
126. Waters A SD, Beattie P, Ezughah F, Lewis-Jones S. Severity stratification of Children's Dermatology Life Quality Index (CDLQI) scores. Br J Dermatol; 2010.
127. Hongbo Y, Thomas CL, Harrison MA, Salek MS, Finlay AY. Translating the science of quality of life into practice: What do dermatology life quality index scores mean? J Invest Dermatol. 2005;125(4):659-64.
128. Ali FM, Kay R, Finlay AY, Piguet V, Kupfer J, Dalgard F, et al. Mapping of the DLQI scores to EQ-5D utility values using ordinal logistic regression. Quality of Life

Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation. 2017;26(11):3025-34.
129. Chopra R, Vakharia PP, Sacotte R, Patel N, Immaneni S, White T, et al. Severity strata for Eczema Area and Severity Index (EASI), modified EASI, Scoring Atopic Dermatitis (SCORAD), objective SCORAD, Atopic Dermatitis Severity Index and body surface area in adolescents and adults with atopic dermatitis. Br J Dermatol. 2017;177(5):1316-21.
130. Huang X, Xu B. Efficacy and Safety of Tacrolimus versus Pimecrolimus for the Treatment of Atopic Dermatitis in Children: A Network Meta-Analysis. Dermatology. 2015;231(1):41-9.
131. Levy AL SM, Roberts R. Tacrolimus Cream 0.03% Is Safe and Effective in the Treatment of Mild and Moderate Atopic Dermatitis in Adults
J Allergy Clin Immunol. 2005;115:S103.
132. Czarnowicki T, Malajian D, Khattri S, Correa da Rosa J, Dutt R, Finney R, et al. Petrolatum: Barrier repair and antimicrobial responses underlying this "inert" moisturizer. J Allergy Clin Immunol. 2016;137(4):1091-102 e7.
133. Institute for Clinical and Economic Review. Dupilumab and Crisaborole for Atopic Dermatitis: Effectiveness and Value - Final Evidence Report and Meeting Summary. 2017.
134. Dias S WN, Sutton AJ, Ades AE. NICE DSU Technical Support Document 2: A Generalised Linear Modelling Framework for Pairwise and Network Meta-analysis of Randomised Controlled Trials. Available from http://www.nicedsu.org.uk May 2011.
135. Phillippo DM AA, Dias S, Palmer S, Abrams KP, Welton NJ NICE DSU Technical Support Document 18: Methods for population-adjusted indirect comparisons in submissions to NICE. Available from: http://www.nicedsu.org.uk Dec 2016

136. Dias S SA, Welton NJ, Ades AE. NICE DSU Technical Support Document 3: NICE DSU Technical Support Document 3: Heterogeneity: subgroups, meta-regression, bias and bias-adjustment. Available from http://www.nicedsu.org.uk Sept 2011.
137. Dias S WN, Sutton AJ, Caldwell DM, Lu Guobing, Ades AE. NICE DSU Technical Support Document 4: Inconsistency in Networks of Evidence Based on Randomised Controlled Trials. Available from http://www.nicedsu.org.uk April 2011.
138. Dias S WN, Sutton AJ, Ades AE. NICE DSU Technical Support Document 5: Evidence Synthesis in the Baseline Natural History Model. Available from http://www.nicedsu.org.uk August 2011.
139. Meurer M, Fartasch M, Albrecht G, Vogt T, Worm M, Ruzicka T, et al. Long-term efficacy and safety of pimecrolimus cream 1% in adults with moderate atopic dermatitis. Dermatology. 2004;208(4):365-72.
140. Fowler J, Johnson A, Chen M, Abrams K. Improvement in pruritus in children with atopic dermatitis using pimecrolimus cream 1%. Cutis. 2007;79(1):65-72.
141. Kaufmann R, Bieber T, Helgesen AL, Andersen BL, Luger T, Poulin Y, et al. Onset of pruritus relief with pimecrolimus cream 1% in adult patients with atopic dermatitis: a randomized trial. Allergy. 2006;61(3):375-81.

142. Chapman MS, Schachner LA, Breneman D, Boguniewicz M, Gold MH, Shull T, et al. Tacrolimus ointment 0.03% shows efficacy and safety in pediatric and adult patients with mild to moderate atopic dermatitis. 2005;53(2):S177-S85.
143. Murrell DF, Gebauer K, Spelman L, Zane LT. Crisaborole Topical Ointment, 2% in Adults With Atopic Dermatitis: A Phase 2a, Vehicle-Controlled, Proof-of-Concept Study. J Drugs Dermatol. 2015;14(10):1108-12.
144. Hoeger PH, Lee KH, Jautova J, Wohlrab J, Guettner A, Mizutani G, et al. The treatment of facial atopic dermatitis in children who are intolerant of, or dependent on, topical corticosteroids: a randomized, controlled clinical trial. Br J Dermatol. 2009;160(2):415-22.
145. Hordinsky M, Fleischer A, Rivers JK, Poulin Y, Belsito D, Hultsch T. Efficacy and safety of pimecrolimus cream 1% in mild-to-moderate chronic hand dermatitis: a randomized, double-blind trial. Dermatology. 2010;221(1):71-7.
146. Murrell DF, Calvieri S, Ortonne JP, Ho VC, Weise-Riccardi S, Barbier N, et al. A randomized controlled trial of pimecrolimus cream 1% in adolescents and adults with head and neck atopic dermatitis and intolerant of, or dependent on, topical corticosteroids. Br J Dermatol. 2007;157(5):954-9.
147. Futamura M, Leshem YA, Thomas KS, Nankervis H, Williams HC, Simpson EL. A systematic review of Investigator Global Assessment (IGA) in atopic dermatitis (AD) trials: Many options, no standards. J Am Acad Dermatol. 2016;74(2):288-94.
148. National Institute for Health and Care Excellence. Certolizumab pegol and secukinumab for treating active psoriatic arthritis after inadequate response to DMARDs. 2017.
149. Owen RK, Tincello DG, Keith RA. Network meta-analysis: development of a three-level hierarchical modeling approach incorporating dose-related constraints. Value Health. 2015;18(1):116-26.
150. van Valkenhoef G, Kuiper J, van Valkenhoef MG. Package ‘gemtc’. 2016.
151. Team RC. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. h ttps. www R-project org. 2015.
152. Achana FA, Cooper NJ, Dias S, Lu G, Rice SJ, Kendrick D, et al. Extending methods for investigating the relationship between treatment effect and baseline risk from pairwise meta-analysis to network meta-analysis. Statistics in medicine. 2013;32(5):752-71.
153. Elidel (Pimecrolimus) Cream [Internet]. 2001 [cited 10 Aug 2019]. Available from: https://www.accessdata.fda.gov/drugsatfda_docs/nda/2001/21-302_Elidel.cfm.
154. Higgins JPT GSe. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. 2011.
155. Ades AE CD, Reken S, Welton NJ, Sutton AJ, Dias S. NICE DSU Technical Support Document 7: Evidence Synthesis of Treatment Efficacy in Decision Making: A Reviewer's Checklist. Available from http://www.nicedsu.org.uk May 2011. . 2012.
156. Mitchell AE, Fraser JA, Ramsbotham J, Morawska A, Yates P. Childhood atopic dermatitis: a cross-sectional study of relationships between child and parent factors, atopic dermatitis management, and disease severity. Int J Nurs Stud. 2015;52(1):216-28.
157. Klinnert MD, Booster G, Copeland M, Darr JM, Meltzer LJ, Miller M, et al. Role of behavioral health in management of pediatric atopic dermatitis. Ann Allergy Asthma Immunol. 2018;120(1):42-8 e8.

158. Bennetts M, Whalen E, Ahadieh S, Cappelleri JC. An appraisal of meta-analysis guidelines: how do they relate to safety outcomes? Res Synth Methods. 2017;8(1):64-78.
159. Spina D. PDE4 inhibitors: current status. Br J Pharmacol. 2008;155(3):308-15.
160. Cornell RC, Stoughton RB. Six-month controlled study of effect of desoximetasone and betamethasone 17-valerate on the pituitary-adrenal axis. Br J Dermatol. 1981;105(1):91-5.
161. Castela E, Archier E, Devaux S, Gallini A, Aractingi S, Cribier B, et al. Topical corticosteroids in plaque psoriasis: a systematic review of risk of adrenal axis suppression and skin atrophy. J Eur Acad Dermatol Venereol. 2012;26 Suppl 3:47-51.
162. National Institute for Health and Care Excellence. Tacrolimus and pimecrolimus for atopic eczema 2004 [Available from: https://www.nice.org.uk/guidance/ta82.
163. Garside R, Stein K, Castelnuovo E, Pitt M, Ashcroft D, Dimmock P, et al. The effectiveness and cost-effectiveness of pimecrolimus and tacrolimus for atopic eczema: a systematic review and economic evaluation. Health Technology Assessment (Winchester, England). 2005;9(29):iii, xi-xiii,1-230.
164. Ellis CN, Kahler KH, Grueger J, Chang J. Cost effectiveness of management of mild-to-moderate atopic dermatitis with 1% pimecrolimus cream in children and adolescents 2-17 years of age. Am J Clin Dermatol. 2006;7(2):133-9.
165. National Institute for Health and Care Excellence. NICE DSU TECHNICAL SUPPORT DOCUMENT 5: EVIDENCE SYNTHESIS IN THE BASELINE NATURAL HISTORY MODEL. 2012.
166. Pfizer Inc. Data on file. Analysis of clinical trials AD-301 and AD-302. Table 11.1.1.3.
167. Pfizer Inc. Data on file. Analysis of clinical trials AD-301 and AD-302. Table 44.2.2.5.
168. Breneman D, Fleischer AB, Jr., Abramovits W, Zeichner J, Gold MH, Kirsner RS, et al. Intermittent therapy for flare prevention and long-term disease control in stabilized atopic dermatitis: a randomized comparison of 3-times-weekly applications of tacrolimus ointment versus vehicle. J Am Acad Dermatol. 2008;58(6):990-9.
169. Thomsen SF. Epidemiology and natural history of atopic diseases. European Clinical Respiratory Journal. 2015;2.
170. Office for National Statistics. National life tables: England and Wales 2018 [Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/datasets/nationallifetablesenglandandwalesreferencetables.
171. Ali F. Email Correspondence with Dr. Ali re: Mapping CDQLI to EQ-5D. In: Pfizer Inc, editor. 2019.
172. Janssen MF, Szende A, Cabases J, Ramos-Goñi JM, Vilagut G, König HH. Population norms for the EQ-5D-3L: a cross-country analysis of population surveys for 20 countries. The European Journal of Health Economics. 2019;20(2):205-16.
173. National Institute for Health and Care Excellence. BNF 2019. Available from https://bnfniceorguk/.
174. Social DoHa. Drugs and pharmaceutical electronic market information tool (eMIT). Last updated 8 April 2019. Available from

https://www.gov.uk/government/publications/drugs-and-pharmaceutical-electronic-market-information-emit. Last accessed June 2019.
175. National Health Service. Reference costs 2017/18 [Available from: https://improvement.nhs.uk/resources/reference-costs/. 2018.
176. Personal Social Services Research Unit. Unit Costs of Health and Social Care: Personal Social Services Research Unit; 2018. Available from https://www.pssru.ac.uk/project-pages/unit-costs/unit-costs-2018/.

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Single technology appraisal

Crisaborole for Treating Mild to Moderate Atopic Dermatitis in People Aged 2 Years and Older [ID 1195]

Clarification questions

October 2019

File name	Version	Contains confidential information	Date
ID1195 Crisaborole Atopic dermatitis - clarification letter v1.0 to PM [ACIC]	1.0	Yes	4th October 2019

Notes for company

Highlighting in the template

Square brackets and [redacted] highlighting are used in this template to indicate text that should be replaced with your own text or deleted. These are set up as form fields, so to replace the prompt text in [redacted] with your own text, click anywhere within the highlighted text and type. Your text will overwrite the highlighted section.

To delete grey highlighted text, click anywhere within the text and press DELETE.

Section A: Clarification on effectiveness data

Literature searching and included studies

A1. Company submission (CS), Section B.2.1.1, page 31 and CS, Appendix D1 page 8. Please provide a copy of the search strategies and sources searched by the Global Resource for EczemA Trials (GREAT) database to identify randomised clinical trials of eczema treatments for literature before 2017.

Pfizer Response: The search strategy for the GREAT database is presented in Appendix 1. The GREAT database contains records of RCTs of treatments for established eczema published since the inception of the MEDLINE (1966) and EMBASE (1980) databases and also contains records of systematic reviews of treatments for established eczema from 2000 onwards.

A2. CS, Appendix D1, pages 13-14, Tables D1 and D2.

a) Please confirm that the updated search strategies are the same as those conducted by the GREAT database. If not, please explain the implications of the different strategies used.
b) Please confirm that no randomised controlled trials relevant to the decision problem have been missed.

Pfizer response:

a) The search strategies were designed to update the GREAT database search from 2017 to the search date, some modifications were made to increase sensitivity/specificity for the population, study design filters and animal filters search terms.

b) The GREAT database search is cited as having 94% sensitivity (1). It is expected that the Updated search would have similar sensitivity to the GREAT database search. In order to test the robustness of the Updated searches, we tested the recall (sensitivity) and specificity of the both strategies against the list of included studies. The Updated search strategy that was used for searching Medline and Embase from 2017 onward was as sensitive, and more specific, than the original strategy by the GREAT database.

For full details of this comparison, please see Appendix 2. In brief, the search was conducted in Embase and Medline to align with the GREAT database search strategy for literature from 2000-2016 and the searches were limited to English. The summarised results of the search comparison can be found in **Table 1** below.

Table 1: Results of robustness test between GREAT database search strategy and Updated search strategy

Search	Embase # of hits	Medline # of hits
GREAT Database Search Strategy	5717	5291
Updated Search Strategy	4769	2144

Both the GREAT and Updated search strategies missed three citations that were included in the SLR by a handsearching of previous SLR references:

1. Boguniewicz M; Fiedler VC; Raimer S; Lawrence ID; Leung DY; Hanifin JM. A randomized, vehicle-controlled trial of tacrolimus ointment for treatment of atopic dermatitis in children. Pediatric Tacrolimus Study Group. Journal of Allergy & Clinical Immunology. 102(4 Pt 1):637-44, 1998.

   This trial was found in the GREAT database, however, it was not picked up using the search strategies. Therefore, this study was likely added to the GREAT database via hand searching.

2. Hordinsky M; Fleischer A; Rivers JK; Poulin Y; Belsito D; Hultsch T. Efficacy and safety of pimecrolimus cream 1% in mild-to-moderate chronic hand dermatitis: a randomized, double-blind trial. Dermatology. 221(1):71-7, 2010.

   This trial was identified through recursive searching.

3. Sears HW; Bailer JW; Yeadon A. Efficacy and safety of hydrocortisone buteprate 0.1% cream in patients with atopic dermatitis. Clinical Therapeutics. 19(4):710-9, 1997 Jul-Aug.

   This trial was found in the GREAT database, however, it was not picked up using the search strategies. Therefore, this study was likely added to the GREAT database via hand searching.

The Updated search in Embase also missed the following reference:

Meurer M., Fartasch M., Albrecht G., Vogt T., Worm M., Ruzicka T., Altmeyer P.J., Schneider D., Weidinger G., Braeutigam M. Long-term efficacy and safety of pimecrolimus cream 1% in adults with moderate atopic dermatitis. Dermatology. 208 (4) (pp 365-372), 2004.

However, even though the Updated Embase search missed this reference it was picked up by Medline in the Updated search so still remained in the review.

In addition, the Pfizer UK SLR has been thoroughly cross-referenced with an additional SLR undertaken by Pfizer Global.(2) It is noted that the Pfizer Global search was undertaken independently and used a separate search strategy. All SLR results and data extractions from the UK and Global searches were cross-referenced to ensure all relevant publications were captured and as an additional quality assurance step. In addition, the references of a further 21 previously published SLRs were hand searched to ensure all relevant randomised controlled trials were included.

A3. Priority Question. CS, Section B.2.1. Section B.2.1.1 states that many studies of topical corticosteroids and topical calcineurin inhibitors were excluded either because the population was >20% severe or because a breakdown of severity was not reported. Please clarify whether these studies were checked for any subgroup data reported for mild-to-moderate patients.

Pfizer response: A search for subgroup data was conducted for studies excluded for population reasons. Please see **Table 2** provided which lists the studies that were excluded due to the population having >20% severe AD population and whether they reported subgroup data.

Table 2: Studies excluded due to population >20% severe

Author year	Subgroup data available
Almeyda, J.,Burt, B. W. (1974). Double blind controlled study of treatment of atopic eczema with a preparation of hydrocortisone in a new drug delivery system versus betamethasone 17-valerate *Br J Dermatol*, 91(5), 579-83	No mild and/or moderate disease subgroup data
Bleeker, J. (1975). Double blind comparison between two new topical corticosteroids, halcinonide 0.1% and clobetasol propionate cream 0.05% *Current medical research and opinion*, 3 (4), 225-228	No mild and/or moderate disease subgroup data
Savin, R. C. (1976). Betamethasone dipropionate in psoriasis and atopic dermatitis *Conn Med*, 40(1), 5-7	No mild and/or moderate disease subgroup data
Fisher, M.,Kelly, A. P. (1979). Multicenter trial of fluocinonide in an emollient cream base *Int J Dermatol*, 18(8), 660-4	No mild and/or moderate disease subgroup data
Veien, N. K.,Hattel, T.,Justesen, O.,Norholm, A.,Verjans, H. L. (1984). Hydrocortisone 17-butyrate (Locoid) 0.1% cream versus hydrocortisone (Uniderm) 1% cream in the treatment of children suffering from atopic dermatitis *J Int Med Res*, 12(5), 310-3	No mild and/or moderate disease subgroup data
Reitamo, S., Wollenberg, A., Schopf, E., Perrot, JL., Marks, R., Rusicka, T., Christophers, E., Kapp, A., Lahfa, M., Rubins, A., Jablonska, S., Rustin, M., for the European Tacrolimus Study Group. (2000), Safety and Efficacy of 1 Year of Tacrolimus Ointment Monotherapy in Adults With Atopic Dermatitis *Arch Dermatol,* 136, 999-	No mild and/or moderate disease subgroup data

1006	
Hanifin, J. M.,Ling, M. R.,Langley, R.,Breneman, D.,Rafal, E. (2001). Tacrolimus ointment for the treatment of atopic dermatitis in adult patients: part I, efficacy *J Am Acad Dermatol*, 44(1 Suppl), S28-38	Severe AD at baseline results presented for both studies combined in the intervention group but not for vehicle group
Hanifin, J.,Gupta, A. K.,Rajagopalan, R. (2002). Intermittent dosing of fluticasone propionate cream for reducing the risk of relapse in atopic dermatitis patients *Br J Dermatol*, 147(3), 528-37	No mild and/or moderate disease subgroup data
Paller, A.,Eichenfield, L. F.,Leung, D. Y.,Stewart, D.,Appell, M. (2001). A 12-week study of tacrolimus ointment for the treatment of atopic dermatitis in pediatric patients *J Am Acad Dermatol*, 44(1 Suppl), S47-57	No mild and/or moderate disease subgroup data
Soter, NA., Fleischer, AB., Webster, GF., Monroe, E., Lawrence, I., and the Tacrolimus Ointment Study Group. (2001). Tacrolimus ointment for the treatment of atopic dermatitis in adult patients: Part II, Safety. *J Am Acad Dermatol,* S39-S46	No mild and/or moderate disease subgroup data
Meurer, M.,Folster-Holst, R.,Wozel, G.,Weidinger, G.,Junger, M.,Brautigam, M.,Casm-De- study group (2002). Pimecrolimus cream in the long-term management of atopic dermatitis in adults: a six-month study *Dermatology*, 205(3), 271-7	No mild and/or moderate disease subgroup data for outcomes of interest
Reitamo, S.,Rustin, M.,Ruzicka, T.,Cambazard, F.,Kalimo, K.,Friedmann, P. S.,Schoepf, E.,Lahfa, M.,Diepgen, T. L.,Judodihardjo, H.,Wollenberg, A.,Berth-Jones, J.,Bieber, T.,European Tacrolimus Ointment Study, (2002). Efficacy and safety of tacrolimus ointment compared with that of hydrocortisone butyrate ointment in adult patients with atopic dermatitis *Journal of Allergy and Clinical Immunology*, 109(3), 547-555	No mild and/or moderate disease subgroup data
Luger, T. A.,Lahfa, M.,Folster-Holst, R.,Gulliver, W. P.,Allen, R.,Molloy, S.,Barbier, N.,Paul, C.,Bos, J. D. (2004). Long-term safety and tolerability of pimecrolimus cream 1% and topical corticosteroids in adults with moderate to severe atopic dermatitis *J Dermatolog Treat*, 15(3), 169-78	No mild and/or moderate disease subgroup data
Reitamo, S.,Harper, J.,Bos, J. D.,Cambazard, F.,Bruijnzeel-Koomen, C.,Valk, P.,Smith, C.,Moss, C.,Dobozy, A.,Palatsi, R.,European Tacrolimus Ointment, Group (2004). 0.03% Tacrolimus ointment applied once or twice daily is more efficacious than 1% hydrocortisone acetate in children with moderate to severe atopic dermatitis: results of a randomized double-blind controlled trial.[see comment] *British Journal of Dermatology*, 150(3), 554-62	Presents mEASI scores by moderate subgroup (see Figure 2 and Table 2) Not stratified randomized by AD severity
Reitamo, S.,Ortonne, J. P.,Sand, C.,Cambazard, F.,Bieber, T.,Folster-Holst, R.,Vena, G.,Bos, J. D.,Fabbri, P.,Groenhoej Larsen, C.,European Tacrolimus Ointment Study, Group (2005). A multicentre, randomized, double-blind, controlled study of long-term treatment with 0.1% tacrolimus ointment in adults with moderate to severe atopic dermatitis *Br J Dermatol*, 152(6), 1282-9	No mild and/or moderate disease subgroup data
Fleischer, A. B., Jr.,Abramovits, W.,Breneman, D.,Jaracz, E.,U. S/Canada tacrolimus ointment study group (2007). Tacrolimus ointment is more effective than pimecrolimus cream in adult patients with moderate to very severe atopic dermatitis *J Dermatolog Treat*, 18(3), 151-7	Presents EASI and IGADA scores by moderate subgroup (see Figure 5 and Figure 7) Not stratified randomized by AD severity
Remitz, A., Harper, J., Rustin, M., Goldschmidt, W.F.,	No mild and/or moderate disease

Palatsi, R., Van Der Valk, P.G., Sharpe, G., Smith, C.H., Dobozy, A., Turjanmaa, K. and European Tacrolimus Ointment Study Group (2007). Long-term safety and efficacy of tacrolimus ointment for the treatment of atopic dermatitis in children. Acta dermato-venereologica, 87(1), 54-61.	subgroup data
Mandelin, J.,Remitz, A.,Virtanen, H.,Reitamo, S. (2010). One-year treatment with 0.1% tacrolimus ointment versus a corticosteroid regimen in adults with moderate to severe atopic dermatitis: A randomized, double-blind, comparative trial *Acta Derm Venereol*, 90: 170-174	No mild and/or moderate disease subgroup data
Woods, M. T.,Brown, P. A.,Baig-Lewis, S. F.,Simpson, E. L. (2011). Effects of a novel formulation of fluocinonide 0.1% cream on skin barrier function in atopic dermatitis *J Drugs Dermatol*, 10(2), 171-6	No mild and/or moderate disease subgroup data

Trial design

A4. CS, Section B.2.2, Table 9 and Section B.2.3, Table 11.

a) Please clarify why the crisaborole randomised controlled trials, AD-301 and AD-302, were designed identically rather than as a single study.
b) Please confirm whether there were any differences in design between these 2 trials

Pfizer response

a) Manufacturers seeking approval of a new drug in the United States consequently typically run two identical studies to meet FDA requirements for a new drug application submission. The US Food and Drugs Administration (FDA) requires that manufacturers of a new drug provide 'substantial evidence' of efficacy based on 'adequate and well-controlled investigations,' which FDA has clarified as being at least two (adequate and well-controlled) clinical studies.

b) Studies AD 301 and AD 302 were designed identically and there were no differences between the two trials.

A5. CS, Section B.2.4.3, Figures 4 and 5. Please clarify whether there were any stratification factors used when randomising patients to treatment in AD-301 and AD-302.

Pfizer response Subjects were randomised to crisaborole or vehicle stratified by study centre.

A6. CS, Section B.2.6.1.2, page 56. The company submission states that "The MCID for the CDLQI has been reported as ≥ 2.5-point change from baseline". Please explain the appropriateness of using this minimal clinically important difference (MCID) for the Children's Life Quality Index (CDLQI) for patients aged 2-15 years when it is only validated in patients aged 4-15 years.

Pfizer response: The Children's Life Quality Index (CDLQI) was developed in a cohort of 169 children aged 3-16 years and validated in a cohort of 233 children aged 4-16 years.(3) The CDLQI is self-explanatory and may be completed by the child alone. Ages younger than 4 years require parent or caregiver assistance. However, the CDLQI allows for parent or caregiver assistance for children 4 years and older as well. Thus, the addition of children ages 2 and 3 years does not represent a change in methodology. Moreover, the results of the analysis of mean change in baseline CDLQI in the subgroup of patients 4-15 years treated with crisaborole was nearly identical to that of patients 2-15 years (-4.5 versus -4.6, respectively.) The minimal clinically important difference (MCID) is the smallest amount of change in the assessment instrument score that an individual patient would identify as important and which would result in a change in the patient's management. The MCID for the CDLQI was previously estimated as a 2.5-point or greater change from baseline.(4) Thus, patients treated with crisaborole experienced a change in HRQoL as measured by the CDLQI assessment that was nearly double the threshold that would result in a perceived difference and warrant a change in disease management.

Trial analysis and results

A7. CS, Section B.1.1, Table 1. The company decision problem specifies the following population “Adults and children aged 2 years and older with mild to moderate AD that has not been controlled by topical corticosteroids or where there is a serious risk of important adverse effects from further topical corticosteroid use, particularly irreversible skin atrophy (second-line use post-TCS treatment)”. However, AD-301 and AD-302 did not select patients whose condition was uncontrolled on topical corticosteroids or were at serious risk of adverse effects from topical corticosteroids.

a) Please clarify whether the trial populations would be easier to treat and therefore the trial results may be more favourable than the likely results in the intended population in clinical practice.
b) Please clarify whether this has been explored in any sensitivity analyses within these trials or whether there is any evidence from trials of other atopic dermatitis treatments that address this issue.

Pfizer response:

a) Expert clinical advice indicated that patients who have active mild to moderate AD despite previous TCS, represent a more challenging mild to moderate AD population to treat since these patients may include a proportion of patients who are generally resistant to treatment.

[redacted]

We have additionally included a forest plot showing the odds ratio of a response for crisaborole versus vehicle for patients according to prior TCS therapy exposure, see **Figure 1**. It is evident from [redacted] In the cost-effectiveness analysis an exploratory sensitivity analysis has also been undertaken to consider patients who were TCS naïve vs TCS exposed based on the results from AD 301 and AD 302 trial data. In this analysis, consistent with the clinical trial evidence, [redacted] It is noted that in this exploratory analyses, the overall cost-effectiveness results are more favourable for crisaborole, see Appendix 3 and **Figure 1**.

Figure 1: Odds Ratio ISGA Clear/Almost Clear

[redacted]

A8. CS, Section B.2.4.2, Table 16 and Section B.2.7.1. The company submission states that the statistical analysis of primary, secondary and other endpoints was "… tested between treatment groups using logistical regression with factors for treatment group and analysis centre."

a) Please provide estimates of odds ratios (with 95% confidence intervals) adjusted for centre for each trial as per the primary and secondary analyses.
b) Please comment on whether there was any evidence of a differential treatment effect by centre.

Pfizer response:

a) The estimates provided for primary and secondary endpoint analyses have been stratified by treatment centre.
b) There was no evidence of a differential treatment effect by treatment centre. The p-values for treatment by analysis centre interaction were 0.992 and 0.951 for Study 301, 302 respectively, please see **Table 3**.

Table 3: Pooling Analysis for the Primary Efficacy Endpoint: Success in ISGA at Day 29

	Treatment by Analysis Center P-Value	
	AD-301	**AD-302**
Success in ISGA at Day 29	0.992	0.951

Note: P-Value for interaction term from a logistic regression (with Firth option) test with factors of treatment group, analysis center, and treatment group by analysis center interaction. Values have been adjusted for multiple imputation.

A9. Priority Question. CS, Sections B.2.6 and B.2.7, pages 55 to 64. Some of the trial results are presented separately for AD-301 and AD-302, while for some outcomes and some subgroup analyses, the data are pooled across the 2 trials.

a) Please justify why some outcomes were pooled but not others, and why pooling was acceptable (given that it breaks randomisation).
b) Please justify why a meta-analysis of outcomes across AD-301 and AD-302 was not conducted for all outcomes specified in the NICE final scope.
c) Please state any known reasons why more patients achieved Investigator's Static Global Assessment (ISGA) success or ISGA 0/1 in AD-301 than in AD-302, particularly in the vehicle arms.
d) For the outcome ISGA 0/1 sub-grouped by age (Table 24), please provide separate results for AD-301 and AD-302 for consistency with other subgroups presented.

Pfizer response:

a) This was conducted for FDA purposes - pooled analyses can provide more robust results with larger sample size, especially for subgroup analyses.

b) Pooled analyses were available for the following efficacy endpoints which are broadly consistent with the outcomes specified in the NICE final scope:

- Primary:
    - Success in ISGA (an ISGA score of Clear [0] or Almost Clear [1] with at least a 2-grade improvement from Baseline)

Table 4: Proportion of pooled patients achieving ISGA Success at Day 29

Cohort Stratification	Number of subjects	% patients at Day 29 achieving ISGA Success	
Stratified by Baseline ISGA			
Mild Crisaborole	[redacted]	[redacted]	[redacted]
Mild Vehicle	[redacted]	[redacted]	
Moderate Crisaborole	[redacted]	[redacted]	[redacted]
Moderate Vehicle	[redacted]	[redacted]	

*All *p* values are versus vehicle control
Note: Study cohorts from AD-301 and AD-302 were pooled; ISGA score of ≤1 (clear or almost clear) with a ≥2-grade improvement from baseline

- Secondary (Please see Appendix 4 for data tables)
    - Proportion of subjects with ISGA score of Clear (0) or Almost Clear (1) at Day 29
        - Main Submission, Section B.2.7.1, Table 24-25 (p. 63-64)
    - Time to success in ISGA (defined as a score of Clear [0] or Almost Clear [1] with at least a 2-grade improvement from Baseline)
        - Main Submission, Section B.2.6.1.2; Figure 9 (p. 53)
    - Clinical signs of AD (erythema, induration/papulation, exudation [oozing or crusting], excoriation [evidence of scratching], and lichenification [epidermal thickening]) evaluated globally on a 4-point scale and not by body region

- Main Submission, Section B.2.6.1.2; Figure 11-12 page 54-55

- Time to Improvement in Pruritus (a score of None [0] or Mild [1] of the SPS with at least a 1-grade improvement from Baseline)
  - Main Submission, Section B.2.6.1.2, Figure 10 (p 54)
- Proportion of subjects with Improvement in Pruritus by visit
  - Appendix 4, Table 1
- Change from Baseline in Treatable %BSA
  - Appendix 4, Table 2
- CDLQI, DLQI, and DFI. Scores and changes from Baseline summarized by treatment group
  - CDLQI: Main Submission, Section B.2.6.1.2; Table 19 (p. 56)
  - DLQI: Main Submission, Section B.2.6.1.2; Table 20 (p. 56)
  - DFI: Main Submission, Section B.2.6.1.2; Table 21 (p. 56)
- Subgroup analyses
  - Success in ISGA (an ISGA score of Clear [0] or Almost Clear [1] with at least a 2-grade improvement from Baseline). Day 29, by subgroups of:
    - Sex
      - Appendix 4, Table 3
    - Age
      - Appendix 4, Table 4
    - Ethnicity
      - Appendix 4, Table 5

- Race
    - Appendix 4, Table 6
- Prior AD medication (naïve, corticosteroid [systemic & dermatological preparations] and topical calcineurin inhibitor)
    - Main Submission, Section B.2.7.1; Table 26 (p. 65)

It is acknowledged that across the endpoints [redacted]

d)

Table 5: ISGA 0/1 Rates at 29 Days Per Treatment Arm Results for Adult and Child Subgroup RCT Data

Trial number (Acronym)	Baseline ISGA	Arm	N/ % ISGA 0-1			
			Adults	Children (age groups in years)		
AD-301	Mild	Crisaborole 2%	[redacted]			
AD-301	Mild	Vehicle				
AD-301	Moderate	Crisaborole 2%				
AD-301	Moderate	Vehicle				
AD-302	Mild	Crisaborole 2%				
AD-302	Mild	Vehicle				
AD-302	Moderate	Crisaborole 2%				
AD-302	Moderate	Vehicle				

A10. CS, Section B.2.7. Please provide an analysis by study using a single model that includes all relevant covariates together with interaction terms for treatment by covariate and include continuous baseline variables as continuous covariates using a suitable relationship to response (not necessarily linear).

Pfizer response: We point out that the requested multivariate regression analysis had not been part of the initial submission. The matching adjusted indirect comparison (MAIC) used in Document B Section 2.8.10 relies on a propensity score reweighting of the pooled AD301 and AD302 datasets to match the average baseline characteristics of comparator trials. Document D details the combination of literature reviews, univariate regression analyses, and clinical expert surveys used to identify a list of important effect modifiers and prognostic variables to be used in matching. These variables were

[redacted]

In response to the reviewer request, we have conducted simple generalised linear models on ISGA 0/1 and ISGA success including all the relevant prognostic variables and effect modifiers.

[redacted]

Results are presented below from both the “all relevant covariates” model and “only significant covariates” model.

[redacted]

Table 6 Estimated odds ratios and p-values for ISGA 0/1 binomial-logistic regression analysis pooling AD301 and AD302 datasets

	All relevant covariates		Only significant covariates	
	Estimate (95% CI)	p-value	Estimate (95% CI)	p-value
(Intercept)	[redacted]	[redacted]	[redacted]	[redacted]
study	[redacted]	[redacted]	[redacted]	[redacted]
treatment	[redacted]	[redacted]	[redacted]	[redacted]
age	[redacted]	[redacted]	[redacted]	[redacted]
bsa	[redacted]	[redacted]		
sex	[redacted]	[redacted]		
Prior treatment	[redacted]	[redacted]	[redacted]	[redacted]
white	[redacted]	[redacted]		
isga	[redacted]	[redacted]	[redacted]	[redacted]
treatment:age	[redacted]	[redacted]		
treatment:bsa	[redacted]	[redacted]		
treatment:sex	[redacted]	[redacted]		
treatment:prior treatment	[redacted]	[redacted]		
treatment:white	[redacted]	[redacted]		
treatment:isga	[redacted]	[redacted]		

* P-values that are less than 0.05 “significance” threshold. If a variable is significant as an effect modifier (i.e. treatment by covariate interaction) then it is included as both a prognostic factor and effect modifier in the “only significant covariates” model.

Table 7 Estimated odds ratios and p-values for ISGA success binomial-logistic regression analysis pooling AD301 and AD302 datasets

	All relevant covariates		Only significant covariates	
	Estimate (95% CI)	p-value	Estimate (95% CI)	p-value
(Intercept)	[redacted]	[redacted]	[redacted]	[redacted]
study	[redacted]	[redacted]	[redacted]	[redacted]
treatment	[redacted]	[redacted]	[redacted]	[redacted]
age	[redacted]	[redacted]		
bsa	[redacted]	[redacted]		
sex	[redacted]	[redacted]		
Prior treatment	[redacted]	[redacted]	[redacted]	[redacted]
white	[redacted]	[redacted]		
isga	[redacted]	[redacted]	[redacted]	[redacted]
treatment:age	[redacted]	[redacted]		
treatment:bsa	[redacted]	[redacted]		
treatment:sex	[redacted]	[redacted]		
treatment:prior treatment	[redacted]	[redacted]		
treatment:white	[redacted]	[redacted]		
treatment:isga	[redacted]	[redacted]	[redacted]	[redacted]

* P-values that are less than 0.05 “significance” threshold. If a variable is significant as an effect modifier (i.e. treatment by covariate interaction) then it is included as both a prognostic factor and effect modifier in the “only significant covariates” model.

A11. CS, Section B.2.7.1, Table 25. Please comment on the finding that in all age subgroups of mild disease in AD-301, there is no significant difference in ISGA 0/1 between crisaborole and vehicle, while for the age 7-11 subgroup, there is a higher response in the vehicle arm.

Pfizer response: It is noted that the subgroup results by mild/moderate disease are post-hoc analyses and AD 301 and AD 302 trials did not have any prospective stratification or randomisation allocation quotas by age group or disease severity. The baseline "mild" population the sample size is approximately 39% of the overall pooled study population and is substantially smaller than for the baseline "moderate" population. AN2728-AD-301 and AN2728-AD-302 were not statistically powered for the subgroup analyses and as such, results of a post-hoc analysis should be viewed with caution. Whilst, the primary endpoint, success in ISGA, was defined as an ISGA of "clear" or "almost clear" with at least a 2-grade improvement from Baseline. In the pivotal studies, this means that for a treatment to be considered a success for a "mild" subject, they effectively had to show zero disease in order to achieve "clear" on the ISGA scale. Consequently, it is more difficult to achieve success in a "mild" vs "moderate" subject as the latter only needs to achieve a state of "almost clear" whereas as a mild subject was required to be scored as "clear". For example, to illustrate how stringent the scale was in the mild subjects in regards to the grade of "clear", hence, even if a subject had one remaining lesion with 0.25% Body Surface Area (BSA) (equivalent to approximately a 1/4 handprint in size) with ISGA grade = 1, the subject would qualify as "almost clear" and not "clear" and therefore would not meet the primary endpoint. The secondary endpoint, i.e. proportion of patients who achieved an ISGA score of "clear" (0) or "almost clear" (1) at day 29, which does not require a 2-grade improvement (not as stringent as the primary endpoint), was statistically significant for both the "mild" and "moderate" patient populations.

Nevertheless, it is noted that overall a statistically significantly greater proportion of crisaborole-treated patients achieved success in ISGA than in in other subgroups analysed

Extension study, AD-303

A12. CS, Sections B.2.3 and B.2.4.3. The company submission states that patients entered the extension study, AD-303 if they had completed AD-301 or AD-302 without experiencing a crisaborole treatment-related adverse event or a serious adverse event. Figure 4 indicates that 1398 patients completed AD-301 or AD-302. However, Figure 5 indicates that only 517 patients (37%) entered AD-303. Please comment on:

a) why only 37% of trial completers entered AD-303 and
b) whether these patients were likely to have more favourable results than those who did not enter.

Pfizer response:

a) Study AD 303 was conducted as a long-term safety study. It is noted that not all sites were consequently invited to participate because the study recruited only sufficient patient numbers to appropriately assess safety outcomes.

b) Whilst there is no formal record of the reasons that patients were not enrolled in AD 303 for the participating sites, all subjects at selected study sites were eligible for inclusion. The trials were kept blinded so that patients and caregivers/parents did not know if they were on crisaborole or vehicle in AD 301 and AD 302 and, hence, would not create a selection bias for AD 303 participation. There is consequently no reason to suspect that these patients were more likely to have a favourable result than those who did not participate.

A13. CS, Section B.2.4.3, Figure 6. In the extension study, AD-303, only 52% (271 people of 517) completed the study, while 86 (17%) withdrew and 115 (22%) discontinued for “other” reasons. Please provide the reasons for discontinuation, both for those who withdrew and those who discontinued for other reasons.

Pfizer response: The most frequent reasons for discontinuation from Study AN2728-AD-303 were

[redacted]

[redacted]It is noted that the further details on the reason for withdrawal [redacted]

A14. CS, Section B.2.6.2, pages 62 to 63. In addition to the outcomes presented in Section B.2.6.2, please confirm whether any further effectiveness data is available for AD-303. If available, please provide this data. For example, of all on-treatment periods started, how many led to a response (ISGA 0/1) within 4 weeks?

Pfizer response: The proportion of subjects with ISGA Clear or Almost Clear at each Visit by Age Group from Study AN2728-AD-303 is reported in **Figure 2**.

In Study AN2728-AD-303, the proportion of subjects with an ISGA Clear or Almost Clear response was similar among age groups at most visits and improved over time; 55% of subjects achieved a response of ISGA Clear or Almost Clear at Week 48 and improvement was observed from Baseline with no differences observed across age groups. The indirect assessment of efficacy in the long-term extension

Study AN2728-AD-303 suggests that clinical benefit as assessed by ISGA was sustained with long-term intermittent treatment with crisaborole ointment, 2%.

Figure 2: Subjects with ISGA Clear or Almost Clear at each Visit by Age Group, Study AN2728-AD-303 (Safety Population, Observed Cases)

Network meta-analysis (NMA)

A15. Priority Question. CS, Section B.2.8.2. Selection of key outcome for the network meta-analysis (NMA) is described on page 67.

a) Please provide further justification for selecting ISGA/IGA 0/1 as the only effectiveness measure for use in the NMA and model, given that the NICE technology appraisal of topical calcineurin inhibitors (tacrolimus and pimecrolimus) includes a range of outcome measures (for example, Eczema Area and Severity Index [EASI], Physicians Global Evaluation [PGE], reduction in body surface area [BSA]).

b) Please justify why other outcomes such as clinical signs were not analysed in a NMA.

Pfizer response:

(a) AD 301 and AD 302 did not record EASI outcomes during data collection. It is noted that EASI is a composite (multi-item) score assessing extension and severity of signs of atopic dermatitis and with range 0-72. IGA/ISGA are outcome measures that typically score a more global domain on an ordinal scale (e.g. rating overall eczema severity) and ranges from 0-5 only, without specific assessment of extension of the disease. These differences would render combining these two types of endpoints challenging. Nevertheless, EASI outcomes were considered for potential inclusion in an exploratory NMA, but after full review of the trials that reported EASI outcomes it was evident that the network of evidence could not be increased by considering the additional EASI endpoint.(5-14) The relevant studies that reported EASI outcomes were either not suitable for inclusion for other reasons (5-14) or also reported ISGA/IGA (6, 7, 11, 12) and hence, had already been included in the network of evidence based on ISGA/IGA outcomes. The analysis was consequently not undertaken.

As explained in A16 below, all IGA and ISGA outcomes included in the ISGA 0/1 NMA were static while PGE is dynamic, so PGE could not be merged with IGA/ISGA.

(b)The only clinical signs on which a connected network could potentially be performed was the binary outcome of clear almost clear of pruritus. The network could only be connected if studies purely in facial AD were included; this was conducted as a sensitivity analysis and results provided in Appendix D of the submission

A16. Priority Question. CS, Section B.2.8.2. Selection of key outcome for the NMA is described on page 67. The company submission states that ISGA and IGA were sufficiently similar to be pooled within the NMA. It also states that ISGA is a static measure (assessed at one time-point) but that Physicians Global Evaluation (PGE) was not pooled because it is a dynamic measure (relative to baseline). Please clarify whether the Investigator Global Assessment (IGA) measure used in the comparator studies in the NMA was static or dynamic. If dynamic, please clarify why it was acceptable to pool these with ISGA but not to pool other measures such as PGE.

Pfizer response: The IGA in all comparator studies was static as it was assessed at only one time point.

A17. Priority Question. CS, Sections B.2.8.2 and B.2.8.4. Selection of key outcome for the NMA is described on page 67. The company submission states that the primary outcome of ISGA success in the crisaborole studies (that is, ISGA 0/1 and ≥2-grade improvement) was not consistently reported in comparator trials, so could not be analysed in the NMA. However, section B.2.8.4 and Table 30 indicate that 3 comparator studies report an outcome of ISGA/IGA 0/1 and ≥1-grade improvement. Please clarify whether these data could also be generated for AD-301 and AD-302. If so, please justify why no NMA was provided for this outcome given that these studies would also form a (limited) network.

Pfizer response: Please note that it was stated in Document B p 67 that the key outcome we considered in the NMA was ISGA/IGA 0/1 and that 'this outcome was selected as the key outcome for the NMA rather than ISGA success (the primary outcome in the crisaborole studies) because ISGA success was not **consistently reported in comparator RCTs and could not be analysed appropriately**'. ISGA 0/1 was consequently the outcome which allowed us to put together the most robust network of evidence, which included the largest number of trials, with an outcome that was reported consistently across studies. ISGA 0/1 was also the outcome which was considered the most appropriate endpoint for the economic model (which is based on disease states rather than treatment response). Nevertheless, it is also noted that full results of the NMA on ISGA success have already been reported in Appendix D and are available for review.

A18. Priority Question. CS, Section B.2.8.3. The differences in vehicle across randomised controlled trials in mild to moderate atopic dermatitis is described on page 68. Section B.2.8.3 states that

[redacted] **The company submission also states that**

[redacted] Based on this, the base-case NMA includes a meta-regression for vehicle response.

a) Please explain why the crisaborole base ointment would be likely to have more benefit than other ointments such as the base ointment for tacrolimus.
b) Trial results are usually analysed based on the relative difference between arms. Please justify why it is assumed that the greater difference between intervention and response in the comparator trials than in the crisaborole trials does not reflect a true greater effect of comparators.
c) Given the above, please justify why the NMA with meta-regression for vehicle response was used as the base case.

Pfizer response:

[redacted]

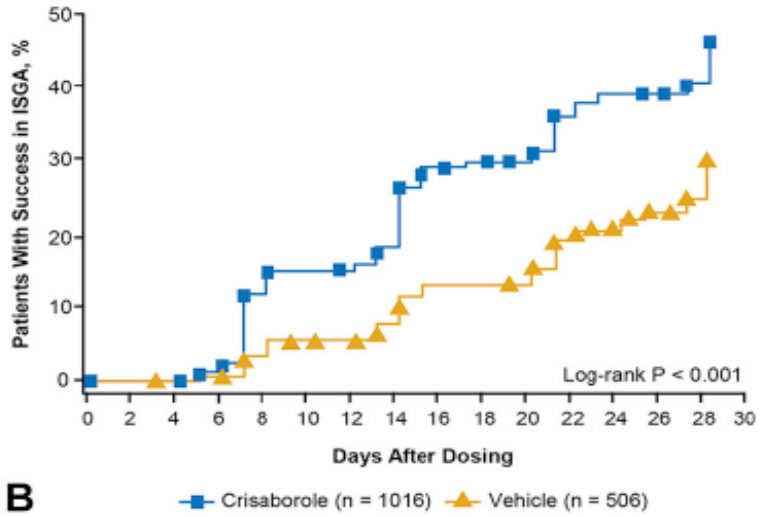
[redacted]

b) This is a result of our answer to the first query. The ‘control’ used across trials in the AD networks, namely the vehicle, are not comparable as they are active treatments. The ointment used in the crisaborole trials has much greater treatment effect than that in the comparator trials. In our base case NMA, we do analyse the relative treatment effects between arms but include a vehicle response meta-regression to adjust for this difference.

c) Given the differences in vehicle across trials, we were motivated to consider a meta-regression for vehicle response. In Document B Table 32, we provide the total residual deviance and deviance information criterion (DIC) for all models with and without vehicle response regression. The DIC and deviance are substantially lower (~5 points difference on both statistics) for vehicle response regressions, strongly justifying our choice to include regression in the base case.

A19. CS, Section B.2.8.4. Please provide evidence in support of a proportional hazards assumption in the NMA.

Pfizer response: There was limited evidence available on which to assess the assumption of proportion hazards. Paller 2016 on AD301/AD302 provide the Kaplan-Meier curve below for ISGA success. There is no evidence of crossing curves or non-proportional hazards. As this multiple timepoint data was only available for AD301/AD302, it would not have been feasible to use non-proportional hazards models such as fractional polynomials or piecewise constant models.



Studies synthesised reported between 4 and 6 weeks, meaning reasonable homogeneity in duration. A sensitivity analysis was provided in Appendix D synthesising only studies reporting at 4 weeks; results were consistent with the base case but shifted somewhat in favour of crisaborole.

A20. CS, Section B.2.8.5. A potential explanation for [redacted] is a lack of evidence, including the use of so-called non-informative prior distributions for model parameters. If this is the case, please incorporate weakly informative prior information about model parameters rather than simplifying the model.

Pfizer response: The only models for which the fixed treatment effect, random class effect, vehicle response regression (FE-RCE-VR) model [redacted]

[redacted]. We used the [redacted] Forest plots in the two subgroups using FE-RCE-VR with related regression coefficients and informative priors are presented in Figure 3 and Figure 5 for children AD and adult AD, respectively. The results presented in Document B, i.e. those using simple fixed effects models and no class effects or vehicle regression, are presented in Figure 2 and Figure 4. A comparison of the mean ranks with the simple fixed effects models is also presented in Table 1.

[redacted]

[redacted]

Figure 3 ISGA/IGA 0/1 Forest plot for children with AD using FE-RCE-VR model with related regression coefficients across treatments and informative prior on regression coefficient.

Figure 4 Subgroup of children with AD. Forest plot of NMA estimated hazard ratios (95% CrI) with P-best (probability crisaborole superior) of ISGA/IGA 0/1 of comparators versus crisaborole. Using simple fixed treatment effects

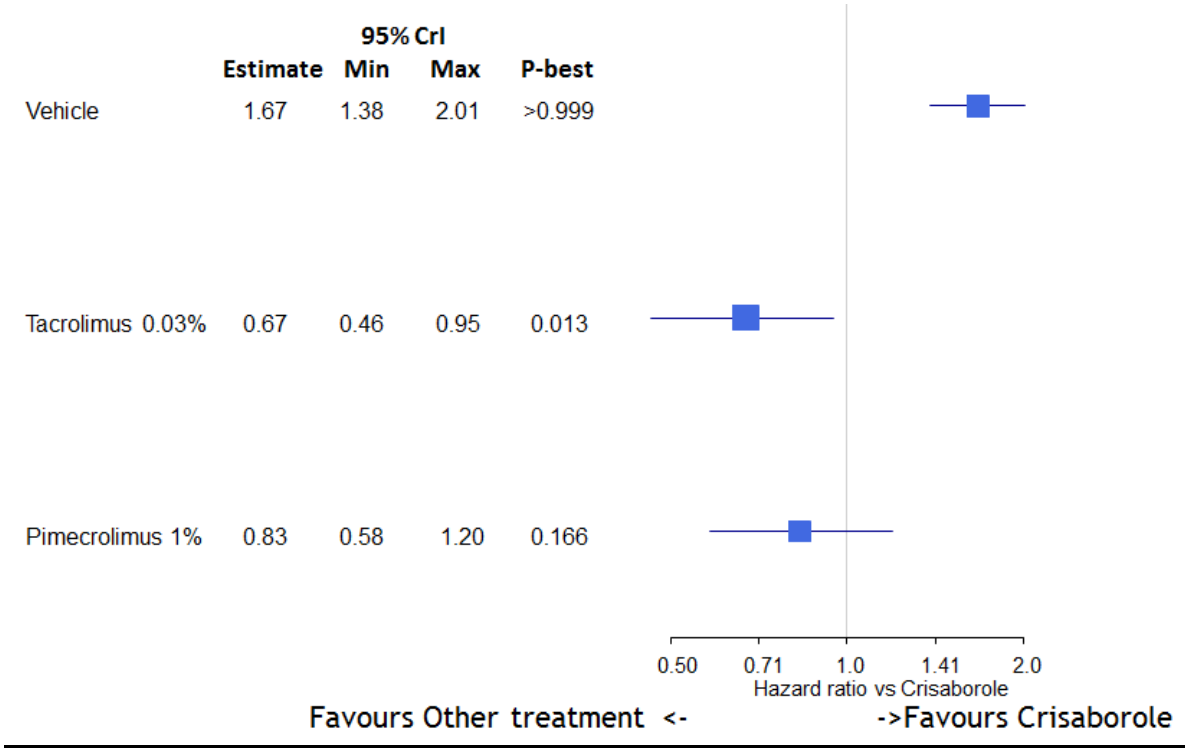


Figure 5 ISGA/IGA 0/1 Forest plot for adults with AD using FE-RCE-VR model with related regression coefficients across treatments and informative prior on regression coefficient

Figure 6 Subgroup of adults with AD. Forest plot of NMA estimated hazard ratios (95% CrI) with P-best (probability crisaborole superior) of ISGA/IGA 0/1 of comparators versus crisaborole. Using simple fixed treatment effects model.

Table 8 Comparison of mean ranks on ISGA/IGA 0/1 outcomes in subgroup analyses using simple fixed effects FE-RCE-VR models with informative priors on the regression coefficients

	Children with AD using simple fixed effects	Children with AD using FE-RCE-VR and informative prior	Adults with AD using simple fixed effects	Adult with AD using FE-RCE-VR and informative prior
Vehicle	[redacted]	[redacted]	[redacted]	[redacted]
Tacrolimus 0.03%	[redacted]	[redacted]	[redacted]	[redacted]
Pimecrolimus 1%	[redacted]	[redacted]	[redacted]	[redacted]
Crisaborole 2%	[redacted]	[redacted]	[redacted]	[redacted]

A21. CS, Appendix D, page 203. Page 203 states "We used [redacted]". Please clarify whether autocorrelation has been assessed when generating [redacted] and that [redacted] are sufficient with which to estimate parameters allowing for autocorrelation and accuracy.

Pfizer response:

[redacted]

A22. CS, Appendix D, page 198, "Fixed and random-effects models" and "Class effect models". It is unlikely that a fixed treatment effect NMA model will be appropriate on the basis that we expect heterogeneity because not all studies follow the same protocol and the choice of a fixed or random effects model should not be based on goodness-of-fit criteria. Unless there is evidence of updating from prior

distributions to posterior distributions, then posterior distributions are unlikely to represent reasonable posterior beliefs because prior distributions do not represent reasonable prior beliefs.

a) Please consider the impact of the choice of prior distributions that are used in the analyses given the model and the amount of sample data available.
b) Please justify the use of a class effect model and the prior distribution used for the between class standard deviation given classes that include only one, two, two and one treatments. (An assumption that the variability between treatments within classes is the same across classes will not overcome the problem that there are few treatment effects estimated within class).
c) Please provide summaries of the posterior distributions of all between study and between class standard deviation parameters.

Pfizer response:

[redacted]. Forest plots are presented in **Figure 7** and **Figure 8**, and mean ranks (with comparison to base case) in **Table 9**. There [redacted]

[redacted]Figure 7 ISGA 0/1 up to 6 weeks using FE-RCE-VR model but with Uniform(0,1) prior on between class standard deviation.

Figure 8 ISGA 0/1 up to 6 weeks using FE-RCE-VR model but with Uniform(0,2) prior on between class standard deviation.

Table 9 Mean ranks for ISGA 0/1 using FE-RCE-VR model and alternative priors on the between class standard deviation

	Mean rank using Uniform(0,5) prior: base case	Mean rank using Uniform(0,1) prior	Mean rank using Uniform(0,2) prior
Vehicle	[redacted]	[redacted]	[redacted]
tacrolimus 0.03%	[redacted]	[redacted]	[redacted]
tacrolimus 0.1%	[redacted]	[redacted]	[redacted]
pimecrolimus 1%	[redacted]	[redacted]	[redacted]
Crisaborole 2%	[redacted]	[redacted]	[redacted]

(b) We do not assume a class effect on crisaborole. The class effect is only TCIs: pimecrolimus 1%, tacrolimus 0.03%, and tacrolimus 0.1%. This gives three treatment contrasts on which to estimate the between class standard deviation. [redacted] we have provided them below for all models explored for the full ISGA 0/1 dataset in **Table 10** and for age/severity subgroups in **Table 11**. [redacted]

Table 10 Table 32 from Document B: Comparison of model fit statistics for ISGA/IGA 0/1. Number of data points 18

Model			**Between studies SD mean (95% CrI)**	**Between class SD mean (95% CrI)**	**Totresdev**	**DIC**
Across studies fixed or random treatment effect	**Across treatment class (vehicle, TCI, PDE4 inhibitors) effect**	**Vehicle response regression**				
Random treatment effect	No class effect	No	[redacted]		[redacted]	[redacted]
Random treatment effect	Fixed class effect	No	[redacted]		[redacted]	[redacted]
Random treatment effect	Random class effect	No	[redacted]	[redacted]	[redacted]	[redacted]
Fixed treatment effect	No class effect	No	[redacted]		[redacted]	[redacted]
Fixed treatment effect	Fixed class effect	No	[redacted]		[redacted]	[redacted]
Fixed treatment effect	Random class effect	No		[redacted]	[redacted]	[redacted]
Random treatment effect	No class effect	Adjusted for vehicle response	[redacted]		[redacted]	[redacted]
Random treatment effect	Fixed class effect	Adjusted for vehicle response	[redacted]		[redacted]	[redacted]
Random treatment effect	Random class effect	Adjusted for vehicle response	[redacted]	[redacted]	[redacted]	[redacted]
Fixed treatment effect	No class effect	Adjusted for vehicle response			[redacted]	[redacted]
Fixed treatment effect	Fixed class effect	Adjusted for vehicle response			[redacted]	[redacted]
Fixed treatment effect	**Random class effect**	**Adjusted for vehicle response**		[redacted]	[redacted]	[redacted]
FE inconsistency – random class effects – adjusted for baseline risk			[redacted]		[redacted]	[redacted]

Table 11 Table 41 of Document B: Assessment of model fit in subgroups of ISGA/IGA 0/1 evidence networks.

Subgroup	Model	Total residual deviance	Number of datapoints	DIC	Between class SD mean (95% CrI)
Moderate AD	FE-RCE-VR	[redacted]	[redacted]	[redacted]	[redacted]
	FE-RCE-VR inconsistency	[redacted]	[redacted]	[redacted]	
Mild AD	FE-RCE-VR	[redacted]	[redacted]	[redacted]	[redacted]
	FE-RCE-VR inconsistency	[redacted]	[redacted]	[redacted]	
Children with AD	FE-RCE-VR	[redacted]			
	Random effects	[redacted]			
	Fixed effects	[redacted]	[redacted]	[redacted]	
	FE inconsistency	[redacted]	[redacted]	[redacted]	
Adults with AD	FE-RCE-VR	[redacted]			
	Random effects	[redacted]			
	Fixed effects	[redacted]	[redacted]	[redacted]	
	FE inconsistency	[redacted]	[redacted]	[redacted]	
Moderate AD in children	FE-RCE-VR	[redacted]	[redacted]	[redacted]	[redacted]
	FE-RCE-VR inconsistency	[redacted]	[redacted]	[redacted]	
Mild AD in children	FE-RCE-VR	[redacted]	[redacted]	[redacted]	[redacted]
	FE-RCE-VR inconsistency	[redacted]	[redacted]	[redacted]	

A23. CS, Appendix D, page 234. There are 3 assumptions that could be made regarding regression parameters for potential treatment effect modifiers: identical, different or related. An assumption that they are identical is a strong assumption. The assumption has been made as a consequence of limited sample data. Please discuss the plausibility of this assumption and consider the alternative approaches including weakly informative prior distributions for model parameters.

Pfizer response:

[redacted]

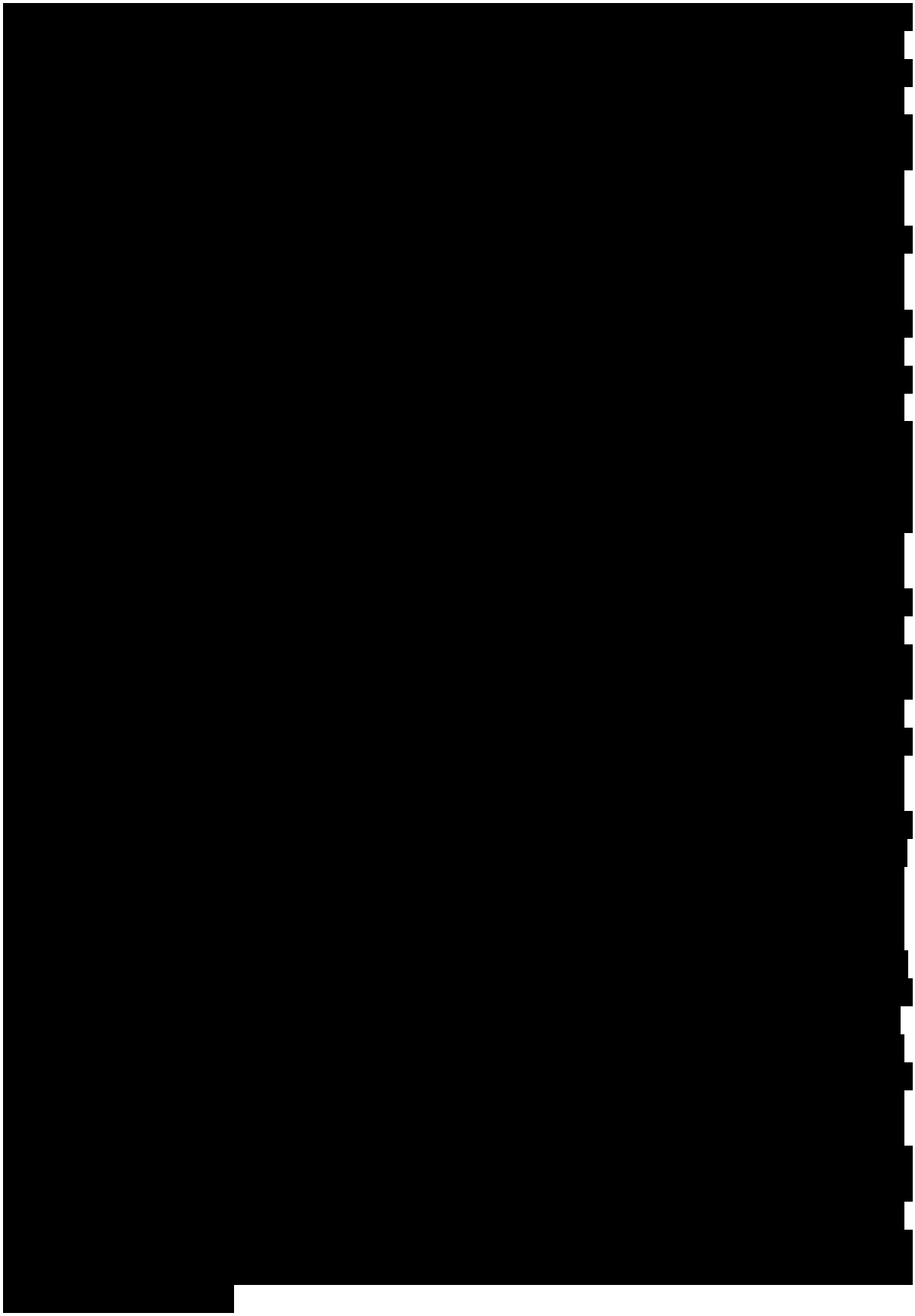


Figure 9 Forest plot of base case FE-RCE-VR NMA with ‘independent’ regression coefficients and informative priors. Estimated hazard ratios (95%

CrI) with P-best (probability crisaborole superior) of ISGA/IGA 0/1 of comparators versus crisaborole.

Figure 10 Forest plot of base case FE-RCE-VR NMA with 'identical' regression coefficients and non-informative priors. Estimated hazard ratios (95% CrI) with P-best (probability crisaborole superior) of ISGA/IGA 0/1 of comparators versus crisaborole.

Table 12 Comparison of mean ranks on ISGA/IGA 0/1 outcomes in overall population using the base case 'identical' regression coefficients with non-informative priors and the alternative 'independent' regression coefficients with informative priors.

	Mean rank from base case FE-RCE-VR NMA with 'identical' regression coefficients and non-informative priors	Mean rank from FE-RCE-VR NMA with 'independent' regression coefficients and informative priors
Vehicle	[redacted]	[redacted]
tacrolimus 0.03%	[redacted]	[redacted]
tacrolimus 0.1%	[redacted]	[redacted]
pimecrolimus 1%	[redacted]	[redacted]
Crisaborole 2%	[redacted]	[redacted]

A24. CS, Appendix D. Please provide summaries of the posterior predictive distributions of treatment effect and baseline response for each outcome.

Pfizer response: The majority of the treatment effects were estimated using fixed treatment effects models, which assume no variation across studies in the observed treatment effect. In a fixed effects meta-analysis, the posterior predictive distribution

is identical to the posterior Fixed treatment effects were used for the FE-RCE-VR (fixed treatment effects, random class effects, with vehicle response regression) model used in the base case ISGA/IGA 0/1 analysis and majority of subgroup and sensitivity analyses.

The only analysis for which random effects were employed is the sensitivity analysis on ISGA/IGA 0/1 up to 6 weeks using a simple random treatment effects model (no class effects or vehicle response regression) presented in Appendix D. We provide a forest plot with mean and 95% credible intervals from both the posterior and posterior predictive distribution for this analysis in Figure 11. The difference is that

Figure 11 ISGA/IGA 0-1 up to 6 weeks, base case, simple random effects model with posterior and posterior predictive means, 95% credible intervals, and Bayesian probabilities of being best.

The baseline response models employed a random effect on the mean response and we have generated the posterior predictive distribution in **Table 13**.

Table 13 Estimated baseline log hazard ratio (i.e. on vehicle) for ISGA/IGA 0/1 base case and subgroups, and for safety outcomes with posterior predictive distribution results alongside posterior distribution results.

	Mean (95% CrI)	Predicted mean (95% CrI)	Mean age	Log hazard ratio age	Mean proportion moderate disease (%)	Log hazard ratio proportion moderate
ISGA/IGA 0/1 base case	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	Subgroup analyses (no adjustment for age or severity as targeted analyses)					
ISGA/IGA 0/1 moderate	[redacted]	[redacted]				
ISGA/IGA 0/1 mild	[redacted]	[redacted]				
ISGA/IGA 0/1 moderate children	[redacted]	[redacted]				
ISGA/IGA 0/1 mild children	[redacted]	[redacted]				
	Safety outcomes (no adjustment for age or severity as not expected to vary)					
Overall AEs	[redacted]	[redacted]				
Overall	[redacted]	[redacted]				

withdrawals	[redacted]	[redacted]				

A25. CS, Appendix D. Please clarify whether it is the posterior predictive distributions of treatment effect and baseline response that have been included in the economic model. If not, please provide results of economic analyses using these distributions.

Pfizer response:

[redacted]. A scenario analysis using the posterior predictive distributions has been presented in Appendix 3 scenario 9.

Section B: Clarification on cost-effectiveness data

Model structure

B1. Priority Question. CS, Section B.3.2.2. The current base-case model assumes that patients unable to receive topical corticosteroids (TCSs) can be offered either topical calcineurin inhibitors (TCIs) or crisaborole and if these fail, then patients move on to subsequent therapies (systemic therapy and / or phototherapy).

a) Please clarify why the model does not allow patients to have TCIs as subsequent therapy when their condition fails to respond to crisaborole.
b) Similarly, please clarify why the model does not explore the use of crisaborole in patients whose condition has failed to respond to TCIs.
c) Please provide an incremental cost-effectiveness analysis comparing the following strategies in the population unable to receive TCSs;
    i) crisaborole only (as currently modelled)
    ii) TCIs only (as currently modelled)

iii) crisaborole followed by TCIs

iv) TCIs followed by crisaborole

In each case, the sequence above should be followed by those subsequent therapies such as systemic and / or phototherapy that are considered relevant to the population.

Pfizer response

a) The model was designed to assess crisaborole versus existing standard care in a second line (post-TCS) indication. A cohort model approach based on disease states has been used in order to keep the modelling approach simple and parsimonious, whilst addressing the key decision problem.

[redacted] The model has consequently not tracked individual patients and has not been specifically designed to evaluate an optimal sequence of therapies (i.e. 1st, 2nd, 3rd and subsequent lines of treatment), but only consider (separately) which treatment would be the most cost-effective in a 1st line (TCS eligible) or 2nd line (post-TCS) indication for mild and moderate patients.

The economic analysis consequently assumes that following failure of a 2nd line treatment with either a TCI or crisaborole, therapy is stepped up and a proportion of the cohort are modelled to receive phototherapy or systemic therapy (proportions for mild and moderate disease have been modelled using evidence from the UK BAD audit data). The same basket of subsequent therapies has been assumed for both TCIs and crisaborole, for simplicity and to avoid bias in modelled costs. This approach has been taken to avoid clouding the decision problem and given the lack of data to inform efficacy assumptions for 3rd line treatment.

b) See above

c) Nevertheless, in order to help NICE evaluate the separate decision problem of what is the optimal sequence of therapies, we have run an additional sensitivity analysis examining a scenario in which crisaborole may be used after TCIs and vice versa. It is noted that we have had to assumed that treatment efficacy is unchanged by the sequence of treatments. In these scenarios crisaborole and TCIs are included in the basket therapies

considered following the second-line treatment. Results are reported in Appendix 3 scenario 1.

B2. Priority Question. CS, Sections B.1.3.3.1 and B.1.3.3.4. TA82 states that "topical tacrolimus and pimecrolimus are not recommended for the treatment of mild atopic eczema" and the NICE guideline for atopic dermatitis in children (CG57) recommends that "systemic treatments and phototherapy are used for the treatment of severe atopic eczema in children when other management options have failed or are inappropriate". Table 6 and Figure 2 of the company submission also suggest that emollients and TCSs are the only treatment options available in mild atopic dermatitis. The NICE final scope also lists TCSs and emollients as the only comparators in mild disease. However, a comparison against emollients in patients where TCSs are not appropriate is not provided in the company submission. Please provide a comparison of crisaborole against emollients in patients with mild AD who have not been adequately controlled by TCSs or where there is a serious risk of important adverse effects from further topical corticosteroid use.

Pfizer response: Section 1.5.1.1 of NICE Clinical Guidance CG57 states "healthcare professionals should use a stepped approach for managing atopic eczema in children. This means tailoring the treatment step to the severity of the atopic eczema. Emollients should form the basis of atopic eczema management and should always be used, even when the atopic eczema is clear." In clinical practice, this recommendation is also applicable to adults (this has been confirmed by our clinical expert). Taking this into account, our assumption is that all patients receiving appropriate clinical care by an HCP should be receiving emollients as a basic therapy plus an additional medical treatment tailored to the severity of their AD. Given the above, a patient with a clinical diagnosis of "mild AD who has not been adequately controlled by TCSs or where there is a serious risk of important adverse effects from further topical corticosteroid use" should be already receiving emollient therapy (basis of atopic eczema management). Furthermore, despite using emollient treatment such a patient would be in a symptomatic state with visible signs, itch, potentially sleep disruption, and decreased quality of life (this assumption has also been confirmed by our clinical expert). The previous understanding of the management of AD in real clinical practice, makes Pfizer believes that even if NICE question is relevant and reasonable from an academic standpoint, it is extremely

unlikely that patients with mild AD who have not been adequately controlled by TCSs or where there is a serious risk of important adverse effects from further topical corticosteroid will be managed with additional emollients only and kept with uncontrolled symptomatic disease. In this specific population the most likely treatment pathway would be to prescribe a different therapy. The previous statements and assumptions are also supported by UK real world data from the British Association of Dermatologists (BAD) National Clinical Audit Programme 2015: Atopic eczema in children (NICE CG57), which has shown that patients with mild AD, as diagnosed at the last appointment, have been treated with more potent therapies, including TCIs. Overall, it is clear emollients are not a relevant comparator in this and due to this, a comparison with emollients alone in patients that have discontinued TCS is deemed inappropriate and has not been undertaken.

B3. Priority Question. CS, Section B.3.2.2. In its model, the company appears to assume that patients who do not experience an adequate response to treatment and who therefore move on to subsequent treatment continue to have the same disease severity as they had at the start of the model. However, this appears to be inconsistent with the subsequent treatments offered. The NICE guideline for the treatment of atopic eczema in children (CG57) states that “Healthcare professionals should consider phototherapy or systemic treatments for the treatment of severe atopic eczema in children when other management options have failed or are inappropriate and where there is a significant negative impact on quality of life.” This would suggest that the disease would need to be classed as severe for patients to progress to systemic therapy and there would need to be a corresponding decrease in quality of life. However, those on systemic therapies in the model have quality of life scores equivalent to mild or moderate disease. Please clarify whether the subsequent treatment states in the model are meant represent the use of systemic treatments in patients with mild to moderate disease or whether it is meant to represent the disease progression to severe atopic dermatitis in a proportion of patients with mild to moderate atopic dermatitis at baseline.

Pfizer response:

We have not assumed disease progression in the cost-effectiveness analysis and have assumed all patients remain in the original disease state for model simplicity and parsimony. UK audit data (BAD National Clinical Audit Programme 2015: Atopic eczema in children NICE CG57) data suggests that in UK clinical practice, a small proportion of mild patients classified as having “mild” disease at the last appointment, receive therapeutic options that are currently NICE recommended for patients with more severe disease, a similar pattern is seen in moderate disease, see Figure 12 and Figure 13. Pfizer recognises that use of these more potent treatment options is not consistent with NICE guidelines, however, in the base case analysis, Pfizer has proposed a model aligned with UK observed clinical evidence. The base case economic model consequently assumes that in the 8.26% of mild patients, who may no longer be suitable for TCS, 13% may receive systemic therapies. This means that overall approximately 1.1% of the total mild patient population have been assumed to receive more potent treatments (i.e. systemic therapies), which appears broadly consistent with the BAD audit data (slightly more conservative). Similarly, the economic model assumes that in the 12.65% of moderate patients, who may no longer be suitable for TCS, 13% may receive systemic therapies. This means that it has been assumed overall approximately 1.6% of the total moderate patient population could receive more potent treatments (i.e. systemic therapies), again which appears broadly consistent with the NICE audit data (again slightly more conservative).

It is noted that UK expert opinion indicated that a treating clinician may also simply re-classify a symptomatic mild patient who had failed TCS (and TCI) as a moderate patient purely due to the failure of standard care therapies and without any change in the objective baseline disease severity based on symptoms and extension/severity of the skin lesions. This reclassification would enable the treating physician to escalate therapy to more potent treatment options. In this scenario, again it would not be assumed that the patient’s clinical symptoms have progressed perse. Whilst, this re-classification, has not been expressly captured in the economic model, this advice supports an assumption regarding escalation of treatment for patients who have failed standard care therapies.

It is further noted that

[redacted]

It is noted that whilst, clinically a patients' disease may be classed as mild to moderate, disease symptoms associated with mild to moderate disease are quite burdensome for the patient and family (particularly in children). The median HRQoL loss in young adults with mild AD are thought to be similar to losses in young adults with chronic sinusitis, while those with moderate AD experienced losses similar to migraine, asthma, upper spinal problems and lower back disorders. The intense itching (pruritis) associated with even mild AD, can result in potentially severe sleep disturbances and associated psychological distress. Children with AD exhibit higher levels of irritability, mood changes, and sleep loss.(17) Sleep disruption, in particular, has a strong negative impact on HRQoL and can precipitate other AD-related co-morbidities.(18, 19) AD-related sleeping problems early in life are associated with an increased risk for subsequent diagnoses of attention deficit hyperactivity disorder (ADHD), and mental health disorders such as anxiety and depression.(18, 20-22). Overall, it is not considered clinically plausible that an 'uncontrolled' mild or moderate AD patient, who has failed previous standard treatments, would be routinely left uncontrolled on emollients alone and it is expected that treatment would be escalated in these patients in order to address the important HRQoL loss that occurs in symptomatic patients.

Nevertheless, in recognition that systemic treatments are not currently NICE recommended in mild patients, a sensitivity analysis has been undertaken which assumes that mild patients, who have failed both TCS and TCIs (pimecrolimus), receive emollient treatment alone. In this scenario, it is assumed that GP visits increase by 50%, due to uncontrolled disease which is not being actively treated, but that dermatologist visits decrease by 50% (since more potent therapies are not being prescribed). The results of this analysis may be found in Appendix 3 Scenario 11.

Figure 12 British Association of Dermatologists National Clinical Audit Programme 2015: Breakdown of treatment regimen by Mild Disease

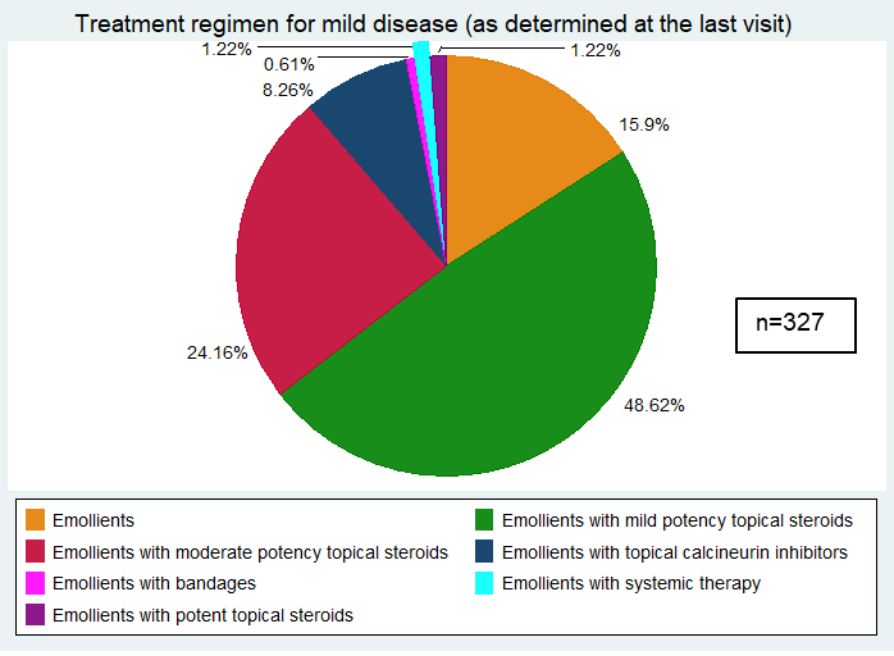
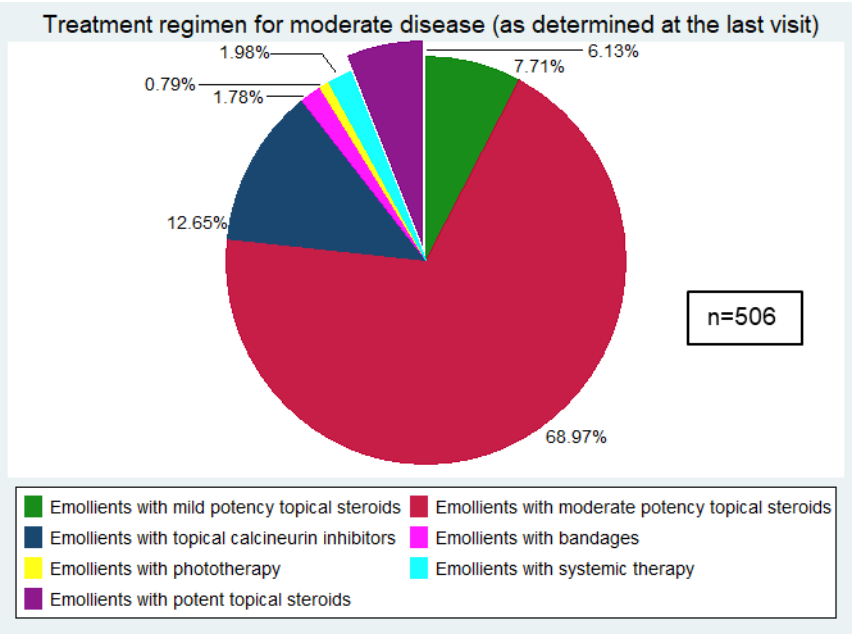


Figure 13 British Association of Dermatologists National Clinical Audit Programme 2015: Breakdown of treatment regimen by Moderate Disease



B4. Priority Question. CS, Section B.3.2.2. Please clarify why patients whose condition does not respond to subsequent therapy are not able to have a partial response, that is, a patient with mild atopic dermatitis receiving subsequent therapy can move either to the controlled state if they respond or they can stay in the moderate state if they do not respond but they are unable to move to the mild state following subsequent therapy.

Pfizer response: The question is incorrect in the statement "a patient with mild atopic dermatitis receiving subsequent therapy can move either to the controlled state if they respond or they can stay in the moderate state if they do not respond". To clarify:

- A patient with mild atopic dermatitis receiving subsequent therapy can move either to the controlled state if they respond or they can stay in the mild state if they do not respond
- A patient with moderate atopic dermatitis receiving subsequent therapy can move either to the controlled state if they respond or they can stay in the moderate state if they do not respond

The assumption that patients with moderate atopic dermatitis receiving subsequent therapy were not eligible for partial response was a simplifying assumption made in the absence of alternative data, as this was not provided in TA82. In order to address this concern, a sensitivity analysis was performed in which patients with moderate atopic dermatitis receiving subsequent therapies are eligible for partial response, with the proportion of non-responders achieving a partial response conservatively assumed to be equal to that of crisaborole, tacrolimus, and pimecrolimus (and reported in B.3.3.1.1 of the original submission). Results are reported in Appendix 3 scenario 2.

B5. Priority Question. CS, Section B.3.2.2. Please explain why outcomes beyond age 18 are not relevant to the committee's decision making, particularly for older children starting treatment with crisaborole or TCIs in adolescence. Please provide scenario analyses exploring 18-year time horizons for older children, for example, those starting treatment at age 12 or 15. Alternatively, please explain why this cannot be done.

Pfizer response: The economic evaluation considers a time horizon of up to age 18 for children and lifetime for adults. Beyond the age of 18, patients are considered adults, and the outcomes for these individuals are evaluated within the adult analyses. This was considered more appropriate than modelling children becoming adults, because adult patients are exposed to a different clinical pathway and are subject to differential rates of efficacy.

A scenario analysis was performed with a starting age of 12 and time horizon of 6 years. Results are reported in Appendix 3 scenario 3.

B6. CS, Section B.3.2.2. Please explain why a starting age of 18 has been assumed for adults. Please clarify the average age of adults in the AD-301 and AD-302 trials. If the model results are not expected to be sensitive to the starting age for adults, please provide scenario analyses exploring the impact of different starting ages in adults to demonstrate this.

Pfizer response: A starting age of 18 has been used for adults in order to capture the costs and effects of treatment over a lifetime. While the average age of adults with AD in the clinical trial may be greater than this, AD typically presents in childhood and the majority of adults would be eligible for treatment at 18. The mean age of adults in AD-301 and -302 has not been reported, however the mean age of adults in AD-303 was 34 years. A scenario analysis has been performed using this as the starting age for adults. Results are reported in Appendix 3 scenario 4.

B7. Please clarify why patients are assumed to have a fixed weight from age 18 onwards. Please conduct a sensitivity analysis exploring the impact of using average weights by age band for patients over the age of 18. The ERG suggests using data on average weight by age band from a national source such as the Health Survey for England.

Pfizer response: Mean patient weight is only used to calculate the dose, and ultimately cost, of subsequent systemic therapies. Consequently, weight is not a significant driver of model outcomes in adults and therefore a constant weight was assumed. This is consistent with the approach in TA82, which assumed constant weight (other economic evaluations did not report weight as a model input) (23). Nevertheless, we have identified the data referred to in the question, which is presented in **Table 14**. These values have been used in the updated base-case analysis.

Table 14: Average weight by age band from the Health Survey for England

Minimum age	Average weight – men (kg)	Average weight – women (kg)	Weight average weight (kg)
18	75.64	65.60	70.05
25	83.98	69.49	75.92
35	87.26	72.38	78.97
45	88.67	75.25	81.20
55	88.01	73.94	80.18

65	85.75	72.01	78.10
75	79.68	67.98	73.17

Response to treatment

B8. CS, Section B.3.3.1, page 125. The company submission states that "the base-case analysis uses estimated response probabilities from the NHM [natural history model] with the age and severity parameters set to their average population values." Please clarify what values were used in the base-case for age and severity (% moderate) and from where these figures were taken. For example, were they based on the average across the two pivotal crisaborole trials, all trials in the network meta-analysis or other sources?

Pfizer response:

[redacted]

B9. Priority question. CS, Section B.3.3.1.1. Please clarify whether patients whose condition partially responds are assumed to respond to the second cycle of treatment with a probability equivalent to the likelihood of response during the first cycle. If so, please describe the trial results which support this assumption. For example, please calculate the proportion of patients with a partial response at 28 days who then responded by 56 days.

Pfizer response: The probability of response in the second cycle in patients achieving partial response was assumed to be the same as the probability of response in the first cycle. This assumption was made in the absence of clinical data to inform alternative assumptions.

Two scenario analyses are presented in which the probability of response in the second cycle in patients achieving partial response is altered. In the first scenario the

probability of response is increased by 20% and in the second it is decreased by 20%. Results are reported in Appendix 3 scenario 10.

B10. CS, Section B.3.3.1 and B.3.3.3. The distributions in Table 56 of mild, moderate and potent TCSs do not appear to match the data in the powerpoint slides available on the British Association of Dermatologists (BAD) website. It is unclear whether the proportion having potent TCS has been taken in error from the data for TCI usage as this gives the reported proportions in Table 56. This error does not appear to have affected the data in Table 63. Please double check the data in Table 56 against the BAD audit data. If incorrect, please correct the model and cost-effectiveness results.

Pfizer response: The proportion of patients receiving mild/moderate/potent TCS is calculated by reweighting the respective proportion reported in the BAD audit data by the total proportion of patients who received TCS (calculated by summing the proportions receiving mild/moderate/potent TCS).

Upon review of the BAD audit data, we agree that the proportion of patients with moderate disease receiving mild/moderate/potent TCS has been calculated incorrectly in the original submission. **Table 15** presents revised estimates. The correct estimates are incorporated within the revised base-case analysis, with results reported in Appendix 3.

Table 15: Proportion of patients receiving mild, moderate and potent TCS in mild and moderate disease

	Mild TCS	Moderate TCS	Potent TCS
% of patients - Mild disease	66%	33%	2%
% of patients - Moderate disease	9%	83%	7%

B11. CS, Sections B1.3.1 and B.3.3.1. The response rate for subsequent therapies is applied as an average response rate across the range of subsequent therapies used. This does not take into account that these therapies will be used sequentially

as shown Figure 1 of the company submission. Please correct this in the model or calculate the size and direction of any bias introduced by this approximation.

Pfizer response: To clarify, the response rate for subsequent therapies is not calculated as a simple naïve average response rate across subsequent therapies; rather, the probability of response used within the economic evaluation is based on a weighted average of response probabilities (see also response to question B12) and the distribution of subsequent therapies (see also responses to question B19 and B41). This probability has therefore been updated in the revised base-case using data presented in response to question B19.

Subsequent therapies are applied as a weighted basket of costs and efficacy. This is assumed to represent the overall use of therapies which might be expected in a cohort of patients who are unable to achieve disease control. It is effectively assumed that at any given time, a proportion of patients would be treated with some form of therapy while symptoms persist. This simplifying approach was adopted in order to permit the development of a Markov cohort model without the requirement to track individual subsequent therapy use over time. See also response to question B15.

B12. CS, Section B.3.3.1.2. Please clarify why the response rates for second-line treatments in Table 59 do not match either the figures or the associated descriptions of the disease controlled mentioned in Table 24 of TA82 (Garside et al. 2005).

Pfizer response: We believe that this is a consequence of rounding within the figures presented in Table 59. The probability of response to UV therapy was assumed to be the same as systemic therapy. A revised table is presented in **Table 16**.

Table 16: Response rates for second-line treatments

	Response rate - Mild disease	Response rate - Moderate disease
Tacrolimus 0.03%	0.385	0.385
Tacrolimus 0.1%	0.374	0.374
Pimecrolimus	0.249	0.249
Systemic treatment	0.7	0.7
UV therapy	0.7	0.7

Footnotes: AD, atopic dermatitis; UV, ultraviolet

Upon review, we have noted that the data for tacrolimus 0.1% and 0.03% were labelled incorrectly within the electronic model. This is updated in the revised base-case version but does not affect the base-case ICER.

B13. CS, Sections B.2.7.1 and B.3.3.1.1. Please clarify why the partial response rates used in the model does not match those reported in Table 27 of the company submission. Please clarify whether it is because Table 27 includes those moving from a score of 2 to 1 who would be considered 'responders' in the model (in addition to those moving from 3 to 2), whereas the rates in section B.3.3.1.1. only includes those moving from a score of 3 to 2.

Pfizer response: The partial response rates reported in B3.3.1.1 of the company submission and used in the electronic model are based on the proportion of patients with moderate disease at baseline in AD-301 and AD-302 who achieved ISGA2, of those who did not achieve ISGA 0-1.

The assertion within the question is correct; Table 27 includes those moving from an ISGA score of 2 to 1 who would be considered responders in the model, and additionally includes patients with mild disease.

Table 17: Shift table of ISGA at Baseline versus at Day 29

		ISGA at Day 29 N (% of total population)				
Treatment	**Baseline ISGA**	**0 - Clear**	**1 - Almost Clear**	**2 - Mild**	**3 - Moderate**	**4 - Severe**
Crisaborole	**0 - Clear**	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	1 - Almost Clear	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	2 - Mild	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	3 - Moderate	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	4 - Severe	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Vehicle	**0 - Clear**	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

	1 - Almost Clear	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	2 - Mild	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	3 - Moderate	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	4 - Severe	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Abbreviations: ISGA, Investigator's Static Global Assessment. Note: Study cohorts from AD-301 and AD-302 were pooled

Treatment withdrawal and subsequent therapies

B14. Priority Question. CS, Section B.3.3.3. Please clarify why the costs for subsequent therapies are only applied in the first cycle of subsequent treatment for patients whose condition responds to subsequent therapy. Preliminary discussions with the ERG's clinical experts suggest that patients may take longer than 4 weeks to achieve a response to systemic therapy and they would then remain on that therapy to achieve a period of stable disease before having their dose gradually reduced to the lowest possible without causing the disease to flare. Please provide evidence that patients whose condition responds to subsequent therapy (systemic or phototherapy) within 4 weeks are then immediately stopped treatment. If no evidence is available, please provide scenario analyses exploring alternative plausible durations of subsequent therapy.

Pfizer response: The statement is correct in that a patient who achieves controlled disease with subsequent therapy is assumed to only receive treatment for one cycle (4 weeks). This approach was adopted as a simplifying assumption. Within the context of a Markov cohort model, implementing the tracking of time in receipt of individual subsequent therapies was considered impractical. Additionally, the data used to inform response rates for systemic therapies are taken from TA82 and represent the best estimate of the proportion of patients achieving disease control with 4 weeks of treatment.

In order to explore the effect of patients continuing to receive treatment beyond 4 weeks, several scenario analyses have been included where patients may remain on therapy beyond flare resolution. According to Garibaldinos et al the average duration of treatment with phototherapy is 3 months.(24) The average treatment duration with various systemic therapies was taken from Taylor K, et al and is presented in **Table 18** below.(25) In the first scenario analysis the cost of subsequent therapy is applied as a one-off cost at the start of subsequent treatment, with treatment durations taken from **Table 18**. However, this approach will underestimate the cost of subsequent treatment, as it will not capture patients going on to require further therapy.
Two further scenarios consider systemic treatments to be taken continuously and phototherapy to be received for 3 months per flare. In the second of these the flare

rate is assumed to be halved, as patients may experience greater disease control with systemic treatments.

Table 18: Average treatment duration and per-cycle probability of discontinuation

Therapy	Average treatment duration	4 week probability of discontinuation
Phototherapy	3 months	0.26
Ciclosporin	5.8 months	0.15
Azathioprine	13.8 months	0.06
Methotrexate	15.1 months	0.06

The results of these scenarios are presented in Appendix 3 scenario 5.

B15. Priority Question. CS, Sections B.3.3.3 and B.3.5.2.

a) Please clarify what happens to patients whose condition does not adequately respond to subsequent therapy. Are they assumed to remain on the same treatment (ciclosporin or phototherapy) until their condition responds or are they assumed to switch between different treatments? If they are assumed to stay on the same treatment, how does this compare with the maximum duration of therapy for subsequent treatments, in particular the nephrotoxicity of ciclosporin?

b) In its model, the company assumes that only 50% of 'non-responders' to subsequent therapies (who remain on subsequent treatment until a response is achieved) have a dermatologist consultation every 4 weeks. Please clarify whether 4 weeks is adequate given the monitoring requirements for ciclosporin.

Pfizer response: Patients who do not adequately respond to subsequent therapy are assumed to receive the basket of subsequent therapies until resolution of atopic dermatitis (children only), or death. This is assumed to represent the overall use of therapies which might be expected in a cohort of patients who are unable to achieve long-term disease control. It is not assumed that individuals would remain on a single therapy indefinitely; rather, it is effectively assumed that individuals would remain on some form of therapy while symptoms persist. This simplifying approach was adopted in order to permit the development of a Markov cohort model without the requirement to track individual subsequent therapy use over time.

Berth Jones et al provide guidelines on monitoring use of ciclosporin in dermatology. They state that in the first 2 months, it is recommended that serum creatinine and blood pressure be measured at fortnightly intervals followed by monthly measurements thereafter. More frequent monitoring is indicated should there be a rise in blood pressure or creatinine. Less frequent monitoring is sometimes acceptable in cases where clinicians are confident that the risks are low (e.g. in otherwise healthy young patients with minimal previous exposure to NSAIDs or other nephrotoxic drugs). Monitoring at 2–3-monthly intervals is adequate when these parameters appear to be stable on long-term treatment after 4 months.

This does correspond to greater monitoring than has been applied in the model base-case, however the progressive nature of the monitoring does not align well with the structure of a Markov model. In the updated base-case, in order to capture the long-term impact systemic therapies can have on monitoring costs, we have assumed that patients would visit a dermatologist upon treatment failure and then every 2 months thereafter, with any additional monitoring occurring in primary care.

B16. Priority Question. CS, Section B.3.3.1.2. The probability of response to subsequent treatment (70%) appears to be applied repeatedly for each cycle of subsequent treatment in those who do not respond to the first cycle.

a) Please clarify how this response rate was calculated from the source study with particular reference to the time period that the response rate relates to.
b) Please clarify whether the source study showed that an additional 70% of 'non-responders' to the previous cycle would respond every 4 weeks (that is, a constant hazard of response over several cycles of treatment).

Pfizer response: Please refer also to the response to question B11. The probability of response to subsequent therapies is calculated as a weighted average of response probabilities taken from TA82 (see also response to question B12) and the distribution of subsequent therapies (see also responses to question B19 and B41). This probability has therefore been updated in the revised base-case using data presented in response to question B19. The response probability is therefore not derived from a single study or source. However, it should be noted that in TA82 they do not make any adjustments to response probabilities due to failing previous lines

of therapy. We recognise that the represents a simplifying assumption. This assumption is conservative for crisaborole, as it assumes no loss of efficacy for subsequent treatment. The calculation of this probability is presented in D25 of the 'transition probabilities' sheet.

B17. Priority Question. CS, Section B.3.5.1.1. In the guidance for dupilumab (TA534), the committee recognised that ciclosporin, methotrexate, azathioprine and mycophenolate mofetil were all part of current clinical management (see section 3.3). Discussions with the ERG's clinical experts suggest that methotrexate is used in preference to ciclosporin because it can be used safely for longer periods.

a) Please justify why the base-case analysis assumes that only ciclosporin is used for systemic treatment.
b) Please conduct sensitivity analyses exploring the impact of incorporating other systemic therapies in addition to ciclosporin within the model.

Pfizer response: Ciclosporin was used in the base-case as a representative cost for consistency with the assumptions of NICE TA82 and because, of the therapies listed, only ciclosporin is licensed in atopic dermatitis (23). Two sensitivity analyses have been performed to explore these assumptions:

1. Using the costs of methotrexate only

2. Basing costs on a weighted basket of systemic therapies, with weights based on the number of patients reported to have been treated with each therapy in the IQVIA-THIN database for the year 2017 (26)

Table 19 presents the distribution of systemic therapies assumed in the second scenario. Weights are derived based on the number of patients receiving any dose or formulation amongst ciclosporin, methotrexate, azathioprine and mycophenolate mofetil. Drug acquisition costs are taken from eMIT (27) where available and the British National Formulary (BNF) (28).

Table 19: Distribution of systemic therapies and respective costs

Treatment	Weight †	Dose	Mg per unit	Units per pack	Cost per pack (£)	Cost per mg	Target mg /day	Cost/day

ciclosporin	[redacted]	1.25 mg/kg twice daily	25	30	£18.37*	£0.02	190	£4.66
methotrexate	[redacted]	10-25 mg once a week	2.5	24	£0.85‡	£0.01	4	£0.05
azathioprine	[redacted]	1-3mg/kg/day	50	56	£1.50‡	£0.00	3	£0.00
mycophenolate mofetil	[redacted]	1-2 g once daily	500	50	£10.27‡	£0.00	2000	£0.82
					Total weighted cost per day			**£0.78**

†Estimated from IQVIA-THIN database for the year 2017 (29)
*Source: British National Formulary 2019
‡Source: eMIT 2019

Results of both analyses are reported in Appendix 3 scenario 6.

B18. Please clarify why the response to subsequent therapy (D26 in the Transition Probabilities sheet) is adjusted in the scenario analyses exploring adjustments to reflect line of therapy, disease severity and age (using the OR in cell I86 of the clinical data sheet) when these adjustments are based on the network meta-analysis which did not include subsequent therapies.

Pfizer response: The adjustments made in the scenario analyses exploring adjustment to reflect line of therapy, disease severity, and age are made using estimates from the respective subgroups in AD-301 and AD-302, not the NMA; this is reported in Table 55 of the original submission. The odds ratios represent the effect for vehicle in the respective subgroup compared to the vehicle population as a whole.

B19. CS, Section B.3.3.3. The British Association of Dermatologists (BAD) audit data (reference 2 in the company submission) which are used to estimate the proportion using different treatments in Tables 62 and 63 reports only the proportion using TCIs and does not report separate figures for pimecrolimus and tacrolimus. Please clarify why an even split between tacrolimus and pimecrolimus is assumed in the model when the NICE guidance for tacrolimus covers a broader group of patients (recommendation for pimecrolimus is limited to face and neck in children aged 2 to 16 years).

Pfizer response: In order to update this assumption for patients with moderate disease, the relative split between pimecrolimus and tacrolimus was estimated from

the IQVIA-THIN analysis, with relative weights based on the number of patients reported to have been treated with each therapy. This was included in the revised base-case. This was performed separately for children and adults. The relative weights for pimecrolimus relative to tacrolimus were 21%:79% and 24%:76% in adults and children, respectively. Revised distributions of therapies used in subsequent lines are reported in **Table 20**, and results of the revised base-case including these data are reported in Appendix 3. A limitation of this analysis is that these data are not severity- or line of therapy-specific. Please note that this exercise was not performed for patients with mild disease, on the basis that tacrolimus is not licensed in this population (see also response to question B41).

Table 20: Probability of starting different treatments having failed the primary therapy in moderate disease (revision of Table 63 in submission)

	Probability - Children	Probability - Adults
Tacrolimus 0.03%	0.38	0.21
Tacrolimus 0.1%	0.24	0.44
Pimecrolimus	0.20	0.17
Systemic therapy	0.13	0.13
Phototherapy	0.05	0.05

B20. CS, Section B.3.3.3. Please clarify whether the data used to estimate the proportion having different subsequent lines of therapy is all taken from the British Association of Dermatologists (BAD) audit in children. If so, please clarify what attempts were made to identify equivalent data for adults.

Pfizer response: The BAD audit was the only available data source identified, and therefore the proportion receiving different subsequent lines of therapy was assumed to be the same between adults and children. Data from the THIN database do show the proportion of patients receiving different therapies, however these are not split by disease severity and do not include data on phototherapy use. Clinical expert advice has indicated that this pathway is not expected to vary by any important degree in adults.

Flare rates

B21. CS, Section B.3.3.2.

a) Please clarify why patients having a flare are assumed to return to their baseline disease severity.
b) Please provide a sensitivity analysis in which a proportion of the flares are more or less severe than the baseline severity.

Pfizer response: The assumption that patients who experience a flare would return to their baseline disease was made in the absence of alternative data. It also of note that as the analysis considers patients with mild and moderate disease to represent separate populations. Therefore, we believe that to include patients with mild disease converting to moderate disease (and conversely from moderate to mild) would render the analysis less useful for decision-making in each population. Additionally, incorporating disease progression into the model would require a far more complex model structure with treatment sequencing incorporated into the model, as such, we do not consider this to be an appropriate scenario for consideration in the model.

[redacted]

B22. CS, Section B.3.6.1. In Table 71, please explain why a log-normal distribution was selected to describe the uncertainty in the annual number of flares.

Pfizer response: The log-normal distribution is a standard distribution used to estimate the uncertainty in a strictly positive parameter not bounded above at 1.(30) This distribution reflects the skew we may expect to see in such data and does not risk generating any negative values, as we may see with the standard normal distribution.

B23. CS, Section B.3.3.2. Please clarify why the number of flares was based on the Pfizer epidemiology report instead of the control arm of the trial conducted for tacrolimus as a maintenance therapy [Wollenberg et al. Proactive treatment of atopic dermatitis in adults with 0.1% tacrolimus ointment. Allergy 2008 Jul;63(7):742-50.]

Pfizer response: The baseline rate of flares was based on the Pfizer epidemiology report on the basis that this might be expected to be more representative of real-world flare rates seen in clinical practice. In addition, Wollenberg et al (31) do not report the mean number of disease exacerbations during the 12-month disease control phase of the trial or mean length of follow-up; rather they present a graph showing grouped numbers of disease exacerbations. Therefore we are unable to use this data directly within the analysis. Additionally the trial was only in adults and included patients with mild, moderate and severe disease but results are not reported by disease severity. This means the Wollenburg data lacks the granularity provided by the Pfizer epidemiology data.

Resolution of atopic dermatitis

B24. CS, Section B.3.3.4. The company submission states that “It is thought around 75% of children outgrow their AD in childhood or early adolescence”. This is used to calculate a rate of response of 0.708% per 4 weeks by converting the 75% response rate to a 4 weekly response rate over 15 years.

a) Please clarify why this rate of resolution is applied life-long in the model rather than being restricted up to age 16.
b) It appears that this error is likely to have minimal impact on the base-case results because the time horizon for children is restricted to 18 years. However, this would significantly affect any scenario analyses exploring a life-time horizon. Please correct the model to limit the rate of resolution to the first 16 years so that alternative time horizons can be explored.

Pfizer response: As discussed in response to question B5, a time horizon up to the age of 18 is considered in the base-case for children, therefore in the base-case the rate of resolution is not applied for life. Although the estimation of the probability of resolution was based on a data point at age 16, it was assumed that resolution in

children would continue to age 18 (in the base-case). The model has been revised to ensure the probability of resolution is not applied beyond the age of 16 in any analysis. Revised results are presented in Appendix 3.

B25. CS, Section B.3.3.4. When calculating the rate of resolution per 4 weeks that is equivalent to 75% achieving resolution by 16 years, please clarify why 15 years is used instead of 14 years given that patients start aged 2.
Pfizer response: For the purposes of this calculation, it was assumed that the statistic applied to patients as they turned 2 until they were no longer 16 (i.e. when they turned 17). Therefore, there were 15 years between these points. A scenario analysis exploring the effect of assuming resolution is achieved in 14 years and does not continue once patients have turned 16 is presented in Appendix 3 scenario 7.

Resource use and costs

B26. CS, Section B.3.5.2. The NICE guideline for the treatment of atopic eczema in children (CG57) states that phototherapy and systemic treatments "should be undertaken only under specialist dermatological supervision by staff who are experienced in dealing with children". Please clarify why patients being managed in secondary care who are receiving subsequent treatments incur costs for both GP and secondary care consultations if their treatment is being managed in secondary care.

Pfizer response: AD is a complex disease and patients require not only medical treatments, but also assessment of psychological and psychosocial wellbeing and quality of life and intensive education about the disease, the treatment options, and the correct way to apply the treatments. This has been highlighted in the NICE guideline for the treatment of atopic eczema in children (CG57):

- Section 1.2.1.1: Healthcare professionals should adopt a holistic approach when assessing a child's atopic eczema at each consultation, taking into account the severity of the atopic eczema and the child's quality of life, including everyday activities and sleep, and psychosocial wellbeing.

- Section 1.6.1.1 Healthcare professionals should spend time educating children with atopic eczema and their parents or carers about atopic eczema and its treatment.

The holistic approach and the time needed for education usually require the joint collaboration of dermatologists, GPs with extended role (GPER) specialised in dermatology, GPs, and nurses.
Additionally, AD is a chronic and relapsing disease with flares that are not predictable and require prompt medical supervision. In the NHS, GPs are the most accessible health care professionals to control this unpredictable worsening of AD, providing flare-control treatments while patients wait for the appointments in secondary care. These potential flares are specially more frequent in more severe disease.

The previously described factors (complexity of the disease with the need of a holistic approach and intensive education and the unpredictable nature of the flares) are probably the main drivers of the utilisation of resources at the secondary and primary care levels by patients with more severe AD.

These double utilisation of resource at the primary and secondary levels at the same time has been also described in the Appraisal consultation document of Dupilumab for treating moderate to severe atopic dermatitis where Sanofi Regeneron presented evidence from a review of the secondary care notes from UK patients showing visits to dermatologists, GPs, and dermatology nurses (slide 28, public presentation, 1st Appraisal Committee meeting – document attached). Additionally, in the Appraisal consultation document (also attached) the following statement is included “One clinical expert explained that, in practice, patients on systemic therapy would generally be seen 3 to 4 times per year by dermatologists and probably more often by GPs, but that the frequency would likely depend on whether the condition was moderate or severe.” (section 3.23, page 20).

The joint collaboration of primary and secondary care in the treatment of patients with more severe atopic dermatitis has also been confirmed by our clinical expert.

B27. CS, Section B.3.5.2, page 136. The company submission states that “It has been assumed that 50% of patients that progress to subsequent therapy will see a dermatologist. However, in Table 70, dermatologist consultations are said to apply to patients “On treatment failure and in 50% of flares thereafter” suggesting that 100% receive a dermatologist consultation on treatment failure. The latter appears consistent with what is actually applied in the model and it would be consistent with the NICE guideline for children (CG57). Please clarify which is the company’s intended assumption.

Pfizer response: The assertion in the question is correct: the assumption in the economic model and reported in Table 70 is the preferred assumption. Specifically, it is assumed that dermatologist consultations occur on treatment failure and in 50% of flares thereafter. As per the response to question B15, this assumption has now been updated.

B28. Priority Question. CS, Section B.3.5.2. The company submission states that all patients have a dermatologist appointment on treatment failure and 50% have a dermatologist appointment during flares. However, in the Excel model, the 50% appears to be applied to all patients having subsequent therapy regardless of whether they started having subsequent treatment due to treatment failure or due to a flare. So the consultation cost is applied to patients entering column Q of the model ‘engine‘ sheets from columns I to N, but the consultation cost is also applied to those in column AD which reflects the proportion residing in the same state during the first cycle after half cycle correction. Please clarify whether there is double counting of the dermatology appointments in those having subsequent therapy because of failure of TCIs or crisaborole due to the application of costs for flares to the same group of patients.

Pfizer response: This is an error in the Excel model. The calculation of dermatologist visits has been updated to reflect the updated assumptions around monitoring discussed in the response to question B15 and patients progressing in a cycle are excluded from the calculation of subsequent dermatologist visits.

B29. CS, Section B.3.5.1.2. In Table 69, the company submission states that the cost per pack of Diprobase cream is £5.32. However, in the Excel model, £6.32 was used. Please clarify which figure is correct.

Pfizer response: The data reported in the model is correct. A 500 gram tube of Diprobase cream is £6.32 according to the current version of the BNF (28).

B30. CS, Section B.3.5.2. In Table 70, please clarify how dermatologists' visits were calculated for both age groups from NHS Reference costs. The ERG checked the source but could not find the stated costs.

Pfizer response: The dermatologist costs detailed in the model are listed in the Total Outpatient Attendances sheet of the NHS reference costs 2017-2018, Service code 330 for adult dermatology (Cell B95) and 257 for paediatric (Cell B57).

Drug dosing

B31. CS, Section B.3.5.1.1. Please clarify whether the average dosing per application shown in Table 68 includes trial participants [redacted] If so, please provide average dosing per application excluding these participants as they are not covered by the draft indication.

Pfizer response: The data presented in Table 68 does include patients with a [redacted] The average dose per application in patients [redacted] in AD-303 is presented for each age group in Table 19. These data have been used to inform a scenario analysis, with results presented in Appendix 3 scenario 8.

Table 21 Drug use per application by age group in patients with [redacted] in AD-303

Age group	Drug used per application (g) (SD)
2-6	[redacted]
7-11	[redacted]
12-17	[redacted]
>=18	[redacted]

B32. CS, Section B.3.5.1.1. Please provide information on the expected average dosing per application for TCIs and TCSs to support the assumption that the amount of treatment used is consistent across all topical treatments.
Pfizer response: The application instructions for crisaborole, TCS and TCIs all state treatments should be spread thinly or sparingly over the affected areas. As such our clinical expert saw no justification for assuming a difference in drug use per application.

Data on mean usage in clinical trials was extracted as part of the systematic literature reviews, however this is inconsistently reported. Approximately a third of trials include some information on usage, however several different outcomes are used, including study days of usage, grams used per day and total drug usage throughout the trial. Since the included trials do not have consistent populations or designs (trials considering prophylactic treatment were most likely to report drug use) a direct comparison across trials is not feasible. However, in the trials that report total or daily drug use the variation in drug use across arms active and control arms was minimal, suggesting little difference is usage patterns according to therapy applied. This is consistent with data from AD 301 and AD 302. **Table 22** contains the relevant extract from the data extraction table.

Table 22: Extract from the data extraction table

Trial ID and Reference				Intervention 1				Intervention 2				Intervention 3			
Trial ID	Trial Name	Author	Year	Drug name	Treatment intended for prophylaxis, prophylaxis + dose intensification for flare or flare treatment	Mean usage (value)	Mean usage units	Drug name	Treatment intended for prophylaxis, prophylaxis + dose intensification for flare or flare treatment	Mean usage (value)	Mean usage units	Drug name	Treatment intended for prophylaxis, prophylaxis + dose intensification for flare or flare treatment	Mean usage (value)	Mean usage units
26	Wahn	Wahn	2002	Pimecrolimus	yes	211.9	study days of usage (69.8%)	control	yes	156	study days of usage (66.3%)				
74	Sigurgeirsson	Sigurgeirsson	2008	pimecrolimus	yes	70.2	days	vehicle	yes	51	days				
79	Murrell	Murrell	2007	pimecrolimus	no	0.67	g (daily)	vehicle	no	0.54	g (daily)				
177	Hoeger	Hoeger	2009	pimecrolimus	no	24.2 (19.5 SD) for head and neck; 45.5 (28.7 SD) in rest of body	g	vehicle	no	26.9 (21.9 SD) for head and neck; 50.7 (39.0 SD) in rest of body	g				
190	Wollenberg	Wollenberg	2008	tacrolimus 0.1%	yes	1.38 (1.59 SD) Median: 0.98 (0.05-9.72)	g/day	vehicle	yes	1.6 (2.47 SD) Median: 0.53 (0.00-11.0)	g/day				
310	Boguniewicz	Boguniewicz	1998	tacrolimus 0.03%	no	94	g	tacrolimus 0.1%	no	86	g	vehicle	no	98	g
816	AD 301	Paller	2016	[redacted]											
816	AD 302	Paller	2016	[redacted]											

Health-related quality of life review

B33. CS, Appendix H. In the review of health-related quality of life (HRQoL) studies (described in Appendix H), please explain why papers were excluded based on the interventions used (eligibility criteria in Table H10). Please clarify why data relating disease severity to quality of life or treatment response to quality of life from a study using a different treatment would not be relevant as a means to validate the HRQoL data derived from the mapping study.

Pfizer Response: Studies were included based on the interventions that aligned with the decision problem. For example, interventions were included if they were topical pharmacological therapies such as emollients, topical corticosteroids (TCSs) and topical calcineurin inhibitors (TCIs). If an intervention was non-pharmacological, a systemic therapy, or a therapy primarily used for infection, the study was not included.

The patient population used to derive the mapping algorithm of the DLQI to EQ-5D utility values represented an international multicentre observational cross-sectional study ranging from benign and malignant skin lesions to chronic inflammatory diseases, with AD representing the second most common diagnosis behind psoriasis.(32) This diverse patient population from a wide range of different European countries gives the mapping algorithm additional strength in terms of universality. However, an algorithm derived from the psoriasis-only population and tested on patients with all other conditions produced results reassuringly similar to the mapping algorithm derived from the full patient population.

B34. CS, Appendix H. In the review of HRQoL studies (described in Appendix H), many studies have been excluded due to not being in a relevant population. However, the reasons why the population is not relevant are not apparent from the titles. For example, Stevens et al. was excluded for not having a relevant population but was used to inform utilities in two of the models described in included papers (NICE TA82 & CADTH crisaborole). Please clarify whether Stevens et al. was inappropriately excluded.

- [**Stevens, K. J.,Brazier, J. E.,McKenna, S. P.,Doward, L. C.,Cork, M. J.** (2005). The development of a preference-based measure of health in children with atopic dermatitis *British Journal of Dermatology*, 153(2), 372-7]

Please check all excluded studies to identify any reporting relevant utility values that could be used to externally validate those used in the economic model.

Pfizer response: The PICOS table included in Appendix H was from a previous draft. Please find the correct PICOS table for the HRQoL SLR below. The Stevens et al. paper was incorrectly categorized as a population exclusion as it was considered a parent/caregiver study. However, as it does not present utility values for parents/caregivers but rather asks parents/givers to provide proxy values for children, therefore, it should have been included on the basis of the PICOS criteria. The utility values presented in the study were nevertheless captured in the HTA search (ref to TA82 and CADTH). However, the utilities were not deemed appropriate for use in the model. As described in Table 53 of the CS, utilities in TA82 come from a range of sources, not all of which are patient-reported. It was determined in the development of the cost-effectiveness model that using mapped data from the clinical trials was more in line with the reference case. Values used in TA82 are used in a scenario analysis in the model.

Category	Inclusion Criteria	Exclusion Criteria
Population	• Children (2 years and older) and adults with a clinical diagnosis of AD (or 'eczema') o All diagnostic criteria for AD and all scales/scores used in assessing disease severity are eligible for inclusion o Data will be sought for the main population (the combined mild to moderate population) and stratified subgroups (mild population, or moderate populations), if reported separately	o Healthy volunteers o Pediatric patients (<2 years)
Intervention or comparators	• Topical pharmaceutical treatments for atopic dermatitis	• Non-pharmacologic treatment for atopic dermatitis in all study arms • Hormone creams • Systemic treatments • Treatments intended primarily for infections (e.g. antibiotics) • Multidisciplinary/combination care (e.g. acupuncture coupled with an emollient)
Outcomes[1]	• HRQoL outcomes (measured via generic or disease-specific tools)	

Category	Inclusion Criteria	Exclusion Criteria
	○ Children's Dermatology Life Quality Index (CDLQI) ○ Dermatology Life Quality Index (DLQI) ○ Patient-Oriented Eczema Measure (POEM) ○ EuroQoL Five Dimensions Questionnaire (EQ-5D) ○ EQ-5DY ○ HUI2 ○ SF36 ○ Presence of mapping algorithm (yes/no)	
Study designs	• All studies that report HRQoL utility values for patients	• Studies reporting utility values for parents/caregivers • Studies solely validating existing HRQoL tools • Studies indirectly estimating QoL outcomes using pooled data • Case reports, expert opinion articles, editorials, letters, narrative (non-systematic reviews) • Commentaries and editorials • Studies with sample size <20 patients (or fewer than 10 per arm) • SLRs and secondary sources of utility values (e.g., model-based economic evaluations) are excluded but will be identified for the purpose of recursive searches
Language	• English abstracts of foreign publications will be considered. Studies printed in a foreign language will be flagged, and their inclusion decided on in conjunction with Pfizer.	• Journal articles and conference abstracts without English full-text
Temporal Limit	01 January 1990 – current	NA
Countries/global reach	None specified	

To ensure no further studies were incorrectly categorised as population excludes, all studies excluded on the basis of population were re-screened by a single reviewer. The re-screening exercise identified one additional paper (Drake et al) which was incorrectly excluded:

Drake, L.,Prendergast, M.,Maher, R.,Breneman, D.,Korman, N.,Satoi, Y.,Beusterien, K. M.,Lawrence, I. (2001). The impact of tacrolimus ointment on health-related quality of life of adult and pediatric patients with atopic

dermatitis *Journal of the American Academy of Dermatology*, 44(1 Suppl), S65-72

This paper, however, does not contain utilities appropriate for use in the cost-effectiveness model. Indeed, the utilities were not considered appropriate for use in TA82. As described in TA82, “Drake and colleagues report separately on adults (aged16 and over), children (aged 5-15) and toddlers (aged 2-4) for those treated with tacrolimus and those treated with vehicle. […] However, only combined categories for those affected “very much”, “a lot” and “a little” compared to those affected “not at all” are reported, so it is not possible to assess the level of change over time.”

All of the other re-screened papers were appropriately excluded. An annotated reference list noting the details of each study exclusion is provided in Appendix 5.

Health-related quality of life

B35. Priority Question. CS, Section B.3.4.5. Please provide more information on the process used to calculate the utility scores in Table 66 from the Dermatology Life Quality Index (DLQI) and Children's Dermatology Life Quality Index (CDLQI) scores. The ERG presumes that the process was as follows:

- Take individual DLQI (or CDLQI in children) scores from both AD-301 and AD-302 from a particular time point in the trial.
- Calculate individual EQ-5D scores using the individual's gender and age at baseline their DLQI (or CDLQI in children).
- Group patients according to their ISGA score at the same time point that their DLQI (or CDLQI in children) score was measured.
- Calculate average EQ-5D scores in each ISGA group using i) only data from adults in the base-case, ii) separate groups for children and adults for column 3 of Table 66, iii) all data regardless of age for column 6 of Table 66.

Please confirm if the above description is accurate. Please clarify:

a) From which trial time point the DLQI scores used in the mapping were taken.
b) What age was applied in the mapping algorithm when calculating scores for children.
c) Whether any adjustment was made to convert CDLQI scores to equivalent DLQI scores.
d) Whether the data in Table 66 show the average EQ-5D scores from the mapping of DLQI/CDQI to EQ-5D in patients having an ISGA score of 0 to 1. If not, please provide the average mapped scores for patients having an ISGA score of 0 to 1 for each scenario presented in Table 66. Please also provide further details on how the figures in Table 66 for ISGA scores of 0 to 1 were obtained.

Pfizer response:

a) [redacted]The individual patient age was applied in the mapping algorithm (scores were mapped using individual patient data).

b) Patient age (which featured in the Ali algorithm) was the only factor used to adjust CDLQI outcomes.
c) The column labelled “Mean EQ-5D index Value AD 301 and AD 302” show the average EQ-5D scores from the mapping of DLQI/CDQI to EQ-5D from AD 301 and AD 302. The row labelled controlled disease (ISGA 0-1) represent EQ-5D values for patients having an ISGA score of 0 to 1.

B36. CS, Section B.3.4.5. The company submission states that “Patients in the ISGA 0-1 health state are assumed to revert to population utility norms for their age group in the UK”. For the purposes of calculating utility multipliers, please clarify whether this means that the average EQ-5D scores measured in patients with an ISGA score of 0 to 1 are assumed to reflect average population health utility in a patient of that age.

Pfizer response: We can confirm that the assertion in the question is correct; patients with an ISGA score of 0 to 1 are assumed to reflect average population health utility in a patient of that age. This approach allowed the natural decline in HRQoL utility values which occurs in older patients to be captured appropriately.

B37. Priority Question. CS, Section B.3.4.2. The company submission states that [redacted] Please clarify if any literature searches were undertaken to identify [redacted]

Pfizer response: A search was undertaken by Pfizer to identify equivalent mapping algorithms for CDLQI and no studies were identified. It is noted, however, when compared with the adult DLQI, questions 1, 2, 4-6 and 10 (out of ten questions) are substantially the same.(3) Questions related to an adult context such as related to shopping, looking after the home, problems with partners or close friends, and to sexual difficulties, are replaced by questions concerning friendships, adverse

comments and sleep in the CDLQI. As such, the questions in the CDLQI mirror those of the DLQI within a childhood context.(3)

B38. Priority Question. CS, Section B.3.4.2. Please clarify how the mapping algorithm by Ali et al. (2017) was identified from the literature. Please explain the rationale for choosing to use the ordinal logistic regression (OLR) instead of the linear regression in Ali et al. to map from DLQI to EQ-5D, given that the linear regression was described as having better predictive accuracy.
Pfizer response: The Ali et al. (2017) paper was identified in the screening phases of the HRQoL SLR as a potential mapping study. We used the Ali et al reported algorithm, which used OLR to map outcomes. It is noted that both the linear model, which was developed by Currie and Conway, 2007 and the OLR model, were reported to have good predictive qualities. Overall, the Ali algorithm was considered a robust and recent source for mapping DLQI to EQ-5D estimates in a UK AD population, although it is acknowledged that other mapping algorithms could feasibly have been used.

B39. CS, Section B.3.4.2. The ERG identified 5 alternative mapping algorithms from the Health Economics Research Centre (HERC) database of mapping studies (*Dakin H, Abel L, Burns R, Yang Y. 2018. Review and critical appraisal of studies mapping from quality of life or clinical measures to EQ-5D: an online database and application of the MAPS statement. Health and Quality of Life Outcomes. 16:31* and *Dakin H, Abel, L, Burns R, Yang Y. 2019. HERC database of mapping studies, Version 7.0 DOI: 10.5287/bodleian:bxBdRNwBJ*. These are listed below:

- Blome C, Beikert FC, Rustenbach SJ, Augustin M. (2013) Mapping DLQI on EQ-5D in psoriasis: transformation of skin-specific health-related quality of life into utilities. Arch Dermatol Res, 305(4), 197-204.
- Currie CJ, Conway P. (2007) Evaluation of the association between EQ5D utility and dermatology life quality index (DLQI) score in patients with psoriasis. Value Health. 10 (6), A470-1 (Abstract PSK11).
- Davison NJ, Thompson AJ, Turner AJ, Longworth L, McElhone K, Griffiths CEM, Payne K. (2018) Generating EQ-5D-3L Utility Scores from the Dermatology Life Quality Index: A Mapping Study in Patients with Psoriasis. Value Health, 21(8), 1010-1018.
- Heredi E, Rencz F, Balogh O, Gulacsi L, Herszenyi K, Hollo P, Jokai H, Karpati S, Pentek M, Remenyik E, Szegedi A, Brodszky V. (2014) Exploring the relationship

between EQ-5D, DLQI and PASI, and mapping EQ-5D utilities: a cross-sectional study in psoriasis from Hungary. Eur J Health Econ, 15 Suppl 1, 111-119.

- Norlin JM, Steen Carlsson K, Persson U, Schmitt-Egenolf M. (2012) Analysis of three outcome measures in moderate to severe psoriasis: a registry-based study of 2450 patients. Br J Dermatol, 166(4), 797-802.

Please explain why the algorithm by Ali et al. was preferred to these alternative mapping algorithms identified from the Oxford database.

Pfizer response: The algorithm by Ali et al, 2017, was a validated mapping algorithm identified for mapping DLQI to EQ-5D, based on a UK study population which included atopic dermatitis patients. The remaining mapping algorithms identified above are in psoriasis patients (with the exception of Currie and Conway 2007) and two studies were also not in UK patients.(33, 34) The patient sample used in the Ali et al study (n=4010), was also one the largest out of the suggested studies. Overall the Ali algorithm was considered a robust data source for mapping DLQI to EQ-5D estimates in a UK AD population, although it is acknowledged that other mapping algorithms could feasibly have been used.

Miscellaneous

B40. CS, Appendix M. Please provide a full set of results for outcomes surveyed by the Pfizer epidemiology report mentioned in page 4 of Appendix M.

Pfizer response: To avoid duplication of results being presented, we refer you to reference **2017-06-05 AD Epi Report UK (Kantar).** The results for the requested outcomes below can be found in this reference.

Excel model

B41. Excel model and CS, Section B.3.3.3. Table 62 does not appear to relate to the data inputs in cells C74 to D78 of the clinical data sheet in the Excel model. Instead, the data in the clinical data sheet shows that pimecrolimus is used in 87% of mild disease with the remaining 13% receiving systemic therapy. Please confirm which data is the company's preferred assumptions.

Pfizer response: The data reported in Table 62 assumed a proportion of patients with mild disease were treated with tacrolimus, however this is incorrect because tacrolimus is not licensed in mild disease. The preferred assumptions are those included in the model and are reported in **Table 23** for completeness.

Table 23: Probability of starting different treatments having failed the primary therapy in mild disease (revision of Table 62 in submission)

	Probability - Children	Probability - Adults
Tacrolimus 0.03%	0.00	0.00
Tacrolimus 0.1%	0.00	0.00
Pimecrolimus	0.87	0.87
Systemic therapy	0.13	0.13
Phototherapy	0.00	0.00

B42. Excel Model. Model 'Engine' sheets, cells H2, H4 & H6. Please clarify why the row number in the INDEX formula differed between sheets for comparators 1 and 2 versus comparators 3 and crisaborole (A1-2 versus A1-3).

Pfizer response: The row numbers in the comparator 3 and crisaborole sheets should also be A1-3. However, this error only affects the flare rates for tacrolimus in the maintenance scenarios and is not applicable to crisaborole, thus this change does not impact any of the base case results.

B43. Excel Model. VBA, Module 1, Sub pop_age. When considering the children age group (Range("pop_age").Value = 1), D11 referring to tacrolimus 0.03% (and not D12 referring to tacrolimus 0.1%) in the 'Cell links' sheet was set to FALSE. The same was noted in the adult age group where D12 referring to tacrolimus 0.1% (and not D13 referring to pimecrolimus) in the same sheet was set to FALSE. Please clarify the use of such code lines and adjust for errors (if any).

Pfizer response: These lines of VBA are remnants of an earlier version of the model and can now be removed. Please note they do not affect the model results as the correct values are set later in the subroutine.

B44. Excel Model. VBA, Module 1, Sub NMA_selection. Please clarify why Range ("NMA").Value was equal to 4 and not 2.

Pfizer response: Upon review we agree that this conditional statement should use the value 2 rather than 4. This has been updated in the revised version of the electronic model but does not affect the results of any analysis.

Section C: Textual clarification and additional points

C1. Priority Question. Please provide full reference pack including:

- clinical study reports for AD-301 and AD-302
- all papers cited in the submission and a RIS file to allow the references to be uploaded into reference manager software such as Endnote

Pfizer Response: Please find the clinical study reports and RIS file in the reference pack.

C2. CS, Section B.2.8.4. In Table 29, please clarify if the outcome timepoint for Chapman et al. 2005 was stated wrongly as 46 weeks instead of 6 weeks as in the company submission.

Pfizer response: Chapman 2005 was stated incorrectly. The timepoint should be 6 weeks, not 46 weeks.

C3. CS, Section B.2.8.4 and Appendix D. In Table 29, please clarify if ISGA 0/1 was reported at 29 days or at 43 days (as in Table D14 in Appendix D) for Kempers 2004.

Pfizer response: Day 29 is correct in **Section B.2.8.4, Table 29.** In **Appendix D, Table D14** this should be corrected from 43 days to Day 29.

References

1. Nankervis H, Maplethorpe A, Williams HC. Mapping randomized controlled trials of treatments for eczema--the GREAT database (the Global Resource of EczemA Trials: a collection of key data on randomized controlled trials of treatments for eczema from 2000 to 2010). BMC dermatology. 2011;11:10.
2. Fahrbach K, Tarpey J, Bergrath E, Hughes R, al e, editors. A Systematic Literature review (SLR) and Network Meta-Analysis (NMA) of Crisaborole 2% Ointment for the Treatment of Mild-to-Moderate Atopic Dermatitis (AD). ISPOR 2020; 2020.
3. Lewis-Jones MS, Finlay AY. The Children's Dermatology Life Quality Index (CDLQI): initial validation and practical use. Br J Dermatol. 1995;132(6):942-9.
4. Salek MS, Jung S, Brincat-Ruffini LA, MacFarlane L, Lewis-Jones MS, Basra MK, et al. Clinical experience and psychometric properties of the Children's Dermatology Life Quality Index (CDLQI), 1995-2012. Br J Dermatol. 2013;169(4):734-59.
5. Murrell DF, Calvieri S, Ortonne JP, Ho VC, Weise-Riccardi S, Barbier N, et al. A randomized controlled trial of pimecrolimus cream 1% in adolescents and adults with head and neck atopic dermatitis and intolerant of, or dependent on, topical corticosteroids. Br J Dermatol. 2007;157(5):954-9.
6. Chapman MS, Schachner LA, Breneman D, Boguniewicz M, Gold MH, Shull T, et al. Tacrolimus ointment 0.03% shows efficacy and safety in pediatric and adult patients with mild to moderate atopic dermatitis. 2005;53(2):S177-S85.
7. Paller AS, Lebwohl M, Fleischer AB, Jr., Antaya R, Langley RG, Kirsner RS, et al. Tacrolimus ointment is more effective than pimecrolimus cream with a similar safety profile in the treatment of atopic dermatitis: results from 3 randomized, comparative studies. J Am Acad Dermatol. 2005;52(5):810-22.
8. Meurer M, Fartasch M, Albrecht G, Vogt T, Worm M, Ruzicka T, et al. Long-term efficacy and safety of pimecrolimus cream 1% in adults with moderate atopic dermatitis. Dermatology. 2004;208(4):365-72.
9. Hoeger PH, Lee KH, Jautova J, Wohlrab J, Guettner A, Mizutani G, et al. The treatment of facial atopic dermatitis in children who are intolerant of, or dependent on, topical corticosteroids: a randomized, controlled clinical trial. Br J Dermatol. 2009;160(2):415-22.
10. Hanifin JM, Ellis CN, Frieden IJ, Folster-Holst R, Stein Gold LF, Secci A, et al. OPA-15406, a novel, topical, nonsteroidal, selective phosphodiesterase-4 (PDE4) inhibitor, in the treatment of adult and adolescent patients with mild to moderate atopic dermatitis (AD): A phase-II randomized, double-blind, placebo-controlled study. J Am Acad Dermatol. 2016;75(2):297-305.
11. Abramovits W, Fleischer AB, Jr., Jaracz E, Breneman D. Adult patients with moderate atopic dermatitis: tacrolimus ointment versus pimecrolimus cream. J Drugs Dermatol. 2008;7(12):1153-8.
12. Eichenfield LF, Lucky AW, Boguniewicz M, Langley RG, Cherill R, Marshall K, et al. Safety and efficacy of pimecrolimus (ASM 981) cream 1% in the treatment of mild and moderate atopic dermatitis in children and adolescents. J Am Acad Dermatol. 2002;46(4):495-504.
13. Wahn U, Bos JD, Goodfield M, Caputo R, Papp K, Manjra A, et al. Efficacy and safety of pimecrolimus cream in the long-term management of atopic dermatitis in children. Pediatrics. 2002;110(1 Pt 1):e2.

14. Boguniewicz M, Fiedler VC, Raimer S, Lawrence ID, Leung DY, Hanifin JM. A randomized, vehicle-controlled trial of tacrolimus ointment for treatment of atopic dermatitis in children. Pediatric Tacrolimus Study Group. J Allergy Clin Immunol. 1998;102(4 Pt 1):637-44.
15. Gesslein B. Humectants in Personal Care Formulation: A Practical Guide. Conditioning Agents for Hair and Skin: CRC Press; 1999. p. 95-6.
16. Schachner LA, Lamerson C, Sheehan MP, Boguniewicz M, Mosser J, Raimer S, et al. Tacrolimus ointment 0.03% is safe and effective for the treatment of mild to moderate atopic dermatitis in pediatric patients: results from a randomized, double-blind, vehicle-controlled study. Pediatrics. 2005;116(3):e334-42.
17. Alanne S, Nermes M, Soderlund R, Laitinen K. Quality of life in infants with atopic dermatitis and healthy infants: a follow-up from birth to 24 months. Acta Paediatr. 2011;100(8):e65-70.
18. Simpson EL. Comorbidity in Atopic Dermatitis. Curr Dermatol Rep. 2012;1(1):29-38.
19. Dogan DG, Canaloglu SK, Kivilcim M, Kum YE, Topal E, Catal F. Sleep patterns of young children with newly diagnosed atopic dermatitis. Postepy Dermatol Alergol. 2017;34(2):143-7.
20. Yaghmaie P, Koudelka CW, Simpson EL. Mental health comorbidity in patients with atopic dermatitis. J Allergy Clin Immunol. 2013;131(2):428-33.
21. Riis JL, Vestergaard C, Deleuran MS, Olsen M. Childhood atopic dermatitis and risk of attention deficit/hyperactivity disorder: A cohort study. J Allergy Clin Immunol. 2016;138(2):608-10.
22. Paller AS LT, Herbert AA, Zaenglein AL, Silverberg JI, Tallman AM, Tan H, Ports WC, Zielinski MA. . Crisaborole Ointment Improves Global Atopic Dermatitis Severity Across Patients with Varying Baseline Characteristics: Pooled Results from Two Phase 3 Trials. International Symposium on Atopic Dermatitis; Utrecht, The Netherlands2018.
23. Garside R, Stein K, Castelnuovo E, Pitt M, Ashcroft D, Dimmock P, et al. The effectiveness and cost-effectiveness of pimecrolimus and tacrolimus for atopic eczema: a systematic review and economic evaluation. Health Technology Assessment (Winchester, England). 2005;9(29):iii, xi-xiii,1-230.
24. Garibaldinos T. Phototherapy to treat eczema. In: (BAD) TBAoD, editor.
25. Taylor K, Swan DJ, Affleck A, Flohr C, Reynolds NJ, Dermatology UKTRNi, et al. Treatment of moderate-to-severe atopic eczema in adults within the U.K.: results of a national survey of dermatologists. Br J Dermatol. 2017;176(6):1617-23.
26. Pfizer Inc. Data on file. Analysis of clinical trials AD-301 and AD-302. MO-48 Table 2.
27. Social DoHa. Drugs and pharmaceutical electronic market information tool (eMIT). Last updated 8 April 2019. Available from https://www.gov.uk/government/publications/drugs-and-pharmaceutical-electronic-market-information-emit. Last accessed June 2019.
28. National Institute for Health and Care Excellence. BNF 2019. Available from https://bnfniceorguk/.
29. Pfizer Inc. Data on file. Atopic Dermatitis Prevalence - UK.
30. Briggs A, Claxton K, Sculpher M. Decision modelling for health economic evaluation: Oxford University Press; 2006.
31. Wollenberg A, Barbarot S, Bieber T, Christen-Zaech S, Deleuran M, Fink-Wagner A, et al. Consensus-based European guidelines for treatment of atopic

eczema (atopic dermatitis) in adults and children: part I. J Eur Acad Dermatol Venereol. 2018;32(5):657-82.
32. Ali FM, Kay R, Finlay AY, Piguet V, Kupfer J, Dalgard F, et al. Mapping of the DLQI scores to EQ-5D utility values using ordinal logistic regression. Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation. 2017;26(11):3025-34.
33. Blome C, Beikert FC, Rustenbach SJ, Augustin M. 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.
34. Heredi E, Rencz F, Balogh O, Gulacsi L, Herszenyi K, Hollo P, et al. Exploring the relationship between EQ-5D, DLQI and PASI, and mapping EQ-5D utilities: a cross-sectional study in psoriasis from Hungary. Eur J Health Econ. 2014;15 Suppl 1:S111-9.

NICE National Institute for Health and Care Excellence

Patient organisation submission

Crisaborole for treating mild to moderate atopic dermatitis in people aged 2 years and older [ID1195]

Thank you for agreeing to give us your organisation's views on this technology and its possible use in the NHS.

You can provide a unique perspective on conditions and their treatment that is not typically available from other sources.

To help you give your views, please use this questionnaire with our guide for patient submissions.

You do not have to answer every question – they are prompts to guide you. The text boxes will expand as you type.

Information on completing this submission

- Please do not embed documents (such as a PDF) in a submission because this may lead to the information being mislaid or make the submission unreadable
- We are committed to meeting the requirements of copyright legislation. If you intend to include **journal articles** in your submission you must have copyright clearance for these articles. We can accept journal articles in NICE Docs.
- Your response should not be longer than 10 pages.

About you	
1.Your name	[redacted]

2. Name of organisation	National Eczema Society
3. Job title or position	Head of Services
4a. Brief description of the organisation (including who funds it). How many members does it have?	National Eczema Society is the UK charity for people of all ages living with eczema and those who care for them, in order to improve their quality of life. We support with information and advice about eczema and its management and treatment, which we deliver through our website, social media, publications and nurse-supported Helpline. We are the campaigning voice for people with eczema, and raise awareness of the needs of people with eczema with healthcare professionals, teachers and the government. We are funded by membership fees, donations from the public and organisations, and our corporate partners (pharmaceutical companies that sell products or services for people with eczema). We have approximately 2,600 members.
4b. Do you have any direct or indirect links with, or funding from, the tobacco industry?	No.
5. How did you gather information about the experiences of patients and carers to include in your submission?	National Eczema Society operates a nurse-supported Helpline service, responding to telephone and email enquiries from people affected by eczema who are seeking advice either on their own behalf or for a loved one. The calls and e-mails we receive give us a valuable insight into the experiences of people living with eczema and the many challenges they face. We also gain insights from the conversations and insights shared by people with eczema on our busy social media channels.

Living with the condition	
6. What is it like to live with the condition? What do carers experience when caring for someone with the condition?	Atopic eczema is a chronic dry skin condition. Its major symptom is itchiness, which can be intense and unbearable. Constant scratching causes the skin to split and bleed, and leaves it open to infection. Even when the eczema is mild to moderate (as opposed to severe), when it is not well-controlled it can have a significant impact on quality of life. In the UK, one in five children and one in twelve adults has eczema. Constant itchiness is one of the most challenging aspects of eczema; it can result in reduced social interaction and inability to work and study. In addition to the pain and discomfort brought about by scratching, itchiness often makes sleeping extremely difficult. Lack of sleep can compromise someone's ability to concentrate at work and school/university and carry out everyday tasks effectively. It also damages personal relationships - as can itchiness alone. Eczema can have a significant negative impact on the whole family. People who are constantly itchy and/or have eczema on visible areas of their body can feel extremely self-conscious about their condition and appearance, and reluctant to leave their home. The resulting problems of social isolation are particularly challenging and damaging for children and adolescents. Eczema management is time-consuming. In addition to applying topical treatments at least twice a day, and every few hours when the skin is very dry, people who scratch a lot overnight may have to wash their bedding every day to remove blood and skin flakes. People who have a mental health condition (e.g. depression) as a result of their eczema, or in addition to it, often find it difficult to manage both conditions effectively. Even people who haven't been diagnosed with a mental health condition can find daily eczema management onerous and dispiriting.

	Caring for a child or adult with eczema can be time-consuming and exhausting, both physically and emotionally. Carers may need to apply topical treatments to the person in their care multiple times a day, try to distract them when they are itchy, provide emotional support and take them to regular GP or hospital appointments. Carers' ability to sleep is compromised when the person in their care is unable to sleep because of itchiness. Carers often need to get up several times during the night to apply emollient, and comfort the person for whom they are caring. Lack of sleep for carers, as for people with eczema, can lead to their experiencing a diminished ability to concentrate at work and other activities, and carry out tasks effectively.
Current treatment of the condition in the NHS	
7. What do patients or carers think of current treatments and care available on the NHS?	Many patients and carers consider the current treatments for eczema available on the NHS to be limited in number and effectiveness. Many patients are reluctant to use topical corticosteroids – even those of mild and moderate potencies – because of concerns about the adverse effects of using topical corticosteroids, notably skin thinning.
8. Is there an unmet need for patients with this condition?	At present, the only topical treatments for eczema are emollients, topical corticosteroids of different potencies and topical calcineurin inhibitors. Adherence to current treatments is often sub-optimal. Many people are reluctant to use topical corticosteroids – even those of mild and moderate potencies – because of concerns about corticosteroids' potential to thin the skin (some 72.5% of respondents in a study said they worried about using topical corticosteroids on their own or their child's skin – Charman C.R., Morris A.D. & Williams H.C. (2000) 'Topical corticosteroid phobia in patients with atopic eczema', British Journal of Dermatology 142(5):931-6). While these patient concerns do not reflect the scientific evidence, they nevertheless affect patient behaviour and compliance. Additionally, many GPs are reluctant to prescribe topical corticosteroids of the appropriate strength needed to bring eczema flare-ups under control, reflecting both patient concerns and their own clinical cautiousness, which compounds the sub-optimal use of the main current treatment to reduce inflammation in eczema. Some patients and clinicians also have concerns about using topical

	calcineurin inhibitors due to the potential increased risk of skin cancer that was reported when the treatment was first introduced. Crisaborole, as a non-steroidal topical treatment, is likely to be more acceptable to many patients than topical corticosteroids. Crisaborole also does not have a potential association with skin cancer risk, like topical calcineurin inhibitors. Crisaborole therefore has the potential to help meet an unmet need for patients with eczema, especially as it has been shown to reduce the major debilitating symptom of itchiness.
Advantages of the technology	
9. What do patients or carers think are the advantages of the technology?	Crisaborole can be used on both eczema that is flaring and as part of a longer-term disease control/maintenance regime. Unlike topical corticosteroids, Crisaborole may be safe to use on a long-term, continuous basis. This would make it easier to use than topical corticosteroids, as patients wouldn't have to remember to stop the treatment for specific lengths of time. In addition, Crisaborole can be applied safely to the skin anywhere on the face and body. Higher potencies of topical steroids are not often prescribed for use on the face, and, when they are prescribed, people with eczema tend to be (understandably) wary of using them, as the risk of skin thinning increases when higher potencies of topical steroids are used on delicate areas of skin. Crisaborole would be another option for people with eczema on delicate areas of skin. Many people with eczema are finding it difficult to obtain any leave-on emollient on prescription at all following the introduction of NHS England's 'Guidance for which over the counter items should not routinely be prescribed in primary care' (2018). Crisaborole may help to alleviate eczema symptoms in people who are struggling to access leave-on emollient on prescription, or obtain it in sufficient quantities. The introduction of Crisaborole would broaden patient choice, which is vital given the limited treatment options for the condition at present, and increase the likelihood that patients with mild to moderate eczema would find a treatment that is effective for them and to which they would be willing to adhere.

Disadvantages of the technology	
10. What do patients or carers think are the disadvantages of the technology?	One disadvantage of the technology is that it is unlikely to work effectively for everyone eligible to use it. We understand that Crisaborole has several potential side effects in the form of allergic reactions at or near the application site, and application site pain, such as burning or stinging.
Patient population	
11. Are there any groups of patients who might benefit more or less from the technology than others? If so, please describe them and explain why.	Patients with mild to moderate eczema who have concerns about and are unwilling to use topical corticosteroids of any level of potency are more likely to benefit from this technology. This group of patients, given their reluctance to use topical corticosteroids, have fewer treatments available to which they will adhere, so are likely to benefit from the introduction of a new topical treatment. Patients with moderate eczema whose symptoms have not responded effectively to higher potency topical corticosteroids, or who must use higher potency topical corticosteroids regularly as weekend therapy due to frequent flares, are likely to benefit from the introduction of a new topical treatment, offering them a new option.
Equality	
12. Are there any potential equality issues that should be taken into account when	N/A

considering this condition and the technology?	
Other issues	
13. Are there any other issues that you would like the committee to consider?	N/A
Key messages	

15. In up to 5 bullet points, please summarise the key messages of your submission:

- The treatment options for eczema currently available on the NHS are limited. The introduction of Crisaborole would broaden patient choice by providing an additional treatment option for people with mild to moderate eczema to treat inflammation and itch.
- Adherence to current topical treatments is often sub-optimal, largely due to patients' concerns about using corticosteroids. Crisaborole, as a non-steroidal topical treatment that also reduces itch, would be more acceptable than topical corticosteroids for many patients, who are therefore likely to accept and adhere to treatment as prescribed and manage their eczema effectively.
- Crisaborole may be safe to use on a long-term, continuous basis as part of an inflammation maintenance regime. This would make it easier to use than topical corticosteroids, as patients would not have to start and stop treatments to manage periodic flares.

Thank you for your time.

Please log in to your NICE Docs account to upload your completed submission.

.....

Your privacy

The information that you provide on this form will be used to contact you about the topic above.

☐ **Please tick this box** if you would like to receive information about other NICE topics.

For more information about how we process your personal data please see our privacy notice.

.....

Crisaborole for Treating Mild to Moderate Atopic Dermatitis in People Aged 2 Years and Older [ID 1195]: A Single Technology Appraisal

Produced by School of Health and Related Research (ScHARR), The University of Sheffield

Authors Sarah Davis, Senior Lecturer, ScHARR, University of Sheffield, Sheffield, UK

Edith Poku, Research Fellow, ScHARR, University of Sheffield, Sheffield, UK

John W. Stevens, Reader, ScHARR, University of Sheffield, Sheffield, UK

Katy Cooper, Senior Research Fellow, ScHARR, University of Sheffield, Sheffield, UK

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

Ruth Wong, Information Specialist, ScHARR, University of Sheffield, Sheffield, UK

Carolyn Charman, Consultant Dermatologist, Royal Devon and Exeter Hospital, Exeter, UK

Correspondence Author Sarah Davis, Senior Lecturer, ScHARR, University of Sheffield, Sheffield, UK

Date completed 14/01/2020 [version revised post FAC]

Source of funding: This report was commissioned by the NIHR HTA Programme as project number 129371.

Declared competing interests of the authors

Carolyn Charman declared that she provides ad hoc consultancy, presentations and teaching for Galderma (hydrocortisone), GlaxoSmithKline (betamethasone valerate, clobetasone butyrate, fluticasone propionate, hydrocortisone) and Leo Laboratories (betamethasone valerate, fusidic acid, hydrocortisone, tacrolimus). None of the other authors have any conflicts of interest to declare.

Acknowledgements

We would like to thank Paul Tappenden, ScHARR, for providing comments on the draft report and Andrea Shippam and Gill Rooney, Programme Managers, ScHARR, for providing administrative support and in preparing and formatting the report. A second clinical expert who preferred not to be named as an author also provided clinical advice. They declared that they had no conflicts of interest.

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:

Davis S., Poku E., Stevens J.W., Cooper K., Metry A., Wong R., Charman C. Crisaborole for Treating Mild to Moderate Atopic Dermatitis in People Aged 2 Years and Older [ID 1195]: A Single Technology Appraisal. School of Health and Related Research (ScHARR), 2019.

Contributions of authors

Katy Cooper and Edith Poku summarised and critiqued the clinical effectiveness data reported within the company's submission. John Stevens critiqued the evidence synthesis reported in the company submission. Sarah Davis and Andrew Metry critiqued the health economic analysis submitted by the company and conducted the ERG exploratory analyses. Ruth Wong critiqued the company's search strategy. Carolyn Charman provided clinical expert advice. All authors were involved in drafting and commenting on the final report.

Standard copyright statement on the front page of the ERG/AG report:

CONTENTS

List of Tables

List of Figures

List of Boxes

Abbreviations

AD	Atopic dermatitis
AAD	American Academy of Dermatology
AE	Adverse event
BAD	British Association of Dermatologists
BSA	Body surface area
BSC	Best supportive care
CADTH	Canadian Agency for Drugs and Technologies in Health
CDLQI	Children's Dermatology Life Quality Index
CG	Clinical Guideline
CODA	Convergence Diagnosis and Output Analysis
CS	Company's submission
DALY	Disability-adjusted life year
DIC	Deviance Information Criterion
DLQI	Dermatology Life Quality Index
DSA	Deterministic sensitivity analysis
EASI	Eczema Area and Severity Index
EADV	European Academy of Dermatology and Venereology
EMA	European Medicines Agency
ERG	Evidence Review Group
ESPD	European Society for Paediatric Dermatology
FDA	Food and Drug Administration
GP	General practitioner
HOME	Harmonising Outcome Measures in Eczema
HRQoL	Health-related quality of life
HTA	Health Technology Assessment
ICD	International Congress of Dermatology
ICER	Incremental cost-effectiveness ratio
IGA	Investigator's Global Assessment
IGADA	Investigator Global Atopic Dermatitis Assessment
ISAD	International Symposium on Atopic Dermatitis
ISGA	Investigator's Static Global Assessment
ISPOR	International Society for Pharmacoeconomics and Outcomes Research
ITT	Intention-to-treat

LYG	Life year gained
MAIC	Matching-adjusted indirect comparison
MRU	Medical resource use
NHM	Natural history model
NHS	National Health Service
NHS EED	National Health Service Economic Evaluation Database
NICE	National Institute for Health and Care Excellence
NMA	Network meta-analysis
NR	Not reported
PBAC	Pharmaceutical Benefits Advisory Committee
PDE4	Phosphodiesterase 4
PGA	Physician's Global Assessment
POEM	Patient-Oriented Eczema Measure
PSA	Probabilistic sensitivity analysis
PSGA	Physician's Static Global Assessment
PSS	Personal Social Services
PSSRU	Personal Social Services Research Unit
QALY	Quality-adjusted life year
RCT	Randomised controlled trial
SCORAD	Scoring Atopic Dermatitis index
SLR	Systematic literature review
SMC	Scottish Medicines Consortium
STA	Single Technology Appraisal
TCI	Topical calcineurin inhibitor
TCS	Topical corticosteroid
THIN	The Health Improvement Network
WCD	World Congress of Dermatology

1 SUMMARY

1.1 Critique of the decision problem in the company's submission

The intervention is consistent with the NICE scope: crisaborole (20mg/g i.e. 2%). Crisaborole is a topical ointment which controls inflammation by inhibiting phosphodiesterase 4 (PDE4), thereby inhibiting cytokine production. Crisaborole is currently under assessment by the European Medicines Agency (EMA). The proposed marketing authorisation for crisaborole is a treatment for mild to moderate atopic dermatitis (AD) in adults and children aged 2 years and older with ≤40% body surface area (BSA) affected (CS, page 14(1)). Crisaborole is intended to be applied twice daily to a maximum of 40% BSA, for up to 4 weeks per treatment course, and to be discontinued if symptoms persist after 3 consecutive 4-week courses.(1)

The population in the company decision problem is adults and children aged 2 years and older with mild to moderate AD which is consistent with the final NICE scope.(2) However, the company's submission (CS)(1) focusses on a second-line (post-topical corticosteroid [TCS]) population, i.e. people uncontrolled on TCS or with serious risk of important adverse events (AEs) from TCS (particularly irreversible skin atrophy). Scenario analyses for a first-line (TCS eligible) population are also presented in the economic section.

The comparator in the second-line population with moderate AD, topical calcineurin inhibitors (TCIs), i.e. tacrolimus and pimecrolimus, is consistent with the final NICE scope for moderate AD.(2) The comparator specified in the NICE scope for patients with mild AD is emollients in combination with TCSs.(2) It is therefore unclear what the appropriate comparator is for patients with mild AD in the post-TCS population and whether this should be emollients alone. In the second-line post-TCS population, the company does not provide a comparison of crisaborole against emollients alone, but instead compares against TCIs, which are not recommended by NICE for patients with mild AD. The company's justification is that TCIs are used in clinical practice in some patients with mild AD and that use of emollients alone would not be appropriate for patients with uncontrolled AD. The ERG's clinical advisors felt that in patients with uncontrolled mild AD symptoms, it would be unethical to step down treatment to emollients alone where TCSs have failed or where patients were at risk of adverse events from further TCS use.

The comparators in the scenario analysis considering the first-line population are TCSs for both mild and moderate AD, which is consistent with the final NICE scope.(2) A minor deviation from the scope is that the company assumes that a mix of mild, moderate and potent TCSs are used for both mild AD and moderate AD, whereas the scope specifies mild to moderate TCSs for mild AD and high potency TCSs for moderate AD.(2) This is considered reasonable as clinical experts advised that a

short course of a higher potency TCS is sometimes necessary. In addition, the clinical experts noted that the use of a range of potencies in both mild and moderate AD may reflect the difficulty in assigning a single severity score to a patient when the severity may vary across body sites and may fluctuate over time.

The outcomes in the CS are generally consistent with the NICE final scope,(2) except that data on the effectiveness of crisaborole in maintaining remission of symptoms or preventing flares is not available due to the short-term nature of the key clinical trials.

The ERG notes that the CS does not provide any evidence for several potential subgroups listed in the final NICE scope including: people with atopic dermatitis affecting the hands; and people with atopic dermatitis affecting sensitive areas (such as the face, neck and flexures). At the request of the ERG, the company provided data on ISGA success subgrouped by race and ethnicity for the crisaborole trials (clarification response to question A9, Appendix 4, Table 5 and Table 6). However, no NMA is conducted for any of these subgroups as the company state that such data were not consistently reported across comparator trials (CS Table 1).

1.2 Summary of clinical effectiveness evidence submitted by the company

The two 4-week crisaborole trials (AD-301 and AD-302) showed a statistically significant effect of crisaborole over vehicle ointment (the placebo control) for success on the Investigator's Static Global Assessment (ISGA; success defined as score of 0/1 and ≥2-grade improvement), ISGA score 0/1, pruritus, clinical signs and health-related quality of life.

Within subgroups for mild and moderate AD, the evidence of effectiveness was less consistent in mild AD. The difference in ISGA success between crisaborole and vehicle (pooled across the two trials) was [redacted] For ISGA 0/1, no pooled data across trials were presented; however in moderate AD [redacted] (AD-301: [redacted]; AD-302: [redacted]), while in mild AD [redacted] (AD-302: [redacted] (AD-301: [redacted] Subgroup analyses for mild and moderate AD were not reported for other clinical outcomes.

Safety outcomes were reported in the two crisaborole trials and in a 48-week single-arm extension study (AD-303); the majority of treatment-emergent AEs associated with crisaborole were mild and moderate and not considered to be related to the study drug.

The company performed a single systematic review to identify studies of clinical effectiveness and safety for all approved topical treatments for mild to moderate AD. The network meta-analysis (NMA) for ISGA/IGA 0/1 included the two trials of crisaborole, plus seven trials of TCIs (tacrolimus 0.03%, tacrolimus 0.1% and pimecrolimus 1%). No trials of TCS were identified which had a population ≥80% mild to moderate AD and reported relevant outcomes. Nineteen trials of TCS and TCIs were excluded from the review as >20% of the trial population had severe AD, or severity was not reported.

Responses were higher in the vehicle arms of the crisaborole trials than in the vehicle arms of the TCI trials, which the CS states [redacted]. Therefore, the company's preferred base case NMA model adjusted for vehicle response using meta-regression, and an unanchored matching-adjusted indirect comparison (MAIC) was also undertaken. In the company base case NMA (fixed treatment effect, random class effect, adjustment for vehicle response, i.e. FE-RCE-VR model) and in the unanchored MAIC, [redacted] in the company's simple random effects NMA, with no class effect and no vehicle response adjustment, [redacted]

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

The methods of the systematic review of effectiveness appear robust. A range of indirect comparison approaches were presented, including NMA using various models as well as a MAIC. There is uncertainty regarding the most appropriate comparator for mild AD in a second-line (post-TCS) population. The crisaborole trials were only 4 weeks duration and only compare against vehicle, not against TCIs or TCSs (though ongoing trials are comparing to TCIs and TCSs). The CS does not present an NMA of crisaborole against TCS, since no TCS trials were identified in a ≥80% mild to moderate population and that reported relevant outcomes.

Of the different indirect comparison approaches, the ERG prefers a simple random effects NMA with no class effect and no vehicle response adjustment, although it accepts that the assumption of transitivity (that is if the effect of B>A, and the effect of C >B, then the effect of C must be >A) may be violated if the vehicles used in the crisaborole and comparator studies are different, and that indirect comparisons may be biased if there is an imbalance in the distribution of treatment effect modifiers in studies comparing different pairs of treatments. Nevertheless, it is unclear whether the

assumptions made in the more complex vehicle response adjusted model preferred by the company are reasonable. The ERG has particular reservations regarding the assumption that the relationship between the population baseline response and relative treatment effect should be the same across treatments.

Although the MAIC circumvents the need to adjust for vehicle response, it provides inferences relative to the populations defined by the comparator studies, which may not be useful. In addition, an unanchored MAIC assumes that all prognostic factors and treatment effect modifiers have been accounted for, although this is an untestable assumption. For these reasons, the ERG considers that caution should be exercised when making inferences based on the results of the MAIC and prefers the simple random effects NMA over the MAIC. Overall, we have limited information about which of crisaborole or TCIs is the more effective, and the comparative effectiveness of crisaborole and TCIs is therefore uncertain.

1.4 Summary of cost effectiveness evidence submitted by the company

The population in the company's base case analysis is a second-line (post-TCS) population, i.e. people uncontrolled on TCS or with serious risk of important adverse effects from TCS (particularly irreversible skin atrophy). A scenario analysis is provided which examines first-line use of crisaborole in which the comparator in each population is TCSs.

Separate analyses are provided according to age and severity of AD. In children with mild AD and in adults with mild AD, the comparator is pimecrolimus. In children with moderate AD, the comparators are pimecrolimus and tacrolimus 0.03%. In adults with moderate AD, the comparators are tacrolimus 0.03% and tacrolimus 0.1%. Costs and QALYs are estimated over a life-time horizon for adults with a starting age of 18 years. In children, the starting age is 2 years but the model horizon is limited to age 18. Costs and QALYs are discounted at 3.5% and the analysis takes an NHS and PSS perspective, although no PSS costs are included.

The economic analysis uses a state-transition model structure with health states defined according to AD severity (mild, moderate or controlled) and whether the patient is receiving the primary therapy (crisaborole or TCIs in the base case) or subsequent treatment which consists of systemic therapy with immunosuppressants and/or phototherapy, depending on the AD severity. In the scenario for TCS-naïve patients, primary therapy is either crisaborole or TCS and subsequent therapy includes TCIs in addition to systemic therapy and phototherapy. Subsequent therapy is modelled as a '*basket*' of therapies, with a proportion receiving each of the various subsequent therapies rather than sequential use of each subsequent therapy being modelled explicitly. In addition, for children with AD the model includes a "*resolved*" disease state to reflect the fact that the disease resolves over time in a proportion of children.

The probability of response to primary therapy, and therefore a transition from the starting AD severity level (ISGA of 3 for moderate AD and ISGA of 2 for mild AD) to the controlled health-state, is based on response rates from the company's preferred NMA (fixed treatment effect, random class effect, adjustment for vehicle response model) which incorporated an adjustment for vehicle response rates. Response, which is assessed after 4 weeks of treatment, is defined as achieving an ISGA score of 0 (clear) or 1 (almost clear), which is assumed to be equivalent to controlled disease. The response rates for subsequent treatments are taken from a previous appraisal (TA82 - Tacrolimus and pimecrolimus for atopic eczema(3)), as are the response rates of TCS used in the TCS-naïve scenario. Patients with moderate disease receiving primary therapies can also experience a partial response leading to a transition to the mild AD state, which is based on the probability of achieving an ISGA score of 2 in the two crisaborole trials (AD-301 and AD-302) and is assumed to be equivalent for TCIs. Patients on either primary or subsequent therapy can also experience a flare resulting in a transition from the controlled state to their previous AD severity level. Flare rates were based on data from an epidemiological study due to the short term nature of the two crisaborole trials.

The utility values for the mild and moderate AD states were estimated by mapping from the DLQI to the EQ-5D using data in adults from the AD-301 and AD-302 trials. The utility values for controlled AD are assumed to be equivalent to population norms. Utility multipliers for mild and moderate AD, relative to controlled AD, were estimated from the trial data for adults and were also applied to children in the base case. In addition to treatment costs for primary and subsequent therapies, the model includes resource use related to GP and dermatologist consultations. Patients with AD are assumed to experience mortality rates equivalent to the general population regardless of treatment. No adverse events are explicitly included in the model. The QALY gains in the base case model are therefore solely derived from reducing the delay in the time taken to achieve a response and move to the controlled disease health state.

The company's economic analysis suggests that crisaborole dominates TCIs (i.e. is less costly but generates more QALYs) in all four populations (adults / children with mild / moderate AD) for the base case which considers second-line (post-TCS) use of crisaborole. This is because treatments that are more effective at achieving a response generate a cost saving by reducing the need for subsequent therapies in addition to generating a QALY gain from earlier disease control. In the scenario analysis considering first-line use (TCS-naïve), TCSs dominate crisaborole in all four populations. The company's sensitivity and scenario analyses demonstrate that the cost-effectiveness results are sensitive to changes in the estimates of the relative treatment effectiveness of crisaborole and TCIs. In particular, when crisaborole is assumed to have equivalent efficacy to TCIs, the base case results are reversed and it is dominated by TCIs.

1.5 Summary of the ERG's critique of cost effectiveness evidence submitted

The CS focuses on a second-line population where TCSs are not considered to be a relevant comparator. This population was not specified in the final NICE scope.(2) However, the ERG notes that a secondary analysis comparing against TCSs is provided within the CS.

The comparator for mild AD in the base case second-line population is TCIs which are not recommend by NICE for mild AD. The company provided audit data in children showing that TCIs are used in clinical practice in a minority of children with mild AD. The comparator in the NICE scope for mild AD is emollients in combination with TCSs. Therefore, it is unclear from the NICE scope whether emollients alone is a relevant comparators for patients with mild AD who have failed to achieve an adequate response to TCSs or who are at risk of adverse events from further TCS use. The company did not provide a comparison against emollients alone in patients with mild AD. However, the ERG's clinical advisors felt that stepping down to emollients alone would not be standard practice in patients with symptoms of uncontrolled mild AD who cannot be treated with TCSs.

The cost-effectiveness results are highly dependent on whether the company's preferred NMA which was adjusted for vehicle response is considered preferable to the simple unadjusted NMA. This is because, relative to its comparator, any treatment that improves response rates achieves both a QALY gain from reducing the delay in achieving symptom control and a cost saving from reducing the need for subsequent therapies. Given that the cost savings associated with the reduced need for subsequent therapies are a key driver of cost-effectiveness, the ERG is concerned with the simplistic manner in which subsequent therapies have been modelled. The ERG considers that the company's model fails to properly characterise the duration of subsequent therapy required to achieve the response rate assumed in the model and the possibility that patients may try multiple lines of subsequent therapy to achieve a response. In addition, the subsequent treatments assumed for patients with mild to moderate AD are not consistent with the treatment pathway recommended by NICE Clinical Guideline (CG) 57,(4) in which systemic treatments and phototherapy are only listed for patients with severe AD. In addition, the model structure also does not allow patients to experience more severe AD symptoms than they experienced at baseline.

1.6 ERG commentary on the robustness of evidence submitted by the company

1.6.1 Strengths

The methods of the systematic review for effectiveness appear robust. Relevant crisaborole trial results are presented. A range of indirect comparison approaches were presented, including NMA using various models as well as a MAIC.

The economic model included estimates of utility obtained my mapping from disease specific measures of health-related quality of life, measured in trials AD-301 and AD-302, to the EQ-5D.

1.6.2 Weaknesses and areas of uncertainty

There is uncertainty regarding the most appropriate comparator for mild AD in a second-line (post-TCS) population. The crisaborole trials were only 4 weeks duration and only compare against vehicle, not against TCIs or TCSs (though ongoing trials are comparing to TCIs and TCSs).

There is no NMA of crisaborole against TCS since no TCS trials were identified in a ≥80% mild to moderate population and that reported relevant outcomes [redacted].

Within the company's preferred vehicle response adjusted NMA model, it is unclear whether the assumptions are reasonable, particularly the assumption that the relationship between the population baseline response and relative treatment effect should be the same across treatments. The MAIC only provides inferences relative to the populations defined by the comparator studies, and it is unclear whether all prognostic factors and treatment effect modifiers have been accounted for. This means that the relative effectiveness of TCIs and crisaborole is highly uncertain.

The key area of uncertainty in the economic model relates to uncertainty in the relative effectiveness of TCIs and crisaborole. The ERG notes that the cost-effectiveness results are being driven by the reduced need for subsequent therapies in patients who achieve a response to the first modelled treatment and the ERG is concerned that the cost savings modelled may not be accurate given the simplistic manner in which subsequent therapies have been modelled.

1.7 Summary of exploratory and sensitivity analyses undertaken by the ERG

The ERGs exploratory analysis which used the median HRs from the ERG's simple random effects NMA demonstrates the sensitivity of the cost-effectiveness results to the uncertainty in the relative effectiveness; crisaborole was found to be dominated by TCIs in this exploratory analysis whereas crisaborole dominated TCIs in the exploratory analyses using the company's vehicle adjusted NMA (except in two scenarios for adults patients with moderate AD where the ICER compared with tacrolimus 0.1% was above £30,000 per QALY using the company's vehicle adjusted NMA). However, the ERG notes that generally the absolute difference in costs and QALYs between crisaborole and TCIs is small across the majority of the scenarios explored.

[redacted]

The other conclusion from the ERG's exploratory analyses is that the cost-savings and QALY gains achieved by using a more effective primary therapy are sensitive to the assumptions made regarding the time taken to achieve an adequate response to subsequent therapies, the duration of subsequent therapy and the need for monitoring during subsequent therapies.

2 BACKGROUND

2.1 Critique of company's description of underlying health problem

The ERG believes that the company's description of the underlying health problem is broadly appropriate and relevant to the population under consideration as specified in the final NICE scope.(2)

Disease presentation

The CS describes AD, also referred to as atopic eczema, as a chronic inflammatory condition that presents with persistent, relapsing or recurrent skin manifestations. There is variation between individuals in terms of susceptibility, onset, presentation and course of AD.(5)

AD may affect any part of the body and tends to show age-related predilections for specific regions such as the face, scalp, neck, hands, upper body and body folds of the elbows and knees. AD can present in differing patterns; for example, AD affecting the creases of the neck, wrists, elbow and knee joints (flexural pattern) or presenting as coin-sized inflamed lesions on the extremities (discoid pattern) or several raised hair follicles on the scalp (follicular pattern) are more commonly seen than AD affecting the face.(6) Additionally, AD in people with skin of colour, i.e. those of African, Afro-Carribean or Asian ethnicity, may present patterns of the condition which are different from those seen in Caucasians.(7) Atypical presentations include affected skin in the front of the elbows and knees (reverse flexural pattern) or red papules or bumps on the scalp, chest or abdomen (papular pattern). (7)

The disease course in AD is unpredictable and people with mild or moderate AD may have periods of remission followed by periods of active AD (CS, p.18(1)). These periods of "*increased disease activity*" are described as flares. The CS states that there is little consistency in the definition of disease flares across clinical studies and this makes it difficult to compare the rate of flares quantitatively between patients.(1) The ERG's clinical advisors agreed that this was a limitation of the current evidence base. This is supported by a recent narrative review by Abuabara *et al*. which states that the, *"frequency and duration of active disease periods remains poorly understood"*.(8)

The most common day-to-day symptoms are dry skin, cracked or raw skin and itching (see CS, Figure 1(1)). These symptoms are associated with a psychological and psychosocial burden as pruritus (itching) can lead to sleep disruption and the presence of visible skin lesions can affect self-esteem and social interactions (CS, p. 21(1)).

Prevalence

Prevalence and incidence estimates of AD in children and adults vary considerably between studies.(2) (9) Differences in genetic, environmental and methodological factors including study design, data collection methods, and diagnostic criteria may explain these differences.(10) AD is commonly seen in people with allergies (including asthma) or those with a family history of similar conditions (such as asthma and allergic rhinitis).(11)

Onset in the first few years of life is common (child-onset or early-onset) but AD may occur for the first time in adulthood (adult-onset or late-onset).(12) The majority of children (around 75%) outgrow their AD in childhood or early adolescence but a proportion have persistent symptoms in adulthood (p130, CS).(1) In addition, around 1 in 4 adult AD patients are said to have adult-onset AD (CS, p. 17,(1)) and some adults present with more severe or persistent AD.(13)

Prevalence estimates of AD presented in the CS were comparable to rates presented in the NICE scope which states that, in the UK, an estimated 1 in 5 children (20%) have AD compared with 1 in 12 adults (8.3%).(2) The CS (p. 17) estimates the prevalence of AD in children as between 5.3% and 23.1%, and the prevalence of AD in adults as 5.1% in those aged 18 to 74 years and 8.7% in those aged 75 to 99 years. Prevalence estimates presented in the CS(1) were obtained from a published study that analysed data from The Health Improvement Network (THIN) covering the duration 1994 to 2013 to estimate age-specific prevalence rates of '*active (as defined by diagnosis and treatment codes applied by physicians) atopic eczema'*.(14) THIN is a computerised database that holds anonymised information of over 9 million patients seen in an estimated 500 UK general practices and is regarded as a sufficiently representative of patients seen in seen primary care.(15)

Disease severity assessment

The population is described as patients aged 2 years and over with mild to moderate AD. Measures of AD severity are required both to define the initial population and to assess treatment effectiveness. The ERG notes that the Harmonising Outcome Measures in Eczema (HOME) initiative(16) has published guidance suggesting that AD trials should measure four core types of outcome: 1) clinical signs; 2) patient-reported symptoms; 3) long-term control, and 4) quality of life.

In terms of clinical signs, the CS(1) (p. 18) describes a range of AD severity measures, but states that "*there is no consensus on which scoring instruments should be employed to measure AD severity in clinical trials*". One type of measure can be classed as "*global assessment*" and has many names including Investigator's Global Assessment (IGA), Investigator's Static Global Assessment (ISGA) and Physician's Global Assessment (PGA). These provide a snapshot of disease severity on a scale

ranging from 4 to 7 points. While there is variability between IGA instruments, they are quick and simple to use, and the US Food and Drug Administration (FDA) mandates their use in AD trials.(17)

Other measures based on clinical signs include the Eczema Area and Severity Index (EASI), and the Scoring Atopic Dermatitis index (SCORAD). The ERG notes that the HOME initiative specifies that clinical signs should be measured using EASI.(16) Clinical advisors to the ERG noted that a benefit of the EASI tool is that it provides a total score that takes into account the body regions affected, the area affected in each region and the clinical signs (erythema, oedema/papulation, excoriation, lichenification) that are present in each region. A more detailed discussion of the various tools for measuring disease severity can be found in Section 4.2.3.

The HOME initiative(16) and clinical advisors to the ERG also advise that a holistic patient assessment should include not just a measure of clinical signs as covered by the ISGA/EASI scores but also a measure of patient-reported symptoms such as the Patient-oriented Eczema Measure (POEM), and a measure of quality of life such as the Dermatology Life Quality Index (DLQI) score or Children's Dermatology Life Quality Index (CDLQI) score. The use of these measures in clinical practice as part of a holistic assessment is also supported by NICE CG57 (Atopic eczema in under 12s: diagnosis and management).(4)

The ERG's clinical advisors acknowledged the difficulty of assigning a single severity classification to a patient, whilst taking into account fluctuations in disease severity across different body sites and at different time points. NICE CG57 states that AD of differing severity can coexist in the same child and, in this case, each area should be treated independently.(4) This corresponds with the company's view that "*a patient may have co-occurring mild and moderate lesions, requiring different treatment regimens"*. (CS, p. 25 (1)). Therefore, the ERG notes that the population may be interpreted in clinical practice as including patients with areas of mild to moderate AD who also have severe AD elsewhere, which is being managed separately.

2.2 Critique of company's overview of current service provision

Stepped approach to AD treatment

The CS(1) describes the recommendations for treatment in NICE CG57 (Atopic eczema in under 12s: diagnosis and management). The step-wise treatment approach recommended by NICE in CG 57(4) is summarised in Table 1. It is noted that NICE does not currently offer any guidance for the diagnosis and management of adults (or people aged 12 and over) with mild to moderate AD. However, one clinical advisor to the ERG stated that it was reasonable to apply the recommendations for treatment of children with AD to affected adults.

Table 1: Stepped approach to treatment of childhood AD recommended in NICE Clinical Guideline 57

Mild AD	Moderate AD	Severe AD
Emollients	Emollients	Emollients
Mild-potency topical corticosteroids	Moderate potency topical corticosteroids	Potent topical corticosteroids
	Topical calcineurin inhibitors	Topical calcineurin inhibitors
	Bandages	Bandages
		Phototherapy
		Systemic therapy

The CS(1) states that mild to moderate AD should be managed at the primary care level but they also describe audit data from the British Association of Dermatologists (BAD) which show that 29% and 44% of patients managed in secondary care have mild or moderate AD, respectively. The ERG notes that these figures do not quantify the proportion of patients with mild to moderate disease appropriately managed in primary care because no information is available from this audit on the numbers not referred to secondary care. The ERG notes that data from the same audit show that dermatologists considered that the referral to secondary care has been appropriate in 92% of cases, suggesting the frequency of unnecessary referrals to secondary care is low.

Use of topical corticosteroids (TCS)

Whilst the stepped approach in NICE CG57 recommends that mild potency TCSs should be used for mild AD and moderate TCSs for moderate AD, additional recommendations in the NICE CG57 state *"… If a mild or moderately potent topical corticosteroid has not controlled the atopic eczema within 7–14 days …. In children aged 12 months or over, potent topical corticosteroids should then be used for as short a time as possible and in any case for no longer than 14 days."* (NICE CG57, section 1.5.3.6)(4)

Therefore, the use of potent steroids is permitted for short periods where mild or moderately potent TCSs have failed.

The ERG's clinical advisors stated that whilst the use of TCSs in primary care is fairly consistent with the NICE CG57, primary care physicians are often reluctant to step treatment up to more potent TCSs. In contrast, dermatologists in secondary care would be more likely to treat patients with mild to moderate AD with short courses of potent TCSs if an adequate response had not been achieved with moderate potency TCSs.

The CS(1) states that "*prolonged or mis-use of TCSs can result in serious side-effects*" and goes on to describe both local adverse effects and the risk of systemic adverse effects when TCSs are used in excess over large body surface areas. They state that up to 73% of patients and carers report "*steroid phobia*" (p26 of CS).(1) The ERG's clinical advisors described how patients, carers and primary care physicians can be overly cautious regarding the use of TCSs because of the perceived risks of side-effects, in particular skin thinning; but that specialist dermatologists consider the risk of skin thinning to be low when TCSs are prescribed appropriately and used as directed, especially following adequate patient education. In addition, they advised that a short course of more potent TCSs is often a better option than persisting with low potency TCSs because it achieves rapid relief of symptoms and brings the AD under control for a longer period by breaking the scratch-itch cycle thereby avoiding the need for prolonged use of milder potency TCSs. The CS states that long-term TCS use may be associated with dyspigmentation in patients with skin of colour. The ERG's clinical advisors stated that skin pigmentation is rarely affected by mild to moderate potency TCSs, but topical calcineurin inhibitors (TCIs) may sometimes be offered as an alternative to TCSs when the potential for dyspigmentation is a concern, particularly for AD on the face.

Use of topical calcineurin inhibitors (TCIs)

The CS states that TCIs may be prescribed for patients who have failed to respond to TCSs, or for whom TCSs are inadvisable due to the affected area of the body, or concerns over low-adherence to TCS use (CS, p. 27(1)). The licensed indications and NICE recommendations for TCIs are provided in Table 2. The ERG's clinical advisors stated that pimecrolimus 1% and tacrolimus 0.03% were considered comparable in potency to mild and moderate TCSs such as hydrocortisone 1% and clobetasone butyrate, whilst tacrolimus 0.1% was considered more comparable to moderate-to-potent TCSs. The ERG's clinical advisors confirmed that, in keeping with the recommendations in NICE TA82(3), TCSs were generally used prior to TCIs, as per Figure 1, and TCIs would be reserved for those patients where there were particular concerns regarding further use of TCSs.

The CS states that concerns regarding the potential risk of skin malignancy with TCIs may lead to poor adherence, that TCIs are associated with skin-burning that may be sufficient to justify discontinuation and that TCIs may transiently worsen AD on acutely inflamed skin (CS, p. 27(1)). The ERG's clinical advisors agreed that TCIs were an option for patients who were unable to use TCSs. They agreed that TCIs may exacerbate a flare and they advised that it was often better to use TCSs for a few days to control the acute inflammation before introducing a TCI.

Table 2: UK licensed and recommended topical calcineurin inhibitors for mild to moderate AD (reproduced from CS Table 7)

	Severity	Adult	Children
Licensed	**Mild**	Pimecrolimus 1%	Pimecrolimus 1%
	Moderate	Pimecrolimus 1% Tacrolimus 0.03% Tacrolimus 0.1%	Pimecrolimus 1% Tacrolimus 0.03%
NICE recommended	**Mild**	None	None
	Moderate	[a]Tacrolimus 0.03% [a]Tacrolimus 0.1%	[b]Pimecrolimus 1% [a]Tacrolimus 0.03%

[a]Tacrolimus only recommended by NICE for second-line use when uncontrolled on TCS or serious risk of AEs from TCS (particularly irreversible skin atrophy)
[b]Pimecrolimus only recommended by NICE for use on the face and neck, and only second-line when uncontrolled on TCS or serious risk of AEs from TCS (particularly irreversible skin atrophy)

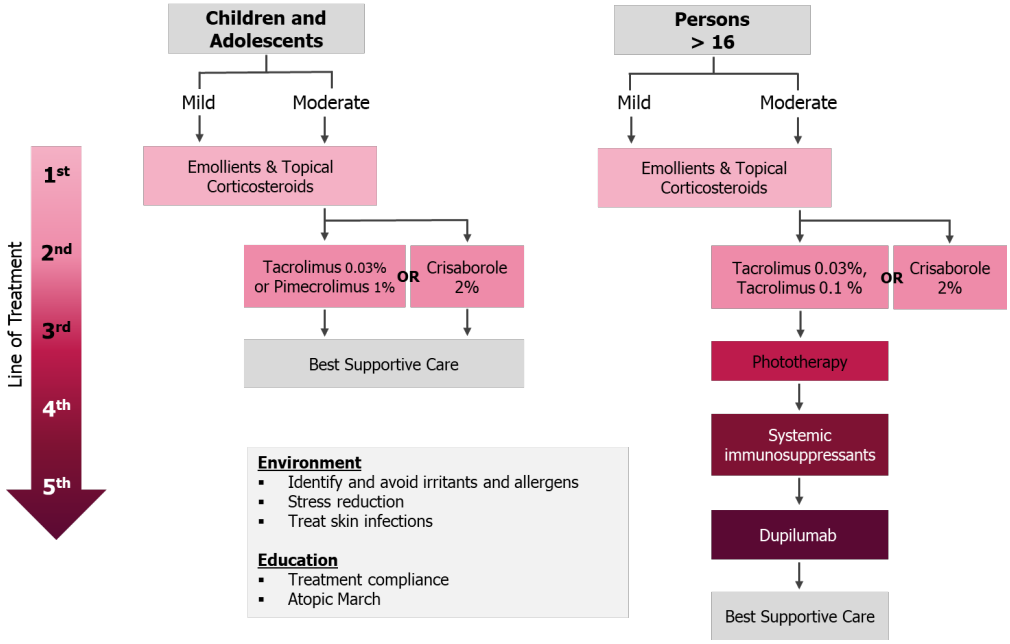
Other interventions

Bandages are also included in NICE CG57 as part of the stepped approach presented in Table 1. The CS states that dry bandages and medicated dressings should only be used with emollients and TCSs for short-term (7-14 days) treatment of flares or areas of chronic lichenified AD, and that TCIs should not be used under bandages and dressings (CS, p26(1)). The role of bandages in the treatment pathway is not discussed further in the CS. The ERG notes that wet wraps were not included in the model for TA82 due "*to a lack of clarity about where wet-wrapping fits in the overall treatment pathways as well as lack of data about costs and effectiveness*." The company's economic model does not include bandages as subsequent treatments but this was considered reasonable given that they were not included in the model for TA82.(3)

Indication for crisaborole

The company describes its proposed reimbursement indication for crisaborole as adults and children aged 2 years and older with mild to moderate AD that has not been controlled by TCS or where there is a serious risk of important adverse effects from further TCS use, particularly irreversible skin atrophy (CS, p. 27(1)). Therefore, the proposed positioning in the treatment pathway for crisaborole is as an alternative to TCIs for second-line use post-TCS treatment (see Figure 1 - reproduced from Figure 2 of the CS).(1) The ERG's clinical advisors suggested this was a reasonable position for crisaborole given that clinicians would continue to use TCSs as first-line treatment where possible. The ERG's clinical advisors also expressed concerns whether crisaborole was sufficiently potent to use when moderate potency TCSs had failed, as they expected crisaborole to be similar in effectiveness to mild TCSs.

Figure 1: Proposed positioning of crisaborole in the reimbursed treatment pathway [reproduced from CS Figure 2]



Note: Figure reflects reimbursed treatment options. Although pimecrolimus 1% is licensed for use in persons > 16, it is not reimbursed. Tacrolimus 0.1% is only licensed for the treatment of adults with moderate-to-severe disease. It is noted for simplicity bandages have not been included in the treatment pathway above.

Subsequent treatments

Figure 1 also shows the subsequent treatments that would follow second-line treatment with either TCIs or crisaborole. It shows a sequence of phototherapy, followed by systemic immunotherapy, followed by dupilumab and then best supportive care (BSC) for adults. However, for children, the only option in Figure 1 after second-line treatment with TCIs or crisaborole is BSC. The ERG notes that this is inconsistent with the stepped care approach from NICE CG57 which shows systematic therapy and phototherapy as later treatment options for severe AD in children (see Table 1)(4). Clinical advisors to the ERG further noted that treatment escalation to systemic immunosuppressants and phototherapy was generally restricted to patients with uncontrolled severe AD or a subgroup of moderate AD patients with severe clinical presentations. The BAD audit data show that in secondary care systemic treatments are used in a small proportion (under 2%) of children with mild to moderate AD, but they are used around 23% of children with severe AD. The ERG's clinical advisors confirmed that whilst the licensed indications for systemic treatments such as methotrexate, ciclosporin, azothioprine or mycophenolate mofetil do not cover children, they are used off-license in children with severe AD by specialist dermatologists. Dupilumab is also not licensed in children under 12.

Overall, the ERG considers that the description of current service provision in the CS inadequately describes the treatment pathway in children following second-line treatment with TCIs. The ERG also considers that the CS does not properly represent the option for using a short course of potent TCSs in children and adults with mild to moderate AD who have not responded to a mild or moderately potent TCS when secondary infection or other triggers of flares have been excluded.

3 CRITIQUE OF COMPANY'S DEFINITION OF THE DECISION PROBLEM

The decision problem is summarised in Table 3.

3.1 Population

Company decision problem: The population in the company decision problem is generally consistent with the NICE scope,(2) i.e. adults and children aged 2 years and older with mild to moderate atopic dermatitis (AD). The CS focusses on a second-line (post-TCS) population, i.e. people uncontrolled on TCS or with serious risk of important adverse effects from TCS (particularly irreversible skin atrophy). However, scenario analyses for a first-line (TCS eligible) population are also presented.

Clinical trial evidence: The crisaborole clinical trials are consistent with the NICE scope in that they included adults and children with mild to moderate AD (although only 14% of participants were adults). However, the trials did not select participants who were uncontrolled on TCS (or at serious risk of AEs from TCS). Although the company has presented *post hoc* analyses which explored the impact of prior usage of TCSs on clinical effectiveness within the AD-301 and AD-302 trials, the ERG notes that having had prior TCS use is not the same as having uncontrolled disease on TCS.

3.2 Intervention

The intervention is consistent with the NICE scope: crisaborole (20mg/g i.e. 2%). Crisaborole is a topical ointment which controls inflammation by inhibiting phosphodiesterase 4 (PDE4), thereby inhibiting cytokine production. Crisaborole is currently under EMA assessment. The proposed marketing authorisation for crisaborole is as a flare-based treatment for mild to moderate AD in adults and children aged 2 years and older with ≤40% body surface area (BSA) affected. Crisaborole is intended to be applied twice daily to a maximum of 40% BSA, for up to 4 weeks per treatment course, and to be discontinued if symptoms persist after 3 consecutive 4-week courses.

The ERG notes that if the final licensed indication is consistent with the proposed indication, then this would effectively restrict the relevant population for the decision problem to those with <40% BSA affected. The ERG's clinical advisors noted that the majority of patients with mild to moderate AD would have <40% BSA affected.

Clinical trial evidence: Evidence for effectiveness of crisaborole consists of two identical 4-week RCTs of crisaborole vs. vehicle ointment in adults and children 2 years of age or older with mild to moderate AD (AD-301 and AD-302). Patients in the RCTs could continue into a 48-week single-arm extension study (AD-303), which reported on crisaborole exposure, subsequent therapy use and

safety, but did not report effectiveness outcomes. In the clinical trials, BSA ranged from 5 to 95% and [redacted] (proportion estimated by ERG from Table 28 of the CS(1)).

3.3 Comparators

Comparators for first-line treatment (TCS eligible)

Company decision problem: For first-line (TCS eligible) patients with mild and moderate AD, the comparators in the company decision problem are generally consistent with the NICE scope, i.e. emollients plus TCS. In terms of TCS potency, the NICE scope specifies mild to moderate potency TCS for mild AD and high-potency TCS for moderate AD.(2) However, the CS (Section B.3.3.1) assumes a mix of TCS potencies ranging from mild to high potency for both mild and moderate AD.(1) The mix of TCS potencies was based on data from the British Association of Dermatologists (BAD) audit in children(18); these data were considered to be a reasonable reflection of current practice by clinical advisors to the ERG. The ERG's clinical experts noted that short courses of potent TCSs may be used in patients with mild to moderate AD when an adequate response has not been achieved by moderate potency TCSs. In addition, it is often difficult to classify a patient into a single disease severity when the severity of symptoms may fluctuate or vary between body areas and therefore patients classified as having mild to moderate AD overall may have areas of more severe AD that requires a more potent treatment.

Clinical evidence: The first-line population required comparison of crisaborole against TCS. However, the crisaborole trials compared against vehicle ointment. The CS(1) (Section B.3.3.1) states that no trials of TCS could be included in the network meta-analysis (NMA) as no TCS trials were identified in mild to moderate AD which reported the outcomes of interest. Therefore, for the comparison of crisaborole vs. TCS in the company model, transition probabilities from TA82(3) were applied. Given a lack of data in the relevant population, these transition probabilities from TA82(3) were not based on trial outcomes in patients with mild to moderate AD but were instead based on assumptions informed by clinical experts. On the other hand, there are two ongoing trials comparing crisaborole vs. TCS (NCT03539601) and crisaborole vs. crisaborole plus TCS (NCT04008784) (See Section 4.2.9).

Comparators for second-line treatment (post-TCS)

The main comparators in the CS for a second-line (post-TCS) population are the topical calcineurin inhibitors (TCIs) tacrolimus and pimecrolimus.(1) Tacrolimus is licensed for moderate AD, and is recommended by NICE for moderate AD in adults and children in a second-line (post-TCS) population. Pimecrolimus is licensed for mild to moderate AD in adults and children, and its NICE recommendation is restricted to moderate AD on the face and neck in children in a second-line (post-TCS) population.

Company decision problem (mild AD): For second-line (post-TCS) treatment of mild AD, there are no NICE-recommended treatments. The CS compares crisaborole against pimecrolimus because it is licensed for mild AD. However, pimecrolimus is not recommended by NICE for mild AD in either adults or children. Clinical advisors to the ERG indicated that few patients with mild AD would not respond to TCS and so the proportion of patients with mild AD requiring post-TCS treatment would be small. The BAD audit data indicates that 8% of mild AD patients receive TCIs,(18) and this is given by the company as justification for including a comparison against pimecrolimus in mild AD. The CS does not include tacrolimus as a comparator for mild AD, since tacrolimus is not licensed or recommended for mild AD. Emollients in combination with TCS are listed as the comparator for mild AD in the final NICE scope.(2) However, it is unclear from the scope whether emollients alone is a relevant comparator for those patients with mild AD who are not eligible to receive TCSs.(2) The CS does not provide a comparison of crisaborole against emollients alone in mild AD. The company's justification for this is that they consider it extremely unlikely that patients whose symptoms were uncontrolled on TCSs would receive only emollients and data from the BAD audit show that some patients with mild AD receive TCIs. The ERG's clinical experts agreed that it was unlikely that patients with uncontrolled symptoms would be managed with emollients alone if further treatment with TCSs was not an option.

Company decision problem (moderate AD): For second-line (post-TCS) treatment of moderate AD, comparators are broadly consistent with the NICE scope, i.e. TCIs in line with their NICE recommendations (adults: tacrolimus 0.1% and 0.03%; and children: tacrolimus 0.03% and pimecrolimus 1%).(1) However, the ERG notes that whilst NICE guidance recommends pimecrolimus (1%) only for use on the face and neck in children, the CS(1) does not differentiate between children with AD on the face and neck and children with AD on other areas of the body in their economic analysis.

Clinical evidence: The second-line population required comparison of crisaborole against TCIs. Given that the crisaborole trials compared against vehicle ointment, an NMA was conducted which included crisaborole trials and TCI trials in mild to moderate AD with vehicle as the reference treatment.

3.4 Outcomes

Company decision problem: The company decision problem (CS, Table 1(1)) erroneously states that all outcomes listed in the NICE scope (except time to relapse) are used in the economic model. The decision problem shown here (Table 3) has been adapted by the ERG to reflect the outcomes used in each stage of the CS.

Clinical evidence: Of the outcomes in the NICE scope, the crisaborole clinical trials report disease severity (ISGA scores), symptom control (pruritus and clinical signs), adverse effects and health-related quality of life. Clinical advisors to the ERG indicated that other AD severity measures such as EASI or SCORAD may provide more granularity of response than ISGA; however, these outcomes were not recorded in the crisaborole trials. There were no data on disease-free period or time to relapse, as the RCTs were only 4 weeks in duration. However, there is an ongoing 52-week RCT of maintenance treatment with crisaborole vs. vehicle (NCT04040192) (See Section 4.2.9).

The NMA to compare crisaborole vs. TCIs assessed one effectiveness outcome (disease severity: proportion achieving ISGA/IGA score of 0/1) as well as safety outcomes.

Outcomes used in company model: The economic model incorporated the following outcomes from the clinical trials: treatment response (proportion achieving ISGA/IGA score of 0/1); partial response in moderate AD (proportion with 1-point improvement from ISGA 3 to 2); health-related quality of life; and drug use per application. The CS states that pruritus is not directly included in the model because it is correlated with ISGA.(1) The CS states that adverse effects are not incorporated in the model because the main effect of crisaborole is application site pain (which is short-term and difficult to capture disutility), while long-term adverse effects (such as possible risk of lymphomas associated with TCIs) cannot be incorporated because of a lack of long-term data on crisaborole.(1) The model includes an estimate of the number of flares per year based on an observational study and not data on time to relapse from the crisaborole RCTs. The base case cost-effectiveness analysis does not include any data on prevention of relapse, although a scenario analysis includes an estimate of the reduced risk of flares in patients receiving maintenance therapy with tacrolimus as a scenario analysis.

Subgroups:

The CS presents data on a number of subgroups consistent with the NICE scope, including adults and children, mild and moderate AD, adults with mild and moderate AD, children with mild and moderate AD, as well as prior AD treatment use (yes/no) and % BSA affected.(1) Cost-effectiveness analyses are presented for the first-line (TCS eligible) and second-line (post-TCS) populations. Separate cost-effectiveness analyses are provided for:

- adults with mild AD
- adults with moderate AD
- children with mild AD
- children with moderate AD.

The company's base case analysis uses outcomes from the company's preferred NMA (i.e. fixed treatment effect, random class effect and adjusted for vehicle response) but not adjusted for age (adult or child) or severity (mild or moderate). It therefore provides an estimate of the response rates for a population with an average and severity distribution that matches the trials included in the NMA. However, sensitivity analyses are provided in which the response rates applied in the model are further adjusted for both age and severity. A sensitivity analysis is also provided in which the response rates are adjusted according to prior AD treatment.

The CS states that there were insufficient data (for crisaborole and/or comparators) to present data for the following subgroups specified in the NICE scope: people with different skin colour, AD affecting the hands, and AD affecting sensitive areas (face, neck and flexures).(1) The ERG notes in particular that cost-effectiveness analyses are not provided for the subgroup of children with moderate AD on the face and neck, which is the only group in which NICE recommends pimecrolimus.

Table 3: Decision problem (adapted from CS Table 1)

	Final scope issued by NICE(2)	Decision problem addressed in the company submission (adapted by ERG)	Rationale if different from the final NICE scope (adapted by ERG)
Population	• People aged 2 years and older with mild to moderate AD	Adults and children aged 2 years and older with mild to moderate AD • second-line patients post-TCS (uncontrolled on TCS or serious risk of important AEs from TCS, particularly irreversible skin atrophy) are presented as the target population in the CS • first-line (TCS eligible) population also presented as a scenario analysis	Company states that "*Whilst crisaborole has been demonstrated to be safe and effective for the first-line treatment of adults and children aged 2 years and older with mild to moderate AD, it is not anticipated that crisaborole will be recommended in the UK for first-line treatment (TCS eligible patients), due to the very low cost of topical corticosteroids*" (p10, CS)(1)
Intervention	• Crisaborole ointment (20mg/g)	• Crisaborole ointment (20mg/g)	Consistent with NICE scope
Comparator(s)	**Mild atopic dermatitis:** • Emollients + mild to moderate potency TCS **Moderate atopic dermatitis:** • High potency TCS • TCIs	**Mild atopic dermatitis:** • First-line (TCS eligible): - Emollients + mild to moderate potency TCS • Second-line (post-TCS): - TCIs: Adults and children: pimecrolimus 1% **Moderate atopic dermatitis:** • First-line: - Emollients + moderate to high potency TCS • Second-line (post-TCS): - TCIs: Adults: tacrolimus 0.1%, tacrolimus 0.03% Children: tacrolimus 0.03%, pimecrolimus 1%	**Mild atopic dermatitis:** • First-line: Mostly consistent with NICE scope. CS assumes a mix of TCS potencies ranging from mild to high potency based on BAD audit data in children.(18) • Second-line: No NICE-recommended treatments; BAD audit data indicates 8% of mild AD patients receive TCIs.(18) Therefore, CS compares against pimecrolimus which is licensed but not recommended for mild AD (tacrolimus is not licensed for mild AD). Emollients alone are not considered a relevant comparator by the company as BAD audit data(18) show that patients with mild AD are in practice treated with TCIs. **Moderate atopic dermatitis:** • First-line: Mostly consistent with NICE scope. CS assumes a mix of TCS potencies ranging from mild to highly potent based on BAD audit data in children.(18) • Second-line: Mostly consistent with NICE scope. The ERG notes that the comparison with pimecrolimus 1% should be restricted to children with AD on the face and neck.

	Final scope issued by NICE(2)	Decision problem addressed in the company submission (adapted by ERG)	Rationale if different from the final NICE scope (adapted by ERG)
Outcomes	Outcome measures include: • Disease severity • Symptom control • Disease free period / maintenance of remission • Time to relapse / prevention of relapse • Adverse effects of treatment • Health-related quality of life	**Clinical evidence from crisaborole trials:** • Disease severity (ISGA scores) • Symptom control (pruritus, clinical signs) • Adverse effects of treatment • Health-related quality of life **NMA to compare crisaborole vs. TCIs:** • Disease severity (ISGA/IGA 0/1) • Adverse effects of treatment **Trial outcomes included in economic model:** • Treatment response based on disease severity (proportion achieving ISGA/IGA 0/1) • Partial response in moderate AD based on a 1 point improvement from ISGA 3 to 2 • Health-related quality of life • Drug use per application	Of the outcomes in the scope, the following were not covered in the CS, as the RCTs were only 4 weeks in duration and the extension study did not report them: • Disease free period / maintenance of remission • Time to relapse / prevention of relapse The economic model only used trial data on disease severity (ISGA), HRQoL and drug use per application. The CS states that pruritus is captured as it is correlated with ISGA. The CS states that adverse effects are not incorporated in the model because the main effect of crisaborole is application site pain (short-term and difficult to capture disutility), while long-term adverse effects of TCIs (such as possible lymphoma risk) cannot be incorporated due to lack of long-term data on crisaborole. Data on time to relapse and prevention of relapse were based on other clinical studies and not outcomes from the crisaborole clinical trials.
Economic analysis	• Cost-effectiveness should be expressed as incremental cost per quality-adjusted life year • Time horizon should be sufficiently long to reflect any differences in costs or outcomes between technologies • Costs should be considered from an NHS and Personal Social Services perspective	Cost effectiveness model has been designed in line with the NICE reference case criteria: • Outcomes are expressed in terms of incremental cost per quality-adjusted life year • Time horizon: Lifetime for adults, up to age 18 for children • Costs have been considered from an NHS and Personal Social Services perspective	The company states that a lifetime horizon is consistent with the final scope. However, a shorter horizon of up to 18 years is used in children. The company states that this is because once patients reach 18 they are considered to be adults and adults are exposed to a different clinical pathway and are subject to different efficacy estimates. A detailed comparison of company's cost-effectiveness analysis against the NICE reference case is provided in Table 30.

	Final scope issued by NICE(2)	Decision problem addressed in the company submission (adapted by ERG)	Rationale if different from the final NICE scope (adapted by ERG)
Subgroups	If evidence allows, the following subgroups will be considered: • Adults and children • Mild and moderate AD • Different skin colour • AD affecting the hands • AD affecting sensitive areas (face, neck and flexures) • People for whom therapies have been inadequately effective, not tolerated or contraindicated	**Subgroups presented for:** • Adults and children • Mild and moderate AD • Adults with mild and moderate AD • Children with mild and moderate AD • Prior AD treatment use (yes/no) • % BSA affected **Cost-effectiveness analyses presented for the following populations, but using intention to treat (ITT) trial data, not subgroup data:** • Second-line (post-TCS) population • First-line (TCS eligible) In both of these populations separate cost-effectiveness analyses are presented for; • Children with mild AD • Children with moderate AD • Adults with mild AD • Adults with moderate AD	Insufficient clinical evidence to present separate subgroup analyses for: • Different skin colour (not consistently reported across comparator trials) • AD affecting the hands (not consistently reported across comparator trials) • AD affecting sensitive areas, i.e. face, neck and flexures (not available for crisaborole) • People for whom therapies have been inadequately effective, not tolerated or contraindicated

4 CLINICAL EFFECTIVENESS

4.1 Critique of the methods of review(s)

4.1.1 Searches

The company performed one clinical effectiveness search to identify all clinical and safety studies of treatments in mild to moderate atopic dermatitis.

The Global Resource for Eczema Trials (GREAT) was searched in March 2019 for systematic reviews and randomised controlled trials from inception until September 2017. Update searches by the company for RCTs in electronic databases were undertaken to cover the period between January 2017 until March 2019. Multiple electronic and online databases were searched: MEDLINE and MEDLINE in Process [via Ovid], Embase [via Ovid], CINAHL [via EBSCO], Cochrane Library including Cochrane Database of Systematic Reviews and Cochrane Central Register of Controlled Trials [via Wiley], Database of Abstracts of Reviews of Effects [via CRD], NHS Economic Evaluation Database [via CRD], Health Technology Assessment Database [via CRD], EconLit [via EBSCO], LILACS and Epistemonikos [Epistemonikos Foundation].

The company's updated searches in MEDLINE, Embase, CINAHL, were restricted to English language only. According to the Cochrane Handbook,(19) it is best practice not to apply language restrictions in the search strategy to prevent the risk of language bias.

Both the conference proceedings database in Embase [via Ovid] and seven conference websites were searched by the company between 2015-2018 (American Academy of Dermatology, British Association of Dermatologists, European Academy of Dermatology and Venereology, European Society for Pediatric Dermatology, International Congress of Dermatology, International Symposium on Atopic Dermatitis, and World Congress of Dermatology).

Several HTA websites (NICE, SMC, CADTH, PBAC, Gertner Institute, AHRQ, ICER and HOME) including the FDA.gov were searched to cover the period 2014 and May 2019. One trials registry was searched (clinicaltrials.gov). Supplementary searches by the company included scanning of bibliographies of included studies, and hand searching FDA submissions (CS Appendix D(20) page 9).

The ERG requested a copy of the search strategies and sources searched within the GREAT database to identify trials of eczema treatments for literature before 2017.(21) The GREAT database search strategies in MEDLINE and Embase (from 2000 onwards) are provided in Appendix 1 (clarification response, question A1(21)). In the update search from 2017 onwards, the company's search strategies differed from

the GREAT database search strategies in terms of keywords and MeSH /Embtree headings, RCT search filters and limits applied (contrast with CS Appendix D1, Tables D1 and D2, p. 13-14,(20)).

The company investigated the impact of the different strategies by comparing the sensitivity and specificity of the search against the included studies retrieved (clarification response, question A2). The comparisons revealed that the total number of records retrieved and required to screen in the update search were less by comparison to the GREAT database searches. The company confirmed that the updated searches did not compromise the search sensitivity even though it was more specific (clarification response, question A2(20)). The company added that only three of the included studies were not retrieved by both GREAT MEDLINE or Embase search strategies (added to the database via hand searching) or in the company's updated searches and were identified via other search methods (e.g. hand searching, recursive searching). The ERG considers that the company search strategies are sufficiently comprehensive to retrieve important citations relating to all eligible studies. The company's comparisons of the GREAT and updated searches raise the question of the purpose and value of approach to searching additional databases beyond MEDLINE and Embase.

4.1.2 Inclusion criteria

Relevant RCTs were identified using pre-specified inclusion criteria presented in Appendix D (Table D 10, p. 24(20)). The criteria for selection of studies are summarised in Table 4. Measures of health-related quality of life were not listed as relevant outcomes. However, the ERG notes that the inclusion criteria (Appendix D, Table D 10(20)) were broadly appropriate to identify suitable RCTs to inform the decision problem. Two independent reviewers selected relevant studies in a two-staged process, with disagreements resolved by discussion between reviewers or referral to a third reviewer. Included studies and excluded studies with reasons are listed in Appendix D (p.38 to 163(20)). Whilst the ERG considers the approach to selection process to be broadly acceptable, there was limited information in the CS(1) about the SLR process.

Table 4: Selection criteria for RCTs included in the SLR and NMA (adapted from Appendix D, Table D10)

PICOS	Inclusion	Exclusion
Population	● Children (2 years and older) and adults with a clinical diagnosis of mild to moderate AD (≥80% of trial population) - All diagnostic criteria for AD and all scales/scores used in assessing disease severity are eligible for inclusion	● More than 20% of the trial population are: - Healthy volunteers - Paediatric patients (<2 years) Severe disease: trials focusing solely on subjects with severe disease
Intervention or comparators	● Crisaborole ointment, 2% (Eucrisa®, Staquis®, formerly AN-2728) ● TCSs (mildly potent; moderately potent and potent), as a monotherapy or in combination (e.g. with salicyclic acid 2 to 3%) ● Topical Calcineurin Inhibitors (TCIs) (e.g. tacrolimus 0.03%/ 0.1%, pimecrolimus 1%) ● Best supportive care ● Alitretinoin (in people with AD affecting the hands) ● Other Phosphodiesterase-4-inhibitor: OPA-15406 ● Calcineurin inhibitors: SB011 ● Placebo/vehicle control (may be used as a comparator)	● Trials comparing only interventions/ comparators not approved in the UK ● Non-pharmacologic treatment/ doses for atopic dermatitis in all trial study arms ● Systemic treatments ● Trials that include only the comparison of an emollient vs. a non-treatment placebo ● Treatments intended primarily for infections (e.g. antibiotics)
Outcomes	1. AD severity: ISGA scale - ISGA success: Proportion with ≥2 grades of improvement to clear (0) or almost clear (1) - ISGA improvement: Proportion with clear (0) or almost clear (1) - IGSA 0 to 1 = PGE of 90% or better Other commonly used scales (in case ISGA is not available) e.g. - EASI/ mEASI - SCORAD - Investigator/Physician Global Assessment scales including: IGA, PGA, PGE - Other AD severity scales 2. Pruritus severity (all scales and definitions will be included) - Based on numerical rating scale - Proportion with ≥1 grade reduction to	Trials without outcomes of interest to this review

PICOS	Inclusion	Exclusion
	none or mild - Percent change 3. Disease exacerbations/flares - Worsening signs/symptoms from baseline and/or initiation or change of treatment - Switch to rescue/other treatment 4. Safety outcomes: - Overall/serious/ cutaneous AEs - Total withdrawals - Withdrawals due to AEs	

Abbreviations: AE, adverse events; AD, atopic dermatitis; EASI, Eczema Area and Severity Index; mEASI, modified EASI; IGA, Investigator Assessment scale; ISGA, Investigator Static Assessment scale; PGA, Physician('s) Global Assessment PGE, Physicians global evaluation; SCORAD, Scoring Atopic Dermatitis

[1] Various measures of endpoints were considered—mean differences, hazard and odds ratios, and relative risks—based on data availability.

4.1.2.1 Included studies

Of the relevant included studies, two identical pivotal double-blinded vehicle-controlled randomised studies of crisaborole and a related extension study contributed data to evaluate the clinical effectiveness of crisaborole compared to a vehicle. The remaining studies informed an indirect comparison to compare crisaborole 2% topical ointment with TCS, tacrolimus 0.1%, tacrolimus 0.03%, and pimecrolimus 1% in mild to moderate AD_in the relevant population (CS, Section B.2.1(1)). The company noted that no relevant RCTs of TCS were identified.(1) Several studies of TCSs and TCIs were excluded from the SLR because the study populations consisted of >20% with severe AD or because severity levels were not available. The ERG checked whether attempts were made to identify subgroup analyses relevant to the decision problem (e.g. subgroups of mild to moderate AD) within such trials during study selection (clarification response, question A3(21)). The company's response confirmed that a check for subgroup analyses was undertaken and provided a table of the nineteen studies excluded due to >20% of the population having severe AD (reproduced in Appendix 1). Furthermore, the company's clarification response to the ERG (clarification response, question A2(21)) stated that the results of the systematic review had also been cross-checked with other systematic reviews.(20) The ERG, therefore, considers the '*full set of RCTs*' presented by the company to be representative of available relevant evidence.

4.1.3 Critique of data extraction

Two independent, blinded reviewers extracted data from included trials into a pre-specified Microsoft Excel form. Subsequently, both reviewers resolved differences or involved a third reviewer to achieve

consensus. The ERG considered the data extraction strategy reported by the company as a reliable methodology to minimise error and improve reliability of data. While no information about piloting the data extraction form was reported by the company, tables and figures with extracted data presented in the CS(1) and Appendix D were transparent and comprehensive. The ERG did not identify any inaccuracies of key data relating to study and participant characteristics and outcomes of interest; however, the ERG was unable to check all extracted data.

4.1.4 Quality assessment

Summaries of quality assessment of included studies based on the Cochrane Risk of Bias Tool for quantitative intervention studies(19) and the NICE quality assessment tool(22) were presented in Appendix D Table D20 (p. 176) and Table D21 (p.177).(20) The pivotal studies, AD-301 and AD-302 were assessed as RCTs of good methodological quality. However, no further information was available for how studies were rated (e.g. number of reviewers involved) or reasons for scoring items of study quality.

4.1.5 Evidence synthesis

Crisaborole data was obtained from two 28-day, double-blinded, vehicle-controlled RCTs of identical design conducted solely in the USA (AD-301: NCT02118766; AD-302: NCT02118792(1, 23-25)). The company explained that the identical studies were not conducted as a single study as a way "*to meet FDA requirements for a new drug application submission. The US Food and Drugs Administration (FDA) requires that manufacturers of a new drug provide 'substantial evidence' of efficacy based on 'adequate and well-controlled investigations,' which FDA has clarified as being at least two (adequate and well-controlled) clinical studies*." The company also stated that "*there were no differences in methodology between the trials"* (clarification response, question A4(21)).

A single-arm, open-label, 48-week extension study (AD-303) provided long-term crisaborole safety data. A subset of eligible patients who had completed either the AD-301 or the AD-302 studies were enrolled into AD-303.(1)

Crisaborole data for the two RCTs (AD-301 and AD-302) were combined across the two trials for some outcomes but not for other outcomes and subgroup analyses. The ERG had some concern with pooling the data. The company stated that pooled analyses were undertaken for FDA purposes to increase robustness of available data by increasing the sample size (clarification response, question A9(21)) Although the ERG would have preferred a fixed effect meta-analysis (the studies followed the same protocol), inferences are likely to be similar given that the sample sizes in the two studies were also similar. The company reported outcomes relevant to the decision problem based on pre-specified,

exploratory and *post hoc* analyses of results from AD-301 and AD-302. Analyses were mainly based on an intent-to-treat population (all subjects who were randomised and dispensed study drug). The company noted that reported estimates of *"primary and secondary endpoint analyses have been stratified by treatment centre and"* that no differential treatment effect by treatment centre was observed (clarification response, question A8(21)).

No formal meta-analysis of the two crisaborole RCTs was undertaken. An NMA was undertaken to compare crisaborole and comparator treatments (see Sections 4.3and 4.4).

4.2 Critique of trials of crisaborole, their analysis and interpretation

The company conducted an SLR and NMA to inform the decision problem.(1) No relevant study of TCSs were identified, mainly because >20% had severe AD or severity breakdown was not reported. One TCS study met the inclusion criteria for the SLR, but was excluded from the NMA because it did not report outcomes of interest. Only relevant studies evaluating crisaborole and relevant comparators (pimecrolimus 1%; tacrolimus 0.03% or 0.1%) were included in the SLR and NMA. Crisaborole studies are described in this section; comparator studies are presented in Section 4.3.

4.2.1 Crisaborole studies: AD-301 and AD-302

Two 4-week, identically designed, randomised, double-blind, vehicle-controlled studies, AD-301: NCT02118766; AD-302: NCT02118792,(1, 23-25) met the inclusion criteria. The aim of the RCTs was to establish superiority of crisaborole ointment, 2%, to vehicle ointment for treatment of the severity, signs, and detrimental impact on the quality of life of mild to moderate AD in subjects 2 years and older (CS Table 16: Summary of statistical analyses, p. 47 to 48).(1) Clinical advisors to the ERG noted that an adequate study duration for assessing response in patients with AD would be four to six weeks but longer to assess relapses and remissions. The primary outcome was the Investigator's Static Global Assessment (ISGA) success defined as an ISGA score of 0 (clear) or 1 (almost clear) with ≥ 2 point grade improvement. AD-301 and AD-302 were conducted in the USA and involved 47 and 42 centres, respectively. Patients were enrolled into AD-301 and AD-302, if they were 2 years old or more; had a clinical diagnosis of AD based on the Hanifin and Rajkacriteria(26); a baseline ISGA score of mild (2) or moderate (3) AD and ≥ 5% treatable body surface area (BSA) involvement. Key study exclusion criteria were: the presence of active skin infection; previous use of systemic corticosteroids or biologic treatment within 28 days or treatment with TCS or TCIs within 14 days of study entry. Permitted or concomitant treatments were antihistamines, inhaled corticosteroids or topical retinoids for patients on '*stable regimens*' (continual use ≥14 days prior to study entry) and topical application of '*bland emollients'* to dry skin around treatable AD-affected areas. Clinical visits for assessments were scheduled on baseline/day1 and days 8, 15, 22, and 29.(1)

A total of 1,522 patients aged 2 years or more with mild to moderate AD were randomised in a 2:1 ratio to crisaborole (AD-301, n=503; AD-302, n=513) and vehicle (AD-301, n=256; AD-302, n=250) and received the study drug (intent-to-treat population).(1) A further five patients initially allocated to the active treatment group did not receive crisaborole and were not included in the intent-to-treat population (AD-301, n=4; AD-302, n=1).(1) Of the 1522 patients, 209 (14%) were adults and 1313 (86%) were children.(1) Summaries of study characteristics of crisaborole studies, AD-301 and AD-302, adapted from the CS (Tables 9 to 13)(1) and Appendix D (Table D16)(20) are provided in Table 5. The participant flow in the RCTs reported in Appendix D (CONSORT DIAGRAM for AD-301 & AD-302 trials)(20) is shown in Figure 2.

Figure 2: Participant flow in crisaborole RCTs (reproduced from CS Appendix D, Figure D6)

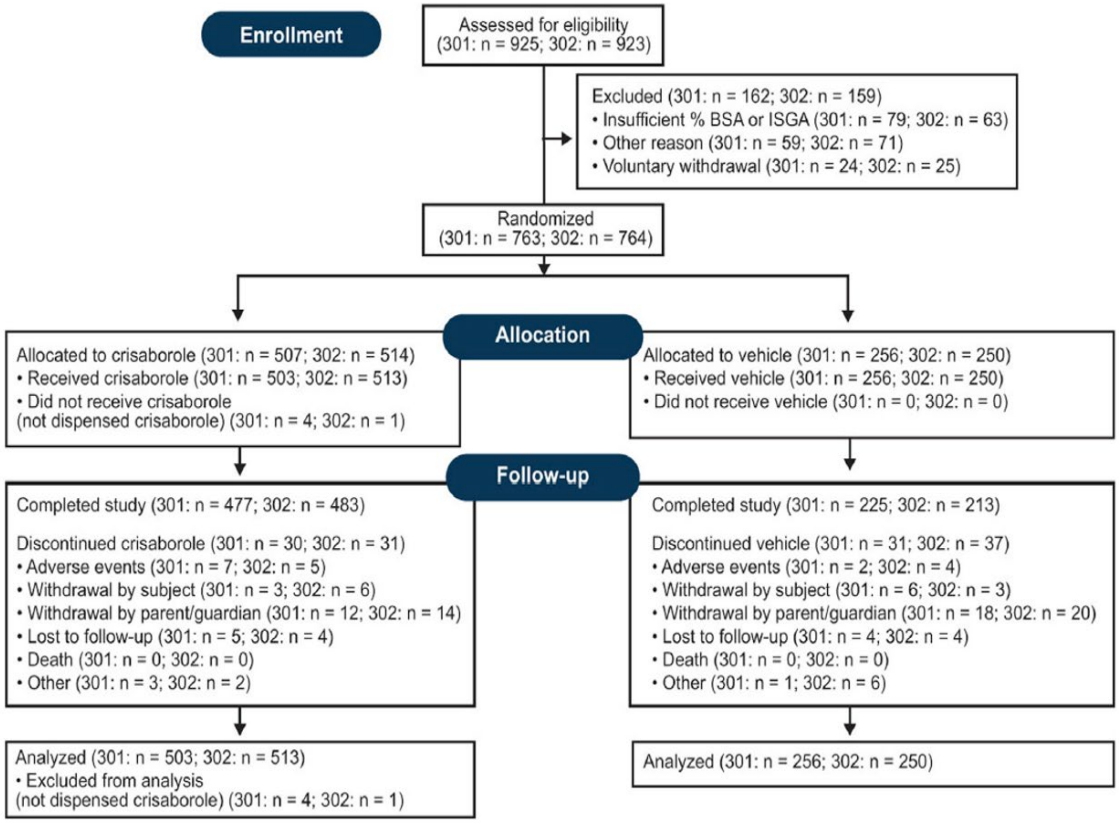


Table 5: Study characteristics of crisaborole studies (adapted from CS, Tables 9 to 13)

Study	**AD-301:** NCT02118766 (n=763)	**AD-302:** NCT02118792 (n=764)	**AD-303** (n=517)
Study setting	USA, n= 47 sites	USA, n=42 sites	Number, not reported (selected sites)
Study design	Double-blinded, vehicle-controlled Phase 3 RCTs		Single-arm, open-label, extension study
Study duration	28 days		48 weeks
Assessments	Baseline/day1 and days 8, 15, 22, and 29		Baseline, every 28 days
Population	● Patients ≥2 years of age ● Hanifin and Rajka(26) clinical criteria for the diagnosis of AD ● Had disease over at least 5% of their total body surface area excluding the scalp ● Baseline disease severity ranging from mild (2) or moderate (3), according to the ISGA Exclusion criteria ● Previous use of biologic therapy or systemic corticosteroids within 28 days or topical corticosteroids or topical calcineurin inhibitor use within 14 days ● Patients with active skin infections		Patient with mild to moderate AD, ≥ 2 years of age, who participated in AD-301 or AD-302
Intervention(s)	Crisaborole ointment 2%, topical application to AD lesions, excluding areas on the scalp, twice daily (n=1,016)		ISGA ≥ 2 patient enters '*on treatment period for 28 days'* Crisaborole ointment 2%, topical application to AD lesions, excluding areas on the scalp, twice daily ISGA=0 or 1, patient enters 'off treatment period for 28 days: Observation, only Stopping rule: No improvement in ISGA after 3 months of continuous treatment
Comparator(s)	Vehicle ointment topical application to AD lesions, excluding areas on the scalp, twice daily, n=506		Not applicable
Numbers randomised (n)	Crisaborole, n=503 Vehicle, n=256	Crisaborole, n=513 Vehicle, n= 250	Number enrolled and analysed: n=517
Primary outcome(s)	- ISGA success: The proportion of patients achieving an ISGA score of 0 (clear) or 1 (almost clear) with ≥2-point improvement from baseline at Day 29[§]		- Treatment exposure - Adverse effects of treatment stratified by treatment period age[§]

Study	AD-301: NCT02118766 (n=763)	AD-302: NCT02118792 (n=764)	AD-303 (n=517)
Secondary outcome(s)	- Time to Achieve ISGA Success[§] - % patients with ISGA Score of Clear (0) or Almost Clear (1) - Change from baseline in signs of AD (erythema, exudation, lichenification, excoriation, papulation/induration)[§] based on a 4-point global scale and not by affected body sites. - Change from baseline in pruritus severity[§] - Time to improvement in pruritus§ (defined as a pruritus score of none (0) or mild (1) with at least one-grade improvement from baseline/days - Treatment-Emergent Adverse Events (AEs) and serious adverse events (SAEs)[§]		
Outcomes based on post-hoc analyses	- Change from baseline in Children's Life Quality Index (CLQI); Dermatology Life Quality Index (DLQI); Dermatitis Family Impact (DFI) Score[§] - % patients with early improvement in pruritus[§] - % patients who experienced pruritus symptom improvement at day[§] - % reduction in pruritus severity over the first 6 days of treatment[§] - Correlations between early improvement in pruritus, based on baseline demographics and disease characteristics, and HRQoL AD signs and sleep scores[§]		- Change from baseline in Children's Life Quality Index (CLQI); Dermatology Life Quality Index (DLQI); Dermatitis Family Impact (DFI) Score[§]
Related publications	Paller 2016(23) and Callendar 2019(27) (efficacy and safety) Yosipovitch 2018(28) (pruritus); Simpson 2018(29) (health-related quality of life)		Eichenfield 2017(30)

[§]Reported outcomes specified in the decision problem

Abbreviations: AD, atopic dermatitis; BSA, body surface area; CLQI, Children's Life Quality Index; DLQI: Dermatology Life Quality Index; DFI: Dermatitis Family Impact; HRQoL, health-related quality of life; ISGA, Investigator Static Global Assessment Scale; TCS, topical corticosteroid; TCIs, topical calcineurin inhibitor

The company's decision problem specifies the following population of interest: "*Adults and children aged 2 years and older with mild to moderate AD that has not been controlled by topical corticosteroids or where there is a serious risk of important adverse effects from further topical corticosteroid use, particularly irreversible skin atrophy (second-line use post-TCS treatment)"*.(1) The ERG notes that AD-301 and AD-302 were not restricted to patients whose condition was uncontrolled on TCSs or were at serious risk of adverse effects from TCSs. This could have led to more favourable efficacy outcomes in the clinical studies than might be expected in a post-TCS population because the eligible study population was easier to treat. The ERG was unable to verify

from the company the extent to which subgroups of the optimised population were included in AD-301 and AD-302. However, the company confirmed that expert clinical advice suggested that some patients could be resistant to previous treatment and therefore more *'challenging'* to treat (clarification response, question A7 (21)). A subgroup analysis of patients who were TCS-naïve or had received prior TCSs is presented in the CS (Table 26, p. 66(1)); however, the ERG notes that having received prior TCSs is not the same as being uncontrolled on TCSs.

4.2.1.1 Statistical analyses

A summary of statistical analyses of crisaborole RCT outcomes is presented in CS, Table 16(1) and Appendix D, Table D22 p.190.(21) The number of patients included in RCTs was adequate, considering the reported sample size calculation.(1) The intention-to-treat (ITT) population, used in analysing efficacy outcomes, was defined as all those who were randomised and dispensed the study treatment, regardless of dropping out of the study.

4.2.2 Crisaborole non-RCT: AD-303

The company also presented evidence derived from a single-arm, open-label, extension study (AD-303) which enrolled eligible patients who had completed AD-301 or AD-302.(1) Patients who completed AD-301 and AD-302 at a subset of study sites were enrolled into AD-303 (n= 517) if they had not experienced a crisaborole treatment-related adverse event (AE) or a serious AE. AD-303 was designed to assess safety of crisaborole over a period of 48 weeks, divided into 28-day cycles. Treatment protocols were determined by AD severity according to ISGA score at the beginning of each 28-day cycle. Patients with ISGA of ≥2 (mild or worse) had topical applications of crisaborole 2% twice daily over affected lesions for 28 days (on-treatment period). Patients with ISGA of 0 or 1 (clear or almost clear) were observed for the duration of the treatment cycle but did not receive treatment (off-treatment period). Crisaborole treatment was discontinued if ISGA scores did not improve after three consecutive treatments (i.e. 3 cycles or 3 months).(1) Overall, there were 12 possible on-treatment periods in AD-303. Study characteristics of the crisaborole trials are presented in Table 5.

The CS stated that patients entered the extension study, AD-303, if they had completed AD-301 or AD-302 without experiencing a crisaborole treatment-related adverse event or a serious adverse event.(1) Figure 4 (CS, p. 49) indicates that 1,398 patients completed AD-301 (crisaborole, n=477; vehicle, n=225) or AD-302 (crisaborole, n=483; vehicle, n=213).(1) However, Figure 5 (CS, p. 49) indicates that only 517 patients (37%) entered AD-303.(1) The company stated that only specific study sites participated in this long-term safety study. The reason being that "*only sufficient patient numbers to appropriately assess safety outcomes"* were needed (Clarification response, A12).(20) No formal documentation for eligibility into the AD-303 was available.

In the extension study, AD-303, only 52% (271 people of 517) completed the study, while 86 (17%) withdrew and 115 (22%) discontinued for "*other*" reasons, although all 517 participants were included in the analysis. According to a clarification response, question A13(1, 21) [redacted]."(21) The company also clarified that, "[redacted]" (21)

4.2.3 Classification of AD severity in AD studies and in crisaborole studies

There is currently no consensus for classifying AD severity or assessing short-term or long-term outcomes in affected patients.(31) Patient eligibility in the pivotal crisaborole studies was based on the 5-point ISGA score. Clinical advisors to the ERG suggested that AD severity may be based on differing clinical presentations in line with the HOME initiative and related recommendations.(16, 32) . The company presented a number of scoring instruments or scales for assessing the severity of AD in clinical trials. These scales provide a global assessment of AD severity based on self-reported and physician-assessed symptoms and signs. Furthermore, assessment may be completed at a single time-point (using a static scale) or in relation to reference time-point (using a dynamic scale).(1) Table 6 is a summary of the instruments presented in the CS for measuring AD severity in clinical trials.(1) The ERG has highlighted items assessed by each instrument.

Table 6: Items assessed in scales used in clinical trials (adapted from CS, Section B 1.3.1)

Items assessed		Eczema Area and Severity Index (EASI) scale	Investigator's Global Assessment (IGA) scale	Physician's Global Assessment (PGA) scale	Investigator's Static Global Assessment (ISGA) scale	Scoring Atopic Dermatitis index (SCORAD) scale
Clinician's assessment						
	Region of body affected	√	N/A	N/A	N/A	√
	BSA affected	√	√	√	√	√
	Erythema	√	√	√	√	√
	Oedema/swelling	√	√	√	√	√
	Excoriations	√	√	√	√	√
	Papulations	√	√	√	√	√
	Lichenification/infiltration	√	√	√	√	√
	Oozing/crusting	√	√	√	√	√
	Dryness	√*	N/A	N/A	N/A	N/A
Patient-reported symptoms						
	Pruritus	N/A	N/A	N/A	N/A	√
	Sleep disturbance	N/A	N/A	N/A	N/A	√
	Dryness	N/A	N/A	N/A	N/A	√

* Assessed only if dryness is due to eczema
Abbreviations: AD: atopic dermatitis; BSA, CS, company's submission; body surface area; EASI, Eczema Area and Severity Index; ERG, evidence review group; N/A, not applicable or assessed

The company described the Investigator's Global Assessment (IGA)/Physician's Global Assessment (PGA) as "*a simple and intuitive instrument, which provides a subjective evaluation of overall disease severity*".(1) The company also stated the following:(1)

- IGA/PGA is available in many versions (static and dynamic forms). The dynamic version measures improvement in relation to baseline severity whilst the static version of the IGA/PGA, "*the Investigator's Static Global Assessment (ISGA), measures the physician's impression of the disease at a single time point"*
- IGA/PGA applies varying definitions/ descriptors and assessment criteria with up to a 6-point severity rating from '*clear*' to '*severe*'.
- IGA/PGA provides information mainly about skin-related symptoms and does not focus on the impact of AD indicators or co-existing conditions.
- IGA/PGA may not reflect modest changes in severity due to its narrow ordinal scale.

The ERG noted that a systematic review of IGA scales reported various scores (4 to 7-point ratings).(33) Table 7 shows a range of rating of specific scales whilst Table 8 compares ISGA scores used in crisaborole studies with NICE recommendations for assessing AD severity.(1) The existing

lack of standardisation of AD scales could potentially hinder satisfactory interpretation or comparisons of AD severities across trials and real-world settings.

Table 7: Scales for global assessment of AD severity

Name of instrument	Rating
Investigator's Static Global Assessment (ISGA) scale	5-point scoring system
Eczema Area and Severity Index (EASI) scale	Weighted scores*: 0 to 72 0 (no eczema); 7.1-21 (moderate); 21.1-50 (severe); 50.1-72 (very severe)
Investigator's Global Assessment (IGA) scale	4 to 7-point score system
Physician's Global Assessment (PGA) scale	5 or 6-point scoring system
Scoring Atopic Dermatitis index (SCORAD) scale	Weighted scores ranging from 0 to 103
Patient-Oriented Eczema Measure (POEM)	Weighted scores ranging from 0 to 28 0 to 2 (clear or almost clear) 3 to 7 (mild eczema) 8 to 16 (moderate eczema) 17 to 24 (severe eczema) 25 to 28 (very severe eczema)
*Weighting is based on affected body region (head and neck; trunk; upper extremities and lower extremities); percentage of body surface of region with eczema and lesion severity.	

Table 8: Comparison between ISGA scores used in crisaborole studies and NICE recommendations for assessing AD severity

	Investigator's Static Global Assessment (ISGA) Scale in crisaborole RCTs *Rating: 5 levels of AD severity*		**NICE Holistic assessment of severity, quality of life, and psychosocial wellbeing** *Rating: 4 levels of AD severity*			
Score	**Grade**	**Definition**	**Skin/ physical severity**	**Definition**	**Impact on HRQoL and psychosocial wellbeing**	**Definition**
0	**Clear**	Minor residual hypo/hyperpigmentation; no erythema or induration/papulation; no oozing/crusting	**Clear**	Normal skin, no evidence of active atopic dermatitis	**None**	No impact on quality of life
1	**Almost Clear**	Trace faint pink erythema, with barely perceptible induration/papulation and no oozing/crusting				
2	**Mild**	Faint pink erythema, with mild induration/papulation and no oozing/crusting	**Mild**	Areas of dry skin, infrequent itching (with or without small areas of redness)	**Mild**	Little impact on everyday activities, sleep and psychosocial wellbeing
3	**Moderate**	Pink-red erythema with moderate induration/papulation with or without oozing/crusting	**Moderate**	Areas of dry skin, frequent itching, redness (with or without excoriation and localised skin thickening)	**Moderate**	Moderate impact on everyday activities and psychosocial wellbeing, frequently disturbed sleep
4	**Severe**	Deep or bright red erythema with severe induration/papulation and with oozing/crusting	**Severe**	Widespread areas of dry skin, incessant itching, redness (with or without excoriation, extensive skin thickening, bleeding, oozing, cracking and alteration of pigmentation)	**Severe**	Severe limitation of everyday activities and psychosocial functioning, nightly loss of sleep

Abbreviations: AD, atopic dermatitis; HRQoL, health-related quality of life

Table of Contents

4.2.4 Interventions in crisaborole studies

The intervention, topical application of crisaborole, 2%, matches the decision problem. Patients aged 2 years and over with mild and moderate AD in AD-301 (n =503) and AD-302 (n =513) were instructed to apply crisaborole 2% topical ointment to treatable AD lesions as a thin layer, twice daily, from baseline until day 28 of the study. The CS states that application of treatments to scalp lesions was not allowed because of likely dissatisfaction related to the use of an ointment.(1)

Patients in the comparator group received a vehicle treatment developed by the study sponsors, Anacor Pharmaceuticals.(23) [redacted] The company further indicated that older trials may use vehicles which could be irritating to the skin. The CS states that these differences in vehicle responses, taken together, may make the relative benefits of therapies in older trials artificially high, compared to the relative benefits in the crisaborole trials.(1) Differences in vehicles and responses across included RCTs were presented in CS, Appendix D (p. 173(21)).

Clinical advisors to the ERG indicated that ointment-based vehicles tend to demonstrate greater responses than cream-based treatments. However, they specified that high placebo effects were common in eczema trials because of the emollient effect of vehicles, and because treatment adherence and compliance with emollient use are generally better in a controlled trial setting compared with real-world clinical scenarios. This notion is supported by Paller *et al.* 2016,(23) a publication of the findings of AD-301 and AD-302 which stated that, "*The significant efficacy of crisaborole versus vehicle was noted, despite the strong 'vehicle effect' observed in these studies, which is a common phenomenon in AD clinical studies that compare active therapeutics with their emollient bases".(1)*

4.2.5 Crisaborole studies: Patient characteristics

Overall, 1,016 patients were randomised (2:1 ratio) to, and received, crisaborole 2% topical ointment and 506 patients were randomised to a vehicle treatment in the AD-301 and AD-302 multicentre studies conducted in the USA. The pooled study population consisted of 33% aged 2 to 6 years; 29% aged 7 to 11 years; 24% aged 12 to 17 years and 14% who were more than 18 years old. The mean age across treatment groups in both studies ranged between 11.8 and 12.6 years (range, 2 to 79 years). The study population comprised of 61% Caucasians and 28% people of black or African origin. Patients with moderate AD (61%) made up a larger percentage of the study population compared to those with mild AD (39%). Mean baseline % BSA affected, pruritus score, condition-specific HRQoL

scores were comparable across treatment group for AD-301 and AD-302. A summary of baseline characteristics of the crisaborole RCTs and non-RCT as reported in the CS is shown in Table 9.

The ERG did not note any significant imbalances in baseline characteristics in the crisaborole RCTs. Clinical advisors to the ERG suggested that factors which are likely to influence the response to AD treatment include severity, age, affected body sites, clinical presentation of AD (e.g. discoid or follicular AD in people of Black or Asian origin), co-existing conditions and allergies.

Age distribution in the trial population is broadly comparable to that of AD patients in the UK. However, distribution of ethnicities in the trial population may not be consistent with those who present with AD in England. Although, the ERG is unable to comment on the impact of these differences in the UK population, a recently published *post hoc* analyses of patients included the crisaborole studies reported that efficacy (ISGA success: white: 33.5% vs. 22.3%, $p < 0.001$; non-white: 30.0% vs. 21.3%, $p < 0.05$; Hispanic/Latino: 35.4% vs. 18.2%, $p < 0.01$; not Hispanic/Latino: 31.3% vs. 22.8%, $p < 0.01$) and safety outcomes were not affected by race and ethnicity.(27)

Table 9: Baseline characteristics of patients: AD-301, AD-302 and AD-303 (adapted from CS, Table 14 and Table 15)

	AD-301		AD-302		AD-303	
Baseline characteristic	**Crisaborole (N=503)**	**Vehicle (N=256)**	**Crisaborole (N=513)**	**Vehicle (N=250)**	**Baseline characteristic**	**Crisaborole (N=517)**
Age, (years)					**Age, (years)**	
Mean	12.0	12.4	12.6	11.8	Mean	11.7
Range	2-65	2-63	2-79	2-79	Range	2-72
Age groups, %					**Age groups, %**	
2-6 y	32.3	30.5	33.7	37.2	2-11 y	59.6
7-11 y	30.8	28.5	26.7	28.4		
12-17 y	24.1	26.2	24.6	22.8	12-17 y	28.2
≥18 y	12.9	14.8	15.0	11.6	≥18 y	12.1
Sex, %					**Sex, %**	
Male	43.5	44.1	45.0	44.8	Male	40.8
Female	56.5	55.9	55.0	55.2	Female	59.2
Ethnicity, %					**Ethnicity, %**	
Hispanic or Latino	25.0	25.8	14.4	14.0	Hispanic or Latino	15.9
Not Hispanic or Latino	75.0	74.2	85.6	86.0	Not Hispanic or Latino	84.1
Race, %					**Race, %**	
American Indian or Alaska Native	1.6	1.2	0.6	0.8	American Indian or Alaska Native	0.2
Asian	5.2	6.6	5.1	4.0	Asian	5.4
Black or African American	27.4	23.8	28.7	31.2	Black or African American	29.4
Native Hawaiian or Pacific Islander	0.0	1.6	1.4	1.6	Native Hawaiian or Pacific Islander	0.2
White	61.2	63.3	60.2	57.6	White	60.9
Other	4.6	3.5	4.1	4.8	Other	3.9
Baseline ISGA, %					**Treatment in AD-301 or AD-302, %**	
Mild (2)	39.0	36.3	38.4	40.0	Crisaborole ointment	69.1
Moderate (3)	61.0	63.7	61.6	60.0	Vehicle	30.9
% BSA					**Patients in each 12-week period, n**	
Mean	18.8	18.6	17.9	17.7	Week 1-12	482
Range	5-95	5-90	5-95	5-90	Week 13-24	428
					Week 25-36	368
					Week 37-48	226
Severity of Pruritus Scale, %					Not reported	Not reported
N	446	223	457	218		
0 – None	3.8	5.8	3.9	2.8		

Table of Contents

	AD-301		AD-302		AD-303	
Baseline characteristic	**Crisaborole (N=503)**	**Vehicle (N=256)**	**Crisaborole (N=513)**	**Vehicle (N=250)**	**Baseline characteristic**	**Crisaborole (N=517)**
1 – Mild	25.8	28.7	24.9	25.2		
2 – Moderate	35.4	33.6	37.9	42.2		
3 – Severe	35.0	31.8	33.3	29.8		
DLQI					Not reported	Not reported
N	95	52	97	40		
Mean (SD)	9.6 (6.37)	9.5 (6.52)	9.7 (6.24)	9.5 (6.52)		
Median	9.0	8.0	8.0	8.0		
Range	1-27	0-27	0-26	0-27		
CDLQI					Not reported	Not reported
N	393	199	404	204		
Mean (SD)	9.7 (6.19)	9.1 (6.54)	9.0 (5.77)	8.9 (5.48)		
Median	9.0	7.0	8.0	8.0		
Range	0-28	0-30	0-28	0-27		
DFI					Not reported	Not reported
N	431	214	431	217		
Mean (SD)	8.5 (6.63)	7.5 (6.66)	7.7 (6.57)	8.0 (5.65)		
Median	7.0	6.0	6.0	7.0		
Range	0-30	0-30	0-30	0-24		

Abbreviations: AD: atopic dermatitis; BSA: body surface area; CDLQI: Children's Life Quality Index; DLQI: Dermatology Life Quality Index; DFI: Dermatitis Family Impact; ISGA: Investigator's Static Global Assessment score

Table of Contents

4.2.6 Crisaborole RCTs: Efficacy outcomes

4.2.6.1 ISGA success: Proportion of patients achieving ISGA score 0 or 1 and ≥2-grade improvement

The primary outcome for the crisaborole RCTs was ISGA success, defined as an ISGA score of clear (0) or almost clear (1) with at least a 2 grade improvement from baseline/day 1. As mentioned, earlier, there are various rating scales for the ISGA/IGA scores. Assessment of the ISGA score in the crisaborole trials was based on the 5-point (0 to 4) ratings as described in Table 8. A statistically significantly higher proportion of crisaborole-treated patients achieved ISGA success at day 29 (AD-301 32.8% vs 25.4%, p=0.038; AD-302 31.4% vs 18.0%, p<0.001) compared to those who received the vehicle (Figure 3).(1, 23) A similar trend was observed when outcomes were assessed according to AD severity in the pooled AD-301 and AD-302 population (Table 10); however, there was no statistically significant difference in the group with mild AD when the crisaborole group was compared with the vehicle group ([redacted] vs [redacted], [redacted]).(20) It is noted, however, that that crisaborole studies were not powered to detect a significant result in subgroup analyses by disease severity.

Figure 3: Percentage of patients with ISGA success at day 29. (Reproduced from CS, Figure 7)

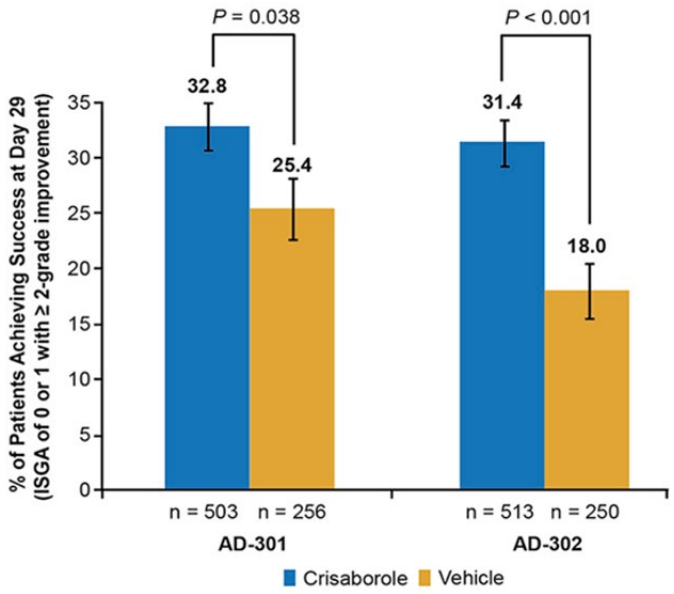


Table 10: Percentage of patients achieving ISGA success according to AD severity. Reproduced from Table 4 (clarification response, question A9)

Trial number (Acronym)	Arm	Mild			Moderate		
		N subjects	% achieving ISGA Success		N subjects	% achieving ISGA Success	
AD-301 and AD-302 pooled	Crisaborole	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
AD-301 and AD-302 pooled	Vehicle	[redacted]	[redacted]		[redacted]	[redacted]	

4.2.6.2 ISGA 0/1: Proportion of subjects achieving an ISGA score of clear (0) or almost clear (1)

Pooled analysis of the proportion of subjects achieving an ISGA score of clear (0) or almost clear (1) was presented as a secondary efficacy outcome. This was the outcome used in the NMA and in the cost-effectiveness analysis. Compared with patients treated with the vehicle, a greater proportion of patients achieved an ISGA score of clear (0) or almost clear (1) at day 29 (AD-301 51.7% vs 40.6%, p=0.005; AD-302 48.5% vs 29.7%, p<0.001) (Figure 4). A similar trend was observed when outcomes were assessed separately for AD-301 and AD-302 according to age and AD severity (Table 11). However, [redacted] for adults with mild AD (AD-301, [redacted]) or moderate AD (AD-301, [redacted] for specific subgroups of children with mild AD (age: 2-<7 years[redacted] 7-<12 years: [redacted] and 12-<18: [redacted] (age: 2-<7 years, [redacted]; 7-<12 years: [redacted]) in AD-301. The ERG notes that the CS only reports p-values and not between-group differences or confidence intervals, which provides little information about the range of plausible effects.

Figure 4: **Percentage of patients with ISGA score of clear (0) or almost clear (1). (Reproduced from CS, Figure 8)**

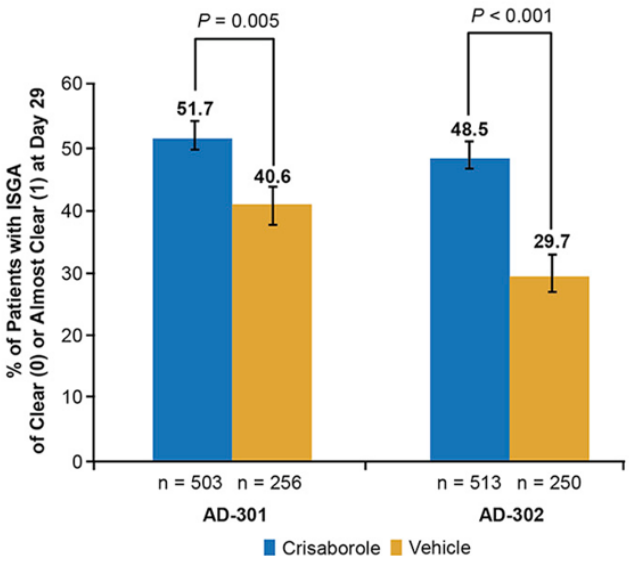


Table of Contents

Table 11: Percentage of patients with ISGA 0/1 scores by AD severity and age (reproduced from CS Table 25)

Trial number (Acronym)	Arm	Total No. Patients/ % ISGA 0-1					
		Mild (All)	Mild Children (age groups in years)	Mild Adults	Moderate (All)	Moderate Children (age groups in years)	Moderate Adults
AD 301							
AD-301	Crisaborole 2%	[redacted]	2-<7: [redacted] 7-<12: [redacted] 12-<18: [redacted]	[redacted]	[redacted]	2-<7: [redacted] 7-<12: [redacted] 12-<18: [redacted]	[redacted]
AD-301	Vehicle	[redacted]	2-<7: [redacted] 7-<12: [redacted] 12-<18: [redacted]	[redacted]	[redacted]	2-<7: [redacted] 7-<12: [redacted] 12-<18: [redacted]	[redacted]
AD-301	p-values	[redacted]	2-<7: [redacted] 7-<12: [redacted] 12-<18: [redacted]	[redacted]	[redacted]	2-<7: [redacted] 7-<12: [redacted] 12-<18: [redacted]	[redacted]
AD 302							
AD-302	Crisaborole 2%	[redacted]	2-<7: [redacted] 7-<12: [redacted] 12-<18: [redacted]	[redacted]	[redacted]	2-<7: [redacted] 7-<12: [redacted] 12-<18: [redacted]	[redacted]
AD-302	Vehicle	[redacted]	2-<7: [redacted] 7-<12: [redacted] 12-<18: [redacted]	[redacted]	[redacted]	2-<7: [redacted] 7-<12: [redacted] 12-<18: [redacted]	[redacted]
AD-302	p-values	[redacted]	2-<7: [redacted] 7-<12: [redacted] 12-<18: [redacted]	[redacted]	[redacted]	2-<7: [redacted] 7-<12: [redacted] 12-<18: [redacted]	[redacted]

Table of Contents

4.2.6.3 ISGA Score: Time to success in ISGA score

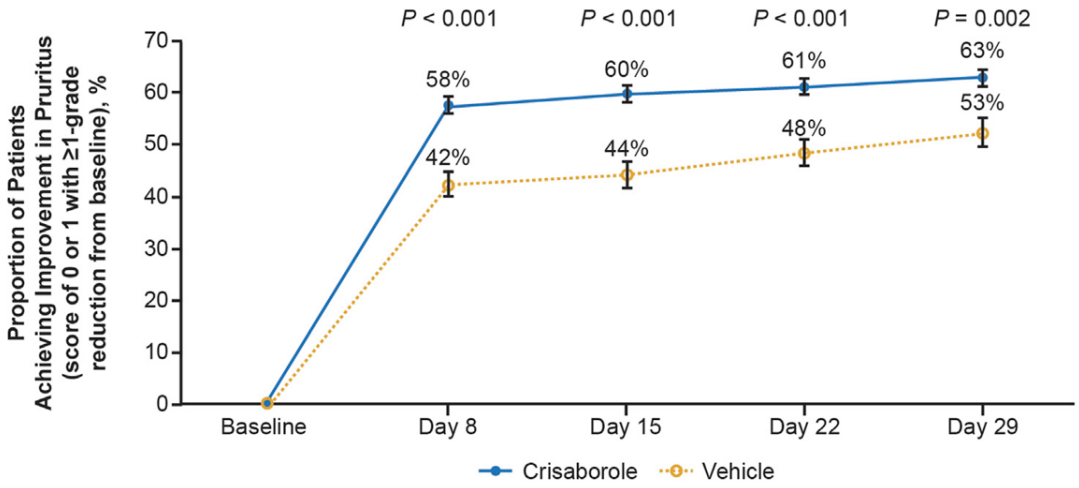
A pooled analysis using Kaplan-Meier and log-rank methods showed that crisaborole-treated patients achieved ISGA success earlier that those in the vehicle group (p<0.001) (CS, Figure 9, p 54(1)).

4.2.6.4 Other measures of AD severity

The company presented outcomes on pruritus score and clinical signs of AD. Criteria for assessments of pruritus severity and AD presentation, were presented in a supplementary file of the Paller *et al.* 2016 publication.(23) Severity of pruritus was rated on a 4-point scale ((0), none to (3), severe). Improvement in pruritus was defined as none or mild with at least a 1-grade improvement from baseline/day 1.(1) Signs of atopic dermatitis were also evaluated on a 4-point scale from none ((0), none to (3), severe).

A statistically significant greater percentage of patients in the crisaborole group achieved improvement in pruritus (Figure 5) in a shorter length of time (pooled data, 1.37 vs 1.70 days, p = 0.001) compared with those in the vehicle group.(1) A statistically significant greater proportion of crisaborole-treated patients had improvement in severity and several clinical signs of AD at day 29 compared with patients applying the vehicle (p<0. 002-0.001 and p<0.008-0.001, respectively; see CS, Figure 12 and Figure 11(1)).

Figure 5: Percentage of patients (AD-301 + AD-302) experiencing improvement in pruritus. (Reproduced from CS, Figure 10)



4.2.6.5 Health-related quality of life

A *post hoc* mapping of condition-specific HRQoL outcomes onto the EQ-5D-3L was performed to obtain preference-based EQ-5D estimates at day 29.(1) In this section, only evidence relating to the condition-specific outcomes will be presented briefly. Health-related quality of life outcomes were obtained from exploratory analysis(29). HRQoL "*was assessed using the Children's Dermatology*

Life Quality Index (CDLQI) (2 to15 years), the Dermatology Life Quality Index (DLQI) (≥16 years) and the Dermatitis Family Impact Questionnaire (DFI) (parents/caregivers/family of patients aged 2 to 17 years)".(29) Details of items of each tool is available in the publication by Simpson *et al.* 2018.(29) Scores of each tool were summarised descriptively and change from baseline/day 1 was noted. The minimal clinically important difference (MCID) for a change from baseline is ≥ 3.3 for the DLQI and ≥ 2.5 for the CDLQI. The latter is validated in patients aged 4 years and over. Therefore, patients younger than 4 were assisted by a parent or caregiver. The DFI scale evaluates ten domains in which AD has a negative impact on the HRQoL of parents, caregivers, and families of patients aged 2 to 17 years. There is no recommended MCID for this scale. A reduction in the score indicates improvement in HRQoL. A summary of HRQoL scores are presented in CS, Tables 19, 20 and 21 (p. 56).(1) Overall, more crisaborole-treated patients compared with those receiving vehicle experienced a clinically significant improvements in HRQoL from baseline to day 29 (CDLQI; age 2 to 15; 61.7% vs 52.1%, *p*=0.003; DLQI age ≥ 16 years, 53.9% vs 41.5%, *p*=0.083). Change from baseline to day 29 for DFI scores was -3.7 in the crisaborole group vs -2.7 in the vehicle group (p=0.003).(1)

4.2.7 Efficacy outcomes in AD-303

The main efficacy outcomes (CS, Section 2.6.2.1, p. 62 to 63) reported in the single-arm, open-label extension study AD-303 were subsequent therapy use ("*defined as the need for concomitant nonconcurrent use of low- to mid-potency TCS or TCI"*) and treatment exposure to crisaborole (CS, Table 23(1)).

4.2.7.1 Subsequent therapy use, AD-303

The proportion of patients aged 2 to 11 years; 12 to 17 years and aged ≥ 18 who required subsequent treatment were 22.4%; 26% and 12.7%, respectively. On average, 150 treatment days preceded the need for subsequent treatment. The mean number of days on TCSs and TCIs was 21.4 (n=155) and 24.2 (n=6).(1)

4.2.7.2 Treatment exposure, AD-303

The study population underwent an average of 6.2 on-treatment periods (a treatment period was defined as 28 days), including the initial pivotal trials.(1) In the long-term safety (LTS) study (AD-303), 396 patients were exposed to crisaborole treatment for 6 months and 271 patients were exposed to crisaborole treatment for 12 months(1). The mean number of drug applications was similar across age groups (347.7-349.4) while the mean amount of drug applied per application was 2.40 grams (g) for patients 2-11 years of age, 2.29 g for patients 12-17 years of age, and 2.10 g for patients ≥ 18 years of age, respectively.(1)

4.2.7.3 ISGA 0/1, AD-303

The clarification response to question A14 provides data on ISGA 0/1 in extension study AD-303. The proportion of subjects with ISGA 0/1 improved over time, with approximately 13% having ISGA 0/1 at study entry and 55% having ISGA 0/1 at Week 48 (clarification response Figure 2).

4.2.8 Safety outcomes (AD-301; AD-302; AD-303)

Short-term and long-term outcomes were reported in the 28-day identically designed crisaborole RCTs (AD-301 and AD-302) and the 48-week extension study (AD-303). Following both short-term and long-term use, crisaborole was associated with mild (51.2%), moderate (44.6%) and severe (4.2%) treatment-emergent adverse events (TEAEs) and the majority (93.1%) were not considered to be related to the study drug. (CS, Section B.2.9). From the RCTs and non-RCT, the most commonly reported TEAEs were atopic dermatitis (exacerbation, worsening and flare-ups) (11.2%) and upper respiratory tract infection (10.3%). The most frequently reported treatment-related AEs were atopic dermatitis (3.1%), application-site pain (burning and/or stinging) (2.3%) and application-site infection (1.2%). Within the 28-day RCTs, application site pain occurred in 4.4% of crisaborole-treated patients and 1.2% of vehicle-treated patients. There were no deaths in any of the three studies. Overall, the company stated that crisaborole demonstrated a favourable safety profile in the population of interest. Information on withdrawals is presented in Section 4.3.6.

4.2.9 Ongoing studies of crisaborole

Ongoing studies in mild to moderate AD, identified from ClinicalTrials.gov, are listed in Table 12.Some key effectiveness studies with 4 weeks or longer follow-up include the following:

- An RCT to assess long-term maintenance and reduction in flares with crisaborole vs. vehicle over 52 weeks in adults and children (planned N=700, NCT04040192)
- An RCT to assess steroid-reducing effects of crisaborole vs. either vehicle or emollient over 12 weeks in children (planned N=60, NCT03832010)
- An observational two-arm study to compare crisaborole alone vs. crisaborole plus TCS over 8 weeks in adults and children (planned N=16, NCT04008784)
- An RCT to compare crisaborole vs. either vehicle, TCS or pimecrolimus over 4 weeks in adults and children (planned N=600, NCT03539601)
- An RCT to compare crisaborole vs. tacrolimus 0.03% over 12 weeks in children (planned N=160, NCT03645057 [ASPIRE]).

Table 12: Ongoing studies of crisaborole

ClinicalTrials.gov Identifier	Dates (planned)	Study question	Design	Duration	N planned	AD severity & population	Age group	Intervention	Comparator	Primary outcomes	Secondary outcomes
NCT04040192	2019-2022	Long-term maintenance with crisaborole	RCT phase 3	52 weeks	700	Mild to moderate AD Responded to crisaborole	Adults & children	Crisaborole 2%	Vehicle	- Flare free maintenance - Safety	- ISGA - EASI - POEM - HRQoL - Pruritus
NCT03832010	2019-2020	Steroid-reducing effects of crisaborole	RCT phase 4	12 weeks	60	Mild to moderate AD	Children	Crisaborole 2% +TCS	a) Vehicle +TCS b) Emollient+TCS	- Reduction in TCS use	- SCORAD - Pruritus - HRQoL
NCT04008784	2019-2020	Crisaborole plus TCS	Observational	8 weeks	16	Mild to moderate AD	Adults & children	Crisaborole 2%	Crisaborole 2% + TCS	- IGA	- Pruritus
NCT03539601	2018-2021	Crisaborole vs TCS or TCI	RCT phase 3B/4	4 weeks	600	Mild to moderate AD	Adults & children	Crisaborole 2%	a) Vehicle b) TCS c) Pimecrolimus	- EASI - Safety	- ISGA - % BSA - Pruritus - HRQoL
NCT03645057 (ASPIRE)	2019-2020	PROs & caregiver burden	RCT	12 weeks	160	Mild to moderate AD	Children	Crisaborole 2%	Tacrolimus 0.03%	- PROs	- EASI - HRQoL - Caregiver burden
NCT03868098	2019-2020	Different application rates of crisaborole	RCT (intra-participant)	2 weeks	30	Mild to moderate AD	Adults	Crisaborole 2%, three application rates	Vehicle (same patient, different body areas)	- Total clinical signs score	
NCT03954158	2019-2019	Efficacy and safety in Japanese patients	RCT phase 2b (intra-participant)	2 weeks	80	Mild to moderate AD Japanese	Adults & children	a) Crisaborole 2% once daily b) Crisaborole 2% twice daily	Vehicle (same patient, different body areas)	- Total clinical signs score	- ISGA - Pruritus - Safety
NCT03250663	2017-2020	Adherence	RCT, phase 1	52 weeks	40	Mild to moderate AD	Adults & children	Crisaborole 2% + SOC	Crisaborole 2% + SOC + treatment response emails	- Adherence to topical therapy	

Abbreviations: AD, atopic dermatitis; BSA, body surface area; EASI, Eczema Area and Severity Index; HRQoL, health-related quality of life; IGA, Investigator's Global Assessment; ISGA, Investigator's Static Global Assessment; POEM, Patient Oriented Eczema Measure; PRO, patient reported outcome; RCT, randomised controlled trial; SCORAD, SCORing Atopic Dermatitis; SOC, standard of care; TCS, topical corticosteroid.

Table of Contents

4.2.10 Critique of included crisaborole trials in the SLR

Two 28-day, identically designed, randomised, double-blind, vehicle-controlled studies, AD-301 and AD-302, met the inclusion criteria.(1) The aim of the RCTs was to establish the superiority of crisaborole ointment, 2%, to vehicle ointment for treatment of the severity, signs, and detrimental impact on the quality of life of mild to moderate AD in subjects 2 years and older (CS Table 16: Summary of statistical analyses, p. 47 to 48).(1) Clinical advisors to the ERG noted that an adequate trial duration for assessing response in patients with AD would be four to six weeks after starting treatment and longer to assess long-term flares and remissions. Both pivotal crisaborole trials (AD-301 and AD-302) and the extension study (AD-303) were multicentre studies conducted solely in the USA. Age distribution in the trial population is broadly comparable to that of AD patients in the UK. However, distribution of ethnicities in the trial population may not be consistent with those who present with AD in England. The 5-point ISGA scale was used to assess AD severity in the crisaborole studies. However, the existing lack of standardisation of AD scales could potentially hinder satisfactory interpretation or comparisons of AD severities across trials and real-world settings. Other AD severity measures such as EASI and SCORAD were not recorded in the crisaborole studies. Health-related quality of life was not measured using the EQ-5D or any of its suitable variants in the pivotal trials. Preference-based estimates of HRQoL were derived by mapping scores of the CDLQI and DLQI. Additionally, the measures of HRQoL used in the trials were developed for patients with skin diseases and may not capture key factors that influence HRQoL in patients aged 2 years or more with mild to moderate AD.

4.3 Critique of trials identified and included in the indirect comparison and/or multiple treatment comparison

4.3.1 Objective of the indirect comparison

Given that the crisaborole trials compared against vehicle ointment, indirect comparisons were conducted to compare crisaborole against other treatments (CS Section B.2.8). Comparison methods included an NMA and an unanchored matching-adjusted indirect comparison (MAIC).

4.3.2 Criteria for inclusion in NMA

Table 13 summarises inclusion criteria for the SLR and NMA. Inclusion criteria for the SLR are taken from CS Appendix D Table D10.(21) Inclusion criteria and final outcome selection for the NMA are not clearly reported in the CS and have been collated by the ERG from CS Section B.2.8 and Appendix D.

Population: The SLR and NMA included adults and children (≥2 years of age), and included only RCTs in which at least 80% of the trial population had mild to moderate AD (not severe AD). Trials restricted to facial AD were excluded from the base case NMA but were included in a sensitivity

analysis. A further sensitivity analysis was undertaken excluding RCTs with an older population (mean ages much higher than in the crisaborole trials).

Intervention and comparators: Inclusion criteria for the SLR and NMA included a range of topical AD treatments licensed in the UK (CS Table D10). The following interventions had relevant studies available for inclusion in the NMA: crisaborole 2%, tacrolimus 0.1%, tacrolimus 0.03%, pimecrolimus 1% and placebo/vehicle as a comparator. The ERG notes that tacrolimus is recommended by NICE for moderate AD in adults and children second-line, while pimecrolimus is only recommended for moderate AD on the face and neck in children second-line. Therefore, no treatments are NICE-recommended for mild AD second-line (i.e. post-TCS) and TCIs are not recommended by NICE for first-line use (i.e. TCS eligible population).

For a first-line (TCS eligible) population, the CS states that it was not possible to compare crisaborole against TCS in the NMA because the SLR did not identify any studies of TCS in a population that was ≥80% mild to moderate which reported relevant outcomes (CS Section B.2.8.1).

Selection of key outcome for NMA: The SLR inclusion criteria (CS Appendix D Table D10) list a range of outcomes including AD severity scores (ISGA/IGA, EASI, SCORAD and others), as well as pruritus, flares, and switching to rescue. The key efficacy outcome considered in the NMA and MAIC is the proportion achieving ISGA/IGA 0/1, i.e. clear (0) or almost clear (1) (CS Section B.2.8.2). The CS states that this outcome was selected rather than ISGA success (the primary outcome in the crisaborole studies, i.e. ISGA 0/1 and ≥2-grade improvement) because ISGA success was only reported in 3 comparator studies and the definition was very different to that for crisaborole; the ERG agrees that this is the case. Clinical advisors to the ERG felt that while ISGA/IGA 0/1 is a reasonable AD severity measure, other severity scores (such as EASI and SCORAD) may provide more granularity of response than ISGA/IGA. However, EASI and SCORAD were not recorded in the crisaborole RCTs. Clinical advisors to the ERG also noted the importance of assessing quality of life, as well as patient symptoms (e.g. via POEM). The POEM measure was not recorded in the crisaborole RCTs, while quality of life (via DLQI and CDLQI) was recorded but was not assessed in the NMA.

ISGA/IGA represent a group of AD severity measures classed as "*global assessment*" measures. There can be some variability between IGA instruments used in different trials. The CS justifies the pooling of ISGA/IGA outcomes in the NMA (CS Section B.2.8.2), stating that this is consistent with earlier NMAs and meta-analyses in AD and that clinical expert opinion judged that ISGA 0/1 and IGA 0/1 were sufficiently similar to analyse together. Given that only one other RCT(34) reported ISGA 0/1, this pooling was also necessary in order to provide a network of studies. Clinical advisors

to the ERG believed that pooling of ISGA and IGA was reasonable. Definitions of ISGA/IGA in the included trials are summarised in Table 17 of this report.

The CS (Sections B.2.8.2 and B.2.8.4) states that the outcome Physician Global Evaluation (PGE) ≥90% was not merged with ISGA/IGA 0/1 because PGE≥90% was based on a continuous scale and was measured as change from baseline rather than at a single time-point. This resulted in exclusion of one tacrolimus study(35) reporting PGE≥90%. Clinical advisors to the ERG thought that this was reasonable.

Pruritis: No NMA was reported in the CS for pruritis. The CS (Section B.2.8.4) states that a connected network for pruritis 0/1 could only be constructed for studies of pimecrolimus in facial AD. Although studies in facial AD were excluded from the base case NMA, there is a sensitivity analysis of ISGA/IGA 0/1 including studies in facial AD. However, although Appendix D alludes to an NMA of pruritus, no NMA results for pruritis are reported. The CS (Section B.3.4.4) states that pruritus is not directly included in the model because it is correlated with ISGA.

Quality of life: No quality of life outcomes were included in the NMA.

Safety outcomes: Of the safety outcomes listed in the SLR (CS Table D10), NMAs were undertaken for overall AEs, overall withdrawals and withdrawals due to AEs (CS Section B.2.9.4).. The CS states that no comparator data on serious AEs were available, and that application site pain could not be included in an NMA as it was not reported by key studies linking TCIs to vehicle.

Study design and follow-up duration: The SLR and NMA included only RCTs. The base case NMA only included studies reporting outcomes at greater than 7 days and less than 8 weeks (CS Section B.2.8.4). All studies included in the base case reported outcomes at 4-6 weeks' follow-up. NMA sensitivity analyses were undertaken to include a) studies reporting at 4 weeks only and b) studies reporting outcomes up to 14 weeks. Regarding the study duration for the crisaborole studies and the base case NMA, clinical advisors to the ERG suggested that 4-6 weeks is relatively short and that longer-term trials would be useful.

Subgroup analyses: NMA subgroup analyses for ISGA/IGA 0/1 were undertaken for: moderate AD, mild AD, children, adults, moderate AD in children, and mild AD in children. The CS states that there was no comparator data for mild AD in adults, and no connected network for moderate AD in adults (only reported in one comparator study of tacrolimus vs. pimecrolimus).

Table 13: Criteria for inclusion in SLR and NMA (adapted from CS Table D10 for SLR, and collated from CS text for NMA)

Criteria	Inclusion criteria (SLR)	Exclusion criteria (SLR)	Additional exclusions (NMA)	Final inclusion in NMA (available data)	Insufficient data for NMA
Participants	• Adults & children (≥2 years) • Mild to moderate AD (≥80% trial population)	>20% trial population: • Healthy volunteers • Children <2 years • Severe AD	• Facial AD only (included in sensitivity analysis)	• Adults & children (≥2 years) • Mild to moderate AD (≥80% trial population)	• N/A
Intervention and Comparators	• Crisaborole • TCS • TCIs: tacrolimus, pimecrolimus • Best supportive care (emollients, TCS, rescue therapies) • Alitretinoin • Other PDE4 inhibitor (OPA-15406) • Other calcineurin inhibitors (SB011) • Placebo/vehicle (as comparator)	• Treatments not approved in UK • Non-pharmacologic treatments • Systemic treatments • Trials of only emollient vs. placebo	• OPA-15406 (not licensed)	• Crisaborole • TCIs: tacrolimus, pimecrolimus • Placebo/vehicle (as comparator)	• TCS (no studies in mild to moderate AD with relevant outcomes)
Outcomes	Efficacy: • ISGA/IGA 0/1 • ISGA/IGA success • Other AD severity scores (PGE, PGA, EASI, SCORAD, others) • Pruritus • Flares • Switching to rescue Safety: • Overall AEs • Overall withdrawals • Withdrawals due to AEs • Serious AEs • Application site pain/ irritation	• Studies not reporting relevant outcomes	• N/A	Efficacy: • ISGA/IGA 0/1 Safety: • Overall AEs • Overall withdrawals • Withdrawals due to AEs Sensitivity analyses of ISGA/IGA 0/1: • Different statistical models • Incl. studies in facial AD • Excl. older patient RCTs • Up to 14 weeks	Efficacy: • ISGA/IGA success: no comparator data • Other AD severity scores: not recorded in crisaborole trials • Pruritus 0/1: unclear why not analysed (in facial AD) Safety: • Serious AEs: no comparator data • Application site pain: no connected

Table of Contents

Criteria	Inclusion criteria (SLR)	Exclusion criteria (SLR)	Additional exclusions (NMA)	Final inclusion in NMA (available data)	Insufficient data for NMA
					comparator data
Follow-up for outcome reporting	• Any duration	• N/A	• Less than (or equal to) 7 days or greater than 8 weeks	• Base case inclusion: >7 days and up to 8 weeks • All base case studies reported data at 4-6 weeks • Sensitivity analysis: Up to 14 weeks	• N/A
Study design	• RCTs	• Studies designs other than RCTs	• N/A	• RCTs	• N/A
Subgroups	• Mild and moderate AD • Adults and children	•		For ISGA/IGA 0/1: • Moderate AD • Mild AD • Children • Adults • Moderate AD in children • Mild AD in children	• Moderate AD in adults • Mild AD in adults

Abbreviations: AD, atopic dermatitis; AE, adverse event; EASI, Eczema Area and Severity Index; IGA, Investigator's Global Assessment; ISGA, Investigator's Static Global Assessment; N/A, not applicable; NMA, network meta-analysis; PDE4, phosphodiesterase-4; PGA, physician's global assessment; RCT, randomised controlled trial; SCORAD, SCORing Atopic Dermatitis; SLR, systematic literature review; TCI, topical calcineurin inhibitor; TCS, topical corticosteroid.

4.3.3 Studies included in and excluded from the NMA

Included studies: Studies included in the NMA are shown in Table 14. Nine studies were included in the NMA of effectiveness (ISGA/IGA 0/1): two studies of crisaborole vs. vehicle,(23) three studies of tacrolimus 0.03% vs. vehicle,(34, 36, 37) 1 study of pimecrolimus vs. vehicle,(38) and 3 studies of tacrolimus (0.03% or 0.1%) vs. pimecrolimus.(39-41) One additional TCI study (of tacrolimus vs. vehicle) was included in the safety NMAs.(35) Seven studies were conducted in the USA while three studies did not report the country in which they were undertaken.

Excluded studies (and studies included in sensitivity analyses): Nine studies were included in the SLR but excluded from the NMA; these are shown in Table 15 with reasons for exclusion. Seven studies of pimecrolimus were excluded: two had only 7-day follow-up;(42, 43) one had only 24-week follow-up;(44) one had no usable outcomes;(45) and three were restricted to facial AD and were excluded from the base case NMA but included in a sensitivity analysis (shown in CS Appendix D Figure D28 p250). One study of the topical PDE4 inhibitor OPA-15406 was excluded as this treatment is not licensed.(26) One study of TCS (hydrocortisone buteprate 0.01%) was excluded as no usable outcomes were reported.(46)

In addition, several further studies were excluded at the SLR stage. Nineteen studies of TCS, tacrolimus and pimecrolimus were excluded at the SLR stage due to the population being >20% severe with no subgroup analyses for mild to moderate AD (Section 4.1.2.1 and CS Appendix D p43-163, clarification response A3). Three pivotal tacrolimus RCTs undertaken for the FDA (FDA 97-0-035, FDA 97-0-036 and FDA 97-0-037) reported outcomes at 14 weeks and so were excluded from the SLR and base case NMA; however, data for the moderate AD subgroups were included in a sensitivity analysis (shown in CS Appendix D Figure D30 p253).

The ERG considers that decisions on inclusion and exclusion for the NMA appear reasonable, but notes that several studies of TCS and TCIs were excluded because the population being >20% severe.

4.3.4 Risk of bias of studies included in the NMA

Risk of bias assessment for RCTs included in the NMA using the Cochrane Risk of Bias Tool and the NICE Quality assessment tool was reported in the CS (CS Table D20 and Table D21). The crisaborole studies were scored as low risk on all criteria, whereas some of the TCI trials scored unclear on some criteria, particularly for details of randomisation and allocation concealment.

Table 14: Studies included in base case NMA (adapted from CS Table 29, Table 31 and Table D13)

Trial name / author	Country	Treatment Drugs			Adults / children	Severity group	Inclusion in NMA, with reason
		Arm 1	Arm 2	Arm 3			
Included in effectiveness and safety NMAs							
Abramovits, 2008(39)	NR	Tacrolimus 0.1%	Pimecrolimus 1%		Adults (≥16y)	Moderate	Yes
AD 301, 2016(23)	USA	Crisaborole 2%	Vehicle		Adults & children (≥2y)	Mild to moderate	Yes
AD 302, 2016(23)	USA	Crisaborole 2%	Vehicle		Adults & children (≥2y)	Mild to moderate	Yes
[a]Chapman, 2005(36)	USA	Tacrolimus 0.03%	Vehicle		Adults (≥16y)	Mild to moderate	Yes
Eichenfield, 2002(38)	NR	Pimecrolimus 1%	Vehicle		Children (1-17y)	Mild to severe	Yes
Kempers, 2004(40)	USA	Tacrolimus 0.03%	Pimecrolimus 1%		Children (2-17y)	Moderate	Yes
Levy, 2005(34)	NR	Tacrolimus 0.03%	Vehicle		Adults (≥18y)	Mild to moderate	Yes
Paller, 2005(41)	USA	Tacrolimus 0.03%	Pimecrolimus 1%		Children (2-15y)	Mild	Yes
Schachner, 2005(37)	USA	Tacrolimus 0.03%	Vehicle		Children (2-15y)	Mild to moderate	Yes
Included in safety NMAs only							
Boguniewicz, 1998(35)	USA	Tacrolimus 0.03%	Tacrolimus 0.1%	Vehicle	Children (7-16y)	Moderate to severe	Efficacy: No, ISGA/ IGA not reported Safety: Yes

[a]Chapman 2005 describes two trials; the trial in adults is listed here under Chapman 2005, while the trial in children is listed under Schachner 2005.

Table of Contents

Table 15: Studies excluded from base case NMA (adapted from CS Table 31 and Table D13)

Trial name / author	Country	Treatment Drugs			Inclusion in base case NMA, with reason	Inclusion in sensitivity analysis
		Arm 1	Arm 2	Arm 3		
Fowler, 2007(42)	USA	Pimecrolimus 1%	Vehicle		No, only 7 day follow-up	
Hanifin, 2016(26)	USA, Poland, Australia	OPA-15406 0.3%	OPA-15406 1%	Vehicle	No, OPA-15406 not licensed	
Hoeger, 2009(47)	Germany, South Korea, Slovakia	Pimecrolimus 1%	Vehicle		No, facial AD only	Yes
Hordinsky, 2010(48)	Germany, Denmark, Canada, Finland, Norway	Pimecrolimus 1%	Vehicle		No, facial AD only	Yes
Kaufmann, 2006(43)	Austria, Canada, Denmark, Hungary, Italy, Norway, USA	Pimecrolimus 1%	Vehicle		No, only 7-day follow-up	
Meurer, 2004(44)	Germany	Pimecrolimus 1%	Vehicle		No, follow-up of 24 weeks was beyond 8 week maximum for NMA	
Murrell, 2007(49)	NR	Pimecrolimus 1%	Vehicle		No, facial AD only	Yes
Sears, 1997(46)	NR	Hydrocortisone buteprate 0.01%	Placebo		No, no useable outcomes	
Wahn, 2002(45)	Europe, USA, Canada, South Africa, Austalia	Pimecrolimus 1%	Control		No, no useable outcomes	

4.3.5 Study and participant characteristics of studies included in the NMA

Participant and study characteristics for studies included in the base case NMA are shown in Table 16.

Severity: The two crisaborole trials(23) included patients with mild to moderate AD (around 60% moderate). Across the seven TCI trials in the effectiveness NMA, two trials(39, 40) had more moderate patients than the crisaborole trials (100% in both), while three trials(36, 37, 41) had fewer moderate patients (0%, 37%, 39%), one had a similar proportion of moderate patients (60%) as well as 10% severe patients,(38) and one did not report this.(34)

Age group: The crisaborole trials recruited adults (14%) and children (86%), with a mean age of 12 years. Across the seven TCI trials in the effectiveness NMA, three were in adults(34, 36, 39) and four were in children.(37, 38, 40, 41) A sensitivity analysis was undertaken excluding two trials(36, 39) in adults with mean ages much higher than in the crisaborole trials (shown in CS Appendix D Figure D29 p251).

BSA, sex and ethnicity: Mean BSA affected was 18-19% in the crisaborole trial arms and ranged from 10-26% in the TCI trial arms. The percentage of male patients was 44-45% in the crisaborole trial arms and 35-52% across TCI trial arms. The percentage of Caucasian patients was 58-63% in the crisaborole trial arms and ranged from 24-75% across TCI trial arms.

Timepoint of outcome measurement: The crisaborole trials measured outcomes at 29 days (4 weeks).**(23)** Across the seven TCI trials in the effectiveness NMA, four measured outcomes at 4 weeks(34, 37, 38, 40) and three at 6 weeks.(36, 39, 41)

Definition of ISGA/IGA outcome: All trials included in the effectiveness NMA used a measure of investigator/physician global assessment (Table 17), and all were static measures i.e. measured at a single timepoint (rather than improvement from baseline). All trials reported the percentage of patients achieving 0 (clear) or 1 (almost clear). However, the crisaborole trials(23) measured this on a 5-point scale (clear to severe) while six TCI trials(36-41) used a 6-point scale (clear to very severe) and one(34) did not report the number of scale points. It is not clear to the ERG whether this difference in scales would be likely to affect results.

Overall view on study characteristics: Overall, the ERG consider that the trials included in the NMA were broadly similar in that they considered a population with ≥80% mild to moderate AD. However, some potential treatment effect modifiers (such as age and severity) varied between trials.

Table 16: Participant and study characteristics for studies included in base case NMA (adapted from CS Table D16 and Table D18)

Trial name / author	Arm	Time point for outcomes	Adults / children	Age (mean/ median)	Severity group	Severity			Males	Caucasian	BSA
						Mild	Moderate	Severe			
Included in effectiveness and safety NMAs											
Abramovits, 2008(39)	Tacrolimus 0.1%	6 weeks	Adults (≥16y)	40	Moderate	-	100%	-	44%	44%	16%
	Pimecrolimus 1%			38		-	100%	-	39%	52%	17%
AD 301, 2016(23)	Crisaborole 2%	4 weeks	Adults & children (≥2y)	12	Mild to moderate	39%	61%	-	44%	61%	19%
	Vehicle			12		36%	64%	-	44%	63%	19%
AD 302, 2016(23)	Crisaborole 2%	4 weeks	Adults & children (≥2y)	13	Mild to moderate	38%	62%	-	45%	60%	18%
	Vehicle			12		40%	60%	-	45%	58%	18%
[a]Chapman, 2005(36)	Tacrolimus 0.03%	6 weeks[c]	Adults (≥16y)	39[b]	Mild to moderate	61%	39%	-	35%	75%	10%
	Vehicle			38[b]		65%	35%	-	35%	70%	10%
Eichenfield, 2002(38)	Pimecrolimus 1%	4 weeks	Children (1-17y)	7	Mild to severe	30%	60%	10%	52%	55%	26%
	Vehicle			7		32%	57%	11%	46%	49%	26%
Kempers, 2004(40)	Tacrolimus 0.03%	4 weeks[d]	Children (2-17y)	8	Moderate	-	99%	1%	44%	44%	NR
	Pimecrolimus 1%			8		-	99%	1%	44%	63%	NR
Levy, 2005(34)	Tacrolimus 0.03%	4 weeks	Adults (≥18y)	NR	Mild to moderate	89%		NR	NR	NR	NR
	Vehicle			NR					NR	NR	NR
Paller, 2005(41)	Tacrolimus 0.03%	6 weeks	Children (2-15y)	7	Mild	100%	-	-	47%	46%	14%
	Pimecrolimus 1%			6		100%	-	-	43%	42%	14%
Schachner, 2005(37)	Tacrolimus 0.03%	4 weeks	Children (2-15y)	7	Mild to moderate	61%	39%	-	47%	65%	12%
	Vehicle			7		60%	40%	-	47%	71%	13%
Included in safety NMAs only											
Boguniewicz, 1998(35)	Tacrolimus 0.03%	5 weeks	Children (7-16y)	10	Moderate to severe	-	88%	12%	42%	24%	18%
	Tacrolimus 0.1%			11		-	86%	14%	43%	38%	16%
	Vehicle			10		-	73%	27%	41%	27%	20%

[a]Chapman 2005 describes two trials; the trial in adults is listed here under Chapman 2005, while the trial in children is listed under Schachner 2005. [b]Median age. [c]Chapman 2005: incorrectly stated in CS Table D14 as 4 weeks; corrected to 6 weeks in clarification response. [d]Kempers 2004: Incorrectly stated in CS Table D14 as 43 days (6 weeks); corrected to 29 days (4 weeks) in clarification response.

Table of Contents

Table 17: ISGA/IGA 0/1 outcome measures for studies included in base case NMA

Trial name / author	Arm 1	Arm 2	IGA name	Definition of response on ISGA/IGA scale	Number of ISGA/IGA scale points
Abramovits, 2008(39)	Tacrolimus 0.1%	Pimecrolimus 1%	IGADA	0 (clear) or 1 (almost clear)	6-point (clear to very severe)
AD 301(23)	Crisaborole 2%	Vehicle	ISGA	0 (clear) or 1 (almost clear)	5-point (clear to severe)
AD 302(23)	Crisaborole 2%	Vehicle	ISGA	0 (clear) or 1 (almost clear)	5-point (clear to severe)
[a]Chapman, 2005(36)	Tacrolimus 0.03%	Vehicle	IGADA	0 (clear) or 1 (almost clear)	6-point (clear to very severe)
Eichenfield, 2002(38)	Pimecrolimus 1%	Vehicle	IGA	0 (clear) or 1 (almost clear)	6-point (clear to very severe)
Kempers, 2004(40)	Pimecrolimus 1%	Tacrolimus 0.03%	IGA	0 (clear) or 1 (almost clear)	6-point (clear to very severe)
Levy, 2005(34)	Tacrolimus 0.03%	Vehicle	PSGA	0 (clear) or 1 (almost clear)	(number of scale points not reported)
Paller, 2005(41)	Tacrolimus 0.03%	Pimecrolimus 1%	IGADA	0 (clear) or 1 (almost clear)	6-point (clear to very severe)
Schachner, 2005(37)	Tacrolimus 0.03%	Vehicle	IGADA	0 (clear) or 1 (almost clear)	6-point (clear to very severe)

IGA: Investigator's Global Assessment; IGADA: Investigator Global Atopic Dermatitis Assessment; ISGA: Investigator's Static Global Assessment; PSGA: Physician's Static Global Assessment.

Table of Contents

4.3.6 Effectiveness results of trials included in NMA

ISGA/IGA 0/1

The percentage of patients with ISGA/IGA 0/1 within trials included in the NMA are shown in Table 18 (based on CS Appendix D Table D14). Proportions achieving ISGA/IGA 0/1 were as follows:

- Two trials(23) of crisaborole (49%, 52%) vs. vehicle (30%, 41%)
- Three trials(34, 36, 37) of tacrolimus 0.03% (39%, 41%, 49%) vs. vehicle (23%, 19%, 32%)
- One trial(38) of pimecrolimus (31%) vs. vehicle (12%)
- One trial(39) of tacrolimus 0.1% (45%) vs. pimecrolimus (31%)
- Two trials(40, 41) of tacrolimus 0.03% (32%, 47%) vs. pimecrolimus (24%, 41%).

Table 18: Results of trials included in NMA: ISGA/IGA 0-1 (adapted from CS Table D14)

Trial number (Acronym)	Adults / children	Severity group	Timepoint (weeks)	Crisaborole 2%	Tacrolimus 0.03%	Tacrolimus 0.1%	Pimecrolimus 1%	Vehicle
AD 301, 2016(23)	Adults & children (≥2y)	Mild to moderate	week 4	260/503 (52%)				104/256 (41%)
AD 302, 2016(23)	Adults & children (≥2y)	Mild to moderate	week 4	249/513 (49%)				74/250 (30%)
[a]**Chapman, 2005**(36)	Adults (≥16y)	Mild to moderate	week 6[b]		74/152 (49%)			48/148 (32%)
Levy, 2005(34)	Adults (≥18y)	Mild to moderate	week 4		17/44 (39%)			10/44 (23%)
Schachner, 2005(37)	Children (2-15y)	Mild to moderate	week 4		65/158 (41%)			31/159 (19%)
Eichenfield, 2002(38)	Children (1-17y)	Mild to severe	week 4				83/267 (31%)	16/136 (12%)
Abramovits, 2008(39)	Adults (≥16y)	Moderate	week 6			44/98 (45%)	28/90 (31%)	
Kempers, 2004(40)	Children (2-17y)	Moderate	week 4[c]		22/69 (32%)		17/70 (24%)	
Paller, 2005(41)	Children (2-15y)	Mild	week 6		97/207 (47%)		88/216 (41%)	

[a]Chapman 2005 describes two trials; the trial in adults is listed here under Chapman 2005, while the trial in children is listed under Schachner 2005. [b]Chapman 2005: incorrectly stated in CS Table D14 as 4 weeks; corrected to 6 weeks in clarification response. [c]Kempers 2004: Incorrectly stated in CS Table D14 as 43 days (6 weeks); corrected to 29 days (4 weeks) in clarification response.

Table of Contents

4.3.7 Safety results of trials included in NMA

Withdrawals and adverse events

Data for overall adverse events within trials included in the NMA are shown in Table 19, and data for overall withdrawals and withdrawals due to adverse events are shown in Table 20 (based on CS Appendix D Table D17). No results were reported in the CS for the Boguniewicz (1998)(35) trial which was included in the safety NMAs only. The CS notes that the analysis of overall adverse events is subject to high variability in reporting between trials. Within each trial, the percentage with overall adverse events was similar between arms (including between active intervention and vehicle arms). Overall withdrawals within each trial tended to be higher in the vehicle group than the active treatment group. Withdrawals due to adverse events within each trial were either higher in the vehicle group or similar between groups.

Table 19: Results of trials included in NMA: adverse effects (adapted from CS Table D17)

Trial number (Acronym)	Adults / children	Severity group	Time point (weeks)	Overall AEs				
				Crisaborole	Tacrolimus 0.03%	Tacrolimus 0.1%	Pimecrolimus 1%	Vehicle
AD 301, 2016(23)	Adults & children (≥2y)	Mild to moderate	week 4	147/502 (29%)				50/252 (20%)
AD 302, 2016(23)	Adults & children (≥2y)	Mild to moderate	week 4	150/510 (29%)				79/247 (32%)
[a]**Chapman, 2005**(36)	Adults (≥16y)	Mild to moderate	week 6[b]		142/310 (46%)			161/307 (52%)
Levy, 2005(34)	Adults (≥18y)	Mild to moderate	week 4		24/44 (55%)			21/44 (48%)
Schachner, 2005(37)	Children (2-15y)	Mild to moderate	week 4		58/158 (37%)			72/159 (45%)
Eichenfield, 2002(38)	Children (1-17y)	Mild to severe	week 4				117/267 (44%)	58/136 (43%)
Abramovits, 2008(39)	Adults (≥16y)	Moderate	week 6			32/98 (33%)	21/90 (23%)	
Kempers, 2004(40)	Children (2-17y)	Moderate	week 4[c]		59/70 (84%)		61/71 (86%)	
Paller, 2005(41)	Children (2-15y)	Mild	week 6		32/208 (15%)		36/217 (17%)	

Table 20: Results of trials included in NMA: withdrawals (adapted from CS Table D17)

Trial number (Acronym)	Adults / children	Severity group	Time point (weeks)	Overall withdrawals					Withdrawals due to AEs				
				Crisaborole	Tacrolimus 0.03%	Tacrolimus 0.1%	Pimecrolimus 1%	Vehicle	Crisaborole	Tacrolimus 0.03%	Tacrolimus 0.1%	Pimecrolimus 1%	Vehicle
AD 301, 2016(23)	Adults & children (≥2y)	Mild to moderate	week 4	30/507 (6%)				31/256 (12%)	7/507 (1%)				2/256 (1%)
AD 302, 2016(23)	Adults & children (≥2y)	Mild to moderate	week 4	31/514 (6%)				37/250 (15%)	5/514 (1%)				4/250 (2%)
[a]**Chapman, 2005**(36)	Adults (≥16y)	Mild to moderate	week 6[b]		50/310 (16%)			101/307 (33%)		6/310 (2%)			21/307 (7%)
Levy, 2005(34)	Adults (≥18y)	Mild to moderate	week 4		NR			NR		NR			NR
Schachner, 2005(37)	Children (2-15y)	Mild to moderate	week 4		29/158 (18%)			61/159 (38%)		7/158 (4%)			12/159 (8%)
Eichenfield, 2002(38)	Children (1-17y)	Mild to severe	week 4				34/267 (13%)	30/136 (22%)				5/267 (2%)	4/136 (3%)
Abramovits, 2008(39)	Adults (≥16y)	Moderate	week 6			18/98 (18%)	21/90 (23%)				2/98 (2%)	2/90 (2%)	
Kempers, 2004(40)	Children (2-17y)	Moderate	week 4[c]		3/70 (4%)		13/71 (18%)			1/70 (1%)		5/71 (7%)	
Paller, 2005(41)	Children (2-15y)	Mild	week 6		47/209 (22%)		56/217 (26%)			1/209 (0.5%)		11/218 (5%)	

4.3.8 Vehicle response in the crisaborole and comparator trials

Company view on differing vehicle responses: [redacted] The CS states that experts have indicated that responses in the vehicle arms of the crisaborole trials were greater than for most topical treatments, which may be because in older trials the vehicle may be irritating and induce dermatitis (the CS does not state how old such trials might be expected to be; the TCI trials in the NMA were published between 2002 and 2008 while the crisaborole trials were published in 2016). This difference in vehicle response is the reason given in the CS for using a base case NMA model which adjusts for vehicle response using meta-regression, and also for using a MAIC.

The vehicle compositions of crisaborole, tacrolimus and pimecrolimus are shown in Table 21. The CS (Section B.2.8.3) states that [redacted]

Clinical advisors to the ERG noted that [redacted] Clarification question A18 therefore asked why a difference in response might be expected between the crisaborole and tacrolimus vehicles, given that both are ointments. The response to question A18 states that [redacted]

Response rates in vehicle arms of trials included in NMA: Six trials in the NMA had a vehicle arm (Table 18). Vehicle responses were [redacted] and [redacted] in two crisaborole trials (vehicle is an ointment); [redacted] in three tacrolimus trials (vehicle is an ointment); and [redacted] in one pimecrolimus trial (vehicle is a cream). [redacted]

Clinical information on difference between vehicles: As noted above, clinical advisors to the ERG noted that ointments (such as crisaborole and tacrolimus) may give greater response rates than creams (such as pimecrolimus). A Cochrane review published in 2017(50) (77 trials, 66003 participants) noted beneficial effects of most emollients, but did not find reliable evidence that one emollient is better than another. A recent BMJ summary(51) by UK specialists also stated that there is little evidence to recommend one type of emollient over another, but noted that ointments may be more effective for more severe dryness and require less frequent application, while lighter creams and lotions may be more cosmetically acceptable as they are less greasy than ointments. This article also states that older, cheaper emollients may be as effective as newer, more expensive ones.

Table 21: Comparison of ingredients used for topical interventions included in the NMA (reproduced from CS Table D19)

Base Components	Crisaborole Ointment	Tacrolimus Ointment	Pimecrolimus Cream
White petrolatum	[redacted]		
Paraffin/white wax			
Mineral oil			
Propylene glycol			
Propylene carbonate			
Mono-di-glycerides			
Triglycerides			
Citric acid			
Oleyl alcohol			
Benzyl alcohol			
Cetyl alcohol			
Stearyl alcohol			
Sodium cetostearyl sulphate			

Base Components	Crisaborole Ointment	Tacrolimus Ointment	Pimecrolimus Cream
Butylated hydroxytoluene	[redacted]		
Edetate calcium disodium			
Sodium Hydroxide			

4.4 Critique of the indirect comparison and/or multiple treatment comparison

4.4.1 Overview of results from different models of NMA and MAIC

Comparison of crisaborole against other treatments was undertaken using NMA (various models) and an unanchored MAIC. The appropriateness of each of these models is discussed in the sections below.

The methods for the NMA and MAIC, and the results of the NMA sensitivity analyses, are presented in Appendix D of the CS. The main NMA results, NMA subgroup analyses by age and severity, and MAIC results are presented in Section B.2.8 of the CS.

Table 22 summarises the results for IGSA/IGA 0/1 for the different models of NMA and MAIC. These include:

- Company results for the ERG's preferred simple NMA model (random treatment effect, no class effect, no vehicle response adjustment);
- ERG-replicated results for the ERG's preferred simple NMA model (random treatment effect, no class effect, no vehicle response adjustment)
- Company's preferred base case model (fixed treatment effect, random class effect, adjustment for vehicle response [FE-RCE-VR]);
- Unanchored MAIC.

The NMAs accounted for differences in duration of follow-up between studies by assuming that the times to response followed an exponential distribution, with the estimate of treatment effect being a hazard ratio (HR). No adjustment for follow-up duration was made when conducting the unanchored MAIC, with the estimate of treatment effect being an odds ratio (OR). Results are presented in Table 22 for crisaborole relative to each of the comparators. In this case, a HR (or OR) greater than 1 indicates that crisaborole is more effective than the comparator. The absolute response rates used in the base case economic model are obtained by combining the HRs with the natural history model presented in Table 33 of the CS. When using the MAIC, the response rates for crisaborole vary for each comparator. For reference, the absolute response rates for the company's base case NMA and MAIC can be seen in Tables 57 and 58 of the CS. (NB: the response rates for crisaborole differ between Table 58 and Table 71 of the CS and the rate in Table 71 was applied in the model.)

[redacted] **Table 22: ISGA/IGA 0/1: Comparison of the results of the simple NMA (company and ERG), vehicle-adjusted NMA, and unanchored MAIC**

Comparison vs. crisaborole	Simple random effects NMA:[a] Company Results - random treatment effects - no class effect - no vehicle response adjustment (CS Fig D25 p247, Table 32)[d]		Simple random effects NMA:[a] ERG Results - random treatment effects - no class effect - no vehicle response adjustment		Company base case model (FE-RCE-VR) - fixed treatment effect - random class effects - vehicle response adjusted (CS Figure 19, Table 32)		Unanchored MAIC (CS Fig 30)	
	HR[c] 95% CrI	SD[b]	HR[c] 95% CrI	SD	HR[c] (95% CrI)		OR[c] (95% CI)	
Vehicle	[redacted]							
Tacrolimus 0.03%								
Tacrolimus 0.1%								
Pimecrolimus 1%								
[redacted]								

[a] Random treatment effect, no class effect, no vehicle response adjustment
[b] SD based on the ERG's analysis using the company's prior distributions
[c] HRs / ORs above 1 favour crisaborole. The ERG notes that the results for the NMA and the separate MAICs for each comparator are not directly comparable as they are estimated in different populations (see section 4.4.4)
[d]Note that a figure numbering error in CS Appendix D means that several figures have the same number.

4.4.2 Overview of NMA sensitivity analyses, subgroup analyses and additional outcomes

NMA sensitivity analyses

Appendix D, Page 242 of the CS describes six sensitivity analyses of the key outcome (IGSA/IGA 0/1). Results of these are summarised here in Table 23. The ERG notes that the results of the sensitivity analyses vary as to whether the median estimates favour crisaborole, favour TCIs, or indicate little difference between treatments.

- ISGA/IGA 0/1 up to 6 weeks using a simple random treatment effects model (random treatment effect, no class effect, no vehicle response adjustment). This is the model corresponding to the results presented in Table 22. In this model, results differ from the company base case and the median estimates favour TCIs over crisaborole.
- ISGA/IGA 0/1 up to 6 weeks with no class effects but using fixed treatment effects and vehicle response regression. Results were similar to the company base case.

- ISGA/IGA 0/1 up to 4 weeks and no adjustment for study duration (with class effects and vehicle response regression). Results were similar to the company base case.
- ISGA/IGA 0/1 up to 6 weeks including 3 studies in facial AD (using FE-RCE-VR model), i.e. including Murrell 2007, Hoeger 2009, and Hordinsky 2010. Results differ somewhat from the company base case, with little difference between crisaborole and TCIs and median estimates favouring TCIs.
- ISGA/IGA 0/1 up to 6 weeks excluding older patient RCTs (using FE-RCE-VR model), i.e. excluding Abramovits 2008 and Chapman 2005 adult study. Results were similar to the company's base case.
- ISGA 0/1 up to 14 weeks including three additional FDA tacrolimus RCTs (using FE-RCE-VR model). Results differ somewhat from the company's base case, with little difference between crisaborole and TCIs.

NMA subgroup analyses

A set of NMA subgroup analyses by age and severity for the key outcome (IGSA/IGA 0/1) are presented in Section B.2.8.8 of the CS. Results were broadly similar to the company base case (FE-RCE-VR model) for the subgroup analyses which used the FE-RCE-VR model. In subgroups with insufficient evidence to use this model, a simple random treatment effects model was used with the results broadly similar to those using the simple NMA model.

NMA additional effectiveness outcomes

Appendix D, Table 23 of the CS describes ten additional outcomes and/or duration of outcome combinations that were analysed using network meta-analysis models, although no results from these analyses are included in the CS.

Table 23: ISGA/IGA 0/1: Results of NMA sensitivity analyses (based on CS Appendix D p242-255)

	Simple random effects NMA (as in Table 22) - random treatment effects - no class effect - no vehicle response adjustment - up to 6 weeks (CS App D	**No class effects** - fixed treatment effect - no class effect - vehicle response adjusted - up to 6 weeks (CS App D Fig D26 p248)	**4 weeks, no offset for study duration** - random class effects - vehicle response adjusted - up to 4 weeks (CS App D Fig D27 p249)	**Including 3 studies in facial AD** - fixed treatment effect - random class effects - vehicle response adjusted - up to 6 weeks (CS App D Fig D28 p250)	**Excluding 2 studies in older patients** - fixed treatment effect - random class effects - vehicle response adjusted - up to 6 weeks (CS App D Fig D29 p251)	**Including 3 FDA tacrolimus studies; up to 14 weeks** - fixed treatment effect - random class effects - vehicle response adjusted - up to 14 weeks

	Fig D25 p247)					(CS App D Fig D30 p253)
Comparison vs. crisaborole						
Vehicle						
Tacrolimus 0.03%						
Tacrolimus 0.1%						
Pimecrolimus 1%						
Total residual deviance						
DIC						

[a] HRs above 1 favour crisaborole.

4.4.3 Safety NMAs

Section B.2.9.4 of the CS describes NMAs undertaken for the following safety outcomes:

- Overall AEs (the CS notes that there was high variability in reporting of this outcome)
- Withdrawals due to AEs
- Overall withdrawals.

Evidence about serious AEs and application site pain did not form a connected network. Safety NMAs were undertaken using the FE-RCE-VR model for overall AEs and overall withdrawals, and a simple random treatment effects model for withdrawals because of AEs.

Results of the safety NMAs are summarised in Table 24. Due to the different models used, and the wide CrIs for most of the estimates, the ERG considers that the direction and magnitude of the safety comparisons between treatments are unclear.

Table 24: Results of NMAs for safety outcomes (based on CS Section B.2.9.4)

	Overall AEs - FE-RCE-VR model - up to 8 weeks (CS Fig 32)	**Withdrawals due to AEs** - simple random treatment effects model - up to 8 weeks (CS Fig 33)	**Overall withdrawals** - FE-RCE-VR model - up to 8 weeks (CS Fig 32)
Comparison vs. crisaborole	**HR[a] 95% CrI**	**HR[a] 95% CrI**	**HR[a] 95% CrI**
Vehicle			
Tacrolimus 0.03%			
Tacrolimus 0.1%			
Pimecrolimus 1%			

Total residual deviance	[redacted]	[redacted]	[redacted]

[a] HRs below 1 favour crisaborole.

4.4.4 Critique of methods for Network Meta-Analysis (NMA)

The effects of treatments on ISGA/IGA 0/1 was assessed using evidence from nine RCTs at Week 4 (including evidence from studies up to Week 6) (CS, Appendix D, Table D14). The company used 12 meta-analysis models to analyse the data, including fixed and random treatment effects models; no class effect, and fixed and random class effects models; and a meta-regression adjusting for baseline response associated with the vehicle (CS, Section B.2.8.5(1)).

Choice of fixed and random treatment effect models

In general, we expect variability about treatment effects between studies because each study uses a different protocol (the exception in this submission being the two crisaborole studies which used the same protocol). Furthermore, relative goodness-of-fit measures do not provide relevant information against prior beliefs about parameter values in Bayesian network meta-analyses. Consequently, the ERG does not consider it appropriate to choose between fixed and random treatment effects using relative goodness-of-fit measures. The ERG's preference is to use a random treatment effects model incorporating plausible prior heterogeneity about treatment effects between studies.

Appropriateness of using a class effect model

There are two classes of treatment:

- TCIs: pimecrolimus 1%, tacrolimus 0.03% and tacrolimus 0.1%
- PDE4-inhibitors: crisaborole

Class effect models aim to borrow strength about relative treatment effects across treatments of the same class and either assumes them to be similar and exchangeable (random effects) or identical (fixed effect). While there may be interest in quantifying the effectiveness and cost-effectiveness of TCIs as a class and the variability about treatment effects within this class, it is not necessary to do so in order to quantify the effectiveness and cost-effectiveness of crisaborole against specific comparator treatments. Furthermore, the ERG has some concern with the implementation of the class effects model and the feasibility of estimating parameters in the model. In particular, it is unclear whether there is sufficient information from the sample data alone to quantify the between-treatments within-class standard deviation for TCIs, a class of treatments comprising three treatment effects. The ERG notes that in Appendix D of the CS (page 198), the company states that there are two treatments per class (i.e. TCIs (pimecrolimus, tacrolimus) and PDE4-inhibitors (crisaborole, OPA-15406)). This is not a correct description of the class effect model in that there are three TCI treatment effects and only one PDE4-inhibitor effect; OPA-15406 was not included in the NMA because it is not a licensed

treatment. The OpenBUGS code is presented in Appendix D of the CS (Page 264) but without the data.

[redacted]

It is possible that a different model was used to analyse the data but the OpenBUGS code for this is not included in Appendix D. In response to clarification question A22, the company clarified

[redacted]

In Section B.2.8.7 of the CS, the company reports that the standard deviation of the random class effect for TCIs (on the log-hazard scale) had mean [redacted] and standard deviation [redacted] The ERG notes that the median of the posterior distribution is a better estimate of the true value than the mean when the posterior distribution is skewed (as expected in this case).

[redacted]

The ERG suggests that not including class in the model would provide a more parsimonious model given the limited number of treatments within each class; otherwise, it would be necessary to incorporate into the model reasonable prior beliefs about the extent of the between-treatments within-class heterogeneity to ensure plausible posterior results.

[redacted]

Impact of so-called non-informative prior distributions

The company's response to clarification question A22 only addressed the issue of the prior distribution used for the between-treatments within-class standard deviation but did not address the prior distributions used for study baselines or the regression parameter for the effect of the baseline response to vehicle. The ERG has concern with the use of so-called non-informative prior distributions that the company has used in the analyses. With a limited number of studies with which

to estimate parameters and prior distributions that do not represent reasonable prior beliefs there may not be sufficient Bayesian updating so that posterior distributions may not represent reasonable posterior beliefs. The ERG has replicated the company's results using the model described by the company as the "*simple random treatment effects model*" model (random treatment effect, no class effect, no vehicle response adjustment) (CS, Appendix D, Page 247) and, in the absence of genuine prior information, re-analysed the data using the following weakly informative prior distributions:

- Baseline hazards and hazard ratios: $N(0, 1000)$
- Between-study standard deviation: $U(0, 2)$

The company's results for the "*simple random treatment effects model*" are presented in Table 32 of the CS and Figure D25 (Page 247) of CS Appendix D (note that a figure numbering error in Appendix D means that multiple figures have the same number).

The ERG's results have been generated using the number of patients with ISGA/IGA 0/1 and crisaborole defined as the reference treatment. To generate hazard ratios greater than one it was necessary for the ERG to invert the hazard ratios from this model; it is not clear how the company generated its results because neither the data used in the analysis nor the OpenBUGS code is provided for this analysis. Table 22 shows that the company's and ERG's results (for a simple random treatment effects model) are similar in terms of point and interval estimates, and goodness-of-fit criteria; the ERG believes that the company's estimate of the between-study standard deviation is the posterior mean rather than the posterior median. However, the company's posterior distributions are skewed and include extreme values, particularly for tacrolimus 0.01% and pimecrolimus 1% as reflected in their posterior standard deviations. The ERG's re-analysis incorporating weakly informative prior information has mitigated the possibility of extreme values in the posterior distributions.

The ERG is unable to assess the impact of the so-called non-informative prior distribution used by the company in its preferred base case model (fixed treatment effect, random class effect, adjustment for vehicle response). Although the conclusions of the company's and ERG's analysis for the simple model are the same based on the results in Table 22, it is important to ensure that posterior distributions represent reasonable posterior beliefs. The ERG considers it likely that the company's posterior distributions of the parameters representing the between-treatments within-class standard deviation and the regression parameter for the effect of the response to vehicle both include implausible values.

Transitivity, treatment effect modifiers and vehicle response

An assumption made when estimating indirect treatment effects in a network meta-analysis is that of transitivity. Transitivity means that if the study-specific population treatment effect of treatment B is greater than that of treatment A, and if the study-specific population treatment effect of treatment C is greater than that of treatment B, then the study-specific population treatment effect of treatment C must be greater than that of treatment A. For transitivity to hold:

- Every treatment in the network should have a fixed definition.
- Treatments not included in each study are missing at random.
- There should be no difference in the distribution of treatment effect modifiers across studies comparing different pairs of treatments.

The company (CS, Section B.2.8.3(1)) notes that the vehicles used across studies are formulated with different excipients. The ERG accepts that lumping treatments that are not sufficiently similar can lead to violation of the transitivity assumption and result in inconsistent estimates of direct and indirect treatment effect. However, inconsistency between direct and indirect estimates of treatment effect can only be assessed for treatments forming feedback loops, which does not apply to crisaborole in this network. Accepting that the vehicles in the crisaborole studies are different to those in the remaining studies would create a disconnected network of evidence with crisaborole and its vehicle control separate to the rest of the evidence base. There are various ways in which such a network of evidence could be analysed, each with their own limitations. The company chose to compare treatments using unanchored matching-adjusted indirect comparisons (See Section 4.4.6). The ERG notes that there is some evidence to suggest that there is inconsistency between indirect and direct estimates of treatment effects with respect to vehicle, pimecolimus 1% and tacrolimus 0.03% in the company's preferred base case model (fixed treatment effect, random class effect, adjustment for vehicle response) (CS, Figure 21(1)). The Bayesian p-value for the comparisons between direct and indirect estimates of treatment effect [redacted] Inconsistency between direct and indirect estimates of treatment effects for treatments forming feedback loops also arises because of an imbalance in the distribution of treatment effect modifiers across studies comparing different pairs of treatments. Differences in the distribution of treatment effect modifiers will also give biased indirect estimates of treatment effect for treatments not forming feedback loops (i.e. all comparator treatments versus crisaborole) but without any way of assessing the bias in the network. The company performed an extensive assessment of potential treatment effect modifiers (see Section 4.4.6 on MAIC). The following baseline characteristics were considered potential treatment effect modifiers (CS, Appendix D, Table 35):

[redacted]In principle, an imbalance in the distribution of any of these characteristics across studies comparing different pairs of treatments could generate biased estimates of indirect treatment effects. The company quantified the effect of the following baseline characteristics, including adjustment for vehicle response (see below), in separate meta-regression models: [redacted] The analyses did not include reasonable prior beliefs about the potential relationship between treatment effect and baseline characteristic and it is not clear whether posterior estimates represent reasonable posterior results. Furthermore, the results may be subject to the ecological fallacy; the company could have investigated the use of a mixed patient-level and aggregate data meta-regression using data from its crisaborole studies. The ERG notes that the DICs are all within five units (CS, Table 37(1)), which is usually considered small, and the ERG suggests that it would be better to use a more parsimonious model in the absence of further evidence.The company (CS, Section B.2.8.3) notes that relatively high response rates to treatment with vehicle in the crisaborole studies was expected and observed given the formulation, and that these were different to the response rates observed in older studies. In response to clarification question A9, the company [redacted]The ERG notes that ignored prognostic factors in non-linear models will produce biased estimates of treatment effect irrespective of whether there is baseline balance.

The ERG notes that six studies provided estimates of the response to treatment with vehicle. To support the company's assertion that the differences are associated with the age of the study the ERG suggests that an analysis could be conducted of the relationship between the baseline response and the date of study. While it is possible that the observed response rates are associated with other baseline

characteristics, the ERG suggests that with only six observations it is difficult to estimate regression parameters in meta-regression models using the sample data alone.

The company attempted to account for differences in baseline responses associated with the vehicle by performing a meta-regression. Although the company's preferred base case model (fixed treatment effect, random class effect, adjustment for vehicle response, i.e. FE-RCE-VR model) was chosen based on goodness-of-fit criteria, it is not clear whether the posterior results represent reasonable posterior beliefs. Alternative meta-regression models would assume that the relationship between treatment effect and response to vehicle is different for each treatment, is similar and exchangeable between treatments or is identical for each treatment. Given that only six studies provided information on the response to vehicle, the meta-regression model assumed that the relationship between treatment effect and response to vehicle was the same for crisaborole, tacrolimus 0.03% and pimecrolimus 1%. Two studies evaluated the effect of crisaborole vs. vehicle, three studies evaluated the effect of tacrolimus 0.03% vs. vehicle and one study evaluated the effect of pimecrolimus 1% vs. vehicle. It is only through the studies comparing tacrolimus 0.03% to vehicle that a linear relationship can be quantified and assumed to hold for crisaborole and pimecrolimus 1%, and it is not possible to assess deviations from linearity within and across treatments. So-called non-informative prior distributions have been used for the uncertain parameters in the meta-regression model, including the regression slope and parameters associated with class effects. Given the complexity of the model, the limited sample data and the use of so-called non-informative prior distributions, it is not clear whether the posterior distributions include extreme values that are implausible. [redacted] The company's preferred base case model (fixed treatment effect, random class effect, adjustment for vehicle response) means that the relative treatment effect depends on the true population response to treatment with vehicle; this is in contrast to a model in which the relative treatment effect is additive on some appropriate scale. Furthermore, the meta-regression is fitted with the effect of the response to vehicle centred on the population mean across all six studies. The interpretation of the relative treatment effects estimated by the company are as the relative treatment effects at the population mean response to vehicle across the six studies included in the analysis. It is not clear whether this estimate of treatment effect is relevant for decision-making.

Model comparison

The company's preferred base case model was chosen based on comparing the Deviance Information Criterion (DIC) of 12 models. Differences in DIC values of up to five are generally considered small

unless the models leads to different conclusions; in these circumstances the model with the smallest DIC is not necessarily the best model for decision-making. The DIC for all 12 models are all within a difference of five (CS, Table 32(1)). The company's preferred base case model (fixed treatment effect, random class effect, adjustment for vehicle response) has a DIC of 124.8 based on the company's results. The ERG's preferred model (random treatment effect, no class effect, no vehicle response adjustment) has a DIC of 129.2 based on the company's results and 129.3 based on the ERG's results. This is a more parsimonious model that does not make strong assumptions about the relationship between baseline risk and treatment effect. Although the ERG's preferred model would be wrong if the vehicles are systematically different, the results do not indicate excessive heterogeneity that might arise as a consequence of differences in response to vehicle between studies.

4.4.5 Baseline Natural History Model

Section B.2.8.6 presents information regarding the baseline natural history model. Results are generated by performing a random effects meta-analysis of the responses to treatment with vehicle for the six studies that used vehicle as the control treatment. The model for the data is a binomial likelihood with a complementary log-log link function to allow for differences between studies in duration of follow-up. The model also incorporates covariates for the mean age and proportion of patients with moderate disease in each study. Results of the company's baseline natural history model are presented in Table 33 of the CS. Although the company has implemented its meta-regression in a standard way, the ERG is concerned with the implied relationship between baseline response (i.e. log-hazard) and the covariates, which is assumed to be linear with respect to mean age and proportion with moderate disease. No information is provided regarding the adequacy of the model in terms of its predictions or the estimate of the between-study standard deviation.

The ERG would prefer not to use the evidence from the six studies to create a baseline natural history model. Rather, the ERG believes that the data sources that best inform the baseline natural history might be cohort studies, a carefully selected subset of the studies included in the meta-analysis, or expert opinion. Furthermore, the company's approach does not specifically address the issue of the definition of the comparator treatment and the expected response in the target population.

4.4.6 Matching Adjusted Indirect Comparison (MAIC)

Details of the methods used by the company to make comparisons between treatments using unanchored MAIC are presented in Appendix D of the CS. The main findings are presented in Section B.2.8.10 of the CS.

An unanchored MAIC involves weighting the data to ensure that the means and standard deviations of continuous covariates of the reweighted patient-level data match the means and standard deviations of

the aggregate data, and similarly for discrete covariates based on proportions, based on potential treatment effect modifiers and prognostic variables.

Potential treatment effect modifiers or prognostic variables were determined through:

- Regression analyses of the patient-level data from the AD-301 and AD-302 crisaborole studies
- Literature reviews of published evidence
- Clinical expert opinion.

For each of 15 covariates, the company used univariable logistic regression analyses using pooled AD-301 and AD-302 data to evaluate whether there was evidence that each covariate was a prognostic variable or treatment effect modifier. The ERG does not believe that this approach is appropriate because it cannot properly control for confounding or inter-correlations between independent variables. Ideally, methods based on comparing full and reduced models should be used to minimise potential bias and control for confounding.(52)

MAIC assumes that all treatment effect modifiers and prognostic factors are accounted for in the analysis, including higher order terms and interactions. This is a strong assumption that leads to an unknown amount of bias in the unanchored estimate.

Furthermore, MAIC assumes that the target population is closer to that represented by the comparator study rather than the population represented by the crisaborole study(s). In practice, the populations represented by the comparator and crisaborole studies may not be representative of the target populations.

The company used a different model to make population adjusted indirect comparisons compared to that used for the NMAs; time was ignored and treatment effects were estimated on the log-odds scale rather than the log-hazard scale.

Histograms of the weights used to generate the effective sample sizes are presented in Appendix D (Figure D33). The ERG notes that the x-axes and y-axes are on different scales and it is difficult to compare these across treatments. Nevertheless, it can be seen that the weights are highly variable in the case of tacrolimus 0.1% and also to some extent for tacrolimus 0.03% resulting in effective sample sizes [redacted] of the crisaborole sample size, respectively. [The ERG notes that there is a discrepancy in the reporting of the effective sample size for tacrolimus 0.03% in Table 42 of the CS and Figure D33 of Appendix D.]

A naive unadjusted indirect comparison includes sampling error and systematic error arising from an imbalance in prognostic factors and treatment effect modifiers, whereas systematic error should be reduced with appropriate use of MAIC. A comparison of naïve unadjusted and adjusted results suggests that there are modest adjustments for systematic error and relatively small impacts on sampling variation in spite of the differences in patient populations (CS, Figure 30(1)).

4.4.7 Summary of ERG's view on indirect comparisons

In summary, on the basis of using a more parsimonious model for models with relatively similar goodness-of-fit, the ERG's preferred model is the NMA with random treatment effects, no class effect, and no vehicle response adjustment. In particular, it is unclear whether the assumptions made in the more complex vehicle response adjusted model preferred by the company are reasonable and apply to all treatment effects. Nevertheless, the ERG accepts that the simple NMA model may also be wrong if the vehicles are systematically different. Although the MAIC circumvents the need to adjust for vehicle response, it provides inferences relative to the populations defined by the comparator studies, which may not be useful for decision-making. In addition, the MAIC assumes that that all prognostic factors and treatment effect modifiers have been accounted for, although this is an untestable assumption. For these reasons, the ERG considers that caution should be exercised when making inferences based on the results of the MAIC.

In the ERG's preferred model (simple random effects NMA with no class effect and no vehicle response adjustment), the median estimates favour TCIs over crisaborole, whereas in the company's preferred base case model (fixed treatment effect, random class effect, adjustment for vehicle response) and in the unanchored MAIC, the median estimates favour crisaborole over TCIs. Therefore, the interpretation of the effectiveness of crisaborole relative to TCIs varies depending on the assumptions made in each model. However, the ERG notes that NMA and MAIC results are not directly comparable because they apply to different populations. Although the ERG prefers a simple random effects NMA on the basis that it is the most parsimonious model with evidence of only moderate heterogeneity between studies, it is difficult to know which model is the most appropriate, and the comparative effectiveness of crisaborole and TCIs is therefore uncertain.

4.5 Additional work on clinical effectiveness undertaken by the ERG

The additional work undertaken by the ERG on clinical effectiveness is the ERG's replication of their preferred simple NMA described in Section 4.4.4 and presented in Table 22 above, and a search for ongoing crisaborole studies. An electronic search was undertaken within the ClinicalTrials.gov database (https://clinicaltrials.gov/, last searched 30 October 2019). The term 'crisaborole' was used in the field,'Other terms', on the home page ('Find a study', https://clinicaltrials.gov/ct2/home). Twenty-nine clinical trial records were retrieved. Subsequently, a filter, which specified study status

as 'recruiting' was, applied to the search records. Of the 11 remaining records, eight ongoing crisaborole studies in children and adults with mild to moderate AD were identified. Table 12 presents a summary of the identified study records.

4.6 Conclusions of the clinical effectiveness section

The two identically designed 4-week crisaborole trials, AD-301 and AD-302, showed a statistically significant effect ($p<0.05$) of crisaborole over vehicle ointment, in the overall trial population, for ISGA success (ISGA 0/1 and ≥2-grade improvement), ISGA 0/1, pruritus, clinical signs and health-related quality of life.

Within subgroups for mild and moderate AD, the evidence of effectiveness was less consistent in mild AD. The difference in ISGA success between crisaborole and vehicle (pooled across the two trials) [redacted] For ISGA 0/1, no pooled data across trials were presented; however in moderate AD [redacted] (AD-301: [redacted]; AD-302: [redacted]), while in mild AD the difference [redacted] (AD-302: [redacted]) but not in the other (AD-301: [redacted]). Subgroup analyses for mild and moderate AD were not reported for other clinical outcomes.

Short-term and long-term safety outcomes were reported in the AD-301, AD-302 and the 48-week extension study, AD-303. There were no reports of deaths in any of the three studies. Overall, crisaborole was associated with mild and moderate TEAEs; the majority of these were not considered to be related to the study drug.

The NMA for ISGA/IGA 0/1 included nine trials of crisaborole, tacrolimus 0.03%, tacrolimus 0.1% and pimecrolimus 1%. No trials of TCS were identified with a population ≥80% mild to moderate AD and reporting relevant outcomes. Nineteen trials of TCS and TCIs were excluded as >20% were severe or severity was not reported.

Responses were higher in the vehicle arms of the crisaborole trials than in the vehicle arms of the TCI trials, which the CS states is because of a greater effect of the crisaborole base ointment. Therefore, the company's preferred base case NMA model adjusted for vehicle response using meta-regression and an unanchored MAIC was also undertaken. In the company base case NMA (fixed treatment effect, random class effect, adjustment for vehicle response) and in the unanchored MAIC, the median [redacted]

[redacted]

The ERG prefers a simple random effects NMA with no class effect and no vehicle response adjustment, although it accepts that the assumption of transitivity may be violated if the vehicles used in the crisaborole and comparator studies are different, and that indirect comparisons may be biased if there is an imbalance in the distribution of treatment effect modifiers in studies comparing different pairs of treatments. Nevertheless, it is unclear whether the assumptions made in the more complex vehicle response adjusted model preferred by the company are reasonable. The ERG has particular reservations regarding the assumption that the relationship between the population baseline response and population relative treatment effects should be the same across treatments.

Although the MAIC circumvents the need to adjust for vehicle response, it provides inferences relative to the populations defined by the comparator studies, which may not be useful for decision making. In addition, the MAIC assumes that that all prognostic factors and treatment effect modifiers have been accounted for, although this is an untestable assumption. For these reasons, the ERG considers that caution should be exercised when making inferences based on the results of the MAIC and prefers the simple random effects NMA over the MAIC.

Overall, we have limited information about which of crisaborole or TCIs is the more effective. Evidence about crisaborole comes from two 4-week trials of crisaborole versus a vehicle which may differ from the vehicles used in the tacrolimus 0.03% and also in the pimecrolimus 1% trials. It is plausible that the simple random effects NMA is the more appropriate model for the data and that TCIs are more effective than crisaborole, although this depends on the assumption that the vehicles are not systematically different and that variability in the response to vehicle depends on study characteristics. It is also plausible that adjusting for vehicle response is appropriate and that crisaborole is the more effective treatment, although this depends on the strong assumption that the relationship between population baseline log-hazard and population log hazard ratio is the same for crisaborole, tacrolimus 0.03% and pimecrolimus 1%. The MAIC provides estimates of relative treatment effect assuming that all relevant prognostic factors and treatment effect modifiers have been accounted for, although inferences depend on the population defined by the comparator trial. Although the ERG prefers the simple random effects NMA on the basis that it is the most parsimonious model with evidence of only moderate heterogeneity between studies, it is difficult to know which model is the most appropriate, and the comparative effectiveness of crisaborole and TCIs is therefore uncertain.

5 COST EFFECTIVENESS

5.1 ERG's comment on company's review of cost-effectiveness evidence

5.1.1 Objective of cost effectiveness review

In March 2019, the company performed a comprehensive and systematic literature search to identify economic evaluations of crisaborole and its comparators and the burden of the disease from 2004 until March 2019 (CS, Appendix G)

5.1.2 Search strategies for cost-effectiveness review

A total of seven electronic and online databases were searched on March 2019: MEDLINE and MEDLINE in Process [via Ovid], Embase [via Ovid], Cochrane Library including Cochrane Database of Systematic Reviews and Cochrane Central Register of Controlled Trials [via Wiley], Database of Abstracts of Reviews of Effects [via CRD], NHS Economic Evaluation Database [via CRD], Health Technology Assessment Database [via CRD], CINAHL [via EBSCO], EconLit [via EBSCO], and Epistemonikos [Epistemonikos Foundation] (see Table 25). The databases searched were restricted to the English language. As mentioned in the Section, 4.1, it is best practice not to apply language restrictions in the search strategy to prevent the risk of language bias.

The Embase conference proceedings database [via Ovid] and seven conference websites were searched by the company between 2015-2018 (American Academy of Dermatology, British Association of Dermatologists, European Academy of Dermatology and Venereology, European Society for Pediatric Dermatology, International Congress of Dermatology, International Symposium on Atopic Dermatitis, and World Congress of Dermatology). One trials registry was searched (clinicaltrials.gov).

Several HTA websites (NICE, SMC, CADTH, PBAC, AHRQ, and ICER) were searched to cover the period between 2004 and May 2019. In addition, the Gertner Institute for Epidemiology and Health Policy Research and the HOME websites were searched. The company carried out hand searching of references of included studies.

In addition to the review of cost-effectiveness studies, the company performed two comprehensive and systematic literature searches to identify: (i) health-related quality of life studies for mild to moderate atopic dermatitis from 1990 until March 2019 (CS, Appendix H) (ii) costs and resource use of treatments (CS, Appendix I) for mild to moderate atopic dermatitis from 2004 until March 2019. These were conducted using the same databases as described for the review of cost-effectiveness studies.

With the exception of terms provided for searches in conference and HTA websites, the search strategies undertaken by the company across all the databases and trials registry were fully reported and provided in Appendix G, H and I of the CS for the three reviews, respectively. The ERG considers that the company search strategies are sufficiently comprehensive to retrieve important citations relating to all eligible studies.

Table 25: Data sources for the systematic review of existing economic studies

Search strategy component	Sources	Date limits
Electronic database searches Key biomedical electronic literature databases recommended by HTA agencies	-MEDLINE® -MEDLINE® In-process -Embase® -The Cochrane Library including National Health Service Economic Evaluation Database (NHS EED) -EconLit® -Health Technology Assessment Database (HTAD)	2004 to 2019
Conference proceedings	-British Association of Dermatologists (BAD) -European Society for Pediatric Dermatology (ESPD) -European Academy of Dermatology and Venereology (EADV) -International Symposium on Atopic Dermatitis (ISAD) -World Congress of Dermatology (WCD) -International Congress of Dermatology (ICD) -American Academy of Dermatology (AAD)	2015 to 2018
Key international HTA websites	-National Institute for Health and Care Excellence (NICE) -Scottish Medicines Consortium (SMC) -Canadian Agency for Drugs and Technologies in Health (CADTH) -Pharmaceutical Benefits Advisory Committee (PBAC) -Institute for Clinical and Economic Review (ICER)	2004 to 2019

5.1.3 The inclusion and exclusion criteria used in the study selection

The inclusion/exclusion criteria used by the company to facilitate study selection are presented in Table 26.

Table 26: Inclusion/exclusion criteria for the economic review (reproduced from CS, Table 10 in Appendix G)

Category	Inclusion criteria	Exclusion criteria
Population (P)	-Age: Children (2 years and older) and adults aged ≥18 years -Gender: any -Race: any -Disease: patients with a clinical diagnosis of mild to moderate AD (≥80% of trial population)	>20% of the trial population were either healthy volunteers, pediatric patients (<2 years old), or patient with severe disease
Intervention (I)	All pharmacological interventions relevant to this review whether taken alone or in combination including: crisaborole, TCSs, TCIs, phototherapy, immunosuppressive therapies, oral steroids, alitretinoin (in people with atopic dermatitis affecting the hands), other PDE4 inhibitors (roflumilast, OPA-15406), and BSC	-Non-pharmacological interventions -Interventions not approved in the UK -Systemic treatments† -Anti-infectives (e.g. antibiotics)
Comparator (C)	-Any pharmacological intervention -Placebo -Best supportive care	None
Outcome (O)	Costs and outcomes in the form of LYGs, QALYs, or DALYs	None
Study design	-All economic evaluation studies based on models -Cost-effectiveness analysis -Cost-utility analysis -Cost-minimisation analysis -Cost-benefit analysis -Cost-consequence analysis -SLRs of economic evaluations	-Case reports/studies -Editorials and other forms of non-systematic reviews
Search timeframe	2004 to 2019	Studies published prior to 2004
Language	No restrictions	None

AD, atopic dermatitis; BSC, best supportive care; DALY, disability-adjusted life year; LYG, life years gained; PDE4, Phosphodiesterase 4; QALY, quality-adjusted life year; TCI, topical calcineurin inhibitor; TCS, topical corticosteroid

† CS stated "*systematic treatments*" but the ERG assume they meant systemic treatments

5.1.4 Findings of the cost effectiveness review

Eight unique studies and eight HTA submissions of relevance to the decision problem were identified. Two HTA submissions included crisaborole as an option and were described in CS Appendix G.(53, 54) Eight evaluations described the reported economic model as a Markov model, three as a cost-utility analysis, one as a discrete event model, decision-analytic model, or cost-minimisation analysis. The structure was not reported in the other two evaluations.

The intended population of the decision problem varied among studies, where most economic models investigated more than one population. Seven studies included both child and adult populations, six an adult population, two a child population, whereas one did not report the age group. Regarding AD severity, six studies involved patients with mild to moderate AD, three involved moderate AD, three involved moderate to severe AD, and three all-severities of AD. One further study did not report the AD severity modelled.

Likewise, the model time horizon varied between the studies; eight reported a 1-year horizon, one reported a 5-year horizon, another reported a 14-year horizon for a population of children. Three evaluations used different horizons for children and adults (14-year and 1-year, 15-year and 1-year, and 360-day and 169-day horizons for children and adults respectively). The time horizon used was not reported in the remaining three studies.

The ERG noted that in three of these models, IGA/ISGA scores were used to define health states and patients were allowed to have an IGA/ISGA score of 4 or 5 indicating severe disease. In two of these studies,(55, 56) the population had mild to moderate AD at baseline. In the third study, the starting severity was not stated but the conclusions suggest that the population under consideration was mild to moderate AD.(53)

5.1.5 Conclusions of the cost effectiveness review

The company claimed that searches did not identify any relevant studies of crisaborole, and as a result it developed a *de novo* health economic model. Although this seems inaccurate and contradicts the SLR results stated in CS Appendix G, the ERG notes that both studies that included crisaborole as a comparator were conducted outside the UK.

Additionally, the ERG noted that there was no discussion of the results of the SLR in the main text of the CS. The results were only reported in the appendix and were not appropriately summarised in either the appendix or the main CS.

5.2 Summary of company's submitted economic analysis

5.2.1 Population

The population included in the company's health economic analysis reflects children and adults aged two years and over with mild or moderate AD. This gives rise to four possible sub-populations:

1- Children (aged 2-17) with mild AD
2- Children (aged 2-17) with moderate AD
3- Adults (aged 18 and over) with mild AD
4- Adults (aged 18 and over) with moderate AD

In its base case, the company considered crisaborole as a second-line treatment for AD that has not been adequately controlled by TCS or where there is a serious risk of important AEs from further TCS use, particularly irreversible skin atrophy. A scenario analysis for TCS-naïve patients was also presented. The model assumed a starting age of 2 years and 18 years for the child and adult patient populations, respectively; whereas the sex distribution reflected AD-301 and 302 trial data (55.7% female).

The model relied on UK growth charts to inform the average weight for each year of age in the child population, whereas a fixed average weight of 76.14 Kg was assumed for the whole adult population. In response to clarification question B7, the company updated the average weight for adults every 10 years as per data from the Health Survey for England.(57)

5.2.2 Interventions and comparators

The technology of interest included in the company's model is crisaborole 2% as [redacted]. TCIs are not recommended by NICE for the treatment of mild AD. However, data from the BAD audit report indicated that 8% of children with mild AD being treated in secondary care received TCIs in the UK. The ERG's clinical experts suggested that this may reflect the difficulty of assigning a single severity level to patients when the severity of symptoms may vary over time and between different sites on the body. Therefore, patients may have co-occurring mild and moderate lesions, requiring different treatment regimens whilst being classified as having mild AD overall. As pimecrolimus is the only licensed TCI in mild AD, the company used pimecrolimus as the TCI comparator of choice in mild AD for both adults and children.

The CS presented both tacrolimus 0.03% and pimecrolimus as comparators for children with moderate AD, and both tacrolimus 0.03% and tacrolimus 0.1% as comparators for adults with moderate AD. The model assumed the use of emollients by all patients at all times.

In addition, the company also introduced a scenario where tacrolimus is used prophylactically between flares to reduce the likelihood of a further flare.

In the scenario analysis for the TCS-naïve population, it was assumed that patients may be treated with mild, moderate or high potency TCSs. The proportions of each potency used were estimated separately for mild and moderate AD based on the BAD audit data as presented in Table 15 of the clarification response (which replaces Table 56 of the CS as this contained an error). The company

selected hydrocortisone 1%, betamethasone valerate 0.025%, and betamethasone valerate 0.1% as the mild, moderate, and potent TCSs of choice respectively.

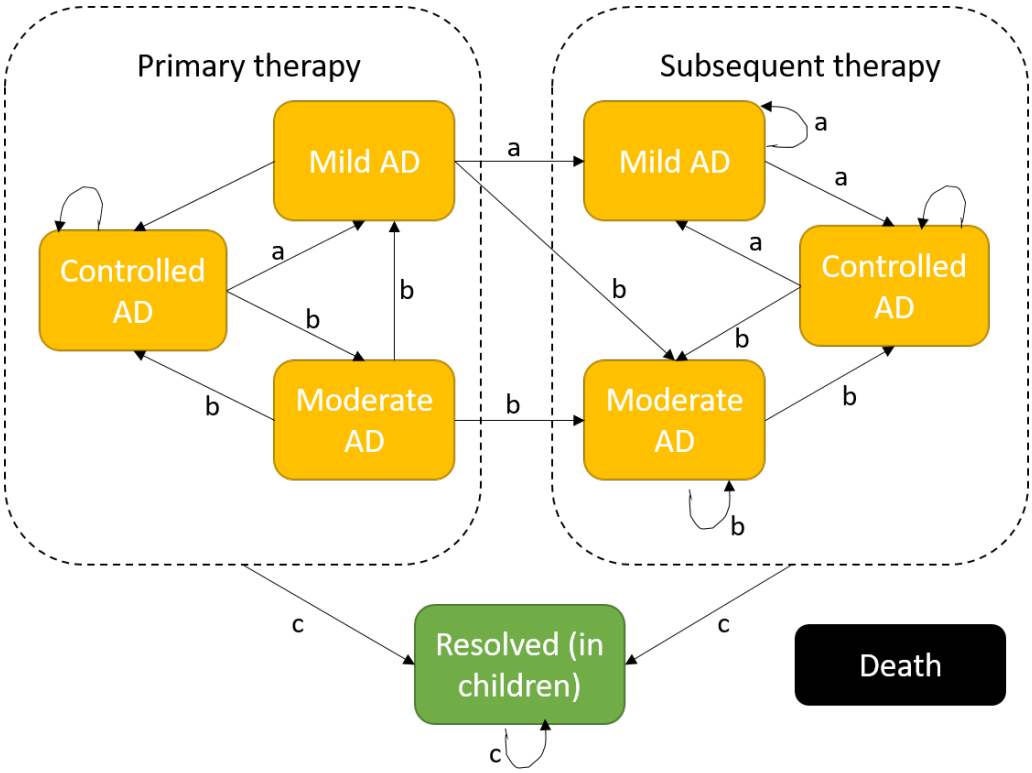
5.2.3 Perspective, time horizon and discounting

The base case model adopts an NHS and Personal Social Services perspective. The base case model uses a lifetime horizon for adults and an age 18 threshold for children. Both costs and QALYs were discounted at 3.5% per annum as recommended by NICE.

5.2.4 Model structure

As part of its submission to NICE, the company developed a fully executable cohort-level state transition model that comprised four mutually exclusive health states based on AD severity as defined by ISGA criteria: (i) moderate flare (ISGA score of 3); (ii) mild flare (ISGA score of 2); (iii) controlled disease (ISGA scores of 0 or 1); (iv) death. The cycle length is 4 weeks as this was deemed sufficient to assess response to treatment. A half-cycle correction is applied. Figure 6 depicts the model structure as understood by the ERG from examining the company's economic model. The ERG noted that Figure 6 is more detailed than the model structure diagram provided in Figure 35 of the CS as the ERG wished to illustrate the use of separate health states for patients receiving primary versus subsequent therapy and the health state representing resolved disease in children.

Figure 6: The company's model structure (NB: allowed transitions differ depending on starting severity as described in text). These are allowed only for a: patients with mild AD; b: patients with moderate AD; c: child patient population



All patients entered the model in either the moderate or mild state on primary therapy (i.e. crisaborole or one of its comparators). Patients who respond to treatment move to the controlled state. In the company's base case, patients who persisted in the same disease severity as the cycle before were directly referred for subsequent therapy; however, the model is flexible enough to incorporate up to 3 cycles of primary treatment for non-responders. In addition, patients on primary treatment starting the model in the moderate health state were allowed to have a partial response and move to the mild health state. Patients whose disease was controlled could later experience a flare and transition to their respective baseline disease severity (i.e. patients with mild AD at baseline who achieved a response and moved to the controlled disease health state could then experience a flare and move back to the mild AD health state at a later time point; patients with moderate disease at baseline who achieved a response could experience a later flare and move back to the moderate AD health state).

Patients failing on primary treatment (i.e. non-responders) and starting subsequent therapy were assumed to move to their baseline AD severity. As with primary treatment, patients receiving subsequent therapy were allowed to persist in their current health state or improve in response to subsequent therapy. However, in contrast to the primary treatment phase, non-responders were allowed to remain on subsequent therapy indefinitely, and no partial response was permitted in the base case. Patients on subsequent therapy could still experience flares moving back to their baseline disease severity at the model start.

As patients who experience a flare always return to their baseline disease severity level, the moderate AD states are not used in the population starting with mild AD. Patients receiving subsequent therapies who started with moderate AD cannot transition to the subsequent therapies mild AD state as partial response to subsequent therapies is not allowed.

Children with AD can experience disease resolution from any treatment phase or severity state, and they then remain in the resolved AD health state until death.

QALYs in the model are accrued according to the time spent in the various AD severity health states which have different utility values (see section 5.2.6). The main costs included in the model are the costs for primary and subsequent treatments and the costs for GP and secondary care visits (see section 5.2.7). No adverse events are included in the model.

5.2.5 Evidence used to inform the company's model parameters

5.2.5.1 *Treatment application*

In the economic model, responders to primary therapy (i.e. crisaborole or TCIs) are assumed to apply the treatment twice daily for the 4-week model cycle (i.e. 56 applications), whereas non-responders were assumed to receive the same daily treatment dose for half a cycle and discontinue it afterwards. The only exception was children receiving tacrolimus 0.03% where responders and partial responders had a total of 49 applications per cycle (twice daily for 21 days then once daily for 7 days). The amount of drug used per application is assumed to be the same for crisaborole, TCIs and TCSs and was based on usage of crisaborole in the AD-301 and AD-302 trials. Treatment costs are based on the cost per gram of treatment used and this implicitly assumes that there is no wastage.

In the scenario analysis for the treatment naïve population, treatment with TCSs is assumed to consist of twice-daily applications for two weeks.

In addition, the company conducted a scenario analysis in which the cost of TCIs was applied for 6 weeks for responders and partial responders to reflect the fact that three of the comparator studies included in the NMA allowed treatment with TCIs for up to six weeks and 6-week outcomes from these studies were included in the NMA. In another scenario analysis, the company assumed a twice-weekly dose for the prophylactic use of tacrolimus. In the scenario analysis regarding TCS-naïve patients, the company assumed that TCSs are applied twice daily for 14 days per cycle for all patients.

In the base case, subsequent therapy was assumed to include either phototherapy or ciclosporin as a systemic therapy. The proportions of patients using each of the possible subsequent therapies were based on data from the BAD audit report (further details on subsequent therapies are provided in

Section 5.2.5.4). In response to clarification question B17, the company conducted a scenario analysis where methotrexate, azathioprine and mycophenolate mofetil were added to the mix with a weight based on proportions treated with each therapy in the IQVIA-THIN database for 2017.

As mentioned in Section 5.2.2, emollients were used by all patients in both the primary and subsequent therapy phases. The model assumes a full pack of 500g of diprobase cream is used by patients with uncontrolled AD per cycle, whilst 250g is used as a maintenance dose by controlled patients in the same duration.

5.2.5.2 *Response to treatment and associated transition probabilities*

In the primary therapy phase, treatment efficacy is determined by the proportion of patients moving to the controlled health state (i.e. achieving an ISGA score of 0/1). The probability of response for patients having treatment with vehicle is estimated from a natural history model (NHM). [redacted] The response rates in the TCI and crisaborole arms of the model are then estimated by applying the HRs from [redacted], reported in Figure 19 of the CS. In the probabilistic sensitivity analysis, the CODA (convergence diagnostics and output analysis) samples from the NHM/NMA are used, whereas in the deterministic analyses, the means of the CODA samples are used to provide midpoint estimates of response. The ERG believes that the company's combined NHM/NMA provides estimates of response for each treatment estimated [redacted]

The company claimed that the NHM cannot be used properly to adjust for age and severity as it is based on few data points and the estimates of the response rates "*become implausible as the target age and severity move away from the mean values observed in the included trials*" (CS, page 124(1)). The age and severity assumed in the NHM are therefore set to [redacted] in the company's base case and not to the average values expected in the target populations. Scenario analyses are presented to explore the effect of allowing the response rates to vary by age and disease severity. In these scenario analyses, response rates for the vehicle arm were adjusted using odds ratios for response to vehicle in the target population (i.e. child or adult, mild or moderate) versus response to vehicle in the population as a whole. These were calculated solely on the vehicle arms of the AD-301 and AD-302

trials (see CS, Table 55(1)). Response rates to primary therapy options in the company's base case, and in the scenario analyses adjusted for age and severity are presented in Table 57 of the CS.

Transition probabilities for partial responders in the moderate subpopulations (i.e. transitioning from moderate to mild health state) were derived from partial response rates reported in AD-301 and AD-302 trials. This represented [redacted] and [redacted] of non-responders for crisaborole and vehicle respectively. In its model base case, the company assumed that the partial response rate for crisaborole could be applied to tacrolimus and pimecrolimus.

Absolute response rates for subsequent therapies were all taken from TA82(3) and presented in Table 59 of the CS. The ERG noted that the rate of response to TCIs as subsequent therapy in the TCS-naïve population (who have TCS or crisaborole as primary therapy and the option of TCIs as subsequent therapy) given in Table 59 were also taken directly from TA82 rather than being based on the company's NMA.

Table 27 shows the transition probabilities (response probabilities) used for primary and subsequent therapies.

Table 27: Transition probabilities used for primary and subsequent therapies

Comparator	Response probabilities in the company's base case model
Primary therapy	
Vehicle	[redacted]
Tacrolimus 0.03%	[redacted]
Tacrolimus 0.1%	[redacted]
Pimecrolimus 1%	[redacted]
Crisaborole	[redacted]
Subsequent therapy	
Ciclosporin	0.7
Phototherapy	0.7
Tacrolimus 0.03%	0.39
Tacrolimus 0.1%	0.37
Pimecrolimus 1%	0.25

5.2.5.3 *Flare rates*

The company's model assumed that patients in the disease-controlled state can experience a flare back to their baseline disease severity. The annual flare rate was extracted from a Pfizer epidemiology report of a survey described in Appendix M. Table 60 of the CS reports the rates per age group and disease severity and the associated cycle transition probabilities. In the base case analysis, the flare rate is assumed to be identical for treatment and comparator arms.

In the scenario analysis where tacrolimus could be used as a maintenance therapy, the reduction in flare rate was informed by a long-term study comparing tacrolimus to vehicle.(58) The flare rate was reduced by a ratio of 0.72, and the model assumed that patients on tacrolimus maintenance therapy received an average of 2.5 applications per week.

5.2.5.4 *Duration of treatment*

Patients who have not responded to primary therapy after four weeks were assumed to discontinue treatment and are referred to start subsequent therapy. Subsequent therapy options included systemic therapy (ciclosporin), or phototherapy in the base case, whereas TCIs was added to the mix in the TCS-naïve scenario for both treatment arms (i.e. patients on TCSs or crisaborole as primary therapy). These data were informed by the BAD audit report where a breakdown of treatment regimen was presented by disease severity in paediatric patients. As the audit does not report the specific TCI used, the company initially assumed equal usage across all doses licensed in the relevant population. Subsequently, in response to the clarification request, the company incorporated data from the IQVIA-THIN database to estimate the breakdown between tacrolimus 0.03%, tacrolimus 0.1% and pimecrolimus. These were estimated separately for adults and children and were applied only in patients with moderate disease. Table 28 shows the subsequent therapy breakdown per sub-population and based on primary therapy received for the company's updated base case that included the data from IQVIA-THIN. It should be noted that the figures for mild AD in the TCS-naïve scenario are based on the figures in the model; these differ from those presented in Table 62 of the CS, as the company confirmed at clarification that the figures in the model were correct. The ERG notes that the IQVIA-THIN data show that tacrolimus 0.1% is being used in children despite the CS submission stating that only the 0.03% dose of tacrolimus is licensed in children (CS, Table 7).

Table 28: Breakdown of subsequent therapies used in the revised company's model

Treatment option	Mild AD		Moderate AD		
	Base case*	TCS-naïve scenario*	Base case*	TCS-naïve scenario (children)	TCS-naïve scenario (adults)
Tacrolimus 0.03%	0%	0%	0%	[redacted]	[redacted]
Tacrolimus 0.1%	0%	0%	0%	[redacted]	[redacted]
Pimecrolimus	0%	87%	0%	[redacted]	[redacted]
Emollients alone	0%	0%	0%	0%	0%
Systemic therapy	100%	13%	71%	13%	13%
Phototherapy	0%	0%	29%	5%	5%

AD, atopic dermatitis; TCS, topical corticosteroid

* for both children and adults

5.2.5.5 Treatment safety

No adverse events (AEs) were incorporated in the model. The CS states that application site pain is the main safety outcome but the company claimed that the disutility of this outcome would be difficult to account for in the model because it is a very-short term event. Lymphoma is one long-term AE possibly associated with TCIs; however, the CS claims that there was not sufficient long-term data for crisaborole to incorporate lymphoma in the model. An ERG clinical advisor noted that untreated AD carries a lymphoma risk and the link between TCI usage and lymphoma may not be causal. The company assumed that disease related AEs (e.g. pruritis) are implicitly captured within the modelled health state because of their correlation with ISGA scores.

5.2.5.6 Resolution of AD

It was assumed in the base case that 75% of children (equating to a 4-week probability of 0.708%) could outgrow their AD in childhood or early adolescence. This was deemed to be irreversible, that is, patients with resolved AD were assumed to remain in the same health state until death or end of time horizon.

5.2.5.7 Mortality

AD was not expected to impact mortality, therefore general population life tables for England and Wales were used to inform the age-specific mortality.

5.2.6 Health-related quality of life

The SLR carried out by the company identified three unique HRQoL studies and nine HTA submissions relevant to the technology appraisal; however, the company opted to use HRQoL data collected in the AD-301 and AD-302 trials. These data were collected via the DLQI and CDLQI questionnaires which were completed at baseline and day 29 in AD-301 and AD-302.

A mapping algorithm published by Ali *et al.* was identified and used for mapping the DLQI scores to the EQ-5D-3L where an ordinal regression model predicted the probability of EQ-5D responses per item based on age, sex and DLQI responses.(59) Results of the mapping algorithm are shown in Table 65 of the CS. The CS states that, [redacted].

The company presented four sets of utility scores; (i) the first set did not combine DLQI and CDLQI responses and mapped each separately to obtain separate utility scores for each age group; (ii) the

second used DLQI responses only to derive the utility scores for both age groups; (iii) the third combined both DLQI and CDLQI responses to get the same utility scores for the disease severity regardless of the age group, and (iv) the fourth set utilised the same scores as in TA82 (3). In addition, the company's model included the functionality to select either the mean values of the EQ-5D scores or to calculate utility multipliers which were estimated relative to the utility score for controlled disease. In the base case analysis utility multipliers were estimated relative to the utility score for controlled disease and therefore controlled patients were assumed to revert back to the population utility norms for their age group in the UK. Table 66 of the CS provides a comparison between the different utility values used in the model for adults and children stratified by diseases severity, whereas Table 67 presents the EQ-5D index values for the UK population stratified by age group.

The ERG noted that in its base case, the company preferred using utility multipliers based on the second set of utility scores which used only values mapped from DLQI.

5.2.7 Resources and costs

The costs and resource use included in the base case model comprised: treatment costs and medical resource use (MRU) in the form of general practitioner (GP) and dermatologist visits. These are discussed in the following sections. The ERG noted that no AE-related or other costs were included.

5.2.7.1 Treatment costs

Crisaborole is available as a 60g ointment tube at a list price of [redacted]. Acquisition costs for TCIs, TCSs, and systemic therapy were derived mainly from the British National Formulary, and are presented in Table 69 of the CS. Phototherapy sessions were costed at £93 per session as stated in NHS reference costs 2017-18.(60)

5.2.7.2 Medical resource use associated with health states

Patients on primary therapy with uncontrolled AD were assumed to see a GP only with a frequency related to disease severity based on the Pfizer epidemiology study mentioned in Section 5.2.5.3. These were estimated at an average of two and three annual visits for mild and moderate disease, respectively. Non-responders were then referred to dermatologists for subsequent therapies and it was assumed that patients on subsequent therapy would see a dermatologist every 2 months (this reflects the revised based case after clarification response). All patients on subsequent therapies (whether controlled or not) were assumed to have GP visits at an increased frequency related to their age group based on data from a retrospective cross-sectional analysis of the THIN database. These were estimated at [redacted] and [redacted] annual visits for children and adults, respectively. Table 70 of the CS shows the different unit costs used for GP and dermatologist visits which were based on PSSRU unit costs and NHS reference costs respectively.(60, 61)

5.2.8 Model validation and face validity check

The company stated that they performed cell-to-cell checks of logic and consistency. In addition, model outputs were compared to those from previous economic analyses (e.g. TA82(3)) which led to modifications. Outputs for the current model version were considered consistent on comparison; however, no supporting evidence was provided.

5.2.9 Cost effectiveness results

Following the clarification process the company submitted a revised version of the model that included updated estimates of the cost-effectiveness of crisaborole. The updates to the base case model included a revised implementation of dermatologist costs in patients having subsequent therapy. There were also two updates to the scenario analysis for TCS-naïve patients in which the data on the proportion of patients receiving each of the TCIs as subsequent therapy and the proportion receiving mild, moderate and potent TCSs as primary therapy were updated. For clarity, the updated data have been presented above and all the results presented in this section and in Section 5.2.10 use the revised model. Table 29 and Table 30 shows the results of the company's base case analysis for the deterministic version of the model for the TCS-experienced and TCS-naïve scenarios, respectively. Results from the probabilistic version of the model were in line with the deterministic results. The absolute costs and QALYs for each intervention for the deterministic analysis were within 5% of those for the deterministic model. There was no disagreement between the deterministic and probabilistic analyses in terms of whether crisaborole dominated TCIs. Based on the probabilistic version of the model, and across various populations and comparisons, crisaborole [redacted] compared to the other comparators in the TCS-experienced population, whereas TCS [redacted] compared to crisaborole in the TCS-naive population.

Probabilistic results for the four subgroups at ICER thresholds of £0 to £100,000 showed that while crisaborole had a probability of generating maximum net benefit (i.e. being the optimal treatment) that varied from 0.10 to 0.20 in all four populations modelled in the TCS-naïve population, it had a probability of generating maximum net benefit that was above 0.80 in all TCS-experienced populations except adults with moderate disease where the probability was 0.5 to 0.6 depending on the threshold applied.

Table 29: The company's base case results (population with TCS prior treatment)

Treatment	Total QALYs	Total Costs	Incremental QALYs*	Incremental Costs*	ICER (£ per QALY gained)
Children with mild AD					
Pimecrolimus	[redacted]	[redacted]			
Crisaborole			[redacted]	[redacted]	Crisaborole dominates pimecrolimus
Children with moderate AD					
Pimecrolimus	[redacted]	[redacted]			
Tacrolimus 0.03%			[redacted]	[redacted]	Dominated by crisaborole
Crisaborole	[redacted]	[redacted]	[redacted]	[redacted]	Crisaborole dominates both TCIs
Adults with mild AD					
Pimecrolimus	[redacted]	[redacted]			
Crisaborole			[redacted]	[redacted]	Crisaborole dominates pimecrolimus
Adults with moderate AD					
Tacrolimus 0.03%	[redacted]	[redacted]			
Tacrolimus 0.1%	[redacted]	[redacted]	[redacted]	[redacted]	Dominated by crisaborole
Crisaborole	[redacted]	[redacted]	[redacted]	[redacted]	Crisaborole dominates both tacrolimus concentrations

AD, atopic dermatitis; ICER, incremental cost-effectiveness ratio; QALY, quality adjusted life year; TCI, topical calcineurin inhibitor
* versus the next least effective comparator

Table 30: The company's base case results (TCS-naive population)

Treatment	Total QALYs	Total Costs	Incremental QALYs*	Incremental Costs*	ICER (£ per QALY gained)
Children with mild AD					
Crisaborole	[redacted]	[redacted]			
TCS			[redacted]	[redacted]	TCS dominates crisaborole
Children with moderate AD					
Crisaborole	[redacted]	[redacted]			
TCS			[redacted]	[redacted]	TCS dominates crisaborole
Adults with mild AD					
Crisaborole	[redacted]	[redacted]			
TCS			[redacted]	[redacted]	TCS dominates crisaborole
Adults with moderate AD					
Crisaborole	[redacted]	[redacted]			
TCS			[redacted]	[redacted]	TCS dominates crisaborole

AD, atopic dermatitis; ICER, incremental cost-effectiveness ratio; QALY, quality adjusted life year

* versus the next least effective comparator

5.2.9.1 Deterministic sensitivity analyses

In its response to the ERG's clarification questions, the company presented a set of tornado diagrams which summarised the ten most influential parameters in terms of impact on incremental net monetary benefit (NMB).(62) Incremental NMB measures the economic value of a given intervention in monetary terms compared with another intervention where a positive value indicates that an intervention is more cost-effective than the comparator at a chosen threshold.

Within the tornado diagrams, most parameters were varied either between the upper and lower bounds of the 95% confidence intervals (e.g. response probabilities on TCIs and crisaborole, utilities, drug use application, and cost per dermatologist visit) or an arbitrary 25% around the mean (response probabilities on TCSs and subsequent therapy and frequency of visits to GPs/dermatologists). In general, the most influential parameters were the treatment response probabilities/rates. In the base case, where patients received either a TCI or crisaborole as a primary therapy, incremental NMB was [redacted], except when response on crisaborole is at the lower bound or response on the comparator is at the higher bound. Conversely, within the TCS-naïve scenario, [redacted], except only when response on crisaborole was at its higher bound. These deterministic sensitivity suggest that the effectiveness of crisaborole relative to comparator treatments is a key area of decision uncertainty.

5.2.10 Scenario analyses

The company undertook several scenario analyses for the TCS-experienced population; these are presented in Appendix 3 of the company's response to the ERG's clarification questions. These scenarios explored: using MAIC response probabilities; adjusting response for line of treatment, age, or severity; using tacrolimus as a maintenance therapy; using alternative sets of utility values; exploring the sequential use of TCIs and crisaborole; incorporating partial response on subsequent therapy; modelling different starting ages and time horizons; assuming longer durations on subsequent therapy; using posterior predictive distributions from the natural history model; altering response probabilities following partial response, and removing systemic treatments from subsequent therapies in mild disease.

In general, most scenarios produced ICERs that were similar to the company's base scenario, with crisaborole dominating TCIs. [redacted] In a scenario whereby patients were assumed to remain on systemic therapy for life and

phototherapy for three months with flare rates halved, TCIs provided more QALYs, however the ICERS ranged between [redacted]. Again, the sensitivity and scenario analyses presented suggest that the relative efficacy of TCIs and crisaborole is a key area of decision uncertainty. This is because when a treatment has a greater response rate it results in marginally greater QALYs from achieving control of symptoms earlier but it also avoids the costs associated with subsequent therapies.

The ERG notes that the NMA results in the scenario [redacted] for moderate AD (53%) do not match those presented in Table 58 of the CS (55% reported) but they do match those reported in Table 71 of the CS.

5.3 Critique of the company's submitted economic evaluation

5.3.1 Methods for reviewing the company's economic evaluation and health economic model

The ERG adopted a number of approaches to explore, interrogate and critically appraise the company's submitted economic evaluation and the underlying health economic model upon which this was based. These included:

- Scrutiny of the company's model and discussion of issues identified amongst the members of the ERG.
- Examination of the correspondence between the description of the model reported within the CS and the company's executable model.
- Re-running the DSA and PSA presented within the CS.
- Where possible, checking the parameter values used in the company's model against their original data sources.
- The use of expert clinical input to judge the credibility of the company's economic evaluation and the assumptions underpinning the model. The ERG had inputs from two experts prior to writing its report.

5.3.2 Adherence of the company's model to the NICE reference case

As shown in Table 31, the company's economic evaluation is generally in line with the NICE reference case.(63)

Table 31: Adherence of the CS to the NICE reference case

Element	Reference case	ERG comments
Type of economic evaluation	Cost-utility analysis with fully incremental analysis	The CS met the NICE reference case.(63)
Time horizon	Long enough to reflect all important differences in costs or outcomes between the technologies being compared	The CS met the NICE reference case for adult patient population for whom a lifetime horizon was adopted.(63) However, for the child population, the model stops simulating their progression when the population reaches 18 years of age.
Synthesis of evidence on health effects	Based on trial outcome data and systematic review	The CS met the NICE reference case.(63) Health outcomes are modelled using the data collected in AD-301 and AD-302 studies. These data were combined with data from trials of the other comparators in an NMA to derive the response-to-treatment probabilities.
Measuring and valuing health effects	Health effects should be expressed in QALYs. The EQ-5D is the preferred measure of HRQoL in adults.	The CS met the NICE reference case.(63)
Source of data for measurement of health-related quality of life	Reported directly by patients and/or carers	The CS met the NICE reference case.(63)
Source of preference data for valuation of changes in HRQoL	Representative sample of the UK population	DLQI/CDLQI data collected in the AD-301 and AD-302 trials were mapped to EQ-5D-3L values. The mapping algorithm used in the company's base case was developed originally to map from DLQI responses. The company used these responses in its base case to derive the utility scores for both age groups (i.e. assumed the same set of utility multipliers for children and adults).
Equity considerations	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	The CS met the NICE reference case.(63)
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	The CS met the NICE reference case.(63)

Element	Reference case	ERG comments
Discount rate	The same annual rate for both costs and health effects (currently 3.5%)	The CS met the NICE reference case.(63)

5.3.3 ERG Critique of the modelling performed by the company

5.3.3.1 Model verification

The ERG checked and verified the implementation of the model and the methods for generating results. During this process, the ERG identified one implementation error concerning dermatologists visits for patients on subsequent therapies, which was addressed by the company in their clarification response to question B28. In addition, there were several coding errors that were addressed in the revised model version (see company response to clarification question B44). The implemented model appears to be generally in line with its description within the CS. Additionally, the company submitted an updated base case and 11 new scenarios in its clarification response. The ERG identified one error in the updated base case. The revised implementation of dermatologist visits for patients on subsequent therapies introduced a new error in that the costs were applied to patients both on and off treatment.

5.3.3.2 Correspondence of the model inputs and the original sources of parameter values

The ERG noticed some discrepancies among the model inputs, parameter values as reported in the CS, and the original data sources. These included response probabilities to subsequent therapy (Table 59 of CS), patient proportions starting different subsequent therapy options in the base case (Table 62 of CS), and patient proportions on different TCS in the TCS-naïve scenario (Table 56 of CS).

In its response to clarification questions (B12, B10, and B41), the company updated the economic model, and aligned the model inputs with data sources (correct values are provided in Tables 15, 16, 23 of the company response to clarification). However, there were still some discrepancies found between the new model inputs and the values stated in the clarification response (e.g. cost and dosage of ciclosporin and dosage of mycophenolate mofetil). The impact of correcting these discrepancies is explored in Section 6.

5.3.4 The main issues identified by the critical appraisal

Generally, the model was adequately described in the company's submission and the implementation of the model was in keeping with the data and structural assumptions presented in the CS. However, the ERG was not satisfied with the quantity and type of structural assumptions used to simplify the model structure, particularly with respect to the modelling of subsequent therapies. In response to the ERG's clarification questions, the company indicated that these assumptions were made to handle the

shortcomings of the state-transition structure used to depict the decision problem. However, the ERG is not convinced that these simplifications adequately capture the clinical pathway. The ERG could identify 9 main issues within the model. These points are summarised in Box 1, with further details provided in the subsequent sections.

Box 1: Summary of the main issues identified within the critical appraisal

Summary of identified concerns within the company's health economic model:

1) Deviation from the NICE scope in terms of the population in the base case scenario which is restricted to patients with an inadequate response to TCSs or where there is a risk of significant adverse events from further use of TCSs
2) Deviation from the NICE scope with regards to the use of pimecrolimus as the comparator for mild AD when it is not recommended by NICE for mild AD
3) Uncertainty in the relative effectiveness of crisaborole and TCIs demonstrated by the differences in the vehicle adjusted NMA used in the company's base case scenario and the ERG's simple random effects NMA
4) Methods use to adjust response rates for age and severity
5) The structural assumptions used to model subsequent therapies failed to capture their sequential use
6) The lack of appropriate estimates for the duration of subsequent therapy in the base case
7) The failure of the model to allow patients to experience relevant health outcomes such as severe AD following non-response
8) The failure of the model to allow patients to experience a flare of different severity to their baseline AD severity
9) The failure of the model to allow patients to experience a partial response to subsequent therapy
10) Poor justification of the assumption that patients will use an equivalent amount of ointment per application and failure to explore the potential impact of deviation from this on the cost-effectiveness results
11) Failure to explore any alternative mapping algorithms and the assumption used to generate utilities for children from a mapping algorithm developed for adults
12) The duration of treatment for non-responders to primary therapy is assumed to be only two weeks even though patients are assumed not move on to receive a subsequent therapy until week 4.

5.3.4.1 The deviation of the implemented base case from NICE final remit/appraisal objective

The ERG recognises the company's intention to submit for an optimised population where TCSs can no longer be used. However, the ERG is also aware that crisaborole is indicated as a first-line treatment and that its marketing authorisation refers to "*mild to moderate atopic dermatitis in people aged 2 years and older*" which prompted the NICE final scope to adopt the same position. Moreover, AD-301 and AD-302, the main trials for crisaborole, were not designed to assess effectiveness in the company's base case population.

The ERG also notes that the company's model failed to account for a lifetime horizon for children and considered no outcomes beyond an age of 18 in children. In response to clarification question B5, the company mentioned that the approach was adopted because children by that age become adults and outcomes could be evaluated through the adult model pathway. However, the ERG is concerned that the current model lacked the flexibility to provide separate cost effectiveness results for the lifetime horizon for child populations.

5.3.4.2 The relevance of the chosen comparators for mild AD

The final NICE scope states that for patients with mild AD, the comparator is "*a combination of emollients and mild to moderate potency TCSs*".(2) However, the base case economic analysis focuses on a second-line population where TCSs have failed or where patients are at risk of adverse events from further TCS use. Therefore, The ERG asked the company (see clarification question B2) to provide a comparison of crisaborole against emollients alone in patients with mild AD who have not been adequately controlled by TCSs or where there is a serious risk of important adverse effects from further TCS use. This was not provided by the company who instead argued that the use of emollients alone was extremely unlikely in patients with uncontrolled systematic disease. The company highlights the use of TCIs in a small proportion of mild patients within the BAD audit data to support this. The ERG's clinical advisors felt that it would be unethical to step down patients to treatment with emollients alone where TCSs have failed or where patients were at risk of adverse events from further TCS use if they had uncontrolled mild AD symptoms. Although they also commented that there is considerable variation in what is considered safe long term use of TCSs between different health care practitioners because of a lack clear evidence on the safety of steroids across different age groups and body sites.

The ERG notes that it is unclear from the final NICE scope whether using emollients alone would be a relevant comparator for patients with mild AD who have not been adequately controlled by TCSs or where there is a serious risk of important adverse effects from further TCS use (see Section 3.3). The ERG notes that TCIs are not recommended by NICE for mild AD, and are not listed in the final NICE

scope as a comparator for mild AD. Therefore, considering pimecrolimus as a comparator in mild AD may not be appropriate, and the base case comparator for such population could be emollients alone. However, the ERG acknowledges that TCIs appear to be being used in a small proportion (8%) of patients with mild AD, based on the data from the BAD audit in children, although this may reflect the difficulty in assigning a single severity score when the severity may vary across body sites and may fluctuate over time.

5.3.4.3 Use of vehicle adjusted NMA

The economic model uses the natural history model and the estimates of relative treatment effect estimated from the company's FE-RCE-VR model which adjusts for vehicle response across the trials included in the NMA. Based on this model, the median response rates for crisaborole (CS, Table 57(1)) are more favourable than for TCIs (48% for crisaborole versus 35%, 40% and 45% for pimecrolimus 1%, tacrolimus 0.03% and tacrolimus 1% respectively). The impact of the company's preferred NMA in the economic model is that crisaborole is estimated to have a greater QALY gain on average in all comparisons against TCIs. However, the ERG notes that in the simple random effects NMA presented by the company, the opposite is true, with TCIs having more favourable median response rates (see Table 22). Although the company has not conducted a scenario analysis using the results of the simple random effects NMA, crisaborole would no longer dominate TCIs if it is on average less effective. In addition, results of the company's scenario analyses, which assumed equal effectiveness for crisaborole and TCIs, suggests that crisaborole would be dominated in any scenario where it was less effective than TCIs. This is because any treatment which improves response rates accrues both additional QALYs from earlier disease control and cost savings from avoiding subsequent treatments. Therefore, the ERG considers the results of the economic model to be highly dependent on whether the adjustment for vehicle response in the company's preferred NMA is appropriate. The ERG prefers the simple random effects NMA as the company's preferred vehicle adjusted NMA relies on the strong and untestable assumption that the same relationship between population baseline risk and population treatment effect applies for each treatment.

5.3.4.4 Adjustment of response rates for age and severity

The ERG notes that the natural history model for response to treatment described in Table 33 of the CS and Section B.3.3.1 of the CS,(1) allows for the response to treatment to be adjusted for age and disease severity. The regression is centred on the mean age and percentage with moderate disease from the trials included in the NMA. This would appear to allow the response rates to be adjusted according to whether the population is children or adults, and whether it is people with mild AD or moderate AD. However, in the base case economic model, the age and percentage with moderate disease are set equal to their mean values from the trials rather than set to appropriate values for the populations being modelled. This means that the estimates of response in the vehicle arms do not

apply to the target populations for the four age/severity subgroups for which separate cost-effectiveness estimates are provided.

A set of sensitivity analyses are provided in the CS in which the response rates are adjusted for age, severity and prior treatment with TCS. However, these adjustments are made not using the company's NHM, but using a separate set of ORs. These ORs were calculated based on the response to vehicle in the target population (i.e. child or adult, mild or moderate) versus response to vehicle in the population as a whole using data solely from the AD-301 and AD-302 trials (see CS, Table 55(1)). It seems inconsistent for the company to develop a NHM that allows adjustment for differences in the ages and severity of AD across the trials included in the NMA and then to use a different method to adjust the response rates to reflect the ages and severity levels in the target populations for the economic modelling.

5.3.4.5 The structural assumption of treatments being non-sequential

In its base case analysis, the company has simplified the decision problem into patients having either TCI or crisaborole as a primary treatment followed by subsequent therapy upon failure of primary treatment. Subsequent therapy was also introduced as a 'basket' of different therapeutic options including systemic therapy and phototherapy (subsequent use of TCIs was also allowed in the scenario analysis for the TCS naïve population). It is assumed that individuals would remain on some form of therapy while symptoms persist and that the proportion of the population receiving any specific subsequent therapy is fixed at any given time.

The ERG believes that this assumption oversimplifies the treatment pathway. Figure 2 of the CS (previously shown as Figure 1 in this report) shows that there is an algorithm of different lines of treatment to be followed sequentially based on the symptoms that a patient develops, and their response to the current line of treatment. NICE CG57(4) (see Table 1) recommends that a stepped approach is used and therefore certain lines of treatment should not be used before others (e.g. phototherapy should be considered before systemic treatments). The ERG does not believe that the company's approach of modelling subsequent therapy as a 'basket' of different therapeutic options is consistent with this stepped approach, as the model does not capture the need for a proportion of patients to try phototherapy before moving on to systemic treatment.

The rate of response to subsequent treatment was calculated as a weighted average of response rates over the available 'basket' of subsequent treatments using the proportions shown in Table 28. This calculation appears to suggest that patients cycle through a 'basket' of subsequent treatment options to achieve a cumulative response rate equivalent to the average response rate based on the proportion remaining on each treatment long-term. Moreover, the probability of response to subsequent therapy

is applied repeatedly every cycle to those who had not previously responded, resulting in a cumulative response rate that was higher than the average response rate. For example, a 70% response rate applied repeatedly over 4 cycles gives a cumulative response rate of 99%. This approach would be more appropriate if patients are being offered a sequence of treatments each with a 70% chance of response. These two approaches appear inconsistent as the use of an average response rate suggests that patients select one of several possible subsequent therapies, but the repeated application of that average response rate is more consistent with the idea of patients try several treatment sequentially in which case, the response rate should be specific to the line of treatment being used.

5.3.4.6 Duration of treatment on subsequent therapy

The company's base case assumed that patients failing on primary treatment could go on to receive a 4-week subsequent therapy which is then discontinued in those whose AD symptoms are controlled. Clinical advice to the ERG suggested that in order to achieve response, patients often require more than 4 weeks of treatment with subsequent therapies, such as systemic immunosuppressants.

In response to clarification question B14, the company provided further details on the average duration of treatment for patients having subsequent therapies (Table 18 of company's response to clarification). These ranged from 3 months on phototherapy to 15.1 months on methotrexate. The company carried out three exploratory scenario analyses to determine how the duration of subsequent therapy affected the ICERs. However, it is the ERG's position that the appropriate treatment durations for subsequent therapy should be reflected in the modelled base case. In addition, the response rates used in the model for subsequent therapies should relate to the time needed to achieve a treatment response. Currently, the response rates are applied at 4 weeks and then the same response rate is repeatedly applied every 4 weeks to non-responders. This would introduce a significant error if the reported response rates related to those achieved after 3 or 6 months of treatment. Appendix I of the CS, which provides a breakdown of costs and QALYs by health states, shows that the additional treatment cost of crisaborole compared with TCIs is being offset by savings from a reduction in use of subsequent therapies. Therefore, the correct modelling of subsequent therapies is important to correctly estimate the cost-effectiveness of crisaborole.

5.3.4.7 Absence from the model of severe AD as a health state

The ERG notes that the population of interest of the current appraisal comprises patients who have only mild or moderate AD. In its response to clarification question B3, the company states that, for the sake of model simplicity and parsimony, the model did not consider disease progression. However, the model should capture the full spectrum of health states which those patients could experience at any time during the model time horizon and this should not be sacrificed for the sake of model simplicity.

Moreover, Table 17 in the clarification responses [redacted] This suggests that the model may not be capturing the real world experience of patients who may fluctuate between having mild to moderate AD and severe AD. It is difficult to say what the impact would be of correcting this but it may be [redacted].

5.3.4.8 Flare severity

The company assumed in its base case that controlled patients could experience only a flare of the same severity as their baseline AD severity. The company, in response to clarification question B21, justified this by absence of relevant data and by assuming that incorporating different levels of flare severity per patient subgroup would render the analysis "*less useful for decision-making in each population*".

However, the ERG is not convinced by the assumption as it does not reflect reality. The model could have been flexible enough to adapt different severity flare levels for modelled patients. However, it is difficult to know what the size and direction on any bias related to this simplification would be.

5.3.4.9 Partial response to subsequent therapy

The company did not consider partial response to subsequent therapy in its base case, and only considered it as a scenario analysis in response to clarification question B4.

The ERG considers this to be a misrepresentation of reality, as patients could respond with different degrees to subsequent therapy, and is concerned this is not reflected in the base case. However, the company's scenario analysis suggests that any bias is likely to be small.

5.3.4.10 Average dosing per application for TCIs and TCSs

The company's base case assumed that drug use per application is consistent for primary treatment based on data from AD-301 and AD-302 trials. It was assumed that drug use per application of TCIs and TCSs is the same as observed for crisaborole in these two trials.

In response to clarification question B32, the company provided the data extraction table used to support its claim of equal patterns of drug usage. However, for many of the studies, insufficient

details are provided to allow the drug usage per application to be calculated from the data provided. The ERG's clinical advisors considered that it was reasonable to assume that a similar amount of product would be required to cover a similar surface area for TCIs, TCSs and crisaborole; however, the ERG notes that any significant difference in actual usage could have a large impact on the relative costs of the treatments and it is unclear how well the company's assumption is supported from the information provided.

5.3.4.11 Selected algorithms for mapping utility values

The company mapped EQ-5D derived utility values from DLQI and CDLQI data collected in AD-301 and AD-302 trials based on a single published algorithm.(59) While the company acknowledged that [redacted], CDLQI scores were mapped using the same algorithm in scenario analyses. It is unclear whether the mapping algorithm which was estimated in patients aged over 18 generates appropriate values in children given that one of the parameters in the algorithm is age. In its base case, the company used the same multipliers generated from mapping DLQI to EQ-5D values, and applied them to the child population thereby avoiding the need to identify an appropriate mapping algorithm for the CDLQI. However, the ERG notes that this approach may not properly reflect differences between adults and children in terms of the impact of disease severity on HRQoL.

The ERG noted that other candidate algorithms were available but not considered in the base case or as a sensitivity analysis.

5.3.4.12 Duration of treatment for non-responders on primary therapy

The current layout of the company's model base case assumed that non-responders would have only two weeks on treatment followed by two weeks without any active therapy before being switched on subsequent therapy. The ERG noted that this failed to reflect usual clinical practice, as clinicians would generally consider keeping patients on treatment for a longer period of time or switching them to another line without waiting for off-treatment periods.

5.3.4.13 Other concerns:

- The model assumes a starting age of 2 for children and 18 for adults but these do not reflect the average ages in the trial populations or in the populations likely to receive crisaborole in clinical practice.
- The model assumes that the proportion of non-responders achieving a partial response is the same for crisaborole and TCIs.

- ORs used for response adjustment in subsequent therapy are the same as those used for primary treatment.
- When TCIs are used as subsequent therapies, the response rates are based on data from TA82(3) rather than the outputs from the NMA

5.4 Exploratory and sensitivity analyses undertaken by the ERG

This section presents the methods of the ERG's exploratory analyses. The ERG notes that issues discussed in Sections 5.3.4.5, 5.3.4.6, and 5.3.4.7 could not be addressed due to the rigidity of the model structure submitted by the company.

The ERG has also rerun the 11 scenario analyses submitted by the company in response to clarification questions. The ERG carried out scenario 1 for children with mild AD which results were not reported in Appendix 3 of the company's response. There were three instances in scenarios 1 (adults with mild disease post-TCS) and 10.2 (both sets of results for tacrolimus 0.03%) where the ERG could not reproduce the company's results. The ERG's results of rerunning these scenarios are reported in Appendix 2.

5.4.1 Applying costs of dermatologist visits to uncontrolled patients only

In response to clarification question B27, the company asserted that patients on subsequent therapy would receive six dermatology consultations per annum on treatment failure and experiencing a flare. However, the current model coding suggests that all patients who have been referred for subsequent therapy (columns AD:AF) are eligible for dermatologist visits even after they achieve disease control and stop receiving subsequent therapy (i.e. cohort in column AD). The ERG amended the model to apply these dermatology consultation only for uncontrolled patients on subsequent therapy (columns AE:AF).

5.4.2 Correcting acquisition costs of subsequent systemic therapy

In its response to clarification question B17, the company presented an updated version for distribution of systemic therapies and respective costs (Table 19 of the clarification response). The ERG noted that some of the Table data still did not match with the updated model. These included units per pack and cost per pack for ciclosporin and units per flare for mycophenolate mofetil. The ERG amended those three inputs to be 750 mg per ciclosporin pack at a cost of £18.37 (instead of 1,500 mg at a cost of £21.80), and 56 g needed of mycophenolate mofetil per flare (i.e, 2 g per day for 28 days instead of 2 g per flare).

5.4.3 Assuming non-responders to stay on treatment for the rest of a cycle

As indicated in Section 5.3.4.12, the company stopped primary treatment for non-responders after two weeks. The ERG amended the model so that all patients receive at least one full cycle of treatment.

5.4.4 Allowing a partial response to subsequent therapies

The ERG adopted the company's scenario 3 (provided in response to clarification question B4) which allowed patients with moderate AD to experience a partial response to subsequent therapies and included it as ERG scenario 4.

5.4.5 Incorporating other systemic treatments in addition to ciclosporin

The ERG adopted the company's scenario 6 (provided in response to clarification question B17), which incorporated methotrexate, azathioprine and mycophenolate mofetil in addition to ciclosporin as subsequent therapies, as ERG scenario 5.

Assuming average dosing per application for participants [redacted] In its response to clarification question B31, the company presented a table showing drug use per application by age group in patients [redacted] as per crisaborole's indication, and conducted a scenario analysis using these data (Appendix 3 of clarification response, scenario 8)

The ERG views this as a more appropriate approach compared to including data of all patients regardless % of affected BSA, and therefore included it in its base case.

5.4.6 Adjusting response on subsequent therapy

The ERG explored a scenario where response rates for subsequent therapy were adjusted to reflect the whole duration on treatment rather than applied repetitively each cycle as was discussed in Section 5.3.4.8. Table 32 presents the duration of treatment for each therapeutic agent, and the corresponding 4-week response probability based on an overall response of 0.7 for all systemic therapy/phototherapy. The 4-week response probability was adjusted then to reflect time to response per treatment.

Table 32: Response rates to subsequent therapy based on overall time to response duration

Therapy	Average treatment duration	Overall Response	Time to response	4-week probability of response
Phototherapy	3 months	0.7	3 months*	0.31
Ciclosporin	5.8 months	0.7	4 weeks†	0.7
Azathioprine	13.8 months	0.7	2 months[64]	0.43
Methotrexate	15.1 months	0.7	10 weeks[65]	0.38

Mycophenolate mofetil	5.8 months	0.7	10 weeks[66]	0.38
* Assumption based on average treatment duration † Assumption based on clinicians' advice				

5.4.7 Correcting phototherapy costs in company's scenario 5

Following the clarification process, the company undertook three different scenario analyses to account for patients continuing to receive subsequent therapy beyond a model cycle of 4 weeks. One of the scenarios explored considering systemic treatments to be taken continuously and phototherapy for three months. In this scenario, the phototherapy costs were adjusted such that the reference cost for phototherapy (£93) was multiplied by 3.26, to reflect 3 months of treatment, but this cost of £303 (=£93 x 3.26) was then applied repeatedly every cycle, which the ERG did not believe was correct. Therefore, in those scenarios where the duration of phototherapy is explicitly modelled as being 3 months (scenarios 7 and 14 of the ERG's exploratory analyses), the ERG has applied the cost of £303 once only to those entering the subsequent treatment health state.

5.4.8 Applying effectiveness data from the simple random effects NMA

Section 5.3.4.3 discussed the implications of using a different NMA to derive effectiveness data in the model. The ERG explored the use of results from the ERG's simple random effects NMA with neither class effect nor vehicle response adjustment as depicted in Table 22 in Section 4.4.1. In its implementation of the NMA, the ERG kept the same baseline risk from the company's FE-RCE-VR model as well as crisaborole's HR in relation to it and applied the HRs for each TCI relative to crisaborole from Table 22. The corresponding 4-week mean response probabilities applied in the model according to the different NMAs are shown in Table 33. The ERG repeated all of the their exploratory and preferred base case scenario using the ERG's simple random effects NMA and presented these as scenarios 8 to 14.

Table 33: Comparison between mean response probabilities derived from the ERG's NMA and those from the company's base case NMA

Comparator	Response probabilities from the simple random effects NMA[a]: ERG Results	Response probabilities from the company base case model (FE-RCE-VR)[b]
Vehicle	[redacted]	[redacted]
Tacrolimus 0.03%	[redacted]	[redacted]
Tacrolimus 0.1%	[redacted]	[redacted]
Pimecrolimus 1%	[redacted]	[redacted]
Crisaborole	[redacted]	[redacted]
[a] Random treatment effect, no class effect, no vehicle response adjustment [b] Fixed treatment effect, random class effects, vehicle response adjusted		

5.5 Conclusions of the cost effectiveness section

The ERG had several concerns regarding the structure of the company's economic model, in particular in terms of its ability to adequately capture the sequential use of subsequent therapies. This is important because the cost-effectiveness results are being largely driven by the cost savings achieved by avoiding the use of subsequent therapies in non-responders.

The ERG's clinical advisors noted that any topical treatment which can reduce the unmet needs of mild to moderate eczema patients could save significant costs by potentially altering the natural course of the disease, reducing progression to severe disease and reducing disease duration. This is because it would interrupt the cycle of a dysfunctional skin barrier serving as a site for allergic sensitisation which can lead patients to progress from childhood AD to subsequent allergic rhinitis and asthma, which is known as the '*atopic march'*.(67) Therefore, effective treatments have the potential to be cost saving in the long-term, but potential long-term benefits were not incorporated in the modelling.

The conclusions of the cost-effectiveness model are highly dependent on the relative effectiveness for crisaborole versus TCIs as whichever treatment has the greater response rate achieves not only a QALY gain from earlier disease control but also a cost-saving from avoiding subsequent treatments which results in it dominating the other treatments. [redacted] this suggests that the conclusions of the cost-effectiveness analysis are very uncertain. Although the conclusions based on the MAIC broadly agree with the result of the vehicle response adjusted NMA, it is important to remember that the results from these two analyses are not directly comparable because the apply to different populations. In addition, the MAIC assumes that that all prognostic factors and treatment effect modifiers have been accounted for, although this is an untestable assumption. For these reasons, the ERG considers that caution should be exercised when making inferences based on the results of the MAIC and the ERG still prefers the simple random effects NMA.

The ERG also notes that whilst the relative effectiveness of crisaborole and TCIs is highly uncertain because of a lack of head-to-head trials and [redacted], there is an on-going trial of crisaborole against pimecrolimus which also includes TCS as a comparator (NCT03539601). There is also an on-going trial of crisaborole against tacrolimus 0.03% (NCT03645057 (ASPIRE)

although this study is open-label and would be unable to contribute to the evidence network because ISGA is not included as an outcome. The ERG believes that the addition of results from NCT03539601 to the NMA model and to the economic model would have the potential to resolve some of uncertainty regarding the relative cost-effectiveness of crisaborole and TCIs. However, the estimated date of study completion is March 2021.

6 IMPACT ON THE ICER OF ADDITIONAL CLINICAL AND ECONOMIC ANALYSES UNDERTAKEN BY THE ERG

The ERG ran all exploratory analyses deterministically. A summary of the exploratory analyses undertaken by the ERG is presented in Table 34 for children with mild AD, Table 35 for children with moderate AD, Table 36 for adults with mild AD, and Table 37 for adults with moderate AD.

The ERGs exploratory analysis which used the median HRs from the simple random effects NMA demonstrates the sensitivity of the cost-effectiveness results to the uncertainty in the relative effectiveness; crisaborole was found to be dominated by TCIs in this exploratory analysis whereas crisaborole dominated TCIs in the exploratory analyses using the company's vehicle adjusted NMA. This is the most important uncertainty identified in the ERG's exploratory analyses. However, the ERG notes that in the majority of the scenarios the absolute difference in costs and QALYs between crisaborole and TCIs is small. The exception to this is the comparison against tacrolimus 0.1% in adults with moderate AD when using the simple random effects NMA (see Table 35). [redacted] The ERG notes that the cost of tacrolimus 0.1% applied in the company's model was £37.82 for 60g based on the 2019 BNF (month not stated). However, the ERG also notes that the NHS Drug Tariff price for tacrolimus 0.1% was £39.90 in September 2019 and then reduced to £38.46 in October 2019. [redacted]

The ERG's correction to the application of dermatology consultation costs demonstrates that the cost-savings achieved by improving the response rate for primary therapy are being driven by the need for monitoring of subsequent therapies by specialist dermatologists. The ERG's exploratory analysis which explored different assumptions regarding the time on treatment for subsequent therapies and the time to achieve treatment response demonstrate that both the cost-savings and the QALYs gained from achieving an improved response rate to primary therapy are sensitive to the assumptions regarding the modelling of subsequent therapies which the ERG felt had been too simplistic in the company's base case.

The other changes made to the company's base case assumptions had a fairly limited impact on the incremental costs and QALYs except in adults with moderate AD, where the ERG's base case (scenario 6) and the exploratory analyses considering alternative assumptions for subsequent therapy (scenario 7) resulted in crisaborole having ICERs of [redacted] respectively when compared with tacrolimus 0.1% (crisaborole still dominated tacrolimus 0.03% in these scenarios).

Table 34: ERG exploratory model results for mild child patients

Analysis	Discounted costs		Discounted QALYS		ICER (crisaborole versus pimecrolimus)
	Crisaborole	Pimecrolimus	Crisaborole	Pimecrolimus	
Company base case	[redacted]	[redacted]	[redacted]	[redacted]	**Crisaborole dominates** [redacted]
1) Applying costs of dermatologist visits to uncontrolled patients only	[redacted]	[redacted]	[redacted]	[redacted]	Crisaborole dominates [redacted]
2) Correcting acquisition costs of subsequent systemic therapy	[redacted]	[redacted]	[redacted]	[redacted]	Crisaborole dominates [redacted]
3) Assuming non responders receive 4 weeks of treatment	[redacted]	[redacted]	[redacted]	[redacted]	Crisaborole dominates [redacted]
4) Assuming partial response on subsequent therapies*	Not applicable to patients with mild AD				
5) Including other options beside ciclosporin in a weighted basket of systemic therapies*	[redacted]	[redacted]	[redacted]	[redacted]	Crisaborole dominates [redacted]
6) [redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Crisaborole dominates [redacted]
ERG base case (scenarios 1 – 5) using the company's preferred NMA	[redacted]	[redacted]	[redacted]	[redacted]	**Crisaborole dominates** [redacted]
7) Adjusting costs and response on subsequent therapy to reflect the whole time on treatment†	[redacted]	[redacted]	[redacted]	[redacted]	Crisaborole dominates [redacted]
Company base case using the ERG's simple random effects NMA results	[redacted]	[redacted]	[redacted]	[redacted]	**Pimecrolimus dominates** [redacted]
8) Applying costs of dermatologist visits to uncontrolled patients only	[redacted]	[redacted]	[redacted]	[redacted]	Pimecrolimus dominates [redacted]
9) Correcting acquisition costs of subsequent systemic therapy	[redacted]	[redacted]	[redacted]	[redacted]	Pimecrolimus dominates [redacted]
10) Assuming non responders receive 4 weeks of treatment	[redacted]	[redacted]	[redacted]	[redacted]	Pimecrolimus dominates [redacted]
11) Assuming partial response on subsequent therapies*	Not applicable to patients with mild AD				
12) Including other options beside ciclosporin in a weighted basket of systemic therapies*	[redacted]	[redacted]	[redacted]	[redacted]	Pimecrolimus dominates [redacted]

Table of Contents

Analysis	Discounted costs		Discounted QALYS		ICER (crisaborole versus pimecrolimus)
	Crisaborole	Pimecrolimus	Crisaborole	Pimecrolimus	
13) [redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Pimecrolimus dominates [redacted]
ERG base case (scenarios 7 – 11) using the ERG's simple random effects NMA results	[redacted]	[redacted]	[redacted]	[redacted]	**Pimecrolimus dominates** [redacted]
14) Adjusting costs and response on subsequent therapy to reflect the whole time on treatment†	[redacted]	[redacted]	[redacted]	[redacted]	Pimecrolimus dominates [redacted]

* The results for these analyses were already presented in Appendix 3 of the company's response to clarification response †In conjunction with the ERG base case mentioned above
ΔC, difference in costs, ΔQ, difference in QALYs; ICER, incremental cost-effectiveness ratio; QALY, quality adjusted life year

Table 35: ERG exploratory model results for moderate child patients

Analysis	Discounted costs			Discounted QALYS			ICER (crisaborole versus tacrolimus 0.03%†)
	Crisaborole	Tacrolimus 0.03%	Pimecrolimus	Crisaborole	Tacrolimus 0.03%	Pimecrolimus	
Company base case	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	**Crisaborole dominates** [redacted]
1) Applying costs of dermatologist visits to uncontrolled patients only	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Crisaborole dominates [redacted]
2) Correcting acquisition costs of subsequent systemic therapy	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Crisaborole dominates [redacted]
3) Assuming non responders receive 4 weeks of treatment	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Crisaborole dominates [redacted]
4) Assuming partial response on subsequent therapies*	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Crisaborole dominates [redacted]
5) Including other options beside ciclosporin in a weighted basket of systemic therapies*	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Crisaborole dominates [redacted]

Analysis	Discounted costs			Discounted QALYS			ICER (crisaborole versus tacrolimus 0.03%†)
	Crisaborole	Tacrolimus 0.03%	Pimecrolimus	Crisaborole	Tacrolimus 0.03%	Pimecrolimus	
6) [redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Crisaborole dominates [redacted]
ERG base case (scenarios 1 – 6) using the company's preferred NMA	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	**Crisaborole dominates** [redacted]
7) Adjusting costs and response on subsequent therapy to reflect the whole time on treatment†	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Crisaborole dominates [redacted]
Company base case using the ERG's simple random effects NMA results	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	**Tacrolimus [redacted] dominates** [redacted]
8) Applying costs of dermatologist visits to uncontrolled patients only	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Tacrolimus dominates [redacted]
9) Correcting acquisition costs of subsequent systemic therapy	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Tacrolimus dominates [redacted]
10) Assuming non responders receive 4 weeks of treatment	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Tacrolimus dominates [redacted]
11) Assuming partial response on subsequent therapies*	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Tacrolimus dominates [redacted]
12) Including other options beside ciclosporin in a weighted basket of systemic therapies*	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Tacrolimus dominates [redacted]
13) [redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Tacrolimus dominates [redacted]

Analysis	Discounted costs			Discounted QALYS			ICER (crisaborole versus tacrolimus 0.03%†)
	Crisaborole	Tacrolimus 0.03%	Pimecrolimus	Crisaborole	Tacrolimus 0.03%	Pimecrolimus	
[redacted]							
ERG base case (scenarios 8 – 13) using the ERG's simple random effects NMA results	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	**Tacrolimus dominates** [redacted]
14) Adjusting costs and response on subsequent therapy to reflect the whole time on treatment†	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Tacrolimus dominates [redacted]

†Tacrolimus 0.03% always dominates pimecrolimus in all scenarios

* The results for these analyses were already presented in Appendix 3 of the company's response to clarification response †In conjunction with the ERG base case mentioned above

ΔC, difference in costs, ΔQ, difference in QALYs; ICER, incremental cost-effectiveness ratio; QALY, quality adjusted life year

Table 36: ERG exploratory model results for mild adult patients

Analysis	Discounted costs		Discounted QALYS		ICER (crisaborole versus pimecrolimus)
	Crisaborole	Pimecrolimus	Crisaborole	Pimecrolimus	
Company base case	[redacted]	[redacted]	[redacted]	[redacted]	**Crisaborole dominates** [redacted]
1) Applying costs of dermatologist visits to uncontrolled patients only	[redacted]	[redacted]	[redacted]	[redacted]	Crisaborole dominates [redacted]
2) Correcting acquisition costs of subsequent systemic therapy	[redacted]	[redacted]	[redacted]	[redacted]	Crisaborole dominates [redacted]
3) Assuming non responders receive 4 weeks of treatment	[redacted]	[redacted]	[redacted]	[redacted]	Crisaborole dominates [redacted]
4) Assuming partial response on subsequent therapies*	Not applicable to patients with mild AD				
5) Including other options beside ciclosporin in a weighted basket of systemic therapies*	[redacted]	[redacted]	[redacted]	[redacted]	Crisaborole dominates [redacted]
6) [redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Crisaborole dominates [redacted]

Analysis	Discounted costs		Discounted QALYS		ICER (crisaborole versus pimecrolimus)
	Crisaborole	Pimecrolimus	Crisaborole	Pimecrolimus	
ERG base case (scenarios 1 – 5) using the company's preferred NMA	[redacted]	[redacted]	[redacted]	[redacted]	**Crisaborole dominates** [redacted]
7) Adjusting costs and response on subsequent therapy to reflect the whole time on treatment†	[redacted]	[redacted]	[redacted]	[redacted]	Crisaborole dominates [redacted]
Company base case using the ERG's simple random effects NMA results	[redacted]	[redacted]	[redacted]	[redacted]	**Pimecrolimus dominates** [redacted]
8) Applying costs of dermatologist visits to uncontrolled patients only	[redacted]	[redacted]	[redacted]	[redacted]	Pimecrolimus dominates [redacted]
9) Correcting acquisition costs of subsequent systemic therapy	[redacted]	[redacted]	[redacted]	[redacted]	Pimecrolimus dominates [redacted]
10) Assuming non responders receive 4 weeks of treatment	[redacted]	[redacted]	[redacted]	[redacted]	Pimecrolimus dominates [redacted]
11) Assuming partial response on subsequent therapies*	Not applicable to patients with mild AD				
12) Including other options beside ciclosporin in a weighted basket of systemic therapies*	[redacted]	[redacted]	[redacted]	[redacted]	Pimecrolimus dominates [redacted]
13) [redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Pimecrolimus dominates [redacted]
ERG base case (scenarios 7 – 11) using the ERG's simple random effects NMA results	[redacted]	[redacted]	[redacted]	[redacted]	**Pimecrolimus dominates** [redacted]
14) Adjusting costs and response on subsequent therapy to reflect the whole time on treatment†	[redacted]	[redacted]	[redacted]	[redacted]	Pimecrolimus dominates [redacted]

* The results for these analyses were already presented in Appendix 3 of the company's response to clarification response †In conjunction with the ERG base case mentioned above
ΔC, difference in costs, ΔQ, difference in QALYs; ICER, incremental cost-effectiveness ratio; QALY, quality adjusted life year

Table of Contents

Table 37: ERG exploratory model results for moderate adult patients

Analysis	Discounted costs			Discounted QALYS			ICER (crisaborole versus tacrolimus 0.1%†)
	Crisaborole	Tacrolimus 0.1%	Tacrolimus 0.03%	Crisaborole	Tacrolimus 0.1%	Tacrolimus 0.03%	
Company base case	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	**Crisaborole dominates** [redacted]
1) Applying costs of dermatologist visits to uncontrolled patients only	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted] per QALY [redacted]
2) Correcting acquisition costs of subsequent systemic therapy	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Crisaborole dominates [redacted]
3) Assuming non responders receive 4 weeks of treatment	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Crisaborole dominates [redacted]
4) Assuming partial response on subsequent therapies*	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Crisaborole dominates [redacted]
5) Including other options beside ciclosporin in a weighted basket of systemic therapies*	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Crisaborole dominates [redacted]
6) [redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Crisaborole dominates [redacted]
ERG base case (scenarios 1 – 6) using the company's preferred NMA	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	**[redacted] per QALY versus tacrolimus 0.1%** [redacted]
7) Adjusting costs and response on subsequent therapy to reflect the whole time on treatment†	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted] per QALY versus tacrolimus 0.1% [redacted]
Company base case using the ERG's	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	**Tacrolimus 0.1%**

Analysis	Discounted costs			Discounted QALYS			ICER (crisaborole versus tacrolimus 0.1%†)
	Crisaborole	Tacrolimus 0.1%	Tacrolimus 0.03%	Crisaborole	Tacrolimus 0.1%	Tacrolimus 0.03%	
simple random effects NMA results							**dominates** [redacted]
8) Applying costs of dermatologist visits to uncontrolled patients only	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Tacrolimus 0.1% dominates [redacted]
9) Correcting acquisition costs of subsequent systemic therapy	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Tacrolimus 0.1% dominates [redacted]
10) Assuming non responders receive 4 weeks of treatment	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Tacrolimus 0.1% dominates [redacted]
11) Assuming partial response on subsequent therapies*	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Tacrolimus 0.1% dominates [redacted]
12) Including other options beside ciclosporin in a weighted basket of systemic therapies*	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Tacrolimus 0.1% dominates [redacted]
13) [redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Tacrolimus 0.1% dominates [redacted]
ERG base case (scenarios 8 – 13) using the ERG's simple random effects NMA results	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	**Tacrolimus 0.1% dominates** [redacted]
14) Adjusting costs and response on subsequent therapy to reflect the whole	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Tacrolimus 0.1% dominates

Analysis	Discounted costs			Discounted QALYS			ICER (crisaborole versus tacrolimus 0.1%†)
	Crisaborole	Tacrolimus 0.1%	Tacrolimus 0.03%	Crisaborole	Tacrolimus 0.1%	Tacrolimus 0.03%	
time on treatment†							[redacted]

†Tacrolimus 0.1% always dominates Tacrolimus 0.03% in all scenarios

* The results for these analyses were already presented in Appendix 3 of the company's response to clarification response †In conjunction with the ERG base case mentioned above

ΔC, difference in costs, ΔQ, difference in QALYs; ICER, incremental cost-effectiveness ratio; QALY, quality adjusted life year

7 END OF LIFE

The company made no claims that crisaborole would meet the end of life criteria as it was assumed that the intervention would not extend life. The ERG concurs with the company's view.

8 OVERALL CONCLUSIONS

8.1 Overall conclusions

There is significant uncertainty regarding the relative effectiveness of TCIs and crisaborole due to a lack of head-to-head trials and uncertainty regarding whether [redacted]. This uncertainty in the clinical effectiveness evidence makes the cost-effectiveness results very uncertain with opposite conclusions regarding the most cost-effective intervention (crisaborole or TCIs) depending on whether or not the estimates of treatment effect are [redacted]. However, the absolute differences in costs and QALYs between crisaborole and TCIs are generally small in both scenarios.

8.2 Implications for research

Further information is needed to quantify the relative effectiveness of crisaborole and TCIs. This need will be partially addressed by the on-going trial of crisaborole against pimecrolimus which also includes TCS as a comparator and reports ISGA (NCT03539601) and the on-going trial of crisaborole against tacrolimus 0.03% which reports EASI but not ISGA (NCT03645057 (ASPIRE)). Limitations to these studies include the open-label design of ASPIRE, and the lack of a direct comparison against tacrolimus 0.1%. Additional information on long-term outcomes such as frequency of flares would also be beneficial, as any increase in the time between flares would affect the estimates of cost-effectiveness.

9 REFERENCES

1. Pfizer Inc. Crisaborole for Treating Mild to Moderate Atopic Dermatitis in People Aged 2 Years and Older [ID 1195]: Company's evidence submission. 2019.
2. National Institute for Health and Care Excellence. Crisaborole for treating mild to moderate atopic dermatitis in people aged 2 years and older. Final scope. 2019.
3. National Institute for Health and Care Excellence. Tacrolimus and pimecrolimus for atopic eczema 2004 [01/11/2019]. Available from: https://www.nice.org.uk/guidance/ta82.
4. National Institute for Health and Care Excellence. Atopic eczema in under 12s: diagnosis and management. 2007.
5. National Eczema Society. What is eczema? London, UK: National Eczema Society; [01/11/2019]. Available from: http://www.eczema.org/what-is-eczema.
6. British Association of Dermatologists. Atopic eczema. London, UK: British Association of Dermatologists,; 2017.
7. National Eczema Society. Atopic London, UK: National Eczema Society; [Available from: http://www.eczema.org/atopic.
8. Abuabara K, Margolis DJ, Langan SM. The Long-Term Course of Atopic Dermatitis. Dermatol Clin. 2017;35(3):291-7.
9. Vos T, Abajobir AA, Abate KH, Abbafati C, Abbas KM, Abd-Allah F, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. The Lancet. 2017;390(10100):1211-59.
10. Mathiesen SM, Thomsen SF. The prevalence of atopic dermatitis in adults: systematic review on population studies. Dermatology online journal. 2019;25(8).
11. Atopic eczema: NHS; 2016 [Available from: https://www.nhs.uk/conditions/atopic-eczema/.
12. Garmhausen D, Hagemann T, Bieber T, Dimitriou I, Fimmers R, Diepgen T, et al. Characterization of different courses of atopic dermatitis in adolescent and adult patients. Allergy. 2013;68(4):498-506.
13. National Institute for Health and Care Excellence. TA81 Frequency of application of topical corticosteroids for atopic eczema. London, UK: National Institute for Health and Care Excellence,; 2005.
14. Abuabara K, Magyari A, McCulloch CE, Linos E, Margolis DJ, Langan SM. Prevalence of Atopic Eczema Among Patients Seen in Primary Care: Data From The Health Improvement Network. Ann Intern Med. 2019;170(5):354-6.
15. Blak BT, Thompson M, Dattani H, Bourke A. Generalisability of The Health Improvement Network (THIN) database: demographics, chronic disease prevalence and mortality rates. Informatics in primary care. 2011;19(4):251-5.
16. Schmitt J, Spuls PI, Thomas KS, Simpson E, Furue M, Deckert S, et al. The Harmonising Outcome Measures for Eczema (HOME) statement to assess clinical signs of atopic eczema in trials. J Allergy Clin Immunol. 2014;134(4):800-7.
17. Chopra R, Vakharia PP, Sacotte R, Patel N, Immaneni S, White T, et al. Severity strata for Eczema Area and Severity Index (EASI), modified EASI, Scoring Atopic Dermatitis (SCORAD), objective SCORAD, Atopic Dermatitis Severity Index and body surface area in adolescents and adults with atopic dermatitis. Br J Dermatol. 2017;177(5):1316-21.
18. BAD-BSPD-NICE. Paediatric National Audit on Atopic Eczema in Children [updated March 201501/11/2019]. Available from: http://www.bad.org.uk/healthcare-professionals/national-audits/clinical-audits/paediatric-eczema.
19. Higgins JPT GSe. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. 2011.
20. Pfizer Inc. Crisaborole for Treating Mild to Moderate Atopic Dermatitis in People Aged 2 Years and Older [ID 1195]: Appendix D, Company's evidence submission. 2019.
21. Pfizer Inc. Crisaborole for Treating Mild to Moderate Atopic Dermatitis in People Aged 2 Years and Older [ID 1195]: Company's response to clarification questions from the ERG. 2019.
22. National Institute for Health and Care Excellence (NICE). The guidelines manual: appendices B-I (Appendix C: Methodology checklist: randomised controlled trials) United Kingdom: NICE; 2012

[Available from: https://www.nice.org.uk/process/pmg6/resources/the-guidelines-manual-appendices-bi-pdf-3304416006853.
23. Paller AS, Tom WL, Lebwohl MG, Blumenthal RL, Boguniewicz M, Call RS, et al. Efficacy and safety of crisaborole ointment, a novel, nonsteroidal phosphodiesterase 4 (PDE4) inhibitor for the topical treatment of atopic dermatitis (AD) in children and adults. J Am Acad Dermatol. 2016;75(3):494-503 e6.
24. Safety and Efficacy of AN2728 Topical Ointment, 2% in Children, Adolescents, and Adults (Ages 2 Years and Older) With Atopic Dermatitis. AD-301: NCT02118766 [13/11/2019]. Available from: https://clinicaltrials.gov/ct2/show/results/NCT02118766.
25. Safety and Efficacy of AN2728 Topical Ointment, 2% in Children, Adolescents, and Adults (Aged 2 Years and Older) With Atopic Dermatitis. AD-302: NCT02118792 [11/11/2019]. Available from: https://clinicaltrials.gov/ct2/show/NCT02118792.
26. Hanifin JM, Ellis CN, Frieden IJ, Folster-Holst R, Stein Gold LF, Secci A, et al. OPA-15406, a novel, topical, nonsteroidal, selective phosphodiesterase-4 (PDE4) inhibitor, in the treatment of adult and adolescent patients with mild to moderate atopic dermatitis (AD): A phase-II randomized, double-blind, placebo-controlled study. J Am Acad Dermatol. 2016;75(2):297-305.
27. Callender VD AA, Stein Gold LF, Lebwohl M, Desai SR, Paller AS, Tan H, Vranic I, Ports WC, Zielinski MA, Tallman AM. . Long-Term Safety of Crisaborole Ointment, 2%, Across Racial and Ethnic Groups. American Academy of Dermatology Annual Meeting; San Diego, CA2018.
28. Yosipovitch G, Gold LF, Lebwohl MG, Silverberg JI, Tallman AM, Zane LT. Early Relief of Pruritus in Atopic Dermatitis with Crisaborole Ointment, A Non-steroidal, Phosphodiesterase 4 Inhibitor. Acta Derm Venereol. 2018;98(5):484-9.
29. Simpson EL, Paller AS, Boguniewicz M, Eichenfield LF, Feldman SR, Silverberg JI, et al. Crisaborole Ointment Improves Quality of Life of Patients with Mild to Moderate Atopic Dermatitis and Their Families. Dermatol Ther (Heidelb). 2018;8(4):605-19.
30. Eichenfield LF, Call RS, Forsha DW, Fowler J, Jr., Hebert AA, Spellman M, et al. Long-term safety of crisaborole ointment 2% in children and adults with mild to moderate atopic dermatitis. J Am Acad Dermatol. 2017;77(4):641-9 e5.
31. Mulick A, Allen V, Williams H, et al. Classifying atopic dermatitis: protocol for a systematic review of subtypes (phenotypes) and associated characteristics. BMJ Open. 2018:8:e023097.
32. Spuls PI, Gerbens LAA, Simpson E, Apfelbacher CJ, Chalmers JR, Thomas KS, et al. Patient-Oriented Eczema Measure (POEM), a core instrument to measure symptoms in clinical trials: a Harmonising Outcome Measures for Eczema (HOME) statement. British Journal of Dermatology. 2017;176(4):979-84.
33. Futamura M, Leshem YA, Thomas KS, Nankervis H, Williams HC, Simpson EL. A systematic review of Investigator Global Assessment (IGA) in atopic dermatitis (AD) trials: Many options, no standards. J Am Acad Dermatol. 2016;74(2):288-94.
34. Levy AL SM, Roberts R. Tacrolimus Cream 0.03% Is Safe and Effective in the Treatment of Mild and Moderate Atopic Dermatitis in Adults. J Allergy Clin Immunol. 2005;115:S103.
35. Boguniewicz M, Fiedler VC, Raimer S, Lawrence ID, Leung DY, Hanifin JM. A randomized, vehicle-controlled trial of tacrolimus ointment for treatment of atopic dermatitis in children. Pediatric Tacrolimus Study Group. J Allergy Clin Immunol. 1998;102(4 Pt 1):637-44.
36. Chapman MS, Schachner LA, Breneman D, Boguniewicz M, Gold MH, Shull T, et al. Tacrolimus ointment 0.03% shows efficacy and safety in pediatric and adult patients with mild to moderate atopic dermatitis. 2005;53(2):S177-S85.
37. Schachner LA, Lamerson C, Sheehan MP, Boguniewicz M, Mosser J, Raimer S, et al. Tacrolimus ointment 0.03% is safe and effective for the treatment of mild to moderate atopic dermatitis in pediatric patients: results from a randomized, double-blind, vehicle-controlled study. Pediatrics. 2005;116(3):e334-42.
38. Eichenfield LF, Lucky AW, Boguniewicz M, Langley RG, Cherill R, Marshall K, et al. Safety and efficacy of pimecrolimus (ASM 981) cream 1% in the treatment of mild and moderate atopic dermatitis in children and adolescents. J Am Acad Dermatol. 2002;46(4):495-504.
39. Abramovits W, Fleischer AB, Jr., Jaracz E, Breneman D. Adult patients with moderate atopic dermatitis: tacrolimus ointment versus pimecrolimus cream. J Drugs Dermatol. 2008;7(12):1153-8.

40. Kempers S, Boguniewicz M, Carter E, Jarratt M, Pariser D, Stewart D, et al. A randomized investigator-blinded study comparing pimecrolimus cream 1% with tacrolimus ointment 0.03% in the treatment of pediatric patients with moderate atopic dermatitis. J Am Acad Dermatol. 2004;51(4):515-25.
41. Paller AS, Lebwohl M, Fleischer AB, Jr., Antaya R, Langley RG, Kirsner RS, et al. Tacrolimus ointment is more effective than pimecrolimus cream with a similar safety profile in the treatment of atopic dermatitis: results from 3 randomized, comparative studies. J Am Acad Dermatol. 2005;52(5):810-22.
42. Fowler J, Johnson A, Chen M, Abrams K. Improvement in pruritus in children with atopic dermatitis using pimecrolimus cream 1%. Cutis. 2007;79(1):65-72.
43. Kaufmann R, Bieber T, Helgesen AL, Andersen BL, Luger T, Poulin Y, et al. Onset of pruritus relief with pimecrolimus cream 1% in adult patients with atopic dermatitis: a randomized trial. Allergy. 2006;61(3):375-81.
44. Meurer M, Fartasch M, Albrecht G, Vogt T, Worm M, Ruzicka T, et al. Long-term efficacy and safety of pimecrolimus cream 1% in adults with moderate atopic dermatitis. Dermatology. 2004;208(4):365-72.
45. Wahn U, Bos JD, Goodfield M, Caputo R, Papp K, Manjra A, et al. Efficacy and safety of pimecrolimus cream in the long-term management of atopic dermatitis in children. Pediatrics. 2002;110(1 Pt 1):e2.
46. Sears HW, Bailer JW, Yeadon A. Efficacy and safety of hydrocortisone buteprate 0.1% cream in patients with atopic dermatitis. Clinical therapeutics. 1997;19(4):710-9.
47. Hoeger PH, Lee KH, Jautova J, Wohlrab J, Guettner A, Mizutani G, et al. The treatment of facial atopic dermatitis in children who are intolerant of, or dependent on, topical corticosteroids: a randomized, controlled clinical trial. Br J Dermatol. 2009;160(2):415-22.
48. Hordinsky M, Fleischer A, Rivers JK, Poulin Y, Belsito D, Hultsch T. Efficacy and safety of pimecrolimus cream 1% in mild-to-moderate chronic hand dermatitis: a randomized, double-blind trial. Dermatology. 2010;221(1):71-7.
49. Murrell DF, Calvieri S, Ortonne JP, Ho VC, Weise-Riccardi S, Barbier N, et al. A randomized controlled trial of pimecrolimus cream 1% in adolescents and adults with head and neck atopic dermatitis and intolerant of, or dependent on, topical corticosteroids. Br J Dermatol. 2007;157(5):954-9.
50. van Zuuren EJ, Fedorowicz Z, Christensen R, Lavrijsen A, Arents BWM. Emollients and moisturisers for eczema. Cochrane Database Syst Rev. 2017;2:CD012119.
51. Ridd MJ, Roberts A, Grindlay D, Williams HC. Which emollients are effective and acceptable for eczema in children? BMJ. 2019;367:l5882.
52. Sun GW, Shook TL, Kay GL. Inappropriate use of bivariable analysis to screen risk factors for use in multivariable analysis. J Clin Epidemiol. 1996;49(8):907-16.
53. (PBAC) PBAC. CRISABOROLE, Ointment containing crisaborole 20 mg per g, 60 g, Staquis®, Pfizer Australia Pty Ltd. Public Summary Document: Nov 2018 PBAC Meeting. 2018.
54. Canadian Agency for Drugs and Technologies in Health. Common Drug Review Report: Crisaborole, 2019. 2019.
55. Coyle D, Barbeau M. Cost effectiveness of Elidel in the management of patients with atopic dermatitis in Canada. J Cutan Med Surg. 2004;8(6):405-10.
56. Ellis CN, Kahler KH, Grueger J, Chang J. Cost effectiveness of management of mild-to-moderate atopic dermatitis with 1% pimecrolimus cream in children and adolescents 2-17 years of age. American journal of clinical dermatology. 2006;7(2):133-9.
57. NHS Digital. Health Survey for England Leeds, UK: NHS Digital,; 2019 [Available from: https://digital.nhs.uk/data-and-information/publications/statistical/health-survey-for-england/2017.
58. Breneman D, Fleischer AB, Jr., Abramovits W, Zeichner J, Gold MH, Kirsner RS, et al. Intermittent therapy for flare prevention and long-term disease control in stabilized atopic dermatitis: a randomized comparison of 3-times-weekly applications of tacrolimus ointment versus vehicle. J Am Acad Dermatol. 2008;58(6):990-9.
59. Ali FM, Kay R, Finlay AY, Piguet V, Kupfer J, Dalgard F, et al. Mapping of the DLQI scores to EQ-5D utility values using ordinal logistic regression. Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation. 2017;26(11):3025-34.

60. National Health Service. Reference costs 2017/18 2018 [Available from: https://improvement.nhs.uk/resources/reference-costs/.
61. Personal Social Services Research Unit. Unit Costs of Health and Social Care: Personal Social Services Research Unit; 2018. [Available from: https://www.pssru.ac.uk/project-pages/unit-costs/unit-costs-2018/.
62. Mailhol C, Lauwers-Cances V, Rance F, Paul C, Giordano-Labadie F. Prevalence and risk factors for allergic contact dermatitis to topical treatment in atopic dermatitis: a study in 641 children. Allergy. 2009;64(5):801-6.
63. NHS Improvement. Reference costs 2018 [Available from: https://improvement.nhs.uk/resources/reference-costs/.
64. Hughes R, Collins P, Rogers S. Further experience of using azathioprine in the treatment of severe atopic dermatitis. Clinical and Experimental Dermatology. 2008;33(6):710-1.
65. Deo M, Yung A, Hill S, Rademaker M. Methotrexate for treatment of atopic dermatitis in children and adolescents. International Journal of Dermatology. 2014;53(8):1037-41.
66. Haeck IM, Knol MJ, ten Berge O, van Velsen SGA, De Bruin-Weller M, Bruijnzeel-Koomen CAFM. Enteric-coated mycophenolate sodium versus cyclosporin A as long-term treatment in adult patients with severe atopic dermatitis: A randomized controlled trial. Journal of the American Academy of Dermatology. 2011;64(6):1074-84.
67. Bantz SK, Zhu Z, Zheng T. The Atopic March: Progression from Atopic Dermatitis to Allergic Rhinitis and Asthma. J Clin Cell Immunol. 2014;5(2):202.

10 APPENDICES

Appendix 1: Table showing excluded studies based on > 20% of population having severe AD.

Adapted from Clarification response, Table 2

Author year	Interventions	Subgroup data available
Almeyda, J.,Burt, B. W. (1974). Double blind controlled study of treatment of atopic eczema with a preparation of hydrocortisone in a new drug delivery system versus betamethasone 17-valerate *Br J Dermatol*, 91(5), 579-83	TCS	No mild and/or moderate disease subgroup data
Bleeker, J. (1975). Double blind comparison between two new topical corticosteroids, halcinonide 0.1% and clobetasol propionate cream 0.05% *Current medical research and opinion*, 3 (4), 225-228	TCS	No mild and/or moderate disease subgroup data
Savin, R. C. (1976). Betamethasone dipropionate in psoriasis and atopic dermatitis *Conn Med*, 40(1), 5-7	TCS	No mild and/or moderate disease subgroup data
Fisher, M.,Kelly, A. P. (1979). Multicenter trial of fluocinonide in an emollient cream base *Int J Dermatol*, 18(8), 660-4	TCS	No mild and/or moderate disease subgroup data
Veien, N. K.,Hattel, T.,Justesen, O.,Norholm, A.,Verjans, H. L. (1984). Hydrocortisone 17-butyrate (Locoid) 0.1% cream versus hydrocortisone (Uniderm) 1% cream in the treatment of children suffering from atopic dermatitis *J Int Med Res*, 12(5), 310-3	TCS	No mild and/or moderate disease subgroup data
Reitamo, S., Wollenberg, A., Schopf, E., Perrot, JL., Marks, R., Rusicka, T., Christophers, E., Kapp, A., Lahfa, M., Rubins, A., Jablonska, S., Rustin, M., for the European Tacrolimus Study Group. (2000), Safety and Efficacy of 1 Year of Tacrolimus Ointment Monotherapy in Adults With Atopic Dermatitis *Arch Dermatol,* 136, 999-1006	Tacrolimus	No mild and/or moderate disease subgroup data
Hanifin, J. M.,Ling, M. R.,Langley, R.,Breneman, D.,Rafal, E. (2001). Tacrolimus ointment for the treatment of atopic dermatitis in adult patients: part I, efficacy *J Am Acad Dermatol*, 44(1 Suppl), S28-38	Tacrolimus	Severe AD at baseline results presented for both studies combined in the intervention group but not for vehicle group
Hanifin, J.,Gupta, A. K.,Rajagopalan, R. (2002). Intermittent dosing of fluticasone propionate cream for reducing the risk of relapse in atopic dermatitis patients *Br J Dermatol*, 147(3), 528-37	TCS	No mild and/or moderate disease subgroup data
Paller, A.,Eichenfield, L. F.,Leung, D. Y.,Stewart, D.,Appell, M. (2001). A 12-week study of tacrolimus ointment for the treatment of atopic dermatitis in pediatric patients *J Am Acad Dermatol*, 44(1 Suppl), S47-57	Tacrolimus	No mild and/or moderate disease subgroup data
Soter, NA., Fleischer, AB., Webster, GF., Monroe, E., Lawrence, I., and the Tacrolimus Ointment Study Group. (2001). Tacrolimus ointment for the treatment of atopic dermatitis in adult patients: Part II, Safety. *J Am Acad Dermatol,* S39-S46	Tacrolimus	No mild and/or moderate disease subgroup data

Author year	Interventions	Subgroup data available
Meurer, M.,Folster-Holst, R.,Wozel, G.,Weidinger, G.,Junger, M.,Brautigam, M.,Casm-De- study group (2002). Pimecrolimus cream in the long-term management of atopic dermatitis in adults: a six-month study *Dermatology*, 205(3), 271-7	Pimecrolimus	No mild and/or moderate disease subgroup data for outcomes of interest
Reitamo, S.,Rustin, M.,Ruzicka, T.,Cambazard, F.,Kalimo, K.,Friedmann, P. S.,Schoepf, E.,Lahfa, M.,Diepgen, T. L.,Judodihardjo, H.,Wollenberg, A.,Berth-Jones, J.,Bieber, T.,European Tacrolimus Ointment Study, (2002). Efficacy and safety of tacrolimus ointment compared with that of hydrocortisone butyrate ointment in adult patients with atopic dermatitis *Journal of Allergy and Clinical Immunology*, 109(3), 547-555	Tacrolimus TCS	No mild and/or moderate disease subgroup data
Luger, T. A.,Lahfa, M.,Folster-Holst, R.,Gulliver, W. P.,Allen, R.,Molloy, S.,Barbier, N.,Paul, C.,Bos, J. D. (2004). Long-term safety and tolerability of pimecrolimus cream 1% and topical corticosteroids in adults with moderate to severe atopic dermatitis *J Dermatolog Treat*, 15(3), 169-78	Pimecrolimus TCS	No mild and/or moderate disease subgroup data
Reitamo, S.,Harper, J.,Bos, J. D.,Cambazard, F.,Bruijnzeel-Koomen, C.,Valk, P.,Smith, C.,Moss, C.,Dobozy, A.,Palatsi, R.,European Tacrolimus Ointment, Group (2004). 0.03% Tacrolimus ointment applied once or twice daily is more efficacious than 1% hydrocortisone acetate in children with moderate to severe atopic dermatitis: results of a randomized double-blind controlled trial.[see comment] *British Journal of Dermatology*, 150(3), 554-62	Tacrolimus TCS	Presents mEASI scores by moderate subgroup (see Figure 2 and Table 2) Not stratified randomized by AD severity
Reitamo, S.,Ortonne, J. P.,Sand, C.,Cambazard, F.,Bieber, T.,Folster-Holst, R.,Vena, G.,Bos, J. D.,Fabbri, P.,Groenhoej Larsen, C.,European Tacrolimus Ointment Study, Group (2005). A multicentre, randomized, double-blind, controlled study of long-term treatment with 0.1% tacrolimus ointment in adults with moderate to severe atopic dermatitis *Br J Dermatol*, 152(6), 1282-9	Tacrolimus	No mild and/or moderate disease subgroup data
Fleischer, A. B., Jr.,Abramovits, W.,Breneman, D.,Jaracz, E.,U. S/Canada tacrolimus ointment study group (2007). Tacrolimus ointment is more effective than pimecrolimus cream in adult patients with moderate to very severe atopic dermatitis *J Dermatolog Treat*, 18(3), 151-7	Tacrolimus Pimecrolimus	Presents EASI and IGADA scores by moderate subgroup (see Figure 5 and Figure 7) Not stratified randomized by AD severity
Remitz, A., Harper, J., Rustin, M., Goldschmidt, W.F., Palatsi, R., Van Der Valk, P.G., Sharpe, G., Smith, C.H., Dobozy, A., Turjanmaa, K. and European Tacrolimus Ointment Study Group (2007). Long-term safety and efficacy of tacrolimus ointment for the treatment of atopic dermatitis in children. Acta dermato-venereologica, 87(1), 54-61.	Tacrolimus	No mild and/or moderate disease subgroup data

Author year	Interventions	Subgroup data available
Mandelin, J.,Remitz, A.,Virtanen, H.,Reitamo, S. (2010). One-year treatment with 0.1% tacrolimus ointment versus a corticosteroid regimen in adults with moderate to severe atopic dermatitis: A randomized, double-blind, comparative trial *Acta Derm Venereol*, 90: 170-174	Tacrolimus TCS	No mild and/or moderate disease subgroup data
Woods, M. T.,Brown, P. A.,Baig-Lewis, S. F.,Simpson, E. L. (2011). Effects of a novel formulation of fluocinonide 0.1% cream on skin barrier function in atopic dermatitis *J Drugs Dermatol*, 10(2), 171-6	TCS	No mild and/or moderate disease subgroup data

Appendix 2: Results for scenario analyses conducted by the company that were either not reported or unreproducible (results shown here are the ERG's rerun of these scenarios)

Scenario 1 (Incorporating TCIs and crisaborole in a sequential pathway)

Results for children with mild AD post-TCS

	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER vs. baseline	ICER incremental
Crisaborole	[redacted]	[redacted]	[redacted]	[redacted]		
Pimecrolimuis	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Crisaborole -> TCIs	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Pimecrolimus -> crisaborole	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Results for adults with mild AD post-TCS

	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER vs. baseline	ICER incremental
Crisaborole -> TCIs	[redacted]	[redacted]	[redacted]	[redacted]		
Crisaborole	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Pimecrolimus	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Pimecrolimus -> crisaborole	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Scenario 10.2 (Reducing response rates by 20% following a partial response)

Results for children with moderate AD post-TCS

	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER vs. baseline	ICER incremental
Crisaborole	[redacted]	[redacted]	[redacted]	[redacted]	-	Dominant
Tacrolimus 0.03%	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominated
Pimecrolimus	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominated

Results for adults with moderate AD post-TCS

	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER vs. baseline	ICER incremental
Crisaborole	[redacted]	[redacted]	[redacted]	[redacted]	-	Dominant
Tacrolimus 0.1%	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominated
Tacrolimus 0.03%	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominated

National Institute for Health and Care Excellence

Centre for Health Technology Evaluation

Pro-forma Response

ERG report

Crisaborole for Treating Mild to Moderate Atopic Dermatitis in People Aged 2 Years and Older [ID 1195]

You are asked to check the ERG report from Centre for Reviews and Dissemination and Centre for Health Economics – York to ensure there are no factual inaccuracies contained within it.

If you do identify any factual inaccuracies you must inform NICE by **5pm on 29 November 2018** using the below proforma comments table. All factual errors will be highlighted in a report and presented to the Appraisal Committee and will subsequently be published on the NICE website with the committee papers.

The proforma document should act as a method of detailing any inaccuracies found and how and why they should be corrected.

Issue 1

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
The cost of tacrolimus 0.1% is currently £39.90. The ICER estimates the ERG have provided are not accurate at the time of writing.	The ERG economic analyses and report text require updating to reflect current tacrolimus pricing.	Accuracy	See response to issue 50

Issue 2

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
The ERG states p3 (also page 16 and 88) [redacted] [redacted]	Pfizer propose a text amendment improve accuracy on page 3, page 16 and page 88. The proposed marketing authorisation for crisaborole is for treatment of mild to moderate atopic dermatitis in patients from 2 years of age with ≤40% body surface area (BSA) affected. Crisaborole can be used twice daily for up to 4 weeks per treatment course. If any signs and/or symptoms persist, or new areas affected with atopic dermatitis appear further treatment course can be used. Crisaborole should be discontinued if signs and/or symptoms persist after 3 consecutive	Accuracy	Page 122 of the company submission (CS) states that "[redacted]" In addition, the CS states on page 123, [redacted]

[redacted]	treatment courses of 4 weeks each or if the signs and/or symptoms worsen during treatment.		[redacted] We therefore do not consider the text to be factually inaccurate based on the information provided by the company.

Issue 3

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Page 4 of the ERG report states: [redacted]	Pfizer proposes the following text amendment [redacted] Furthermore, AN2728-AD-301 and AN2728-AD-302 were not statistically powered for these subgroup analyses and as such, results of post-hoc analyses	The paragraph is misleading since it does not provide context to the results. Notably, that the mild population was smaller, that the trial was not powered to detect significant differences in subgroup analyses by age and disease severity and that in mild patients, the primary outcome was effectively harder to achieve. This context should be provided to allow the reader to draw informed conclusions.	The statement is not factually inaccurate. Page 4 is part of the executive summary and therefore highlights key findings without going into extensive details.

	should be viewed with caution. The primary endpoint, success in ISGA, was defined as an ISGA of "clear" or "almost clear" with at least a 2-grade improvement from Baseline, this means that for a treatment to be considered a success for a "mild" subject, they had to show zero disease in order to achieve "clear" on the ISGA scale. Consequently, it is more difficult to achieve success in a "mild" vs "moderate" subject as the latter only needs to achieve a state of "almost clear".		

Issue 4

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Page 4 of the ERG report states: "The ERG notes that the CS does not provide any evidence for several potential subgroups listed in the final NICE scope including: people with different skin colour; people with atopic dermatitis affecting the hands; and people with atopic dermatitis affecting sensitive	Pfizer proposes that the text be amended to: The ERG notes that the CS does not provide any evidence for several potential subgroups listed in the final NICE scope including: ~~people with different skin colour~~; people with atopic dermatitis affecting the hands; and people with atopic dermatitis affecting sensitive areas (such as the face, neck and flexures)."	Pfizer provided evidence on the subgroup of people with different skin colour in the response to Clarification Question A9.	The text will be revised as follows: "The ERG notes that the CS does not provide any evidence for several potential subgroups listed in the final NICE scope including: people with atopic dermatitis affecting the hands; and people with atopic dermatitis affecting

areas (such as the face, neck and flexures)."			sensitive areas (such as the face, neck and flexures). At the request of the ERG, the company provided data on ISGA success subgrouped by race and ethnicity for the crisaborole trials (clarification response to question A9, Appendix 4, Table 5 and Table 6). However, no NMA is conducted for any of these subgroups as the company state that such data were not consistently reported across comparator trials (CS Table 1)."

Issue 5

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
The summary of critique of clinical effectiveness in Section 1.3 page 5, The overview of NMA results in section 4.4.1 pages 69-71, summary of ERG view on indirect comparisons in	The text of these sections must be amended to reflect the use of FE-RCE-VR as base case NMA model.	The ERG have used a simple random effects NMA which makes no adjustment for differences in vehicle response. Please see rationale below as to why this is not a justifiable choice.	The issue depends on whether the control vehicles associated with the treatments in each study are fundamentally different. If the control vehicles are fundamentally

Section 4.4.7 page 83, and conclusions of indirect comparisons provided in Section 4.6 pages 84-85, are all based the ERG NMA model, which has been selected based on factually inaccurate assessment that both the class effects and vehicle regression are unnecessary. As described in Issue 13 through 23, the correct model to use as base case is the fixed effects, random class effect, vehicle response regression (FE-RCE-VR).		Furthermore, that class effects should be included in order to borrow strength across tacrolimus and pimecrolimus, thus improving robustness. In the “Model Comparison” section on page 80 the authors note that “Differences in DIC values of up to five are generally considered small unless the models leads to different conclusions”. The simple random effects model DIC was 129.2 while FE-RCE-VR was 124.8, which is a difference of 4.4 (CS: Document B, Table 32, page 76). In the pivotal Spiegelhalter 2002 DIC paper1, section 9.2.4 indicate that models that are different by 3-7 points have "considerably less support". In Dias 2018 Section 3.3.2 DIC differences <3 are considered small2. In the Lunn et all 2013 BUGS book a difference of 5-10 was identified as substantial 3, by which a difference of 4.4 is on the edge of being ‘substantial’. Furthermore, DIC and residual deviance consistently favour the vehicle response	different then the network is not connected and the situation is not a baseline risk problem that can be solved using regression on the baseline risk. The statements are not factually inaccurate.

		adjusted models (CS: Document B, Table 32, page 76) so this is not a chance finding due to statistical variation. Statistical support to vehicle response regression is also provided by the large size of the vehicle effect [redacted] These three statistical arguments cannot be ignored when making model selection and thus vehicle response regression must be used. In Section 4.2.2 page 40 and again in Section 4.3.8 page 67 the ERG acknowledge that clinical advice was that ointments will have better response than cream (e.g. pimecrolimus vehicle). As our networks include both pimecrolimus and tacrolimus/crisaborole, it is difficult to say creams have lower response but then not consider adjustment with vehicle response regression.	

		In Section 4.3.8 page 67 the ERG argue that tacrolimus and crisaborole vehicles have some similarities and can't be said to vary systematically; this is based on their both containing white petrolatum and [redacted] and tacrolimus containing propylene carbonate. However, [redacted], it is unknown what proportion of tacrolimus vehicle is white petrolatum. Propylene carbonate is a cyclic carbonate ester derived from propylene glycol. According to the Handbook of Pharmaceutical Excipient 4, propylene carbonate is used mainly as a solvent in topical formulations while propylene glycol contributes in a multifunctional way by acting as a solubiliser, penetration enhancer and humectant[redacted]	

		[redacted] Regardless, the potential for variation due to variation in ingredients, and observed variation in vehicle response (30% and 41% in crisaborole trials compared to 19%, 23%, and 32% in three tacrolimus trials and 12% in one pimecrolimus trial), is the reason vehicle response models should be considered. This variation was supported by both the 2017 Cochrane review and BMJ articles cited by the ERG in Section 4.3.85 6. The strong statistical evidence, repeated above, is justification for using such models. In Section 4.4.6 pages 81-82 and Section 4.4.7 page 83 the ERG have paid insufficient attention to the use of unanchored MAIC as an argument in favour of vehicle-	

		response regression models. The CS did not recommend the unanchored MAIC as an alternative to NMA; the MAIC was a sensitivity analysis to demonstrate the impact of differences in vehicle-response on conclusions. [redacted] This at least supports the use of vehicle-response regression models. In Section 4.3.5 page 60 the ERG note that some potential treatment effect modifiers (such as age and severity) varied between trials. However, they do not connect this to the need for vehicle response regression. As explained in Achana 2013 vehicle/baseline	

		response regression attempts to adjust for imbalance in treatment effect modifiers across trials by assuming their imbalance is represented by differences in vehicle response (which were substantial with 30% and 41% in crisaborole trials compared to 19%, 23%, and 32% in three tacrolimus trials and 12% in one pimecrolimus trial) 7. In Section 4.4.4 page 80, the ERG note “The interpretation of the relative treatment effects estimated by the company are as the relative treatment effects at the population mean response to vehicle across the six studies included in the analysis. It is not clear whether this estimate of treatment effect is relevant for decision-making.” However, the population of the random effects NMA is a mixture over three tacrolimus trials, two crisaborole trials, and a pimecrolimus trial; it is similarly unclear that estimates of treatment effect from this mixed population are relevant for decision-making.	

		In section 4.4.4 page 75, the ERG state “Information on the way class was defined was also missing”. This is not correct as CS Appendix D2 includes a section titled "Class effect models" which describes in detail the nature of the class effects models. In Section 4.4.4. page 75 the ERG state “It is not clear what the motivation is for fitting a class effects model.” The CS Appendix D2 answers that “We explored class effect models for our primary outcome of ISGA/IGA 0/1 (12) in order to borrow strength across treatments in the same class.” This follows Section 8.6.2 of Dias 20182 which explains that when treatments fall into classes with similar modes of action, it is reasonable to assume a relationship between effects of treatments in the same class, thus borrowing strength across treatments. The motivation was to improve the scientific robustness by reducing uncertainty in treatment effect estimates.	

		However, CS: Appendix D, Section 2 Table D38 on page 246 and Table D26 page 248 presented results of a sensitivity analysis with fixed treatment effects, no class effect, and with vehicle response regression; the results were very similar to the base case with random class effect.	

Issue 6

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
In Section 1.3 page 5 the ERG state "For these reasons, the ERG considers that caution should be exercised when making inferences based on the results of the MAIC and prefers the simple random effects NMA over the MAIC." The text is used again in Section 4.6 page 85 and in Section 5.5 page 115.	All three instances should be amended to: "For these reasons, the ERG considers that caution should be exercised when making inferences based on the results of the MAIC and prefers the simple random effects NMA over the MAIC. The MAIC should be viewed as a sensitivity analysis of the impact of differences vehicle in vehicle response on the results of the NMA. Given the impact on results, vehicle response regression is likely necessary in NMA"	The CS did not recommend the unanchored MAIC as an alternative to NMA; the MAIC was a sensitivity analysis to demonstrate the impact of differences in vehicle-response on conclusions. In the unanchored MAIC, removing vehicle altogether had a substantial impact on both point estimates and 95% uncertainty (credible interval in NMA, confidence interval in MAIC) intervals (Compare results of NMA (even with vehicle response regression) in CS:	The statement is not factually inaccurate.

		Document B Figure 19 page 78 to results from MAIC in CS: Document B Figure 30 page 95). This at least supports the use of vehicle-response regression models.	

Issue 7

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Page 9 the following text is misleading The NMA is complicated by the company's claim that the vehicles for the different topical treatments have different compositions and therefore different response rates.	Pfizer proposes the following text amendment: 'The NMA is complicated by the fact that different topical treatments have different compositions and have reported different response rates.'	For accuracy. The vehicles in the NMA included studies do have different ingredients (this is not a company claim). The vehicle of the included studies also report different response rates (this is not company claim).	The text will be revised to: "The NMA is complicated by the fact that the vehicles for the different topical treatments have different compositions, which the company claims may lead to different response rates." We disagree with stating that the vehicles do have different response rates, because it is possible that the difference in vehicle response between trials may be due to other factors such as differing trial populations, as well as random variation (random variation is suggested by the company as the reason for the different vehicle response rates between the two crisaborole trials; clarification response to question A9).

Issue 8

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
p. 9 the following text is misleading The MAIC only provides inferences relative to the populations defined by the comparator studies, and it is unclear whether all prognostic factors and treatment effect modifiers have been accounted for.	Pfizer propose a text amendment: The MAIC only provides inferences relative to the populations defined by the comparator studies. Whilst, the company undertook a systematic literature review and a separate patient level analysis of crisborole trial data to identify prognostic factors as well as enlisting an expert clinical adviser to review included variables, it is nevertheless feasible that not all prognostic factors and treatment effect modifiers have been accounted for.	For accuracy. There is no acknowledgement of the extensive efforts which were undertaken to identify prognostic variables. This context should be provided to allow the reader to draw informed conclusions.	The statement is not factually inaccurate.

Issue 9

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
The page number on p. 10 of the report referencing the following text is incorrect: ”The most common day-to-day symptoms are dry skin, cracked or raw skin and itching (see CS, Figure 1[1]). These symptoms are associated with a psychological and psychosocial burden as pruritus (itching) can lead to	Pfizer propose the page reference number be changed to (CS, p. 22) to accurately reflect the location of this text in the CS	For accuracy	This minor discrepancy does not need correcting.

sleep disruption and the presence of visible skin lesions can affect self-esteem and social interactions (CS, p. 21)"			

Issue 10

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Page 16 of the report states: The ERG notes that this is inconsistent with the stepped care approach from NICE CG57 which shows systematic therapy and phototherapy as later treatment options for severe AD in children (see Table 1)[4].	Pfizer propose amending this text to note this stepped approach is not complete as the CS was written to focus on mild to moderate AD and does not include treatment options for individuals with severe AD as per the decision problem.	For accuracy	This is not a factual inaccuracy.

Issue 11

Description of problem	Description of proposed amendment	Justification for amendment	
On page 19 of the report the following sentence has an incorrectly calculated value: [redacted]	Pfizer propose adjusting the value to [redacted] The calculation is derived as follows: 133 patients with BSA>40% out of a total of 1415 patients = 9.39%	For accuracy	This has been corrected to [redacted]

[redacted]			

Issue 12

Description of problem	Description of proposed amendment	Justification for amendment	
Page 26 of the report states: The Global Resource for Eczema Trials (GREAT) was searched in March 2019 for systematic reviews and randomised controlled trials from inception until 2017	Pfizer propose to adjust the language to state: The Global Resource for Eczema Trials (GREAT) was searched in March 2019 for systematic reviews and randomised controlled trials from inception until September 2017	For completeness	This is not a factual inaccuracy but the additional detail has been added.

Issue 13

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Page 26 of the report states: According to the Cochrane Handbook,[19] it is best practice not to apply language	Pfizer response: The updated guidance from the recently released Cochrane Handbook is as follows: "One particular study focused on the contribution of unpublished studies, including dissertations, and studies in languages other than English, to the results	For accuracy	This is not a factual inaccuracy. Language limits applied during literature searching is not the same as excluding English language articles during study selection. The implications of the former is that, the total number of non-English articles that might be eligible (regardless of impact on findings) are unknown

restrictions in the search strategy to prevent the risk of language bias.	of meta-analyses in reviews relevant to children (Hartling et al 2017). They found that, in their sample, unpublished studies and studies in languages other than English rarely had any impact on the results and conclusions of the review. They did, however, concede that inclusion of these study types may have an impact in situations where there are few relevant studies, or where there are 'questionable vested interests' in the published literature." "Evidence indicates that excluding non-English studies does not change the conclusions of most systematic reviews (Morrison et al 2012, Jiao et al 2013, Hartling et al 2017), although exceptions have been observed for complementary and alternative medicine (Moher et al 2003, Pham et al 2005, Wu et al 2013)." Carol Lefebvre, Julie Glanville, Simon Briscoe, Anne Littlewood, Chris Marshall, Maria-Inti Metzendorf, Anna Noel-Storr, Tamara Rader, Farhad Shokraneh, James Thomas, L. Susan Wieland; on behalf of the Cochrane Information Retrieval Methods Group. Chapter 4: Searching for and selecting studies. Cochrane Handbook for Systematic Reviews of Interventions (Version 6.0). Editors: Julian Higgins and James Thomas. URL: https://training.cochrane.org/handbook/current/chapter-04. Accessed: 11/27/ 2019.		The most recent version of the Cochrane Handbook states that "The best way [to balance the thoroughness of the search with efficiency in the use of time and funds] is to be aware of, and try to minimize, the biases such as publication bias and language bias that can result from restricting searches in different ways." Furthermore, a current recommendation is that, "If searches are restricted by publication status or by language of publication, there is a possibility of publication bias, or language bias (whereby the language of publication is selected in a way that depends on the findings of the study), or both." Lefebvre C, Glanville J, Briscoe S, Littlewood A, Marshall C, Metzendorf M-I, Noel-Storr A, Rader T, Shokraneh F, Thomas J, Wieland LS. Chapter 4: Searching for and selecting studies. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019). Cochrane, 2019. Available from

			www.training.cochrane.org/handbook. Accessed 16 December 2019

Issue 14

Description of problem	Description of proposed amendment	Justification for amendment	
On page 31 of the report the following text has an incorrect cross-reference to the CS: The aim of the RCTs was to [redacted] (CS Table 16: Summary of statistical analyses, p. 47 to 48).	Pfizer propose correcting the page cross-reference to: (CS Table 16: Summary of statistical analyses, p. 47 to 49)	For accuracy	I think the cross-reference is sufficiently accurately to allow the reader to identify the relevant table. No change needed.

Issue 15

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
p.31 the following text is incorrect: Clinical visits for assessments were scheduled on day 2, 6, 8, 15, 22, and 29.1	Pfizer proposes the following text amendment: Clinical visits for assessments were scheduled on Screening, Baseline/Day 1, Day 8, 15, 22 and 29.	For accuracy	This has been corrected to “Baseline/day1 and days 8, 15, 22, and 29”.

Issue 16

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
p. 33 Table 5 Study characteristics of crisaborole studies (adapted from CS, Tables 9 to 13). This table incorrectly states on the row labelled 'Assessments': Baseline, day 2, 6, 8, 15, 22, and 29	Pfizer proposes the following text amendment: Baseline/Day 1, Day 8, 15, 22 and 29.	Accuracy	This has been corrected to "Baseline/day1 and days 8, 15, 22, and 29".

Issue 17

Description of problem	Description of proposed amendment	Justification for amendment	
p.33 Table 5. Primary outcome(s) has not been accurately described.	Pfizer propose the following text amendment: ISGA success: The proportion of patients achieving an ISGA score of 0 (clear) or 1 (almost clear) at Day 29 with >=2-point improvement from baseline	Accuracy. The primary outcome(s) has not been accurately described and does not include a key criteria of ISGA success, (at least a 2 point improvement from baseline).	This has been corrected.

Issue 18

Description of problem	Description of proposed amendment	Justification for amendment	
Page 33 of the ERG report states in Table 5: [redacted]	Pfizer proposes the following amended text: [redacted]	For factual accuracy	This has been corrected in the erratum

Issue 19

Description of problem	Description of proposed amendment	Justification for amendment	ERG Response
Page 33 of the ERG report states in Table 5: ISGA=0, patient enters 'off treatment period for 28 days: Observation, only	Pfizer proposes the following amended text: ISGA of 0 or 1, patient enters 'off treatment' period for 28 days: Observation, only	For factual accuracy	This has been corrected.

Issue 20

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
p.35 The intention-to-treat (ITT)	Pfizer propose the following text amendment	Accuracy.	This has been corrected.

population, used in analysing efficacy outcomes, was defined as all those who were randomised and received the study treatment ….	The intention-to-treat (ITT) population, used in analysing efficacy outcomes, was defined as all those who were randomised and dispensed the study treatment, regardless of dropping out of the study.		

Issue 21

Description of problem	Description of proposed amendment	Justification for amendment	
In Section 4.2.4 page 40 the ERG states "Clinical advisors to the ERG indicated that ointment-based vehicles tend to demonstrate greater responses than cream-based treatments."	This should be amended to "Clinical advisors to the ERG indicated that ointment-based vehicles tend to demonstrate greater responses than cream-based treatments; adjustment is therefore necessary in indirect comparison when comparing cream and ointment based interventions."	This important difference between creams (e.g. pimecrolimus) and ointments (e.g. tacrolimus and crisaborole) must be somehow addressed when conducting network meta-analysis. In the CS vehicle response adjustment was conducted.	This is not a matter of factual accuracy. We are reporting what our clinical advisors said to us and therefore we cannot simply amend it with an additional statement that they did not make.

Issue 22

Description of problem	Description of proposed amendment	Justification for amendment	
Page 40 of the report states two incorrect percentages: Patients with moderate AD (61%) made up a larger percentage of the study population compared to those	Pfizer propose changing the text to accurately reflect the percentages on page 45 of the CS to the following: Patients with moderate AD (62%) made up	For accuracy	If you use the numbers in Table 14 (page 44 to 45 of the CS) to back calculate the numbers of patients with mild and moderate AD in each arm and then calculate

with mild AD (39%). Mean baseline % BSA affected, pruritus score, condition-specific HRQoL scores were comparable across treatment group for AD-301 and AD-302.	a larger percentage of the study population compared to those with mild AD (38%). Mean baseline % BSA affected, pruritus score, condition-specific HRQoL scores were comparable across treatment group for AD-301 and AD-302.		the percentage overall then you get 61.4% and 38.5%. Therefore, the ERG believes they have presented the data accurately to the best of their ability given the data that was provided.

Issue 23

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Table 9 (p.42) should be marked commercial in confidence	Pfizer propose the table should be in turquoise highlight with all text underlined	These data are CiC as indicated in the CS	The data in Table 9 of the ERG report are based on data in Tables 14 and 15 of the CS which were not marked as CIC in the CS. This table has been marked as CIC as requested but the ERG queries whether the whole table needs marking CIC given that much of the data on the base-line characteristics of patients in studies AD-301 and AD-302 have been published by Paller et al. in 2016 (ref 12

			of CS).

Issue 24

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Page 44 of the ERG report states: "A similar trend was observed when outcomes were assessed according to AD severity in the pooled AD-301 and AD-32 population (Table 10); however, there was no statistically significant difference in the group with mild AD when the crisaborole group was compared with the vehicle group ([redacted]."	Pfizer proposes the following amendment: "A similar trend was observed when outcomes were assessed according to AD severity in the pooled AD-301 and **AD-302** population (Table 10). There was no statistically significant difference in the group with mild AD when the crisaborole group was compared with the vehicle group [redacted] It is noted, however, that that crisaborole studies were not powered to detect a significant result in subgroup analyses by disease severity."	For factual accuracy and to amend typo.	We have amended the typo from AD-32 to AD-302. We have also added: "It is noted, however, that that crisaborole studies were not powered to detect a significant result in subgroup analyses by disease severity."

Issue 25

Description of problem	Description of proposed amendment	Justification for amendment

Issue 26

Description of problem	Description of proposed amendment	Justification for amendment	
Page 48 of the report is missing a symbol for quantification in the following text: The minimal clinically important difference (MCID) for a change from baseline is 3.3 for the DLQI and ≥ 2.5 for the CDLQI.	Pfizer propose amending the text to the include a ≥ symbol: The minimal clinically important difference (MCID) for a change from baseline is ≥3.3 for the DLQI and ≥ 2.5 for the CDLQI.	For accuracy	We have corrected this typo

Issue 27

Description of problem	Description of proposed amendment	Justification for amendment	
Page 49 of the report has an incorrect page cross-reference with the CS: A summary of HRQoL scores are presented in CS, Tables 19, 20 and 21 (p. 56).	Pfizer propose amending the text to read: A summary of HRQoL scores are presented in CS, Tables 19, 20 and 21 (p. 57).	For accuracy	We do not believe it is necessary to correct page numbers in references, particularly when referring to named items such as Tables which can be easily located

			by the reader.

Issue 28

Description of problem	Description of proposed amendment	Justification for amendment	
p.51 Table 12: NCT04040192 lists “flare reduction” as a primary outcome	Pfizer propose amending to “flare free maintenance” to accurately reflect the outcome listed on clinicaltrials.gov	For accuracy	This has been corrected.

Issue 29

Description of problem	Description of proposed amendment	Justification for amendment	
p.51 Table 12: NCT03832010 lists: a) vehicle b) emollient as the comparators	Pfizer propose amending the comparator list to include hydrocortisone and triamcinolone as these comparators are also listed on clinicaltrials.gov	For accuracy	All study arms receive TCS (hydrocortisone and triamcinolone) in addition to the treatment under study, since the outcome is TCS reduction. The table has been amended as follows: Intervention: Crisaborole 2% + TCS Comparators: a) Vehicle + TCS b) Emollient + TCS

Issue 30

Description of problem	Description of proposed amendment	Justification for amendment	
Page 53 of the ERG report states: [redacted] "CDLRI" should read "CDLQI"	Pfizer proposes the following amended text: [redacted]	For factual accuracy	Typo has been corrected.

Issue 31

Description of problem	Description of proposed amendment	Justification for amendment	
In Section 4.3.5 page 60 the ERG state "Overall, the ERG consider that the trials included in the NMA were broadly similar in that they considered a population with ≥80% mild to moderate AD. However, some potential treatment effect modifiers (such as age and severity)	This should be changed to "Overall, the ERG consider that the trials included in the NMA were broadly similar in that they considered a population with ≥80% mild to moderate AD. Nevertheless, some potential treatment effect modifiers (such as age and severity) varied between trials and given these differences in effect modifiers, vehicle response regression in	As explained in Achana 2013, vehicle/baseline response regression attempts to adjust for imbalance in treatment effect modifiers across trials by assuming their imbalance is represented by differences in vehicle response (which were substantial with 30% and 41% in crisaborole trials compared to	This is not a matter of factual accuracy.

varied between trials.”	NMA may be necessary.”	19%, 23%, and 32% in three tacrolimus trials and 12% in one pimecrolimus trial). The difference in age and severity is a key reason for employing vehicle response regression and the point needs to be made when discussing differences across trials. Achana FA, Cooper NJ, Dias S, et al. Extending methods for investigating the relationship between treatment effect and baseline risk from pairwise meta-analysis to network meta-analysis. Stat Med 2013;32(5):752-71. doi: 10.1002/sim.5539	

Issue 32

Description of problem	**Description of proposed amendment**	**Justification for amendment**	
In Section 4.3.8 page 67 the ERG states “Clinical advisors to the ERG noted that ointments (such as crisaborole and tacrolimus) may give greater response rates than creams (such as pimecrolimus).”	This should be amended to “Clinical advisors to the ERG noted that ointments (such as crisaborole and tacrolimus) may give greater response rates than creams (such as pimecrolimus); adjustment is therefore necessary in indirect comparison when comparing	This important difference between creams (e.g. pimecrolimus) and ointments (e.g. tacrolimus and crisaborole) must be somehow addressed when conducting network meta-analysis. In the CS vehicle	Again, as per our response to issue 21, we cannot amend what the clinical advisors stated to include additional statements that they did not make to us. This is not a matter of

	cream and ointment based interventions.”	response adjustment was conducted.	factual accuracy.

Issue 33

Description of problem	Description of proposed amendment	Justification for amendment	
In Section 4.3.8 page 67 the ERG states “[redacted].”	This should be amended to “[redacted].”	There is no publicly available information on the contents of tacrolimus base ointment. It would be speculation to say that it has the same percentage, and therefore same efficacy, as crisaborole ointment. This is an important point as it justifies the need for vehicle response regression.	This is not a matter of factual accuracy as [redacted]. However, we have amended the text to say that “[redacted].”

Issue 34

Description of problem	Description of proposed amendment	Justification for amendment	
In Section 4.3.8 page 67 the ERG states “The ERG notes that while tacrolimus doesn't contain propylene glycol, it does contain propylene carbonate. The ERG is unsure whether the crisaborole and tacrolimus	This should be amended to “The ERG notes that while tacrolimus doesn't contain propylene glycol, it does contain propylene carbonate. Propylene carbonate is a cyclic carbonate ester derived from propylene glycol. According to the Handbook of Pharmaceutical Excipient 4, propylene carbonate is used mainly as a	The differences between propylene glycol and propylene carbonate are public knowledge and described in the cited textbook. This is an important point as it justifies the need for vehicle	What we have stated is not factually inaccurate. If the company wished to provide a detailed discussion regarding the composition of the vehicle ointments this should have been provided in their

vehicles would be expected to differ substantially in their emollient effects, based on their composition.”	solvent in topical formulations while propylene glycol contributes in a multifunctional way by acting as a solubiliser, penetration enhancer and humectant. Propylene glycol is a well-known humectant since it contains two hydroxyl groups which attract water molecules by forming hydrogen bonding. Propylene carbonate does not contain any hydroxyl group and may only form weak hydrogen bonding with water involving mainly the carbonyl oxygen. Based on this, propylene glycol is expected to provide better emollient effect than propylene carbonate. The ERG therefore expects the tacrolimus vehicle to have lesser emollient benefit than the crisaborole vehicle.” With reference to textbook Sheskey P, Cook W, Cable C. Handbook of Pharmaceutical Excipients: Pharmaceutical Press 2017.	response regression.	submission to allow the ERG to critique this information.

Issue 35

Description of problem	**Description of proposed amendment**	**Justification for amendment**	
p.69 Table 22. p-value for tacrolimus is an implausible value	Please review p-value	Accuracy.	The ERG is unable to identify the issue.

Issue 36

Description of problem	Description of proposed amendment	Justification for amendment	
In section 4.4.4 page 75, the ERG state “Information on the way class was defined was also missing”.	This should be changed to “Information on the way class was defined was given in CS: Appendix D, Section 2 page 198.”	This information was provided and it is factually incorrect to state otherwise.	This is not factually inaccurate.

Issue 37

Description of problem	Description of proposed amendment	Justification for amendment	
In Section 4.4.4. page 75 the ERG state “It is not clear what the motivation is for fitting a class effects model.”	This should be changed to “As stated in CS: Appendix D, Section 2 page 198, class effects models were explored for the key outcome ISGA/IGA 0/1 in order to borrow strength across treatments in the same class (Dias et al 2018).” Reference should be provided to Section 8.6.2 of the textbook Dias S, Ades A, Welton N, et al. Network Meta-Analysis for Decision-Making: Wiley 2018.	This follows Section 8.6.2 of Dias 20182 which explains that when treatments fall into classes with similar modes of action, it is reasonable to assume a relationship between effects of treatments in the same class, thus borrowing strength across treatments. The motivation was to improve the scientific robustness by reducing uncertainty in treatment effect estimates. However, CS: Appendix D, Section 2 Table D38 on page 246 and Table D26 page 248 presented results of a sensitivity analysis with fixed treatment effects, no class effect, and with vehicle response	This is not factually inaccurate.

		regression; the results were very similar to the base case with random class effect. Dias S, Ades A, Welton N, et al. Network Meta-Analysis for Decision-Making: Wiley 2018.	

Issue 38

Description of problem	**Description of proposed amendment**	**Justification for amendment**	
In Section 4.4.4 page 75 the ERG state "The company did not report the 95% credible interval for the between treatment within class standard deviation".	This should be changed to "The company provided 95% credible intervals for the between treatment within class standard deviation in The discussion on pages 75 and 76 needs to be revised to reflect this information.	Company provided these credible intervals in response to A22 of the clarificatory questions.	The text has been amended as follows: "In response to clarification question A22, the company provided estimates and 95% credible intervals for the between-treatment within class standard deviations. The means of the posterior distributions are indicative of moderate heterogeneity and the uncertainty about them are indicative of extreme heterogeneity between treatments within class. Rather than borrowing strength and suggesting

			similarity of treatment effects within class, the prior information used for the variance parameters has resulted in highly uncertain posterior distributions.”

Issue 39

Description of problem	**Description of proposed amendment**	**Justification for amendment**	
In the “Model Comparison” section on page 80 the authors note that “Differences in DIC values of up to five are generally considered small unless the models lead to different conclusions”.	Recommend changing to “It has been suggested that differences in DIC value of five are important, while those less than 3 (not 5) are not important, provided that the conclusions are robust to choice of model (Dias et al. 2018)” With reference to the Section 3.3.2 of the textbook: Dias S, Ades A, Welton N, et al. Network Meta-Analysis for Decision-Making: Wiley 2018.	The simple random effects model DIC was 129.2 while FE-RCE-VR (fixed treatment effects, random class effect, vehicle response regression) was 124.8, which is a difference of 4.4 (CS: Document B, Table 32, page 76); this is potentially an important difference between the models and favours the the CS base case of FE-RCE-VR. In the pivotal Spiegelhalter 2002 DIC paper, section 9.2.4 indicate that models that are different by 3-7 points have "considerably less support". In Dias 2018 Section 3.3.2 DIC differences <3 are considered small. In the Lunn et all 2013 BUGS book a difference of 5-10 was identified as substantial 3, by which a	This is not factually inaccurate.

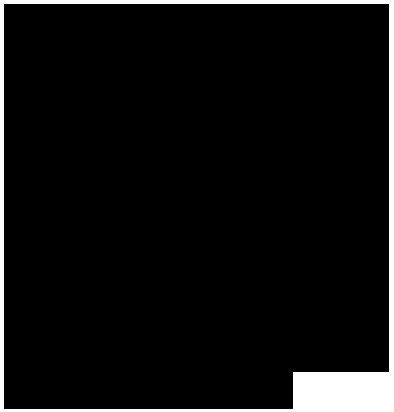
		difference of 4.4 is on the edge of being ‘substantial’. Furthermore, DIC and residual deviance consistently favour the vehicle response adjusted models (CS: Document B, Table 32, page 76) so this is not a chance finding due to statistical variation. Statistical support to vehicle response regression is also provided by the large size of the vehicle effect (regression coefficient of -0.847 on log hazard ratio scale) and the 95% credible interval (-1.229, -0.417) which clearly excluded zero (no effect) (CS: Document B, Table 37, page 81). These three statistical arguments cannot be ignored when making model selection and thus vehicle response regression must be used. References Spiegelhalter DJ, Best NG, Carlin BP, et al. Bayesian measures of model complexity and fit. JRStatist Soc B 2002;64(4):583-639.	

		Dias S, Ades A, Welton N, et al. Network Meta-Analysis for Decision-Making: Wiley 2018. Lunn D, Jackson C, Best N, et al. The BUGS book : a practical introduction to Bayesian analysis. Boca Raton ; London: CRC Press 2013.	

Issue 40

Description of problem	**Description of proposed amendment**	**Justification for amendment**	
In Section 4.4.4 page 80, the ERG state "The interpretation of the relative treatment effects estimated by the company are as the relative treatment effects at the population mean response to vehicle across the six studies included in the analysis. It is not clear whether this estimate of treatment effect is relevant for decision-making."	Text should be added following this paragraph "However, the population of a random effects NMA with no vehicle response regression is also a mixture over three tacrolimus trials, two crisaborole trials, and a pimecrolimus trial. It is similarly unclear that estimates of treatment effect from this mixed population are relevant for decision-making."	This is the basic nature of NMA; it mixes populations and distributions of effect modifiers and it is not known to which population NMA results are relevant. As both vehicle adjusted and vehicle-unadjusted models have questions about the relevance of their populations, this should be not be used as a criteria for selecting vehicle-unadjusted models.	The company appears to have misunderstood the point being made, which relates to the company's estimate of the treatment effects from their baseline-adjusted model. This is not factually inaccurate.

Issue 41

Description of problem	Description of proposed amendment	Justification for amendment	
Page 83 of the ERG report states: 	Pfizer proposes the following text amendment 	For accuracy. This statement is misleading since there is no acknowledgement of the extensive efforts which were undertaken to identify prognostic variables. This context should be provided to allow the reader to draw informed conclusions.	This is not factually inaccurate.

Issue 42

Description of problem	Description of proposed amendment	Justification for amendment	
In Section 4.6 page 85 the ERG write “Although the MAIC circumvents the need to adjust for vehicle response, it provides inferences relative to the populations defined by the comparator studies, which may not be useful for decision	This should be amended to ““However, the population of NMA is also a mixture over three tacrolimus trials, two crisaborole trials, and a pimecrolimus trial. It is similarly unclear that estimates of treatment effect from this mixed population are relevant for decision-making.”	This is the basic nature of NMA; it mixes populations and distributions of effect modifiers and it is not known to which population NMA results are relevant. It is misleading to point out disadvantages of target populations in MAIC without	This is not factually inaccurate.

making."		noting the same issue in NMA.	

Issue 43

Description of problem	Description of proposed amendment	Justification for amendment	
Page 104 of the ERG report states: "The ERG identified one error in the updated base case. The revised implementation of dermatologist visits for patients on subsequent therapies introduced a new error in that the costs were applied to patients both on and off treatment." This is repeated on page 112.	Pfizer propose that this sentence be removed.	This is not an error in the model, rather this assumption was made to reflect the fact that, as highlighted by the ERG, patients may continue to receive treatment beyond achieving disease control. These patients would continue to receive specialist care to monitor their treatment. Applying these costs only to patients in the uncontrolled disease state will likely underestimate the cost of monitoring.	The company's claim here is incorrect. They have introduced column R in their updated model version to account for controlled patients and named it 'Subsequent therapy – Off treatment'. Furthermore, according to the model, the only treatment these patients are receiving is emollients. Therefore, the model correctly applies GP visits to this cohort, but dermatologists visits were also unnecessarily applied to those with controlled disease. Although the company claims that these patients would continue to receive specialist care to monitor their treatment beyond achieving disease control, the ERG does not believe that it is realistic to assume

			that dermatologist appointments would continue for the rest of the patient's lives just because they required a subsequent therapy once.

Issue 44

Description of problem	Description of proposed amendment	Justification for amendment	
Page 108 states: "Subsequent therapy was also introduced as a 'basket' of different therapeutic options including systemic therapy and phototherapy (subsequent use of TCIs was also allowed in the scenario analysis for the TCS naïve population). However, a patient is assumed to be on one of these subsequent therapies until control of symptoms, resolution of AD (in children) or death, and is assumed never to change between subsequent treatments.	Pfizer propose that this statement is amended to read: "Subsequent therapy was also introduced as a 'basket' of different therapeutic options including systemic therapy and phototherapy (subsequent use of TCIs was also allowed in the scenario analysis for the TCS naïve population). **It is assumed that individuals would remain on some form of therapy while symptoms persist and that the proportion of the population receiving any specific subsequent therapy is fixed at any given time.**	As the model does not track individual patients, assuming a fixed proportion of patients on each subsequent therapy is not equivalent to assuming that patients may not switch between therapies.	The text highlighted by the company has been amended as suggested to remove the focus on individual patients but additional text has been added to the following paragraph to make it clear that the use of a 'basket' approach to model subsequent therapies is not compatible with the stepped care approach recommended by NICE. The additional text is as follows, "The ERG does not believe that the company's approach of modelling subsequent therapy as a 'basket' of different therapeutic options is

			consistent with this stepped approach, as the model does not capture the need for a proportion of patients to try phototherapy before moving on to systemic treatment."

Issue 45

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Pages 109-110 of the ERG report state: "This suggests that the model may not be capturing the real world experience of patients who may fluctuate between having mild to moderate AD and severe AD. It is difficult to say what the impact would be of correcting this but it may marginally favour crisaborole given that a slightly higher proportion of patients in the crisaborole arm were in a severe state at the end of the trials."	Pfizer propose that this text be amended to say: This suggests that the model may not be capturing the real world experience of patients who may fluctuate between having mild to moderate AD and severe AD. It is difficult to say what the impact would be of correcting this but it may be [redacted]	Table 11.1.1.3 (Table 17 in the clarification responses) presents the proportion of patients with each ISGA score at day 29 in AD-301 and AD-302 by treatment arm. This shows that [redacted] of patients in the crisaborole arm had ISGA score 4, compared to [redacted] of patients in the vehicle arm, i.e. a lower proportion of patients were in the severe state.	Sorry, the company is correct. The ERG has amended the text to say, "It is difficult to say what the impact would be of correcting this but it may be [redacted]

Issue 46

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Page 113 of the report states: "Table 32 presents the duration of treatment for each therapeutic agent, and the corresponding 4-week response probability based on an overall response of 0.7 for all systemic therapy/phototherapy. The 4-week response probability was adjusted then to reflect time to response per treatment."	Pfizer request that this be updated to include the reference for the time to response with each therapy.	Without references it is difficult to compare the suitability of these times with other sources.	Some of the stated times to achieve response in Table 32 are based on assumptions or expert opinion but we have added references to Table 32 where the duration is based on published evidence.

Issue 47

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Page 113 of the ERG report states: "One of the scenarios explored considering systemic treatments to be taken continuously and phototherapy for three months. However, phototherapy costs were not adjusted for three months."	Pfizer request that these sentences be removed.	In the submitted cost-effectiveness model for this scenario ('Crisaborole AD CEM v4.4 scenario 5.2') in 'Cost data'!G101 the cost of phototherapy is adjusted to reflect 3 months of treatment.	In the original model provided by the company, the reference cost of £93 for phototherapy was applied to the proportion of patients in the uncontrolled subsequent therapy states receiving phototherapy each cycle. This suggests that £93 is the cost of phototherapy per cycle (i.e. per four weeks).

			In company scenario 5.2, the duration of phototherapy is assumed to be 3 months. In the model version named 'Crisaborole AD CEM v4.4 scenario 5.2', this cost of £93 is then multiplied by 3.26 to reflect the number of cycles in 3 months ([3/12]*[52/4]) to give a cost per course of phototherapy of £303.29. However, this cost is then applied repeatedly each cycle. Therefore, the ERG believes that this is an error in company scenario 5.2 as the costs are first multiplied by the number of cycles in a course and then applied every cycle. To correct this error, in those scenarios where the cost of phototherapy is explicitly modelled as 3 months (ERG scenarios 7 and 14), the ERG has applied the cost of £303.29 once only to patients starting subsequent therapy. We accept that the text on

			page 113 is ambiguous so we have amended this to say, “One of the scenarios explored considering systemic treatments to be taken continuously and phototherapy for three months. In this scenario, the phototherapy costs were adjusted such that the reference cost for phototherapy (£93) was multiplied by 3.26, to reflect 3 months of treatment, but this cost of £303 (=£93 x 3.26) was then applied repeatedly every cycle, which the ERG did not believe was correct. Therefore, in those scenarios where the duration of phototherapy is explicitly modelled as being 3 months (scenarios 7 and 14 of the ERG’s exploratory analyses), the ERG has applied the cost of £303 once only to those entering the subsequent treatment health state.”

Issue 48

Description of problem	Description of proposed amendment	Justification for amendment	
Page 115 of the ERG report states: [redacted] ."	Pfizer proposes the following amended text: [redacted]	For factual accuracy	This is not factually inaccurate.

Issue 49

Description of problem	Description of proposed amendment	Justification for amendment	
Page 116 of the ERG report states: [redacted]	Pfizer propose this text be amended to say: [redacted]	This was correct at the time of the company submission, however the cost of tacrolimus 0.1% has now risen to £39.90 and the estimates the ERG have provided are not accurate	The ERG used the cost for tacrolimus provided in the CS which the company states was correct at the time of the company submission. However, the ERG has added a few

[redacted]	[redacted]	at the time of writing.	sentences to make it clear that the cost of tacrolimus 0.1% has changed (it reduced to £38.46 in Oct 2019 NHS Drug Tariff) and that the results are sensitive to changes in the cost of tacrolimus 0.1%. The ERG has also removed the prices from the exec summary text on page 9 so that the statement simply says that the cost differences are small in absolute terms.

Issue 50

Description of problem	Description of proposed amendment	Justification for amendment	
Page 116 of the ERG report states: "The other changes made to the company's base case assumptions had a fairly limited impact on the incremental costs and QALYs except in adults with moderate AD, where the ERG's base case (scenario 6) and the exploratory analyses	Pfizer prose this text be amended to say: "The other changes made to the company's base case assumptions had a fairly limited impact on the incremental costs and QALYs. ~~except in adults with moderate AD, where the ERG's base case (scenario 6) and the exploratory analyses considering alternative assumptions for subsequent therapy (scenario 7) resulted in crisaborole having ICERs of [redacted] respectively when compared with~~	The price of tacrolimus 0.1% has increased to £39.90 since the original submission. Re-running the ERG base-case and scenario 7 for moderate adults with the new price results in crisaborole being dominant in both cases.	The ERG used the cost for tacrolimus provided in the CS which the company states was correct at the time of the company submission and does not consider it necessary to re-run all the results given the small change in price. However, the ERG has added some text on page

considering alternative assumptions for subsequent therapy (scenario 7) resulted in crisaborole having ICERs of [redacted] respectively when compared with tacrolimus 0.1% (crisaborole still dominated tacrolimus 0.03% in these scenarios)."	~~tacrolimus 0.1% (crisaborole still dominated tacrolimus 0.03% in these scenarios)."~~		116 to make it clear that the cost of tacrolimus 0.1% has changed and that the results are sensitive to changes in the cost of tacrolimus 0.1%.

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Draft technical report

Crisaborole for treating mild to moderate atopic dermatitis in people aged 2 years and older

This document is the draft technical report for this appraisal. It has been prepared by the technical team with input from the lead team and chair of the appraisal committee.

The technical report and stakeholder's responses to it are used by the appraisal committee to help it make decisions at the appraisal committee meeting. Usually, only unresolved or uncertain key issues will be discussed at the appraisal committee meeting.

The technical report includes:

- topic background based on the company's submission
- a commentary on the evidence received and written statements
- technical judgements on the evidence by the technical team
- reflections on NICE's structured decision-making framework.

This report is based on:

- the evidence and views submitted by the company, consultees and their nominated clinical experts and patient experts and
- the evidence review group (ERG) report.

The technical report should be read with the full supporting documents for this appraisal.

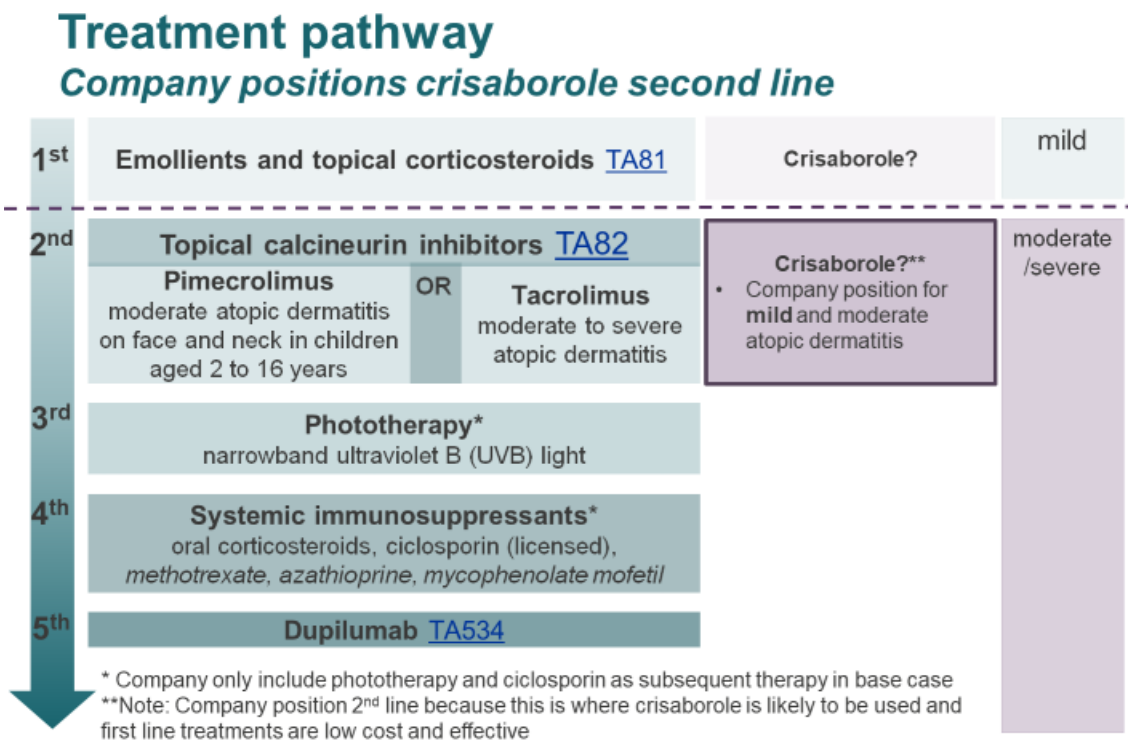
Issue date: January 2020

1. Topic background

1.1 Disease background

Atopic dermatitis (also called atopic eczema)
• Chronic, inflammatory, persistent, relapsing or recurring, immune-mediated skin condition. • Skin may be red/inflamed, thickened/leathery and dry with scaly plaques, bleeding, oozing, cracking, flaking and itching (pruritus) • Can start at any age, onset peaks at infancy
Epidemiology
• Around 1 in 5 children and 1 in 12 adults have atopic dermatitis • Most cases are mild   ○ Company: Severity in children (up to 10 years of age) was estimated to be: 80% mild, 18% moderate and 2% severe
Definition of severity
• Large number of instruments to assess severity such as EASI, POEM, SCORAD • No NICE clinical guideline in adults   ○ CG57 (Atopic eczema in under 12s) recommends a holistic approach considering severity and quality of life • Single measurement may over- or under-estimate severity because of relapsing-remitting nature of condition
Abbreviations: EASI, eczema and severity index; ISGA, Investigator's Static Global Assessment; POEM, patent-oriented eczema measure; SCORAD, scoring atopic dermatitis.

1.2 Treatment pathway



1.3 Technology

Marketing authorisation	Crisaborole is indicated for treatment of mild to moderate atopic dermatitis in patients from 2 years of age with ≤40% body surface area (BSA) affected
Administration	• Crisaborole 20mg/g topical ointment applied as a thin layer twice daily to affected skin areas up to a maximum of 40% BSA, can be used on all skin areas apart from the scalp. • Up to 4 weeks use per treatment course. • Stop if signs and/or symptoms on treated areas persist after 3 treatment courses of 4 weeks or if signs and/or symptoms worsen during treatment. • Administration instructions for children and adults are the same.
Mechanism of action	Non-steroidal small molecule, inhibits phosphodiesterase 4 (PDE4) a regulator of inflammatory cytokines which are proteins involved in the inflammatory process and immune response. It contains a boron atom that helps penetrate the skin.
Additional investigations	Crisaborole does not require special monitoring and no systemic effects have been observed.
Special warnings	• Local skin reactions (burning or stinging) more likely on sensitive skin areas such as the face and neck. • QT prolongation – no cardiac effects in clinical. • Apply to a maximum of 40% BSA: small sample size in trial with >40%.

1.4 Decision problem

	Scope	Company submission deviations from scope
Population	People aged 2 years and older with mild to moderate atopic dermatitis	Adults and children aged 2 years and older with mild to moderate atopic dermatitis in whom **topical corticosteroids (TCS) are contraindicated or not effective** (scenario analyses for first line: crisaborole compared with emollients and topical corticosteroids)

Comparators	For mild atopic dermatitis:
	• Combination of emollients and mild to moderate potency topical corticosteroids
	
	For moderate atopic dermatitis:
	• High potency topical corticosteroids
	• Topical calcineurin inhibitors
		For mild atopic dermatitis:
		• **Topical calcineurin inhibitor (pimecrolimus)**
		• Emollients and mild to moderate potency TCS (in scenario analyses only)
		
		For moderate atopic dermatitis:
		• Topical calcineurin inhibitors:
		  • Adults: tacrolimus 0.1% (1 mg per 1 gram), tacrolimus 0.03% (300 microgram per 1 gram)
		  • Children: tacrolimus 0.03% (300 microgram per 1 gram), pimecrolimus 1% (10 mg per gram)
		• **Combination of emollients and moderate to** high potency topical corticosteroids (in scenario analyses only – company does not position crisaborole 1st line)

	Calcineurin inhibitors, tacrolimus (moderate to severe atopic dermatitis) and pimecrolimus (mild to moderate atopic dermatitis) are licenced for use when TCS are contraindicated or lack of response to TCS. Tacrolimus and pimecrolimus are NOT recommended in mild atopic dermatitis or as first-line treatments for atopic eczema of any severity TA82	
Subgroups	If evidence allows, the following subgroups will be considered: • adults and children • mild and moderate • different skin colour • atopic dermatitis affecting the hands • atopic dermatitis affecting sensitive areas (face, neck and flexures) • people for whom therapies have been inadequately effective, not tolerated or contraindicated.	**Company did not provide the following subgroup analyses listed in the scope**: • People with different skin colour* • People with atopic dermatitis affecting the hands • People with atopic dermatitis affecting sensitive areas (face, neck and flexures) Justification: not enough clinical evidence

Note: text in **bold** differs from the scope
*company provided pooled subgroup data from AD-301 and AD-302 for ethnicity and race for proportion of pooled patients achieving success in Investigator's Static Global Assessment. Network meta-analysis was

not provided as data were not reported consistently across comparator trials.

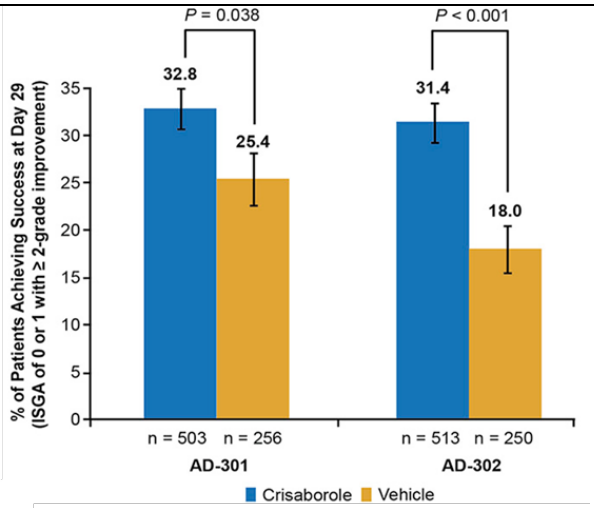
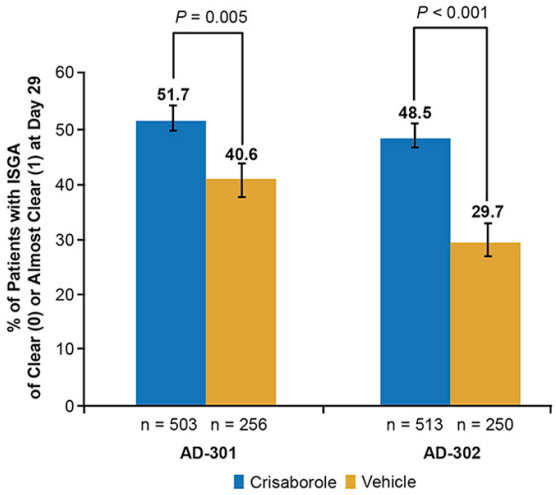
1.5 Clinical evidence

Clinical trial data used in economic model	• AD-301 (n=763) and AD-302 (n=764): identical phase 3 randomised controlled trials, double-blind, 4-weeks, all sites in US, comparing crisaborole with vehicle ointment* (for efficacy data) • AD-303 (n= 517): phase 3 long term (48 week) extension study to assess safety, in people who completed AD-301 or AD-302 (amount of drug in grams per application)
Key results	Primary outcome: Proportion achieving treatment success based on Investigator's Static Global Assessment (ISGA**) score of Clear (0) or Almost Clear (1) **and** at day 29 at least a 2-grade improvement from Baseline/Day 1 [not used in model]: • AD-301: crisaborole, 32.8%; vehicle, 25.4% • AD-302: crisaborole, 31.4%; vehicle, 18.0% Secondary outcome: Proportion achieving an ISGA score of Clear (0) or Almost Clear (1) [used in model]: • AD 301: crisaborole, 51.7%; vehicle, 40.6% • AD-302: crisaborole, 48.5%; vehicle, 29.7%
Comparison	Vehicle adjusted network meta-analysis for crisaborole vs mild topical corticosteroids, moderate topical corticosteroids, tacrolimus (0.03% [children] and 0.1% [adults]), pimecrolimus (1%). *Note: effectiveness of mild and moderate topical corticosteroids based on clinical expert opinion from the tacrolimus and pimecrolimus appraisal (TA82).*
Key result	[redacted]

*Vehicle ointment: ointment that does not contain the active ingredient crisaborole, [redacted], it is not considered a true placebo

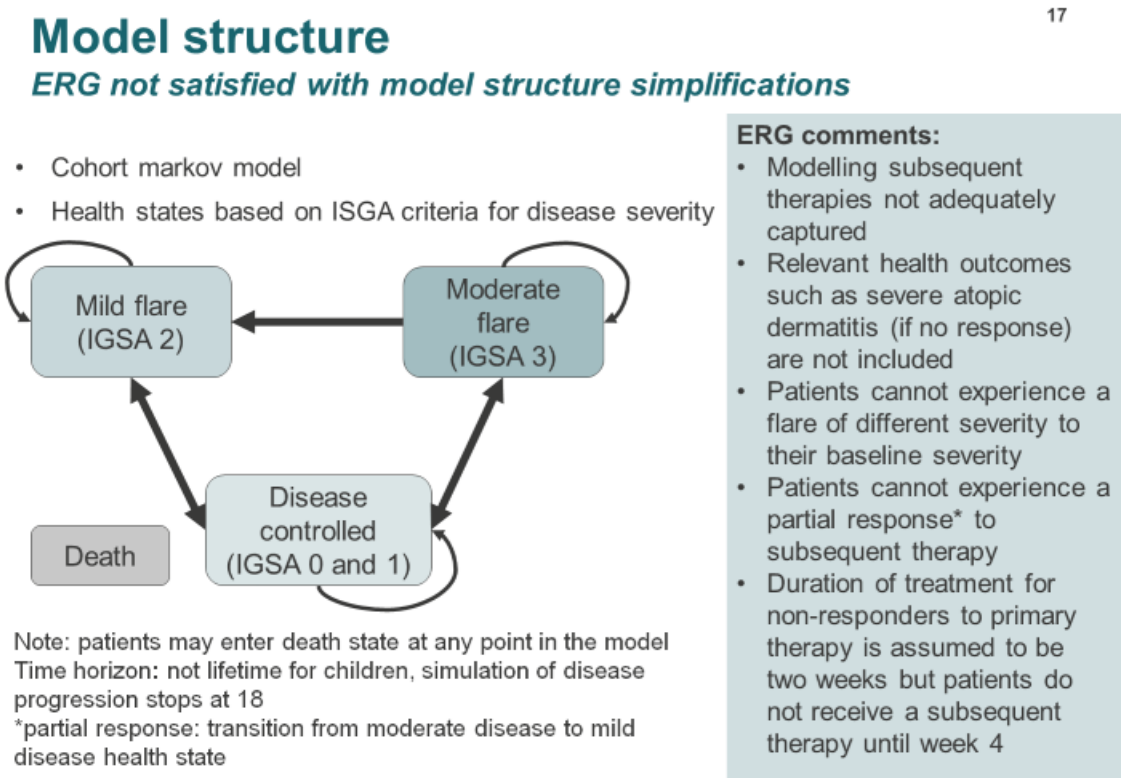
** ISGA is a subjective evaluation of disease severity (5-point scale ranging from clear to severe disease). Success in ISGA was defined as ISGA of Clear or Almost Clear with at least a 2-grade improvement from Baseline/Day 1

1.6 Key trial results

Primary outcome: Proportion of patients achieving ISGA success at Day 29 **Not used in model**	**Secondary outcome: Proportion of patients achieving ISGA score of Clear or Almost Clear at Day 29** **Used in model**
 • A greater proportion of patients treated with crisaborole achieved success in ISGA at day 29 compared with vehicle treated patients in both trials • Primary outcome not used in economic model as not reported consistently across other trials in the network meta-analysis Company submission document B, figure 7	 • A greater proportion of patients treated with crisaborole achieved an ISGA score of clear (0) or almost clear (1) at day 29 compared with vehicle treated patients in both trials • Secondary outcome used in the model: Proportion of patients achieving an ISGA score of Clear (0) or Almost Clear (1) (does not include 2 grade improvement from baseline) Company submission document B, figure 8
Abbreviations: ISGA, Investigator's Static Global Assessment	

1.7 Model structure

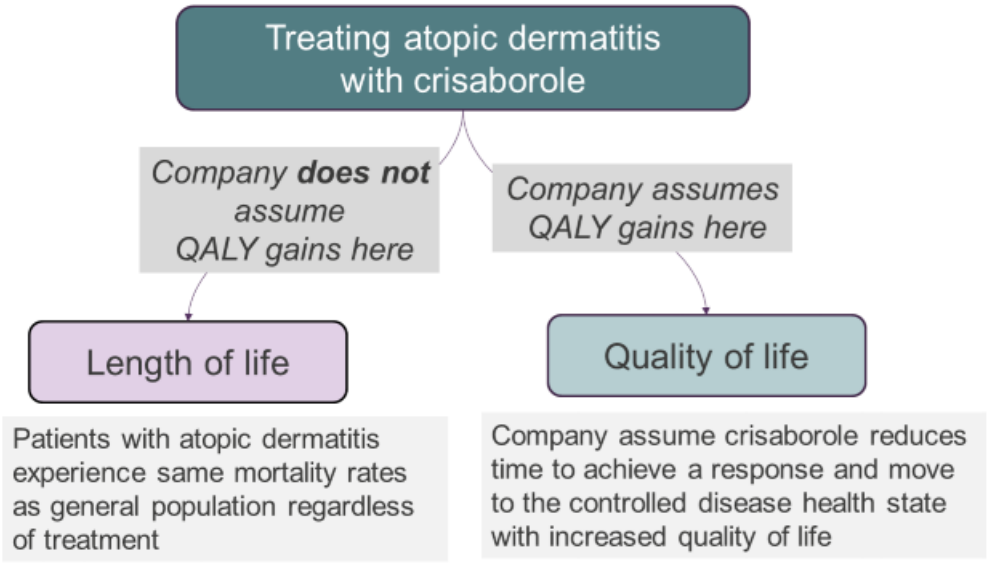
Model	Markov model. 4 health states: mild flare, moderate flare, disease controlled, death
Company ICER	Crisaborole dominates for all four subgroups that have received prior topical corticosteroid therapy (children with either mild or moderate atopic dermatitis, and adults with either mild or moderate atopic dermatitis)
ERG preferred ICER	Mild atopic dermatitis, child: pimecrolimus dominates Moderate atopic dermatitis, child: tacrolimus (0.03%) dominates Mild atopic dermatitis, adult: pimecrolimus dominates Moderate atopic dermatitis, adult: tacrolimus 0.1% dominates
ICER ranges across plausible scenarios	• ICERs are not robust: Changes in costs and QALYs are small, results of the model are sensitive • Crisaborole is dominated in ERG scenarios • Crisaborole is dominated by topical corticosteroids for all subgroups in scenario analyses (note: not proposed position in the treatment pathway).
Abbreviations: ICER, incremental cost-effectiveness ratio	



1.8 Key assumptions in the company's model

Population	Moderate or mild atopic dermatitis with **inadequate response to topical corticosteroids or risk of adverse events with further use** (bold text differs from scope)
Duration Treatment	4 weekly treatment cycles, can stop because of adverse events or no response to treatment Responders: receive 4 weeks of treatment Non responders: discontinue after 2 weeks, receive subsequent therapy at 4 weeks Partial response: a proportion go on to receive a further cycle of treatment if partial response (partial response rate: mild topical corticosteroids [redacted], moderate topical corticosteroids [redacted])
Duration subsequent treatment	Receive subsequent therapy for 4 weeks, discontinue if disease is controlled.
Duration Effect	Company base case: Vehicle adjusted network meta-analysis based on proportion of patients with ISGA 0–1 up to week six. After disease is controlled patients could have a flare and return to their baseline disease severity. Rates are from a report by the company based on a survey and vary based on age and disease severity. Same assumptions applied for treatment and comparator arms.
Quality of life	Trial data from AD-301 and AD-302 collected using dermatology life quality index (DLQI) and mapped to EQ-5D. Mean EQ-5D estimates derived by ISGA state. Company provides scenario analysis for children that maps values from Children's Dermatology Life Quality Index (CDLQI) to EQ-5D.
Abbreviations: EQ-5D, Euroqol 5 dimensions; ISGA, Investigator's Static Global Assessment	

1.9 **Overview of how quality-adjusted life years accrue in the model**



2. Summary of the draft technical report

2.1 In summary, the technical team considered the following:

Issue 1 The company has included pimecrolimus as a comparator for people with mild atopic dermatitis, is this appropriate?

Issue 2 The company's economic model structure does not allow sequential subsequent treatment

Issue 3 The company model does not allow for a partial response on subsequent treatment

Issue 4 The company's model does not take into account the duration of subsequent treatments

Issue 5 Drug use per application should be based on data for the anticipated population for crisaborole

Issue 6 Can the [redacted] and is it appropriate to adjust the relative effectiveness results for [redacted]?

2.2 The technical team recognised that the following uncertainties would remain in the analyses and could not be resolved:

- The possible long-term benefits of crisaborole are unknown, as the efficacy data is based on short term trials (4 weeks).
- There are no head-to-head trials comparing crisaborole with the relevant comparators. The clinical trials compare crisaborole with vehicle ointment [redacted].
- The structure of the company's model precluded the following from being explored fully: sequential subsequent treatments, duration of treatment on subsequent therapy, the potential of atopic dermatitis progressing to severe stage.

2.3 Taking these aspects into account, the ERG's preferred assumptions result in cost-effectiveness results where crisaborole is dominated by all comparators (see tables 4a to 4d). The incremental costs and QALYs are small and are driven by changes in relative effectiveness. The results mostly vary from crisaborole dominating (company base case) to crisaborole being dominated (ERG base case) depending on the source of the relative effectiveness results. It is not possible, within the current model structure, to fully account for sequential treatments.

2.4 The technology is unlikely to be considered innovative.

2.5 No equality issues were identified.

3. Key issues for consideration

Issue 1 – Relevant comparator for people with mild atopic dermatitis

	1 Given the remitting-relapsing nature of atopic dermatitis, how should mild to moderate disease be defined? 2 In clinical practice what treatment would people with mild atopic dermatitis that has not been controlled using topical corticosteroids receive? For example, would people receive topical calcineurin inhibitors, emollients alone, phototherapy, immunosuppressive therapies (azathioprine, ciclosporin, and methotrexate) or oral steroids? 3 Which topical calcineurin inhibitors are used in practice for moderate atopic dermatitis? 4 Is pimecrolimus (1%) used for children with moderate atopic dermatitis in clinical practice? 5 Is crisaborole likely to be used in place of or after treatment with; emollients, topical corticosteroids and topical calcineurin inhibitors? 6 What treatments are used in clinical practice for people with mild to moderate atopic dermatitis who have a steroid phobia?
Background/description of issue	*The company positions crisaborole second line and uses comparators that deviate from the scope.* In its submission **the company** position crisaborole as a treatment for adults and children aged 2 years and older with mild to moderate atopic dermatitis that has not been controlled by topical corticosteroids or where there is a serious risk of important adverse effects from further topical corticosteroid use, particularly irreversible skin atrophy. For adults and children with mild atopic dermatitis the company compares crisaborole with pimecrolimus, which is not recommended by NICE for the treatment of mild atopic eczema or as first-line treatments for atopic eczema of any severity (TA82). The scope defines comparators for mild atopic dermatitis as a combination of emollients and mild to moderate potency topical corticosteroids. The **ERG's** clinical experts advised that it would be unethical to treat people with emollients alone if

	topical corticosteroids had failed. The ERG acknowledges that a small proportion of patients with mild atopic dermatitis (8%) receive topical calcineurin inhibitors based on data from the British Association of Dermatologists audit in children. However, this might be because of the difficulty in recording the data and assigning a single severity score because it may vary by body site or over time. A clinical expert advised that crisaborole could be used for people with mild to moderate eczema who are eligible for corticosteroids but have a steroid phobia and it could benefit people who cannot take topical corticosteroids, cannot tolerate calcineurin inhibitors or it could be used as a steroid sparing treatment on delicate areas such as the face and neck. However, the clinical expert explained that pimecrolimus (1%) is rarely used for people with mild disease, it would only be used on delicate areas such as the face and neck.
Why this issue is important	This is important in determining the relevant comparator.
Technical team preliminary judgement and rationale	The appropriate comparator for people with mild atopic dermatitis that has not responded to topical corticosteroids is unclear. Pimecrolimus is not recommended by NICE for patients with mild atopic dermatitis, and therefore may not be used routinely in clinical practice. This was confirmed by a clinical expert who noted that pimecrolimus is rarely used for mild atopic dermatitis, and would only be used on the face or other delicate areas and only if tacrolimus or ointment preparations are not tolerated. Clinical advice is needed on which treatments are used at 2nd line for people with mild atopic dermatitis that has not been controlled by topical corticosteroids or where there is a serious risk of important adverse effects from further topical corticosteroid use.

Issue 2 – Subsequent therapies

<table>
<tr><td>Questions for engagement</td><td>7 What therapies would people receive in clinical practice after 2nd line treatment with topical calcineurin inhibitors for mild to moderate atopic dermatitis?
8 How is the use of crisaborole likely to change subsequent therapies?</td></tr>
<tr><td>Background/description of issue</td><td>The company's model structure does not allow for subsequent therapies to be captured sufficiently.

In the company's model subsequent therapy is received when a patient does not respond to treatment. Patients are assumed to progress to subsequent therapy if they fail to achieve a response after 1 month of primary therapy, except partial responders who are allowed a second month to achieve a response before progressing to subsequent therapy. All patients that progress to subsequent therapy are assumed to be managed by a dermatologist to initiate subsequent therapy and all receive increased regular dermatologist and GP visits.

The company models subsequent therapy as a 'basket' of therapies. The basket is comprised of a weighted average of costs and benefits of a mixture of therapies, rather than modelling these in a sequence. The model structure does not allow for a treatment sequence to be modelled. The company uses data from the British Association of dermatologists (BAD) UK national audit (table 1) to inform the proportion of patients receiving each therapy.

Table 1: Probability of starting different treatments having failed the primary therapy in mild and moderate disease company base case (Table 28, ERG report)

<table>
<tr><th>Treatment option</th><th>Mild atopic dermatitis</th><th>Moderate atopic dermatitis</th></tr>
<tr><td>Ciclosporin</td><td>100%</td><td>71%</td></tr>
<tr><td>Phototherapy</td><td>0%</td><td>29%</td></tr>
</table>

The ERG is concerned that the model does not allow for sequential modelling of treatments and that</td></tr>
</table>

	the subsequent treatments included in the company's base case are not in line with the treatment pathway in the NICE clinical guideline for atopic eczema in under 12s. The NICE clinical guideline for atopic eczema in under 12s lists systemic treatments and phototherapy for patients with severe atopic dermatitis, not for mild or moderate. The company only included phototherapy and ciclosporin, but the ERG state that methotrexate, azathioprine and mycophenolate mofetil should also be included. The ERG has added these to its base case, which includes phototherapy, ciclosporin, methotrexate, azathioprine and mycophenolate mofetil as subsequent therapies.
Why this issue is important	Cost savings in the model for crisaborole are mostly driven by responders avoiding the use of subsequent therapies. Cost savings and QALY gains are sensitive to changes in the time taken to respond to subsequent therapy, duration of subsequent therapy and the need for monitoring of subsequent therapies.
Technical team preliminary judgement and rationale	The company have simplified subsequent treatment. The model results are driven by the differences in costs of subsequent treatment in non-responders compared with responders to crisaborole. The cost of subsequent treatment for non-responders is greater than the cost of treatment for responders because responders are assumed to stop treatment, so do not incur treatment costs. The ERG's exploratory analyses show that including phototherapy, ciclosporin, methotrexate, azathioprine and mycophenolate mofetil as subsequent therapies decreases the cost savings for the treatment that is more effective (that is for crisaborole when using results from the company's preferred network meta-analysis and for the comparators when using the ERGs preferred network meta-analysis). The technical team would like to see analyses that allowed for sequential modelling of subsequent treatment and adequate modelling of progression of disease to a severe health state. All relevant subsequent therapies should be included in the base case.

Issue 3 – Assuming a partial response on subsequent therapies

Questions for engagement	9 Could someone receiving phototherapy, ciclosporin, methotrexate, azathioprine or mycophenolate mofetil as a 3rd or later line of treatment transition from moderate to mild disease?
Background/description of issue	In the **company's** base case it was assumed that patients could not have a partial response to subsequent therapy (that is, move from moderate to mild disease), however they could have a partial response while on the first treatment received in the model (that is, crisaborole or topical calcineurin inhibitors). The company justified this assumption as a simplifying assumption because there weren't any data about response to the subsequent treatments. The company did a sensitivity analysis where a proportion of non-responders achieved a partial response on subsequent therapy. This proportion was assumed to be the same as for crisaborole, tacrolimus, and pimecrolimus. The **ERG** considers that not including a partial response to treatment in the model is 'a misrepresentation of reality'. The ERG uses the company's scenario that allows non-responders to have a partial response to subsequent therapy in its base case, but this only applies to patients that have moderate disease at baseline.
Why this issue is important	Subsequent therapy assumptions are a key driver of the economic model. Assuming a partial response on subsequent therapy decreased the costs and improved the incremental QALY gain for the intervention which has the higher treatment response (that is for crisaborole when using results from the company's preferred network meta-analysis and for the comparators when using the ERGs preferred network meta-analysis).
Technical team preliminary judgement and rationale	The technical team agree with the ERG's judgement that a partial response to subsequent therapy should be included in the base case.

Issue 4 – Duration of subsequent therapies

Questions for engagement	10 How long does it take for people to respond to phototherapy, ciclosporin, azathioprine,

<table>
<tr><td></td><td>methotrexate, mycophenolate mofetil? Does table 2 (column 4) reflect clinical practice?</td></tr>
<tr><td>Background/description of issue</td><td>In the company's base case non-responders could remain on subsequent therapy indefinitely, but responders discontinue when atopic dermatitis is controlled (ISGA score 0 or 1), which could be as short as one 4-week course.

The ERG is concerned that the duration of subsequent therapy in the company base case is not long enough. Clinical experts advised the ERG that 4 weeks may not be long enough to achieve a response with subsequent therapy.

The response rates in the company's model do not take into account the overall time needed for a response, the company applies the overall response probability for subsequent treatment to each four-week cycle, without adjusting the probability to reflect the model cycle length. In a scenario analysis the ERG adjusted the overall response probability for each subsequent therapy so that it reflected a response probability for the 4 week model cycle (see table 2).

Table 1: Response rates to subsequent therapy based on overall time to response duration
<table>
<tr><th>Therapy</th><th>Average treatment duration (months)</th><th>Overall Response probability (company)</th><th>Time to response</th><th>4-week probability of response (ERG)</th></tr>
<tr><td>Phototherapy</td><td>3</td><td>0.7</td><td>3 months</td><td>0.31</td></tr>
<tr><td>Ciclosporin</td><td>5.8</td><td>0.7</td><td>4 weeks</td><td>0.7</td></tr>
<tr><td>Azathioprine</td><td>13.8</td><td>0.7</td><td>2 months</td><td>0.43</td></tr>
<tr><td>Methotrexate</td><td>15.1</td><td>0.7</td><td>10 weeks</td><td>0.38</td></tr>
<tr><td>Mycophenolate mofetil</td><td>5.8</td><td>0.7</td><td>10 weeks</td><td>0.38</td></tr>
</table></td></tr>
<tr><td>Why this issue is important</td><td>The ERG did a scenario analysis that shows that adjusting the time on subsequent treatment in line with the average time to response, generally increases the cost savings and QALYs gained for the treatment with the highest response depending on the relative effectiveness estimates used in the model (see issue 6), but the absolute change is small.</td></tr>
<tr><td>Technical team preliminary</td><td>The technical team considers that it is appropriate to include the adjustment for subsequent</td></tr>
</table>

judgement and rationale	treatments in the base case.

Issue 5 – Drug use per application

Questions for engagement	11 In clinical practice is the amount of crisaborole used expected to vary based on the amount of body surface area affected? 12 How much drug would be used per application in grams on average for people with [redacted]?
Background/description of issue	The company calculated drug use per application using data from AD-303 in its base case. This includes data from the whole trial population. The marketing authorisation however is expected to state a subgroup based on % body surface area. In response to clarification the company provided a scenario calculated for people with a specific body surface area, in line with the expected marketing authorisation for crisaborole. The ERG states that it is more appropriate to include data for patients likely to receive crisaborole, rather than all patients in the trial and therefore included data for people with [redacted] in its base case.
Why this issue is important	The scenario reflecting the %BSA subgroup did not have a big impact on the ICER. It decreases the cost for the crisaborole arm slightly. However, it is important to include data inputs based on the relevant population throughout the model where possible. Particularly as it is likely that body surface area and drug used per application are correlated.
Technical team preliminary judgement and rationale	Drug use per application should be based on data for the indicated population for crisaborole.

Issue 6 – Network meta-analysis

Questions for engagement	[redacted]
Background/description of issue	[redacted] in **the company's** network meta-analysis. The company state that this is because the vehicles used in the studies are different. The company adjusted the network meta-analysis for the vehicle response using meta-regression and used the vehicle adjusted results in its base case for the relative efficacy of crisaborole. [redacted] The **ERG** states that baseline responses vary from study to study because of the characteristics of patients in the study and it is not possible to tell whether the vehicles in the studies were different. The ERG preferred to use efficacy results from the simple random effects network meta-analysis than the vehicle adjusted analyses or the matching-adjusted indirect comparison (MAIC). However, the company's preferred network meta-analysis assumes that the relationship between the population baseline response and relative treatment effect is the same across treatments, but the ERG's view is that this may not be true. Further, the matching adjusted indirect comparison (MAIC) assumes that all prognostic factors and treatment effect modifiers have been accounted for, but the ERG cannot test this assumption and notes that the MAIC is also defined by the population of comparator treatment. Therefore, it prefers the simple random effect network meta-analysis.
Why this issue is important	The results of the economic model are strongly dependent on the indirect comparison model used to estimate relative treatment effects. Crisaborole tends to be dominant over comparators in analyses using the company's vehicle adjusted network meta-analysis model, which favours crisaborole. Crisaborole tends to be dominated by other comparators when using network meta-analysis results preferred by the ERG, that is the simple random effects model. This is because, a higher response rate in the model leads to QALY gains because patients are spending more time in the controlled

	health state, and cost savings incur because patients are avoiding subsequent treatment. The relative effectiveness is highly uncertain because of the lack of head-to-head trials for crisaborole compared with the relevant comparators.
Technical team preliminary judgement and rationale	There is limited information for adjusting the network meta-analysis for [redacted]. The evidence supporting the vehicle effect adjustment and the network meta-analysis using this method is uncertain. The ERGs approach using the simple effects model may be more appropriate. The technical team would like to see some analyses that assume that crisaborole has the same effectiveness as the comparators, to better understand this uncertainty and these are requested from the company.

4. Issues for information

Tables 4 to 6 are provided to stakeholders for information only and not included in the technical report comments table provided.

Table 4a: ERG preferred assumptions and impact on the cost-effectiveness estimate for mild child patients for crisaborole compared with topical calcineurin inhibitor (pimecrolimus; tacrolimus does not have a marketing authorisation for mild)

Alteration	ERG rationale	Incremental costs	Incremental QALYs	ICER vs pimecrolimus
Company base case	–	[redacted]	[redacted]	Crisaborole dominates
1. Company's base case using ERG's simple random effects network meta-analysis results	The technical team agree with the ERGs approach to modelling relative effectiveness (see issue 6)	[redacted]	[redacted]	Pimecrolimus dominates
Scenarios 2 to 7 use the ERGs preferred simple random effects network meta-analysis				
2. Applying costs of dermatologist visits to uncontrolled patients only	ERG correction, see Table 6	[redacted]	[redacted]	Pimecrolimus dominates
3. Correcting acquisition costs of subsequent systemic therapy	ERG correction, see Table 6	[redacted]	[redacted]	Pimecrolimus dominates
4. Assuming non responders receive 4 weeks of treatment	ERG correction, see Table 6	[redacted]	[redacted]	Pimecrolimus dominates
5. Assuming partial response on subsequent therapies	Not applicable to patients with mild atopic dermatitis			
6. Including other options beside ciclosporin in a weighted basket of systemic therapies	Subsequent therapies in the model should reflect those available in clinical practice. See issue 2.	[redacted]	[redacted]	Pimecrolimus dominates
7. [redacted]	Data based on the anticipated licensed population for crisaborole. See issue 5.	[redacted]	[redacted]	Pimecrolimus dominates
Cumulative impact of the ERG's preferred assumptions on the cost-effectiveness	–	[redacted]	[redacted]	Pimecrolimus dominates

Alteration	ERG rationale	Incremental costs	Incremental QALYs	ICER vs pimecrolimus
estimate				
Note: Crisaborole dominates for all of the above analyses when the Company's preferred network meta-analysis results are applied.				

Table 4b: ERG preferred assumptions and impact on the cost-effectiveness estimate for moderate child patients for crisaborole compared with topical calcineurin inhibitor (tacrolimus 0.03% and pimecrolimus)

Alteration	ERG rationale	Incremental costs	Incremental QALYs	ICER vs tacrolimus 0.03%[a]
Company base case	–	[redacted]	[redacted]	Crisaborole dominates
1. Company's base case using ERG's simple random effects network meta-analysis results	The technical team agree with the ERGs approach to modelling relative effectiveness (see issue 6)	[redacted]	[redacted]	Tacrolimus dominates
Scenarios 2 to 7 use the ERGs preferred simple random effects network meta-analysis				
2. Applying costs of dermatologist visits to uncontrolled patients only	ERG correction, see Table 6	[redacted]	[redacted]	Tacrolimus dominates
3. Correcting acquisition costs of subsequent systemic therapy	ERG correction, see Table 6	[redacted]	[redacted]	Tacrolimus dominates
4. Assuming non responders receive 4 weeks of treatment	ERG correction, see Table 6	[redacted]	[redacted]	Tacrolimus dominates
5. Assuming partial response on subsequent therapies		[redacted]	[redacted]	Tacrolimus dominates

Alteration	ERG rationale	Incremental costs	Incremental QALYs	ICER vs tacrolimus 0.03%[a]
6. Including other options beside ciclosporin in a weighted basket of systemic therapies	Subsequent therapies in the model should reflect those available in clinical practice. See issue 2.	[redacted]	[redacted]	Tacrolimus dominates
7. [redacted]	Data based on the anticipated licensed population for crisaborole. See issue 5.	[redacted]	[redacted]	Tacrolimus dominates
Cumulative impact of the technical team's preferred assumptions on the cost-effectiveness estimate	–	[redacted]	[redacted]	Tacrolimus dominates
[a] tacrolimus dominates pimecrolimus in all scenarios.				

Table 4c: ERG preferred assumptions and impact on the cost-effectiveness estimate for mild adult patients for crisaborole compared with topical calcineurin inhibitor (pimecrolimus; tacrolimus does not have a marketing authorisation for mild)

Alteration	ERG rationale	Incremental costs	Incremental QALYs	ICER vs pimecrolimus
Company base case	–	[redacted]	[redacted]	Crisaborole dominates
1. Company's base case using ERG's simple random effects network meta-analysis results	The technical team agree with the ERGs approach to modelling relative effectiveness (see issue 6)	[redacted]	[redacted]	Pimecrolimus dominates
Scenarios 2 to 7 use the ERGs preferred simple random effects network meta-analysis				
2. Applying costs of dermatologist visits to uncontrolled patients only[a]	ERG correction, see Table 6	[redacted]	[redacted]	Pimecrolimus dominates
3. Correcting acquisition costs of subsequent systemic therapy[a]	ERG correction, see Table 6	[redacted]	[redacted]	Pimecrolimus dominates

Alteration	ERG rationale	Incremental costs	Incremental QALYs	ICER vs pimecrolimus
4. Assuming non responders receive 4 weeks of treatment[a]	ERG correction, see Table 6	[redacted]	[redacted]	Pimecrolimus dominates
5. Assuming partial response on subsequent therapies[a]	Not applicable to patients with mild atopic dermatitis			
6. Including other options beside ciclosporin in a weighted basket of systemic therapies[a]	Subsequent therapies in the model should reflect those available in clinical practice. See issue 2.	[redacted]	[redacted]	Pimecrolimus dominates
7. [redacted]	Data based on the anticipated licensed population for crisaborole. See issue 5.	[redacted]	[redacted]	Pimecrolimus dominates
Cumulative impact of the ERG's preferred assumptions on the cost-effectiveness estimate	–	[redacted]	[redacted]	Pimecrolimus dominates

Table 4d: ERG preferred assumptions and impact on the cost-effectiveness estimate for moderate adult patients for crisaborole compared with topical calcineurin inhibitor (tacrolimus 0.03% and tacrolimus 0.1%)

Alteration	ERG rationale	Incremental costs	Incremental QALYs	ICER vs Tacrolimus 0.1%[a]
Company base case	–	[redacted]	[redacted]	Crisaborole dominates
1. Company's base case using ERG's simple random effects network meta-analysis results	The technical team agree with the ERGs approach to modelling relative effectiveness (see issue 6)	[redacted]	[redacted]	Tacrolimus 0.1% dominates
Scenarios 2 to 7 use the ERGs preferred simple random effects network meta-analysis				
2. Applying costs of dermatologist visits to uncontrolled patients only	ERG correction, see Table 6	[redacted]	[redacted]	Tacrolimus 0.1% dominates

Alteration	ERG rationale	Incremental costs	Incremental QALYs	ICER vs Tacrolimus 0.1%[a]
3. Correcting acquisition costs of subsequent systemic therapy	ERG correction, see Table 6	[redacted]	[redacted]	Tacrolimus 0.1% dominates
4. Assuming non responders receive 4 weeks of treatment	ERG correction, see Table 6	[redacted]	[redacted]	Tacrolimus 0.1% dominates
5. Assuming partial response on subsequent therapies	See issue 3	[redacted]	[redacted]	Tacrolimus 0.1% dominates
6. Including other options beside ciclosporin in a weighted basket of systemic therapies	Subsequent therapies in the model should reflect those available in clinical practice. See issue 2.	[redacted]	[redacted]	Tacrolimus 0.1% dominates
7. [redacted]	Data based on the anticipated licensed population for crisaborole. See issue 5.	[redacted]	[redacted]	Tacrolimus 0.1% dominates
Cumulative impact of the ERG's preferred assumptions on the cost-effectiveness estimate	–	[redacted]	[redacted]	Tacrolimus 0.1% dominates
[a] Tacrolimus 0.1% dominates Tacrolimus 0.03% in all scenarios.				

Table 5: Outstanding uncertainties in the evidence base

Area of uncertainty	Why this issue is important	Likely impact on the cost-effectiveness estimate
Long-term benefits of crisaborole	ERG's clinical experts noted that if a treatment could alter the course of atopic dermatitis in the long-term it could potentially have cost savings by avoiding the development of allergic rhinitis and asthma. However, the economic model does not incorporate these benefits and there is no clinical data on the long-term impact of crisaborole.	The likely impact on the cost-effectiveness estimate is unknown, however if further long-term benefits were proven for crisaborole compared with other treatments this could increase the total QALYs gained for crisaborole.
Structural assumption that treatments are nonsequential	Could not be explored by the ERG because of the rigid structure of the company model	The rigidity of the economic model means that the likely impact on the cost-effectiveness results is unknown. The complexity of different sequences, durations of treatment and treatment benefits means that it is difficult to predict the outcome.
Duration of treatment on subsequent therapy	Could not be explored by the ERG because of the rigid structure of the company model	
Absence from the model of severe atopic dermatitis as a health state	Could not be explored by the ERG because of the rigid structure of the company model	Including severe atopic dermatitis as a health state in the model could help to capture any estimated benefit from progression to severe disease.

Table 6: Other issues for information

Issue	Comments
Comparator costs fluctuate	The cost of the comparators fluctuate over time as there are several treatment options, as well as generic versions, at this point in the treatment pathway. Therefore, the prices from the British National Formulary may not represent those paid by the NHS on any specific day. The Commercial Medicines Unit (CMU) will work with NICE to ensure the prices considered during the appraisal are as up to date as possible. These may change during the course of the appraisal. Analyses will be updated as appropriate.
ERG corrections to company model	The ERG corrected errors in the company's model: • Applied costs of dermatologist visits to uncontrolled patients only, instead of all patients referred to subsequent therapy even after disease control is achieved. • Corrected acquisition costs of subsequent systemic therapy for ciclosporin and mycophenolate mofetil in the model to match the company submission. That is it applied a cost of £18.37 for a 750 mg pack of ciclosporin (instead of £21.80 for a 1,500 mg pack), and 2g of mycophenolate mofetil used per day for 28 days for a flare instead of 2 g per flare. • Amended the model so that non-responders receive treatment for at least one full cycle (4 weeks), instead of 2 weeks. • Duration of time that phototherapy costs are applied for (3 months) corrected. This only applies to scenarios where phototherapy is explicitly modelled (company scenario 5, ERG scenario 7).
Conclusions of the cost-effectiveness analysis are uncertain	The model is driven by the relative clinical effectiveness of crisaborole vs comparators, because the treatment with the best response rate generates a QALY gain and a cost saving (from avoiding the cost of subsequent treatment).

Authors

Amanda Adler

Appraisal committee chair

Jessica Cronshaw

Technical lead(s)

Ellie Donegan

Technical adviser

Melinda Goodall

Associate director

With input from the lead team:

Laura Bojkee

Lead team member

Steve Smith

Lead team member

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Single technology appraisal

Crisaborole for treating mild to moderate atopic dermatitis in people aged 2 years and older [ID1195]

Company response to technical engagement report

February 2020

File name	Version	Contains confidential information	Date
ID1195_Technical Engagement Response_Final_REDACTED	**1.0**	**Yes**	**11 February 2020**

Issue 1 – Relevant comparator for people with mild atopic dermatitis

The appropriate comparator for people with mild atopic dermatitis that has not responded to topical corticosteroids is unclear.

Pimecrolimus is not recommended by NICE for patients with mild atopic dermatitis, and therefore may not be used routinely in clinical practice. This was confirmed by a clinical expert who noted that pimecrolimus is rarely used for mild atopic dermatitis, and would only be used on the face or other delicate areas and only if tacrolimus or ointment preparations are not tolerated.

Clinical advice is needed on which treatments are used at 2nd line for people with mild atopic dermatitis that has not been controlled by topical corticosteroids or where there is a serious risk of important adverse effects from further topical corticosteroid use.

Response: It is important to note that, consistent with the ERG clinical expert advice, we also had feedback from clinical experts that it would be unethical to step down treatment to emollients in mild patients who had failed previous TCS treatment.

While there are no formal NICE guidelines for patients with mild AD who have failed TCS, real world clinical data (i.e., BAD audit data) indicate that a small percentage of mild AD patients may receive TCIs.

Pimecrolimus is the only licensed TCI in mild AD, hence, this appeared to be the most reasonable comparator for analyses in mild AD patients who had failed TCS. Results comparing to tacrolimus were not presented, as tacrolimus is not licensed in this mild AD population. It is noted, however, we would not anticipate any difference in results if these patients were instead assumed to use off-license tacrolimus rather than pimecrolimus, since the cost-effectiveness model, when populated using our base case NMA, showed that crisaborole would be dominant whichever TCI was assumed to be used in mild AD.

Issue 2 – Subsequent therapies

The technical team would like to see analyses that allowed for sequential modelling of subsequent treatment and adequate modelling of progression of disease to a severe health state. All relevant subsequent therapies should be included in the base case.

Response: Comments concerning the use and choice of sequential treatments in the model, and the modelling of disease progression to a severe health state are addressed in turn below.

Sequential use of treatments

We would note that [redacted].

We would also query whether there is strong evidence to show that treatments are used sequentially post-TCS or TCIs in mild and moderate AD, and whether there would be regional variation in treatment patterns (e.g. based on the clinician preference or for example the availability/ease of access to phototherapy).

Nevertheless, we have updated the model to include a treatment sequence where phototherapy is used before systemic therapy, consistent with NICE guidelines and discussions with the ERG on the Technical Engagement teleconference.

The updated analyses show no important changes in base case results and no impact on overall conclusions. Crisaborole dominates TCIs in all post-TCS populations. In summary, these results indicate that this is not an important area of decision uncertainty (see Table 1 to Table 8 in Appendix 1).

Choice of subsequent treatments

Phototherapy and a common systemic treatment (ciclosporin) are modelled as subsequent therapies. The ERG has additionally assumed the potential use of methotrexate, azathioprine and mycophenolate.

The updated model captures the use of these additional systematic therapies, although it is noted that both the previous ERG analyses, and our new analyses, indicate that this alternative mix of treatments has no impact on results or conclusions and crisaborole

continues to dominate TCIs in all post-TCS populations. In summary, these results indicate that this is not an important area of decision uncertainty.

Progression to Severe Health State

In line with comments in the Technical Engagement Report, we have assumed that a proportion of patients who fail TCS or TCI treatment and proceed to subsequent therapies will progress to severe disease [redacted]

The subsequent treatment options used in moderate-or-severe disease have already been captured, so the inclusion of progression to severe disease primarily impacts the HRQoL of patients in our analysis (that is to say, it lowers the HRQoL for those patients whose disease progresses). The inclusion of progression to a severe health state would consequently favour the most effective therapy (i.e. crisaborole), meaning the original base case assumptions were more conservative than the revised analyses presented here.

The updated analyses show no changes in base case results and no impact on overall conclusions. Crisaborole dominates TCIs in all post-TCS populations. In summary, these analyses indicate that this is not an important area of decision uncertainty (see Table 1 to Table 4 in Appendix 1).

Issue 3 – Assuming a partial response on subsequent therapies

The technical team agree with the ERG's judgement that a partial response to subsequent therapy should be included in the base case.

Response: A partial response for crisaborole and TCIs was captured for patients at four weeks due to the treatment patterns of patients treated with these therapies. The assumption was that patients may undertake *another* cycle of treatment if a partial response was observed at 4 weeks. This treatment pattern would not appear to reflect usage of systemic therapies, since patients typically remain on these therapies for longer periods (see Table 2 in Technical Engagement report; Table 32 in ERG report).

Nevertheless, we have updated the model to include a partial response to subsequent therapies as indicated. The updated analyses show no important change in base case results and no impact on overall conclusions. Crisaborole dominates TCIs in all post-TCS populations. In summary, these analyses indicate that this is not an important area of decision uncertainty (see Table 1 to Table 4 in Appendix 1).

Issue 4 – Duration of subsequent therapies

The technical team considers that it is appropriate to include the adjustment for subsequent treatments in the base case.

Response: The ERG has presented a scenario analysis that shows that adjusting the time on subsequent treatment in line with the average time to response generally increases the cost savings and QALYs gained for the treatment with the highest response. This suggests that the CS assumptions are conservative, since inclusion of different durations of subsequent therapies favours the most-effective therapy (i.e. crisaborole).

Nevertheless, we have updated the model to include the ERG's adjustment for subsequent therapies. The updated analyses show no impact on overall conclusions. Crisaborole dominates TCIs in all post-TCS populations. In summary, these analyses show that this is not an important area of decision uncertainty and original base case assumptions were conservative (see Table 1 to Table 4 in Appendix 1).

Issue 5 – Drug use per application

Drug use per application should be based on data for the indicated population for crisaborole.

Response: The inclusion of lower drug use for crisaborole will reduce the costs in all arms rather than increase the costs as indicated in the TE report. The CS submission base case assumption is consequently conservative since lower drug costs would benefit the most-

effective therapy (i.e. crisaborole). It is also noted that we have already provided an analysis with a pricing scenario for patients with a BSA ≤40% in the response to CQs.

In the updated model, we have nevertheless revised the base case assumption so that drug use per application reflects the indicated population BSA ≤40%. The updated analyses show no impact on overall conclusions. Crisaborole dominates TCIs in all post-TCS populations. These analyses show that this is not an important area of decision uncertainty and original base case assumptions were conservative (see Table 1 to Table 8 in Appendix 1).

Issue 6 – Network meta-analysis

There is limited information for adjusting the network meta-analysis for a vehicle effect. However if the vehicles used in the trials in the network meta-analysis are different, the simple random effects model may underestimate the relative effectiveness of crisaborole. If it is assumed that all vehicles are different a network meta-analysis is not appropriate. The evidence supporting the vehicle effect adjustment and the network meta-analysis using this method is uncertain. The ERGs approach using the simple effects model may be more appropriate. The technical team would like to see some analyses that assume that crisaborole has the same effectiveness as the comparators, to better understand this uncertainty and these are requested from the company.

Response: The ERG, technical team, and company agree that there is variation in response on vehicle arms across trials (Table 18 of ERG report, reproduced below); the disagreement is on why this happens and what to do about it.

Given this variation, in choosing a simple random effects NMA, the analyses performed by the ERG and technical team are not aligned with the recommendations of NICE DSU TSD 3. The DIC and regression coefficient provide strong evidence that the difference in response is not due to random variation and must be included in the NMA model.

The ERG and technical team have additionally paid insufficient attention to the unanchored MAIC, which was conducted following NICE DSU TSD 18 recommendations when there are concerns the network is disconnected. Indeed, the technical team acknowledge that the network may not be connected by expressing concern that vehicle is not a true placebo. While the ERG and technical team raise a concern that the unanchored MAIC has not accounted for all prognostic factors and treatment effect modifiers, the criticisms raised are of all MAIC, not specifically of our analysis, and it is not aligned with NICE recommendations

to discount any unanchored MAIC on these grounds. We were rigorous in engaging with clinicians, conducting targeted literature reviews to identify prognostic factors and conducting regression analyses using the crisaborole IPD; this goes beyond what has been done in previous published MAICs. We have provided a detailed response to these points in Appendix 2.

While we strongly dispute the use of the ERG NMA, in response to the TE team's request for additional results, we have nevertheless evaluated additional scenarios in the updated cost-effectiveness model under assumptions that crisaborole has equivalent effects to tacrolimus.

We draw attention to the differences in treatment applications outlined in the SMPCs. The tacrolimus SMPC indicates that in children tacrolimus should be prescribed twice daily for 3 weeks and once daily thereafter (for adults it remains as twice daily). Furthermore, the SMPCs for tacrolimus specify that this therapy may be used for up to 6 weeks. It is also noted that in the CS we conservatively assumed costs for tacrolimus reflected 4 weeks usage, even though, the efficacy data for tacrolimus used in the NMA, captured outcomes for up to 6 weeks tacrolimus treatment.[1-3] The SMPC for pimecrolimus also specifies therapy may be used for up to 6 weeks.

Crisaborole, conversely, reflects 4 weeks efficacy data and costs. This model simplification was undertaken to enable a consistent 4-week modelled cycle length between treatments. Overall, given that TCIs may be used for up to 6 weeks per flare, we would anticipate that the number of drug applications per flare may be higher for TCIs than crisaborole. This may be slightly offset by the reduction in treatment to once daily after 3 weeks for tacrolimus in children.

It is noted that the assumption of 6 weeks treatment use for TCIs was supported by our expert clinical adviser, who indicated that this dosing regimen not only reflected SMPC prescribing information but also reflected local recommendations in the UK NHS. For example, the Shared Care Guidelines for the treatment of atopic dermatitis produced by the Dorset Medicines Advisory Group specify that the Specialist should "assess initial response to treatment and discontinue after 6 weeks if there is no benefit from pimecrolimus or tacrolimus."[4]

Considering all of these factors, we have evaluated scenarios under assumptions of equivalence in efficacy as follows:

1. The number of applications is reduced for tacrolimus 0.03% for children (conservative assumption) with 4 weeks treatment for all other therapies and indications (current base case).
2. The same number of drug applications is used for crisaborole, tacrolimus and pimecrolimus in all populations (4 weeks treatment, all therapies and indications).
3. An increased number of applications for tacrolimus vs crisaborole (6 weeks TCIs, 4 weeks crisaborole) is offset in children by the lower drug usage after 3 weeks for tacrolimus.

The full results of these analyses are presented in Table 9 to Table 18 in Appendix 1.

Response to other aspects of the Technical Engagement Report

Table 4 in the TE report presents the impact of sensitivity analyses performed by the ERG. However, this table provides the impact of cumulative changes in assumptions rather than the one-way analyses to show the impact of each individual change.

The first assumption change is in the efficacy data, which results in crisaborole being dominated. Thus, given that cumulative changes are presented, all subsequent analyses show that crisaborole is dominated, even though changes in these assumptions individually would not impact results. We highlight that the presentation of results in Table 4 misleads the reader and overstates the uncertainty in economic case.

Summary of Pfizer Response

The updated analyses show that the five economic model issues raised do not have an important impact on base case results and do not change economic conclusions. Overall, analyses indicate that these are not important areas of decision uncertainty.

The ERG has favoured a simple random effects NMA which fails to adjust for vehicle response, despite strong scientific/clinical evidence together with methodological and statistical arguments that appear to favour this adjustment. We strongly dispute this approach and note that by choosing a simple random effects NMA, the ERG and TE team have not aligned themselves with the recommendations of NICE DSU TSD 3.[5] Furthermore, while the ERG acknowledge that the network may not be connected by expressing concern that vehicle is not a true placebo, they have paid little attention to the results of the unanchored MAIC, the results of which should be of greater importance in circumstances where there is a potentially disconnected network. We also note the directional consistency of our base case NMA and the MAIC (in contrast to the preferred ERG base case).

[redacted]. The key question is therefore whether crisaborole is at least as effective as this comparator. Combining the NMA model uncertainty and statistical uncertainty, the balance of evidence suggests that crisaborole is not worse and likely to be better than tacrolimus 0.03%. This argument applies equally to pimecrolimus 1% and tacrolimus 0.1%. Under an assumption of equivalence of effects, we can demonstrate that even given a conservative review of the clinical evidence, crisaborole remains a cost-effective therapy. Our analyses show that crisaborole may be either cost-equivalent compared to tacrolimus 0.03% (assuming the same number of treatment applications) or cost saving given that tacrolimus may be used for 6-weeks per treatment cycle compared to 4-weeks per treatment cycle for crisaborole.

Overall the results of the updated base case analyses and sensitivity analyses indicate that crisaborole represents a highly cost-effective treatment for mild to moderate AD in adults and children who have failed prior TCS or TCIs. These conclusions are robust in the face of plausible variation in the assumptions which underpin them.

In addition to being highly cost-effective, crisaborole offers a novel, non-steroidal, mode of action that improves upon the perceived risk-benefit profile of TCSs and TCIs. Crisaborole is not associated with skin atrophy, risk of cancer or other special warnings for human use based on conventional studies of safety and has not been associated with hypopigmentation in non-Caucasian patients. The addition of a topical therapeutic option with a favourable safety profile, good clinical efficacy results, and limited systemic exposure for the treatment of mild-to-moderate AD offers an important additional therapy for AD patients, caregivers, and physicians.

Appendix 1: Updated Cost-Effectiveness Results

The following amendments have been made to the model following the technical report:

- Additional health states have been included to allow the sequential modelling of subsequent therapies
    - Patients will now receive up three cycles of phototherapy per flare before moving on to systemic therapies
    - The cost of phototherapy is incurred in the first cycle, as per the ERGs assumptions
- Additional systemic therapies have been included and the ERGs adjustments for time to response implemented
- Using drug use per application in patients with ≤40% BSA affected
- Including partial response for subsequent therapies
- A proportion of patients that are on treatment with subsequent therapies are modelled as having severe disease
    - This proportion is set to 3.17% in the base-case, which was the proportion of non-responders at day 29 in AD-301/302 that were classified as severe
- The number of clinician visits in subsequent therapy has also been updated based on ERG feedback – patients are now assumed to receive one consultant visit per on-treatment cycle in subsequent therapy
- The cost of tacrolimus 0.1% has been updated to the latest value (£38.46)
- Tacrolimus has been incorporated into comparisons for mild disease

Updated base-case results

Table 1 to Table 4 present the results of the updated analyses. Table 5 to Table 8 present the impact of each of these changes individually.

Table 1: Children with mild disease

	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER vs. baseline	ICER incremental
Crisaborole	[redacted]	[redacted]	[redacted]	[redacted]	-	Dominant
Tacrolimus 0.03%	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominated
Pimecrolimus	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominated

Table 2: Children with moderate disease

	Total costs	**Total QALYs**	**Incremental costs**	**Incremental QALYs**	**ICER vs. baseline**	**ICER incremental**
Crisaborole	[redacted]	[redacted]	[redacted]	[redacted]	-	Dominant

Tacrolimus 0.03%	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominated
Pimecrolimus	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominated

Table 3: Adults with mild disease

	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER vs. baseline	ICER incremental
Crisaborole	[redacted]	[redacted]	[redacted]	[redacted]	-	Dominant
Tacrolimus 0.1%	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominated
Tacrolimus 0.03%	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominated
Pimecrolimus	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominated

Table 4: Adults with moderate disease

	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER vs. baseline	ICER incremental
Crisaborole	[redacted]	[redacted]	[redacted]	[redacted]	-	Dominant
Tacrolimus 0.1%	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominated
Tacrolimus 0.03%	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominated

Table 5: Children with mild disease, one-way analysis to explore the impact of changes

Scenario	Incremental costs	Incremental QALYs	ICER (vs pimecrolimus)
Original base-case	[redacted]	[redacted]	Dominant
Sequential modelling of subsequent therapies	[redacted]	[redacted]	Dominant
Including partial response for subsequent therapies	[redacted]	[redacted]	Dominant
≤40% BSA affected for dosing	[redacted]	[redacted]	Dominant
Including additional subsequent therapies and including the ERG adjustments for time to response	[redacted]	[redacted]	Dominant
Including severe disease for non-responders	[redacted]	[redacted]	Dominant
Adjusting clinician visits	[redacted]	[redacted]	Dominant
Updating the cost of tacrolimus 0.1%	[redacted]	[redacted]	Dominant
Updated base-case (cumulative effect)	[redacted]	[redacted]	**Dominant**

Table 6: Children with moderate disease, one-way analysis to explore the impact of changes

Scenario	Incremental costs	Incremental QALYs	ICER (vs tacrolimus 0.03%)
Original base-case	[redacted]	[redacted]	Dominant
Sequential modelling of subsequent therapies	[redacted]	[redacted]	Dominant
Including partial response for subsequent therapies	[redacted]	[redacted]	Dominant
≤40% BSA affected for dosing	[redacted]	[redacted]	Dominant
Including additional subsequent therapies and including the ERG adjustments for time to response	[redacted]	[redacted]	Dominant
Including severe disease for non-responders	[redacted]	[redacted]	Dominant
Adjusting clinician visits	[redacted]	[redacted]	Dominant
Updating the cost of tacrolimus 0.1%	[redacted]	[redacted]	Dominant
Updated base-case (cumulative effect)	[redacted]	[redacted]	**Dominant**

Table 7: Adults with mild disease, one-way analysis to explore the impact of changes

Scenario	Incremental costs	Incremental QALYs	ICER (vs pimecrolimus)
Original base-case	[redacted]	[redacted]	Dominant
Sequential modelling of subsequent therapies	[redacted]	[redacted]	Dominant
Including partial response for subsequent therapies	[redacted]	[redacted]	Dominant
≤40% BSA affected for dosing	[redacted]	[redacted]	Dominant
Including additional subsequent therapies and including the ERG adjustments for time to response	[redacted]	[redacted]	Dominant
Including severe disease for non-responders	[redacted]	[redacted]	Dominant
Adjusting clinician visits	[redacted]	[redacted]	Dominant
Updating the cost of tacrolimus 0.1%	[redacted]	[redacted]	Dominant
Updated base-case (cumulative effect)	[redacted]	[redacted]	**Dominant**

Table 8: Adults with moderate disease, one-way analysis to explore the impact of changes

Scenario	Incremental costs	Incremental QALYs	ICER (vs tacrolimus 0.1%)
Original base-case	[redacted]	[redacted]	Dominant
Sequential modelling of subsequent therapies	[redacted]	[redacted]	Dominant
Including partial response for subsequent therapies	[redacted]	[redacted]	Dominant
≤40% BSA affected for dosing	[redacted]	[redacted]	Dominant
Including additional subsequent therapies and including the ERG adjustments for time to response	[redacted]	[redacted]	Dominant
Including severe disease for non-responders	[redacted]	[redacted]	Dominant
Adjusting clinician visits	[redacted]	[redacted]	Dominant
Updating the cost of tacrolimus 0.1%	[redacted]	[redacted]	Dominant
Updated base-case (cumulative effect)	[redacted]	[redacted]	**Dominant**

Results assuming equivalence

The differences in costs over the time horizon when assuming equivalence are small. This is in part because these results are averaged over both responders and non-responders and once patients fail to respond to therapy there is no difference in costs. When therapies are assumed to be equivalent the only difference in costs is the cost of primary therapy. Table 9 to Table 14, present the detailed breakdown of results given equivalent efficacy and up to 6 weeks of therapy per treatment cycle for TCIs.

Table 9: Children with mild disease: Equivalent to PIM

	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER vs. baseline	ICER incremental
Pimecrolimus	[redacted]	[redacted]	[redacted]	[redacted]	-	Dominant
Crisaborole	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominated

These results imply that the total incremental cost of treating patients with crisaborole rather than pimecrolimus between the ages of 2 and 18 is £12.
Abbreviations: ICER, Incremental cost-effectiveness ratio; QALY, Quality-adjusted life-year.

Table 10: Children with moderate disease: Equivalent to TAC0.03%

	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER vs. baseline	ICER incremental
Crisaborole	[redacted]	[redacted]	[redacted]	[redacted]	-	Dominant
Tacrolimus 0.03%	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominated
Pimecrolimus	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominated

These results imply that there is no incremental cost associated with treating patients with crisaborole rather than tacrolimus 0.03% between the ages of 2 and 18.
Abbreviations: ICER, Incremental cost-effectiveness ratio; QALY, Quality-adjusted life-year.

Table 11: Children with moderate disease: Equivalent to PIM

	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER vs. baseline	ICER incremental
Tacrolimus 0.03%	[redacted]	[redacted]	[redacted]	[redacted]	-	Dominant
Pimecrolimus	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominated
Crisaborole	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominated

These results imply that the total incremental cost of treating patients with crisaborole rather than pimecrolimus between the ages of 2 and 18 is £23.
Abbreviations: ICER, Incremental cost-effectiveness ratio; QALY, Quality-adjusted life-year.

Table 12: Adults with mild disease: Equivalent to PIM

	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER vs. baseline	ICER incremental
Pimecrolimus	[redacted]	[redacted]	[redacted]	[redacted]	-	Dominant
Crisaborole	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominated

These results imply that the total incremental cost of treating patients with crisaborole rather than pimecrolimus over the course of their lifetime is £12.
Abbreviations: ICER, Incremental cost-effectiveness ratio; QALY, Quality-adjusted life-year.

Table 13: Adults with moderate disease: Equivalent to TAC0.03%

	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER vs. baseline	ICER incremental
Tacrolimus 0.1%	[redacted]	[redacted]	[redacted]	[redacted]	-	Dominant
Crisaborole	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominated
Tacrolimus 0.03%	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominated

These results imply that there is no difference in costs if patients are treated with crisaborole rather than tacrolimus 0.03%.
Abbreviations: ICER, Incremental cost-effectiveness ratio; QALY, Quality-adjusted life-year.

Table 14: Adults with moderate disease: Equivalent to TAC0.1%

	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER vs. baseline	ICER incremental
Crisaborole	[redacted]	[redacted]	[redacted]	[redacted]	-	Dominant
Tacrolimus 0.1%	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominated
Tacrolimus 0.03%	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominated

These results imply that the total incremental cost of treating patients with crisaborole rather than tacrolimus 0.1% over the course of their lifetime is £17.
Abbreviations: ICER, Incremental cost-effectiveness ratio; QALY, Quality-adjusted life-year.

Equivalence scenarios

Table 15 to Table 18 present the results of scenario analyses for the analysis assuming equivalent efficacy. In mild disease crisaborole has been assumed to be equivalent to pimecrolimus. In moderate disease in children all treatments were assumed to be equivalent to tacrolimus 0.03% and in moderate disease in adults all treatments were assumed to be equivalent to tacrolimus 0.1%.

Table 15: Children with mild disease post-TCS, scenarios with equivalent efficacy

Scenario	Total costs - pimecrolimus	Total costs - crisaborole	Incremental cost

Equivalent efficacy	[redacted]	[redacted]	[redacted]
Equivalent efficacy and dosing	[redacted]	[redacted]	[redacted]
Equivalent efficacy and 6 weeks of therapy for TCIs	[redacted]	[redacted]	[redacted]

Table 16: Adults with mild disease post-TCS, scenarios with equivalent efficacy

Scenario	Total costs - pimecrolimus	Total costs - crisaborole	Incremental cost
Equivalent efficacy	[redacted]	[redacted]	[redacted]
Equivalent efficacy and dosing	[redacted]	[redacted]	[redacted]
Equivalent efficacy and 6 weeks of therapy for TCIs	[redacted]	[redacted]	[redacted]

Table 17: Children with moderate disease post-TCS, scenarios with equivalent efficacy

Scenario	Total costs - pimecrolimus	Total costs – tacrolimus 0.03%	Total costs - crisaborole	Incremental cost vs pimecrolimus	Incremental cost vs tacrolimus 0.03%
Equivalent efficacy	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Equivalent efficacy and dosing	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Equivalent efficacy and 6 weeks of therapy for TCIs	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Table 18: Adults with moderate disease post-TCS, scenarios with equivalent efficacy

Scenario	Total costs – tacrolimus 0.1%	Total costs – tacrolimus 0.03%	Total costs - crisaborole	Incremental cost vs tacrolimus 0.1%	Incremental cost vs tacrolimus 0.03%
Equivalent efficacy	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Equivalent efficacy and dosing	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Equivalent efficacy and 6 weeks of therapy for TCIs	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Appendix 2: Issue 6 – Network meta-analysis

Full Response to Issue 6

As noted in the response to Issue 6 above, the ERG, NICE technical team, and company agree that there is variation in response on vehicle arms across trials (Table 18 of ERG report, reproduced below); the disagreement is on why this happens and what to do about it. Regardless of the evidence we provided on differences in vehicle ingredients, the ERG proposed it was due either to differences in baseline characteristics or random variation (issue 7 of Factual Accuracy Check). However, differences in baseline characteristics would suggest potential differences in treatment effect modifiers, which should be explored by meta-regression.[6,a]

[a] Regression coefficient (-0.847 (-1.229, -0.417) on the log hazard ratio scale and 0.44 (0.29, 0.66) on the hazard ratio scale

Vehicle response regression adjusts for differences in multiple effect modifiers simultaneously as their variation is represented by differences in vehicle response.[7] The DIC was 129.3 for the simple random effects and 124.8 for the vehicle response regression, a difference of 4.5, while residual deviances were similar (17.87 and 15.73, respectively, on 18 data points).

We disagree with the ERG's statement that DIC differences below 5 are unimportant; it has been suggested that DIC differences greater than five are important, while those less than 3 (*not 5*) are not important.[8] DIC and residual deviance consistently favour the vehicle response adjusted models [b] (CS: Document B, Table 32, page 76) so this is not a chance finding due to statistical variation. Importantly, the regression coefficient was [redacted]) on the log hazard ratio scale, suggesting that we are at least 95% sure that it is not zero; this is similar in magnitude and certainty to the coefficient in the key example of baseline risk regression of Section 4.4.1 in NICE DSU TSD 3.[5] The DIC and regression coefficient provide strong evidence that the difference in response is not due to random variation and must be included in the NMA model.

In the certolizumab for rheumatoid arthritis example used in Section 4.4.1 of NICE DSU TSD 3,[5] the interaction terms were [redacted] under fixed effects and [redacted] under random effects; this is similar in magnitude, and certainty that it is non-zero, to our regression coefficient of -[redacted] on the log hazard ratio scale. On page 45 of NICE DSU TSD 3[5] they conclude that "Both the fixed and random effects models with covariate have a credible region for the interaction term which is far from zero, suggesting a strong interaction effect between the baseline risk and the treatment effects." They go on to conclude that the striking recommendation that the relation between efficacy and baseline risk that needs to be incorporated into CEA models.

In addition to the above, we believe the ERG and technical team have paid insufficient attention to the unanchored MAIC, which was conducted following NICE DSU TSD 18 recommendations when there are concerns the network is disconnected. The technical team acknowledge that the network may not be connected by expressing concern that vehicle is not a true placebo. Unreported PF and TEM are a criticism of all MAIC, not specifically our analysis, and it is not aligned with NICE recommendations to discount any unanchored MAIC on these grounds. Furthermore, this is also a criticism of unadjusted NMA, such as the

[b] 6 models tested. Random effects and fixed study effects, no/random/fixed class effect.

simple random effects model. Baseline risk adjustment goes someway to adjusting for all PF and TEM as they are represented by variation in baseline response.

Table 22 from ERG report: ISGA/IGA 0/1: Comparison of the results of the simple NMA (company and ERG), vehicle-adjusted NMA, and unanchored MAIC

Comparison vs. crisaborole	**Simple random effects NMA:[a] Company Results random treatment effects no class effect no vehicle response adjustment (CS Fig D25 p247, Table 32)[d]**		**Simple random effects NMA:[a] ERG Results random treatment effects no class effect no vehicle response adjustment**		**Company base case model (FE-RCE-VR) fixed treatment effect random class effects vehicle response adjusted (CS Figure 19, Table 32)**		**Unanchored MAIC (CS Fig 30)**	
	HR[c] 95% CrI	**SD[b]**	**HR[c] 95% CrI**	**SD**	**HR[c] (95% CrI)**		**OR[c] (95% CI)**	
Vehicle	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]		[redacted]	
Tacrolimus 0.03%	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]		[redacted]	
Tacrolimus 0.1%	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]		[redacted]	
Pimecrolimus 1%	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]		[redacted]	
Between-study SD	[redacted]		[redacted]					
Total residual deviance	17.87		17.96		15.73			
DIC	129.2		129.3		124.8			

Specific responses to Questions for Engagement in Issue 6

13 Is it reasonable to assume that the effect of the vehicles used in the trials in the network meta-analysis differ?

14 Is it appropriate to adjust for vehicle effect in the network meta-analysis?

We answer points 13 and 14 at once. We believe that the company and ERG agree that there is variation in response on vehicle arms across trials (see Table 18 of ERG report, reproduced below); the disagreement is on why this happens and what to do about it.

Table 18 of ERG report. Results of trials included in NMA: ISGA/IGA 0-1 (adapted from CS Table D14)

Trial number (Acronym)	**Adults / children**	**Severity group**	**Timepoint (weeks)**	**Crisaborole 2%**	**Tacrolimus 0.03%**	**Tacrolimus 0.1%**	**Pimecrolimus 1%**	**Vehicle**
AD 301, 2016[9]	Adults & children (≥2y)	Mild to moderate	week 4	260/503 (52%)				**104/256 (41%)**
AD 302, 2016[9]	Adults & children (≥2y)	Mild to moderate	week 4	249/513 (49%)				**74/250 (30%)**
[a]**Chapman, 2005**[1]	Adults (≥16y)	Mild to moderate	week 6[b]		74/152 (49%)			**48/148 (32%)**
Levy, 2005[10]	Adults (≥18y)	Mild to moderate	week 4		17/44 (39%)			**10/44 (23%)**
Schachner, 2005[11]	Children (2-15y)	Mild to moderate	week 4		65/158 (41%)			**31/159 (19%)**
Eichenfield, 2002[12]	Children (1-17y)	Mild to severe	week 4				83/267 (31%)	**16/136 (12%)**
Abramovits, 2008[3]	Adults (≥16y)	Moderate	week 6			44/98 (45%)	28/90 (31%)	
Kempers, 2004[13]	Children (2-17y)	Moderate	week 4[c]		22/69 (32%)		17/70 (24%)	
Paller, 2005[2]	Children (2-15y)	Mild	week 6		97/207 (47%)		88/216 (41%)	

[a]Chapman 2005 describes two trials; the trial in adults is listed here under Chapman 2005, while the trial in children is listed under Schachner 2005. [b]Chapman 2005: incorrectly stated in CS Table D14 as 4 weeks; corrected to 6 weeks in clarification response. [c]Kempers 2004: Incorrectly stated in CS Table D14 as 43 days (6 weeks); corrected to 29 days (4 weeks) in clarification response.

We suggested that the variation was primarily due to difference in vehicle ingredient while the ERG proposed it was due either to differences in baseline characteristics or random variation (issue 7 of FAC). However, differences in baseline characteristics would suggest potential differences in treatment effect modifiers, which should be explored by meta-regression.[6,c] Vehicle response regression adjusts for differences in multiple effect modifiers simultaneously as their variation is represented by differences in vehicle response.[7] Individual characteristics (age (mean), severity (% moderate), % BSA (mean), Caucasian (%) males (%)) were explored using meta-regression[d] but the greatest evidence was found for vehicle response regression.

The DIC was 129.3 in simple random effects and 124.8 in the vehicle response regression, while residual deviances were 17.87 and 15.73, respectively, on 18 data points. Importantly, the regression coefficient was -0.847 (-1.229, -0.417) on the log hazard ratio scale, suggesting that we are at least 95% sure that it is not zero. These comparisons are a formal statistical test providing strong evidence that the difference in response is not due to random variation.

We disagree with the ERG's statement that DIC differences below 5 are unimportant; it has been suggested that DIC differences greater than five are important, while those less than 3 (*not 5*) are not important, provided that the conclusions are robust to choice of model.[8] Conclusions in this case are highly model dependent so the DIC difference of 4.5 should not be ignored. In the pivotal Spiegelhalter 2002[14] DIC paper, section 9.2.4 indicate that models that are different by 3-7 points have "considerably less support". In the Lunn et all 2013 BUGS book a difference of 5-10 was identified as substantial, by which a difference of 4.5 is on the edge of being 'substantial'.[15] Furthermore, DIC and residual deviance consistently favour the vehicle response adjusted models (CS: Document B, Table 32, page 76)[e] so this is not a chance finding due to statistical variation.

This was summarised in Table 21 of ERG report and CS Table D19. However, the ERG responded that Pfizer did not include detailed ingredients breakdown in the submission so this could not be critiqued.

[c] Regression coefficient (-0.847 (-1.229, -0.417) on the log hazard ratio scale and 0.44 (0.29, 0.66) on the hazard ratio scale).
[d] The ERG noted that we didn't use individual level meta-regression using the crisaborole IPD. However, such regressions could only explore effects within the crisaborole trial, while our interest was adjustment across trials.
[e] 6 models tested. Random effects and fixed study effects, no/random/fixed class effect.

15 Is it reasonable to use a simple random effects meta-analysis for crisaborole?

No, DIC and residual deviance do not support this selection. The regression coefficient for vehicle response regression should furthermore not be ignored.

16 Is it expected that crisaborole is at least as effective as other comparators in the network?

This repeats our key argument that [redacted] so the key question is whether crisaborole is at least as effective as this comparator. There is at least disagreement about the most appropriate model to employ for network meta-analysis, with DIC and deviance not definitive. Point estimates switch when moving between the simple random effects and vehicle response regression models. Furthermore, under simple random effects the 95% credible intervals for hazard ratios overlap with 1 indicating little or no evidence of a different between crisaborole and comparators. The technical team accept the challenge of comparing efficacy when the vehicle is not a true placebo; the unanchored MAIC [redacted] from Table 22 of ERG report) explored this and indicates further uncertainty. Combining the model uncertainty, questions and vehicle, and statistical uncertainty, the balance of evidence would suggest that crisaborole is no worse than tacrolimus 0.03%. This argument applies equally to pimecrolimus 1% and tacrolimus 0.1%.

Response to technical team position on Issue 6 ('independent' vs 'identical' regression):

That baseline response and relative treatment effects have 'identical' relationship across treatments was a necessary assumption due to limited data. There were only two RCTs on crisaborole versus vehicle (AD301 and AD302), three on tacrolimus 0.03% versus vehicle[1,11,16], three on tacrolimus 0.01% versus vehicle[2,3,13] and only one on pimecrolimus 1% versus vehicle.[12] It was raised in NICE clarification question A23 and in response we explored 'related' and 'independent' regression coefficient models. As suggested by the reviewer, we attempted to overcome identifiability issues by using the regression coefficient of the 'identical' model [redacted] on the log hazard ratio scale) as an informative prior; however, only the 'independent' model converged. The DIC and residual deviance of this 'independent' model was far inferior to the base case; total residual

deviance was 23.21 vs 15.73 in 'independent' model on 18 datapoints while DIC was 131.3 vs 124.8 in 'independent' model. There was also evidence of inconsistency when using the 'independent' coefficients model (inconsistency DIC 129.7 vs 131.3 in consistency model, residual deviance 20.47 vs 23.21 in consistency model).

Figure 9 of NICE clarification questions. Forest plot of base case FE-RCE-VR NMA with 'independent' regression coefficients and informative priors. Estimated hazard ratios (95% CrI) with P-best (probability crisaborole superior) of ISGA/IGA 0/1 of comp (Figure AIC)

Response to technical team position on Issue 6 (concerns about MAIC):

The ERG and technical team have paid insufficient attention to the unanchored MAIC. In cases where there is concern that the evidence network is disconnected, the NICE DSU TSD 18 recommends the use of unanchored MAIC.[17] Our MAIC methodology was fully aligned with the NICE recommendations.

The ERG and technical team raise a concern that the unanchored MAIC has not accounted for all prognostic factors (PF) and treatment effect modifiers (TEM). However, we were rigorous in engaging with clinicians, conducting targeted literature reviews for both PF and

TEM, and conducting regression analyses using the crisaborole IPD; this goes beyond what has been done in previous published MAICs. Our list of potential PF and TEM was extensive and all were adjusted when data were available from comparator studies:

[redacted]We agree that additional unreported PF and TEM may exist, but this is a general criticism of MAIC rather than a specific criticism of the MAIC we conducted. The NICE DSU TSD 18[17] recommends MAIC if networks are disconnected, regardless of this general criticism.

Furthermore, unreported PF and TEM are a criticism of unadjusted NMA, such as the simple random effects model. Baseline risk adjustment goes someway to adjusting for all PF and TEM as they are represented by variation in baseline response.

Figure 4 from Document A of company submission: Forest plot of unanchored MAIC estimated odds ratios (95% CI) of ISGA/IGA 0/1 up to 6 weeks of comparators versus crisaborole—Main Submission, B.2.9.8, Figure 30 (p. 94) (Figure AIC)

Note: ESS is reduced from potential max of 1021.

References

1. Chapman MS, Schachner LA, Breneman D, et al. Tacrolimus ointment 0.03% shows efficacy and safety in pediatric and adult patients with mild to moderate atopic dermatitis. 2005; **53**(2): S177-S85.
2. Paller AS, Lebwohl M, Fleischer AB, Jr., et al. Tacrolimus ointment is more effective than pimecrolimus cream with a similar safety profile in the treatment

of atopic dermatitis: results from 3 randomized, comparative studies. *J Am Acad Dermatol* 2005; **52**(5): 810-22.

3. Abramovits W, Fleischer AB, Jr., Jaracz E, Breneman D. Adult patients with moderate atopic dermatitis: tacrolimus ointment versus pimecrolimus cream. *J Drugs Dermatol* 2008; **7**(12): 1153-8.
4. Dorset Medicines Advisory Group. Shared Care Guidelines for Prescribing Pimecrolimus (ELIDEL®) and Tacrolimus (PROTOPIC®) For Treatment of Atopic Eczema. Available from: https://www.dorsetccg.nhs.uk/Downloads/aboutus/medicines-management/Shared%20Care%20Guidelines/Shared%20care%20guideline%20for%20pimecrolimus%20and%20tacrolimus%20for%20atopic%20eczema.pdf?UNLID=. March 2019
5. Dias S SA, Welton NJ, Ades AE. NICE DSU Technical Support Document 3: NICE DSU Technical Support Document 3: Heterogeneity: subgroups, meta-regression, bias and bias-adjustment. Available from http://www.nicedsu.org.uk Sept 2011.
6. Dias S, Sutton AJ, Welton NJ, Ades AE. Evidence synthesis for decision making 3: heterogeneity--subgroups, meta-regression, bias, and bias-adjustment. *Med Decis Making* 2013; **33**(5): 618-40.
7. Achana FA, Cooper NJ, Dias S, et al. Extending methods for investigating the relationship between treatment effect and baseline risk from pairwise meta-analysis to network meta-analysis. *Statistics in medicine* 2013; **32**(5): 752-71.
8. Dias S, Ades AE, Welton NJ, Jansen JP, Sutton AJ. Network meta-analysis for decision-making: John Wiley & Sons; 2018.
9. Paller AS, Tom WL, Lebwohl MG, et al. Efficacy and safety of crisaborole ointment, a novel, nonsteroidal phosphodiesterase 4 (PDE4) inhibitor for the topical treatment of atopic dermatitis (AD) in children and adults. *J Am Acad Dermatol* 2016; **75**(3): 494-503 e6.
10. Levy AL SM, Roberts R. Tacrolimus Cream 0.03% Is Safe and Effective in the Treatment of Mild and Moderate Atopic Dermatitis in Adults. *J Allergy Clin Immunol* 2005; **115**: S103.
11. Schachner LA, Lamerson C, Sheehan MP, et al. Tacrolimus ointment 0.03% is safe and effective for the treatment of mild to moderate atopic dermatitis in pediatric patients: results from a randomized, double-blind, vehicle-controlled study. *Pediatrics* 2005; **116**(3): e334-42.
12. Eichenfield LF, Lucky AW, Boguniewicz M, et al. Safety and efficacy of pimecrolimus (ASM 981) cream 1% in the treatment of mild and moderate atopic dermatitis in children and adolescents. *J Am Acad Dermatol* 2002; **46**(4): 495-504.
13. Kempers S, Boguniewicz M, Carter E, et al. A randomized investigator-blinded study comparing pimecrolimus cream 1% with tacrolimus ointment 0.03% in the treatment of pediatric patients with moderate atopic dermatitis. *J Am Acad Dermatol* 2004; **51**(4): 515-25.

14. Spiegelhalter DJ, Best N, Carlin B. Bayesian measures of model complexity and fit. *JR Statist Soc B* 2002; **64**(4): 583-639.

15. Lunn D, Jackson C, Best N, Spiegelhalter D, Thomas A. The BUGS book: A practical introduction to Bayesian analysis: Chapman and Hall/CRC; 2012.

16. Levy AL SM, Roberts R. Tacrolimus Cream 0.03% Is Safe and Effective in the Treatment of Mild and Moderate Atopic Dermatitis in Adults *J Allergy Clin Immunol* 2005; **115**: S103.

17. Phillippo DM AA, Dias S, Palmer S, Abrams KP, Welton NJ NICE DSU Technical Support Document 18: Methods for population-adjusted indirect comparisons in submissions to NICE. Available from: http://www.nicedsu.org.uk Dec 2016

NICE National Institute for Health and Care Excellence

Technical engagement response form

Crisaborole for treating mild to moderate atopic dermatitis in people aged 2 years and older [ID1195]

As a stakeholder you have been invited to comment on the technical report for this appraisal. The technical report and stakeholders responses are used by the appraisal committee to help it make decisions at the appraisal committee meeting. Usually, only unresolved or uncertain key issues will be discussed at the meeting.

We need your comments and feedback on the questions below. You do not have to answer every question. The text boxes will expand as you type. Please read the notes about completing this form. We cannot accept forms that are not filled in correctly. Your comments will be summarised and used by the technical team to amend or update the scientific judgement and rationale in the technical report.

Deadline for comments 2 March 2020

Thank you for your time.

Please log in to your NICE Docs account to upload your completed form, as a Word document (not a PDF).

Notes on completing this form

- Please see the technical report which summarises the background and submitted evidence. This will provide context and describe the questions below in greater detail.
- Please do not embed documents (such as PDFs or tables) because this may lead to the information being mislaid or make the response unreadable. Please type information directly into the form.
- Do not include medical information about yourself or another person that could identify you or the other person.
- Do not use abbreviations.
- Do not include attachments such as journal articles, letters or leaflets. For copyright reasons, we will have to return forms that have attachments without reading them. You can resubmit your form without attachments, but it must be sent by the deadline.
- If you provide journal articles to support your comments, you must have copyright clearance for these articles.
- Combine all comments from your organisation (if applicable) into 1 response. We cannot accept more than 1 set of comments from each organisation.
- Please underline all confidential information, and separately highlight information that is submitted under 'commercial in confidence' in turquoise, all information submitted under 'academic in confidence' in yellow, and all information submitted under 'depersonalised data' in pink. If confidential

information is submitted, please also send a second version of your comments with that information replaced with the following text: 'academic/commercial in confidence information removed'. See the Guide to the processes of technology appraisal (sections 3.1.23 to 3.1.29) for more information.

We reserve the right to summarise and edit comments received during engagement, or not to publish them at all, if we consider the comments are too long, or publication would be unlawful or otherwise inappropriate.

Comments received during engagement are published in the interests of openness and transparency, and to promote understanding of how recommendations are developed. The comments are published as a record of the comments we received, and are not endorsed by NICE, its officers or advisory committees.

About you

Your name	[redacted]
Organisation name – stakeholder or respondent (if you are responding as an individual rather than a registered stakeholder please leave blank)	National Eczema Society
Disclosure Please disclose any past or current, direct or indirect links to, or funding from, the tobacco industry.	N/A

Questions for engagement

Issue 1 - Relevant comparator for people with mild atopic dermatitis	
1 Given the remitting-relapsing nature of atopic dermatitis, how should mild to moderate disease be defined?	We agree with the NICE definitions of mild and moderate atopic dermatitis as stated in the guidance 'Atopic eczema in under 12s: diagnosis and management' (2007). The physical severity of mild atopic dermatitis is described as 'Areas of dry skin, infrequent itching (with or without small areas of redness).' The impact of mild atopic dermatitis on quality of life and psychosocial wellbeing is described as follows: 'Little impact on everyday activities, sleep and psychosocial wellbeing'. The guidance acknowledges that even mild atopic dermatitis 'can have a negative impact on psychological and psychosocial wellbeing and quality of life.' The physical severity of moderate atopic dermatitis is described as 'Areas of dry skin, frequent itching, redness (with or without excoriation and localised skin thickening).' The impact of moderate atopic dermatitis on everyday activities, sleep and psychosocial wellbeing is described as follows: 'Moderate impact on everyday activities and psychosocial wellbeing, frequently disturbed sleep.' The treatment options for mild atopic dermatitis are emollients and mild to moderate potency topical corticosteroids. The treatment options for moderate atopic dermatitis are emollients, mild to potent topical corticosteroids, topical calcineurin inhibitors and bandages and wet wraps.
2 In clinical practice what treatment would people with mild atopic dermatitis that has not been controlled using topical corticosteroids receive? For example, would people receive topical calcineurin inhibitors, emollients alone, phototherapy, immunosuppressive therapies (azathioprine, ciclosporin, and methotrexate) or oral steroids?	'Mild' atopic dermatitis that has not been controlled with emollients and appropriately-used topical corticosteroids cannot accurately be defined as 'mild'. If even moderate topical corticosteroids do not effectively control 'mild' atopic dermatitis on delicate areas of skin, or potent topical corticosteroids control 'mild' atopic dermatitis on the body, it does not make sense to describe the atopic dermatitis as 'mild'. Mild atopic dermatitis should respond to emollients and lower potencies of topical corticosteroids. People with mild atopic dermatitis who cannot use topical corticosteroids because they are intolerant to them and/or for whom there is a serious risk of adverse effects, may be offered

	topical calcineurin inhibitors in addition to emollients, or they may be expected to use emollients alone.
3 Which topical calcineurin inhibitors are used in practice for moderate atopic dermatitis?	Pimecrolimus cream (Elidel) and tacrolimus ointment (Protopic) are used in practice for moderate atopic dermatitis.
4 Is pimecrolimus (1%) used for children with moderate atopic dermatitis in clinical practice?	Pimecrolimus (1%) is used for children with moderate atopic dermatitis for whom topical corticosteroids have not worked effectively, those who are intolerant to topical corticosteroids, or those whose eczema is located on areas of skin for which prolonged intermittent treatment with topical corticosteroids may be inappropriate (e.g. the face and neck).
5 Is crisaborole likely to be used in place of or after treatment with; emollients, topical corticosteroids and topical calcineurin inhibitors?	We consider that crisaborole should be offered to patients with mild to moderate eczema who: - are intolerant to topical steroids and/or are at risk of serious adverse effects from using them - have concerns about and are unwilling to use topical corticosteroids of any level of potency. Crisaborole should also be offered to patients with moderate eczema whose symptoms have not responded effectively to higher potency topical corticosteroids, or who must use higher potency topical corticosteroids regularly as weekend therapy due to frequent flares. Crisaborole is likely to be used after treatment with emollients and topical corticosteroids.
6 What treatments are used in clinical practice for people with mild to moderate atopic dermatitis who have a steroid phobia?	People with moderate atopic dermatitis who have a steroid phobia, and who would require more potent topical corticosteroids to control their eczema, may be offered topical calcineurin inhibitors in addition to emollients. People with mild atopic dermatitis who have a steroid phobia are more likely to be offered emollients alone and encouraged to use mild topical corticosteroids.

Issue 2 - Subsequent therapies	
7 What therapies would people receive in clinical practice after 2nd line treatment with topical calcineurin inhibitors for mild to moderate atopic dermatitis?	People with mild atopic dermatitis should not require further therapies after 2nd line treatment with topical calcineurin inhibitors; if they do, their atopic dermatitis is likely to be moderate to severe rather than mild. People for whom topical treatments (i.e. emollients, appropriate potencies of topical corticosteroids and topical calcineurin inhibitors) have not worked effectively would usually be offered phototherapy, oral steroids or immunosuppressant drugs.
8 How is the use of crisaborole likely to change subsequent therapies?	The use of crisaborole may mean that some people who would otherwise have progressed to subsequent therapies would no longer need to progress to these therapies, if crisaborole kept their eczema under control.
Issue 3 - Assuming a partial response on subsequent therapies	
9 Could someone receiving phototherapy, ciclosporin, methotrexate, azathioprine or mycophenolate mofetil as a 3rd or later line of treatment transition from moderate to mild disease?	It is possible for someone receiving phototherapy, ciclosporin, methotrexate, azathioprine or mycophenelate mofetil as a 3rd or later line of treatment to transition from moderate to severe eczema to a mild disease.
Issue 4 - Duration of subsequent therapies	
10 How long does it take for people to respond to phototherapy, ciclosporin, azathioprine, methotrexate, mycophenolate mofetil? Does table 2 (column 4) reflect clinical practice?	We consider that it takes slightly less time for people to respond to ciclosporin and azathioprine than table 2 (column 4) suggests: 1-2 weeks for ciclosporin and 4-5 weeks for azathioprine. We consider that the other timeframes and duration of therapies in table 2 reflect clinical practice.
Issue 5 - Drug use per application	
11 In clinical practice is the amount of crisaborole used expected to vary based on the amount of body surface area affected?	Yes, the amount of crisaborole used is expected to vary based on the amount of body surface area affected.

12 How much drug would be used per application in grams on average for people with [redacted]?	We are not able to answer this question.

NICE National Institute for Health and Care Excellence

Technical engagement response form

Crisaborole for treating mild to moderate atopic dermatitis in people aged 2 years and older [ID1195]

As a stakeholder you have been invited to comment on the technical report for this appraisal. The technical report and stakeholders responses are used by the appraisal committee to help it make decisions at the appraisal committee meeting. Usually, only unresolved or uncertain key issues will be discussed at the meeting.

We need your comments and feedback on the questions below. You do not have to answer every question. The text boxes will expand as you type. Please read the notes about completing this form. We cannot accept forms that are not filled in correctly. Your comments will be summarised and used by the technical team to amend or update the scientific judgement and rationale in the technical report.

Deadline for comments 2 March 2020

Thank you for your time.

Please log in to your NICE Docs account to upload your completed form, as a Word document (not a PDF).

Notes on completing this form

- Please see the technical report which summarises the background and submitted evidence. This will provide context and describe the questions below in greater detail.
- Please do not embed documents (such as PDFs or tables) because this may lead to the information being mislaid or make the response unreadable. Please type information directly into the form.
- Do not include medical information about yourself or another person that could identify you or the other person.
- Do not use abbreviations.
- Do not include attachments such as journal articles, letters or leaflets. For copyright reasons, we will have to return forms that have attachments without reading them. You can resubmit your form without attachments, but it must be sent by the deadline.
- If you provide journal articles to support your comments, you must have copyright clearance for these articles.
- Combine all comments from your organisation (if applicable) into 1 response. We cannot accept more than 1 set of comments from each organisation.
- Please underline all confidential information, and separately highlight information that is submitted under 'commercial in confidence' in turquoise, all information submitted under 'academic in confidence' in yellow, and all information submitted under 'depersonalised data' in pink. If confidential

information is submitted, please also send a second version of your comments with that information replaced with the following text: 'academic/commercial in confidence information removed'. See the Guide to the processes of technology appraisal (sections 3.1.23 to 3.1.29) for more information.

We reserve the right to summarise and edit comments received during engagement, or not to publish them at all, if we consider the comments are too long, or publication would be unlawful or otherwise inappropriate.

Comments received during engagement are published in the interests of openness and transparency, and to promote understanding of how recommendations are developed. The comments are published as a record of the comments we received, and are not endorsed by NICE, its officers or advisory committees.

About you

Your name	[redacted]
Organisation name – stakeholder or respondent (if you are responding as an individual rather than a registered stakeholder please leave blank)	**Centre of Evidence-Based Dermatology (includes Cochrane Skin and UK Dermatology Clinical Trials Network)**
Disclosure Please disclose any past or current, direct or indirect links to, or funding from, the tobacco industry.	**None**

Questions for engagement

Issue 1 - Relevant comparator for people with mild atopic dermatiti	
1 Given the remitting-relapsing nature of atopic dermatitis, how should mild to moderate disease be defined?	Simple - according to the Harmonising Outcome Measures for Eczema (and international group devoted to developing core outcome sets for eczema) **recommended instruments for use in clinical practice ie POEM scale** (filled in by patients and takes 52 secs). See: https://www.nottingham.ac.uk/research/groups/cebd/resources/poem.aspx □ Clear or almost clear = 0 to 2; □ Mild eczema = 3 to 7; □ Moderate eczema = 8 to 16; And: Leshem YA, Chalmers J, Apfelbacher C, Furue M, Gerbens LA, Prinsen CA, Schmitt J, Spuls PI, Thomas KS, Williams HC, Simpson EL; Harmonising Outcome Measures for Eczema (HOME) initiative. Measuring atopic eczema symptoms in clinical practice: The First Consensus Statement from the Harmonising Outcome Measures for Eczema in Clinical Practice Initiative. J Am Acad Dermatol. 2020 Jan 8. pii: S0190-9622(20)30029-3. doi: 10.1016/j.jaad.2019.12.055. [Epub ahead of print] PubMed PMID: 31926221.
2 In clinical practice what treatment would people with mild atopic dermatitis that has not been controlled using topical corticosteroids receive? For example, would people receive topical calcineurin inhibitors, emollients alone, phototherapy,	In my 40 years of practice, I have yet to meet a person with MILD atopic dermatitis who has not been controlled by topical corticosteroids (TCS). The concept of TCS resistance is a nonsense. There are of course some people with severe atopic dermatitis who don't respond, but mild – never. Adherence to TCS can be a problem due to inappropriate fears of side effects, but side effects such as skin thinning are almost impossible in mild disease (the best real world evidence comes from this study where 1 in 1213 had clinically significant thinning with mild to moderate

immunosuppressive therapies (azathioprine, ciclosporin, and methotrexate) or oral steroids?	TCS compared with 0 out of 1205 for topical pimecrolimus: https://pediatrics.aappublications.org/content/135/4/597/tab-e-letters#petite-bit-of-vital-information-still-missing In practice, there are so many TCS preparations out there that it is easy to try a different one if a patient says that their mild eczema does not respond to the one they have been given. I would simply increase the potency too eg to moderate or modern once-daily potent and just use less of it. In the practical world of a person refusing to use mild TCS because of a genuine phobia, then I might use topical tacrolimus or pimecrolimus, but in my practice 99% of patients are happy to use TCS properly (get control and keep control technique) once it is explained to them how to use them safely and confidently with emollients. Emollients alone will not work in this scenario and would be unethical. Phototherapy is a ridiculous suggestion for mild atopic dermatitis.
3 Which topical calcineurin inhibitors are used in practice for moderate atopic dermatitis?	In practice, I find topical pimecrolimus one of the most useless treatments available for atopic dermatitis – on a par with 1% hydrocortisone. At least topical tacrolimus does something, and even then I would tend to use 0.1% rather than 0.03%. For moderate eczema on the face, topical 0.03% tacrolimus may be OK for holding AD in remission once it is cleared initially with a topical corticosteroid
4 Is pimecrolimus (1%) used for children with moderate atopic dermatitis in clinical practice?	No – in my experience it is absolutely useless for moderate AD. It is on a par with mild to moderate TCS as shown in the PETITE study (Sigurgeirsson B, Boznanski A, Todd G, Vertruyen A, Schuttelaar ML, Zhu X, Schauer U, Qaqundah P, Poulin Y, Kristjansson S, von Berg A, Nieto A, Boguniewicz M, Paller AS, Dakovic R, Ring J, Luger T. Safety and efficacy of pimecrolimus in atopic dermatitis: a 5-year randomized trial. Pediatrics. 2015 Apr;135(4):597-606.)

5 Is crisaborole likely to be used in place of or after treatment with; emollients, topical corticosteroids and topical calcineurin inhibitors?	No – the data from the two pivotal trials which compared topical crisaborole against vehicle (ie plain grease) suggests it is even more ineffective than topical pimecrolimus in mild to moderate AD. Just look at the number needed to treat compared with vehicle which amounts to 14 and 8 for the two trials (see Ahmed A, Solman L, Williams HC. Magnitude of benefit for topical crisaborole in the treatment of atopic dermatitis in children and adults does not look promising: a critical appraisal. Br J Dermatol. 2018 Mar;178(3):659-662. doi: 10.1111/bjd.16046. Epub 2017 Dec 3. PubMed PMID: 29205284.). Little wonder the manufacturers did not compare against a standard active comparator like 1% hydrocortisone as I very doubt if it would be any better at all. The thought of topical crisaborole being used AFTER treatment failure with other conventional treatments like TCS ie an even more difficult and severe population is quite ridiculous given the indicative small magnitude of benefit from the two pivotal trials.
6 What treatments are used in clinical practice for people with mild to moderate atopic dermatitis who have a steroid phobia?	We should be cautious about this word phobia as in most cases it is not an *irrational* fear as TCSW can cause skin thinning eg if potent TCS are used on the face. In my clinical practice, 99% of adults and parents who have (legitimate) concerns about using TCS are happy to use them once the evidence of benefits and harms is explained to them and they are given a written care plan that is safe and effective. Very occasionally, I will meet a phobic person, and I would then try topical calcineurin inhibitors (which often fail as they are not so good at inducing remission compared with once daily potent TCS like mometasone or fluticasone)
Issue 2 - Subsequent therapies	
7 What therapies would people receive in clinical practice after 2nd line treatment with topical calcineurin inhibitors for mild to moderate atopic dermatitis?	The idea that topical calcineurin inhibitors are an effective 2nd line treatment for mild to moderate eczema is a myth. I can see why NICE suggested they should be use in such a way to limit costs and overuse in mild cases, but they are simply not needed in this situation. The range of potency of topical corticosteroids is huge and can usually control patients from very mild (1% hydrocortisone), to moderate (clobetasone) and to moderate/severe (mometasone/fluticasone).

	The only place for topical calcineurin inhibitors is on sensitive site eczema such as the face - which is more prone to local side effects such as skin thinning and acne. So in answer to your question, I would just find another topical corticosteroid and spend time with the patient explaining how to use it in adequate quantities and in bursts so that skin thinning is impossible.
8 How is the use of crisaborole likely to change subsequent therapies?	Based on the current evidence, zero – but there is a gullible market out there who might collude with fuelling the notion that topical corticosteriods are “bad” even in mild eczema. Topical corticosteroidophobia is a marketing gift for new relatively ineffective topicals. Maybe crisaborole could have a place if it is cheaper than 1% hydrocortisone, but I would love to see the redacted results of your NMA as I doubt if it would offer any increased efficacy over very mild TCS.
Issue 3 - Assuming a partial response on subsequent therapies	
9 Could someone receiving phototherapy, ciclosporin, methotrexate, azathioprine or mycophenolate mofetil as a 3rd or later line of treatment transition from moderate to mild disease?	Possible but rather unlikely. They might achieve severe to mild status transition with Dupilumab.
Issue 4 - Duration of subsequent therapies	
10 How long does it take for people to respond to phototherapy, ciclosporin, azathioprine, methotrexate, mycophenolate mofetil? Does table 2 (column 4) reflect clinical practice?	They look reasonable. Ciclosporin has the most rapid action. Depends how you define “respond”, but phototherapy around 3 months +, ciclo 6 weeks, aza 3 months+, mtx 3 months+, myco 3months+ (with peaks response around 6 months for the 3 month ones above)
Issue 5 - Drug use per application	
11 In clinical practice is the amount of crisaborole used expected to vary based on the amount of body surface area affected?	What a daft question – of course the amount will increase according to the surface area needed to be covered

12 How much drug would be used per application in grams on average for people with [redacted]?	Sorry – critical question words are redacted, so I cannot answer

NICE National Institute for Health and Care Excellence

Technical engagement response form

Crisaborole for treating mild to moderate atopic dermatitis in people aged 2 years and older [ID1195]

As a stakeholder you have been invited to comment on the technical report for this appraisal. The technical report and stakeholders responses are used by the appraisal committee to help it make decisions at the appraisal committee meeting. Usually, only unresolved or uncertain key issues will be discussed at the meeting.

We need your comments and feedback on the questions below. You do not have to answer every question. The text boxes will expand as you type. Please read the notes about completing this form. We cannot accept forms that are not filled in correctly. Your comments will be summarised and used by the technical team to amend or update the scientific judgement and rationale in the technical report.

Deadline for comments 2 March 2020

Thank you for your time.

Please log in to your NICE Docs account to upload your completed form, as a Word document (not a PDF).

Notes on completing this form

- Please see the technical report which summarises the background and submitted evidence. This will provide context and describe the questions below in greater detail.
- Please do not embed documents (such as PDFs or tables) because this may lead to the information being mislaid or make the response unreadable. Please type information directly into the form.
- Do not include medical information about yourself or another person that could identify you or the other person.
- Do not use abbreviations.
- Do not include attachments such as journal articles, letters or leaflets. For copyright reasons, we will have to return forms that have attachments without reading them. You can resubmit your form without attachments, but it must be sent by the deadline.
- If you provide journal articles to support your comments, you must have copyright clearance for these articles.
- Combine all comments from your organisation (if applicable) into 1 response. We cannot accept more than 1 set of comments from each organisation.
- Please underline all confidential information, and separately highlight information that is submitted under 'commercial in confidence' in turquoise, all information submitted under 'academic in confidence' in yellow, and all information submitted under 'depersonalised data' in pink. If confidential information is submitted, please also send a second version of your comments with that information replaced with the following text: 'academic/commercial in confidence information removed'. See the Guide to the processes of technology appraisal (sections 3.1.23 to 3.1.29) for more information.

We reserve the right to summarise and edit comments received during engagement, or not to publish them at all, if we consider the comments are too long, or publication would be unlawful or otherwise inappropriate.

Comments received during engagement are published in the interests of openness and transparency, and to promote understanding of how recommendations are developed. The comments are published as a record of the comments we received, and are not endorsed by NICE, its officers or advisory committees.

About you

Your name	**Kymmene Dawson**
Organisation name – stakeholder or respondent (if you are responding as an individual rather than a registered stakeholder please leave blank)	
Disclosure Please disclose any past or current, direct or indirect links to, or funding from, the tobacco industry.	

Questions for engagement

Issue 1 - Relevant comparator for people with mild atopic dermatiti	
1 Given the remitting-relapsing nature of atopic dermatitis, how should mild to moderate disease be defined?	**A definition of mild atopic dermatitis may perhaps reflect that the disease is limited to small areas of the body; is manageable with prescribed (or over the counter) emollients and lower potency topical steroid preparations; and may come and go depending on various internal and environmental factors. It would not be expected that a person suffering with mild eczema would need continual supervision by a GP or any referral to a specialist dermatologist. Mild eczema has little negative impact oh the quality of life of the person with the disease and does not prevent them in any aspect of their life.** **Moderate eczema may be a flare up on a person with mild eczema, or may be a longer-term condition managed with higher potency steroid preparations. It would be expected that a person with moderate eczema would be in periodic contact with their GP with potential referral to a dermatologist. There may be some impact on lifestyle but with some due care, this is manageable and mostly non-restrictive.**
2 In clinical practice what treatment would people with mild atopic dermatitis that has not been controlled using topical corticosteroids receive? For example, would people receive topical calcineurin inhibitors, emollients alone, phototherapy, immunosuppressive therapies (azathioprine, ciclosporin, and methotrexate) or oral steroids?	**A patient with mild eczema would certainly receive emollients as a first step and lower potency topical steroid treatment. If this has failed to control the condition, then more potent topical steroids would be the natural progression from this. If topical steroids prove ineffective, then a course of oral steroids may be prescribed.**

3 Which topical calcineurin inhibitors are used in practice for moderate atopic dermatitis?	**Topical calcineurin inhibitors - Pimecrolimus, Tacrolimus**
4 Is pimecrolimus (1%) used for children with moderate atopic dermatitis in clinical practice?	**Unsure**
5 Is crisaborole likely to be used in place of or after treatment with; emollients, topical corticosteroids and topical calcineurin inhibitors?	**It would likely be used after a combination of emollients and topical steroid treatments.**
6 What treatments are used in clinical practice for people with mild to moderate atopic dermatitis who have a steroid phobia?	**Topical calcineurin inhibitors - Pimecrolimus, Tacrolimus**
Issue 2 - Subsequent therapies	
7 What therapies would people receive in clinical practice after 2nd line treatment with topical calcineurin inhibitors for mild to moderate atopic dermatitis?	**Phototherapy would be the next step in the treatment.**
8 How is the use of crisaborole likely to change subsequent therapies?	**Crisaborole will provide a first/second-line alternative to patients who cannot tolerate topical steroid treatments, who have a contraindication with topical steroids, or who have a reluctance to use topical steroids. It will provide the patient with more options based on the severity and location of the affected skin. It is reported that there should be some economic gain through the use of crisaborole since it may reduce the need to progress to third-line treatments such as phototherapy and immunosuppressant drugs.**

Issue 3 - Assuming a partial response on subsequent therapies	
9 Could someone receiving phototherapy, ciclosporin, methotrexate, azathioprine or mycophenolate mofetil as a 3rd or later line of treatment transition from moderate to mild disease?	**It is possible that a person receiving third line treatment may transition from moderate to mild disease, but this may only be temporary and it is highly possible that the disease would flare up or in time, the patient to escalate back to moderate disease.**
Issue 4 - Duration of subsequent therapies	
10 How long does it take for people to respond to phototherapy, ciclosporin, azathioprine, methotrexate, mycophenolate mofetil? Does table 2 (column 4) reflect clinical practice?	Response time will inevitably vary amongst patients. Ciclosporin can take up to 3 months to be effective. Methotrexate can take just a few weeks. Some immunosuppressants are used simultaneously for a 2-4 weeks whilst transitioning from treatment with one immunosuppressant to another.
Issue 5 - Drug use per application	
11 In clinical practice is the amount of crisaborole used expected to vary based on the amount of body surface area affected?	As certain areas of the body are more delicate where the skin is thinner, it would be expected that the advice for application of a generic crisaborole treatment would differ in frequency/amount depending on the body surface area.
12 How much drug would be used per application in grams on average for people with xxxxxxxxxx xxxxxxxxxxxxxxxx?	Unable to answer

Crisaborole for Treating Mild to Moderate Atopic Dermatitis in People Aged 2 Years and Older [ID 1195]: A Single Technology Appraisal – Addendum

Produced by School of Health and Related Research (ScHARR), The University of Sheffield

Authors Sarah Davis, Senior Lecturer, ScHARR, University of Sheffield, Sheffield, UK

Edith Poku, Research Fellow, ScHARR, University of Sheffield, Sheffield, UK

John W. Stevens, Reader, ScHARR, University of Sheffield, Sheffield, UK

Katy Cooper, Senior Research Fellow, ScHARR, University of Sheffield, Sheffield, UK

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

Ruth Wong, Information Specialist, ScHARR, University of Sheffield, Sheffield, UK

Carolyn Charman, Consultant Dermatologist, Royal Devon and Exeter Hospital, Exeter, UK

Correspondence Author Sarah Davis, Senior Lecturer, ScHARR, University of Sheffield, Sheffield, UK

Date completed 19/02/2020

Source of funding: This report was commissioned by the NIHR HTA Programme as project number 129371.

Declared competing interests of the authors

Carolyn Charman declared that she provides ad hoc consultancy, presentations and teaching for Galderma (hydrocortisone), GlaxoSmithKline (betamethasone valerate, clobetasone butyrate, fluticasone propionate, hydrocortisone) and Leo Laboratories (betamethasone valerate, fusidic acid, hydrocortisone, tacrolimus). None of the other authors have any conflicts of interest to declare.

Acknowledgements

We would like to thank Andrea Shippam for formatting the addendum.

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:

Davis S., Poku E., Stevens J.W., Cooper K., Metry A., Wong R., Charman C. Crisaborole for Treating Mild to Moderate Atopic Dermatitis in People Aged 2 Years and Older [ID 1195]: A Single Technology Appraisal- Addendum. School of Health and Related Research (ScHARR), 2019.

Contributions of authors to this addendum

John Stevens critiqued the company's response regarding the evidence synthesis (issue 6). Edith Poku and Katy Cooper critiqued the company's response regarding the duration of treatment for TCIs. Sarah Davis and Andrew Metry critiqued the updated health economic analysis submitted by the company and conducted the ERG exploratory analyses.

Standard copyright statement on the front page of the ERG/AG report:

CONTENTS

List of Tables

List of Figures

Abbreviations

AD	Atopic dermatitis
BAD	British Association of Dermatologists
CEBD	Centre for Evidence-Based Dermatology
CG	Clinical Guideline
CRTE	Company's response to technical engagement
DIC	Deviance Information Criterion
ERG	Evidence Review Group
ICD	International Congress of Dermatology
ICER	Incremental cost-effectiveness ratio
MAIC	Matching-adjusted indirect comparison
NICE	National Institute for Health and Care Excellence
NMA	Network meta-analysis
QALY	Quality-adjusted life year
SmPC	Summary of Product Characteristics
STA	Single Technology Appraisal
TCI	Topical calcineurin inhibitor
TCS	Topical corticosteroid

1 INTRODUCTION

1.1 Technical engagement report

The NICE technical team, in collaboration with the Committee chair and Lead Team, prepared a draft technical engagement report which was sent out for consultation with stakeholders. This report is based on their initial consideration of the company submission, consultee and commentator expert statements and the ERG report. The technical engagement report identified 6 key issues for consideration during technical engagement. These were as follows;

Issue 1: The company has included pimecrolimus as a comparator for people with mild atopic dermatitis (AD), is this appropriate?

Issue 2: The company's economic model structure does not allow sequential subsequent treatment

Issue 3: The company model does not allow for a partial response on subsequent treatment

Issue 4: The company's model does not take into account the duration of subsequent treatments

Issue 5: Drug use per application should be based on data for the anticipated population for crisaborole

Issue 6: Can the [redacted] and is it appropriate to adjust the relative effectiveness results for [redacted]?

The technical team also recognised that there were several other uncertainties in the evidence that would be unlikely to be resolved during technical engagement which were;

- The possible long-term benefits of crisaborole are unknown, as the efficacy data is based on short term trials (4 weeks).
- There are no head-to-head trials comparing crisaborole with the relevant comparators. The clinical trials compare crisaborole with vehicle ointment [redacted]
- The structure of the company's model precluded the following from being explored fully: sequential subsequent treatments, duration of treatment on subsequent therapy, the potential of atopic dermatitis progressing to severe stage.

1.2 Responses to technical engagement

In response to consultation on the technical engagement report, the company (Pfizer) provided a written response to each of the 6 issues identified and also provided a new economic model. The ERG were requested by NICE to provide a critique of the company's response to technical engagement (CRTE). In doing so, the ERG also took into account the response provided by the Centre for Evidence-Based Dermatology (CEBD). Given the limited time allowed for the ERG to provide their

critique, the ERG have focused on issues where the company has provided additional evidence or analyses or where additional evidence has been provided in the CEBD response.

2 RESPONSES TO THE SIX KEY ISSUES

2.1 Issue 1: Comparator for mild atopic dermatitis

The company provided additional evidence on this in the form of cost-effectiveness analyses comparing crisaborole to tacrolimus (0.03% in children and both 0.03% and 0.1% in adults) in patients with mild AD who have failed on topical c tosteroids (TCSs). Previously, in this population, the company had only compared crisaborole to pimecrolimus. The reason given was that pimecrolimus is the only topical calcineurin inhibitor (TCI) with a license in patients with mild AD.

The ERG notes that the submission from the CEBD states that cases of true failure of TCSs in patients with mild AD are not observed in clinical practice and that in practice clinicians would simply try an alternative mild TCS or switch to a more potent TCS. They commented that TCIs, would only be used in the rare situation that patients had genuine steroid phobia, but that 99% of patients would be happy to use TCSs when properly informed about how to use them safely. They stated that they would use either tacrolimus or pimecrolimus in this situation.

The ERG notes that the use of tacrolimus in the population with mild AD would be off-license and the use of either TCI in mild AD would be outside of NICE's recommendations in TA82. But if tacrolimus was considered to be a valid comparator, on the basis that it is preferred over pimecrolimus in current clinical practice, then the company's analysis shows that the cost savings and QALYs gains for crisaborole are smaller when comparing to tacrolimus (0.03% in children or 0.1% in adults) as it is more effective than pimecrolimus.

The company argued that it was unethical to use emollients alone in patients whose mild AD had not responded to TCSs. This was supported in the consultation response from the CEBD.

The ERG's view is that the evidence submitted by the CEBD suggests that there would be few patients with mild AD that would require an alternative to TCSs, but that in those rare cases where one is needed, clinicians would consider using either tacrolimus or pimecrolimus even though neither is recommended by NICE and tacrolimus would be off-license in this population.

2.2 Subsequent therapies

2.2.1 Sequential use of subsequent therapies

The company's updated model submitted in the CRTE assumes that all patients failing second-line treatment, with either TCIs or crisaborole, progress to receive phototherapy and only those failing to achieve a response to phototherapy progress to receive systemic therapy with immunosuppressants. This revised model is more consistent with the stepped care approach for AD management

recommended in NICE Clinical Guideline 57 (see Table 1 of the ERG report) in that phototherapy is assumed to be used prior to systemic therapies. However, the ERG also notes that phototherapy and systemic immunosuppressants are only recommended in the NICE stepped care pathway for severe AD and not for mild or moderate AD. The response from the CEBD also noted that phototherapy was not appropriate for patients with mild AD. This agrees with the ERG's previous comments that phototherapy and systematic therapies are not part of the NICE stepped care approach for mild AD.

The ERG notes the company's comment that there may be regional variation in whether subsequent therapies are used sequentially and whether there would be regional variation in treatment patterns (e.g. based on the clinician preference or for example the availability/ease of access to phototherapy). This agrees with comments by the ERG's clinical experts that uptake of phototherapy would depend on whether it was available locally as it requires frequent attendance at hospital and this would not always be acceptable to patients, particularly if they had to travel long distances.

The ERG would reiterate what it said previously which was that, *"Clinical advisors to the ERG further noted that treatment escalation to systemic immunosuppressants and phototherapy was generally restricted to patients with uncontrolled severe AD or a subgroup of moderate AD patients with severe clinical presentations. The BAD audit data show that in secondary care systemic treatments are used in a small proportion (under 2%) of children with mild to moderate AD, but they are used around 23% of children with severe AD."* In addition, the ERG note the comments from the CEBD when asked to describe treatment options for those failing to response to second line TCIs which was that, *"the range of potency of topical corticosteroids is huge and can usually control patients from very mild (1% hydrocortisone), to moderate (clobetasone) and to moderate/severe (mometasone/fluticasone)."* This suggests that clinicians would not be expecting patients with mild to moderate AD to require treatment with subsequent therapies as it should be possible to achieve a response using TCSs.

The ERG also notes that the updated model still assumes that patients will receive one of several systemic therapies and applies average parameters for cost, efficacy and duration of treatment based on the proportion assumed to receive each of the possible systemic therapies. It therefore does not capture the possibility that patients will try one systemic therapy and then try an alternative systemic therapy if the first does not work. Such an approach would be likely to increase the cumulative efficacy of subsequent therapies and avoid a large proportion of patients failing to achieve a response on subsequent therapies in the long-term as is the case in the current modelling (see section 2.4).

2.2.2 Choice of subsequent therapies

The company's updated model submitted in the CRTE assumes that a mix of possible therapies are available for systemic immunosuppressant therapy including ciclosporin, methotrexate, azathioprine and mycophenolate mofetil. This is consistent with the approach taken in the ERG's preferred base-case analysis, although the ERG note that the company did not incorporate their corrections to the drug costs for ciclosporin and mycophenolate mofetil. The ERG have therefore reapplied these corrections in their exploratory analyses presented in section 5.

2.2.3 Progression to the severe health state

The company's updated model submitted in the CRTE assumes "*that a proportion of patients who fail TCS or TCI treatment and proceed to subsequent therapies will progress to severe disease* [redacted]*).*" To implement this, the company have assumed that this group of patients have a reduced quality of life with no change to resource use or costs. This is because any additional costs required to manage severe disease are assumed to have been captured because these patients are already receiving subsequent therapies such as immunosuppressants which are recommended for severe AD.

The utility multiplier applied in severe disease is [redacted].

The company does not explicitly state the source of this data in their response to technical engagement, however, the data in the model are consistent with the source being the same as the source for mild and moderate AD, i.e. adult EQ-5D values pooled from studies 301 and 302 (based on comparison of the data in the Excel model with data provided in Table 22 of the CS). The ERG note that the mean absolute utility value is therefore based on measurements from [redacted] adult patients with severe AD whereas the values for all other health states are based on [redacted] or more patients. Whilst this value is likely to be associated with considerable uncertainty, the fact that it is only being applied to a small minority of patients suggests that it is unlikely to significantly contribute to decision uncertainty.

Although the company claims that the inclusion of a severe disease state has a limited impact on the results and is therefore not an important area of decision uncertainty, the ERG note that in mild AD, the addition of the severe state increases the incremental QALYs three to four fold (see Table 5 and Table 7 of the CRTE). Therefore, although it does not change the broad conclusion that crisaborole dominates TCIs, when using the company's preferred NMA, it does show that the size of the QALY gains achieved are being driven by the assumptions regarding what happens to patients moving on to subsequent treatments.

2.3 Issue 3: Partial response to subsequent therapies

The company's updated model submitted in the CRTE now incorporates states to allow patients with moderate AD to experience a partial response to subsequent therapies (both phototherapy and systemic immunosuppressants). The company states that this has a limited impact on the cost-effectiveness analysis, although the ERG notes that in patients with moderate AD, applying this change alone was found to approximately halve the incremental QALYs (see Table 6 and Table 8 of the company's response to technical engagement [CRTE]).

The ERG notes that the rate of partial response applied to phototherapies is equivalent to that applied to TCIs and crisaborole. However, the rate of partial response applied to systemic therapies (0.205) is hard coded into the spreadsheet and the source of the value is not described in either the CRTE or in the Excel file itself. Therefore, the ERG cannot confirm the validity of this probability and this introduces some uncertainty into the interpretation of the company's updated model.

2.4 Issue 4: Duration of subsequent therapies

The company's updated model submitted in the CRTE now incorporates rates of response that are adjusted to reflect the ERG's preferred assumptions regarding the duration of time required to achieve a response for each of the subsequent therapies (see Table 32 of the ERG report). However, in the ERG's previous scenario analysis addressing this issue (scenarios 7 and 14 in the ERG's exploratory analyses presented in the ERG report), the ERG also limited the costs of treatment to the duration of treatment to prevent patients accruing costs for unsuccessful subsequent treatments indefinitely. This was done by applying the costs for the whole duration of subsequent treatment at the time of initiation of subsequent treatment. In contrast, in the company's updated model, patients continue to receive the cost of subsequent treatment in each cycle that they remain non-responsive. This results in around 42% of adults remaining on systemic treatment long-term and accruing costs for those systemic treatments without achieving an adequate response. The proportion in children is lower due to the assumption that disease resolves in a proportion of children, but the model still predicts that 21% of children end up on long-term systemic treatment. This does not appear to have clinical face validity given that the data from the BAD audit suggest that in secondary care systemic treatments are used in a small proportion (under 2%) of children with mild to moderate AD. Therefore, the model appears to be over estimating the proportion of patients accruing costs for systemic therapy without achieving a response.

Although the company states that the changes to adjust response times to reflect duration of treatment show no impact on overall conclusions, the ERG notes that the QALY gains approximately double when incorporating both the broader range of subsequent therapies and the appropriate duration of time to response for these subsequent therapies (see Tables 6 to Table 8 of the CRTE). Therefore,

although it does not change the broad conclusion that crisaborole dominates TCIs, when using the company's preferred NMA, it does show that the size of the QALY gains achieved are being driven by the assumptions regarding the effectiveness of subsequent therapies.

2.5 Issue 5: Drug use per application

The ERG agrees with the company's decision to use data on drug use per application which is based on data for the indicated population for crisaborole despite the fact that this has a small impact on the incremental costs.

2.6 Issue 6: Network meta-analysis (NMA)

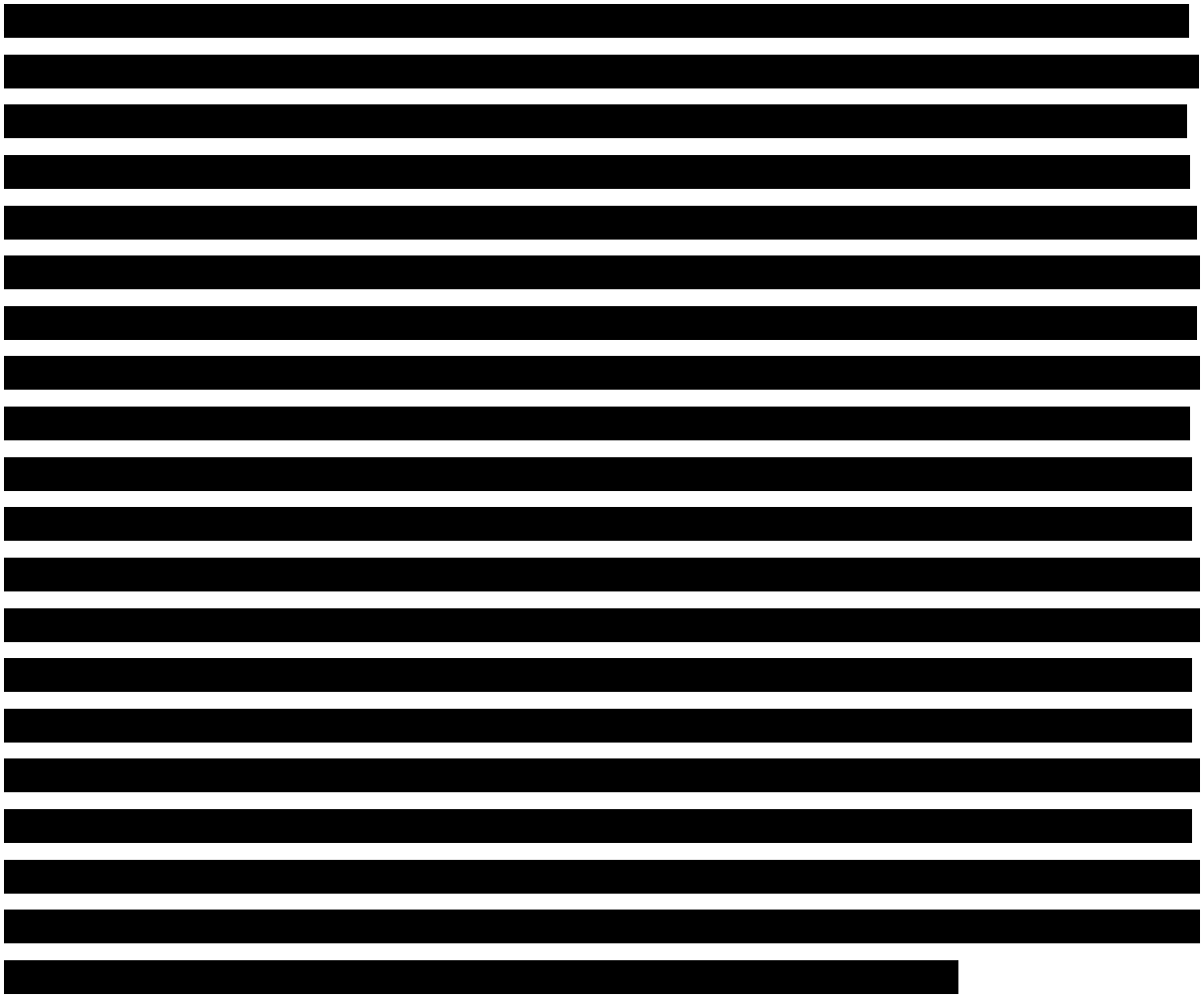
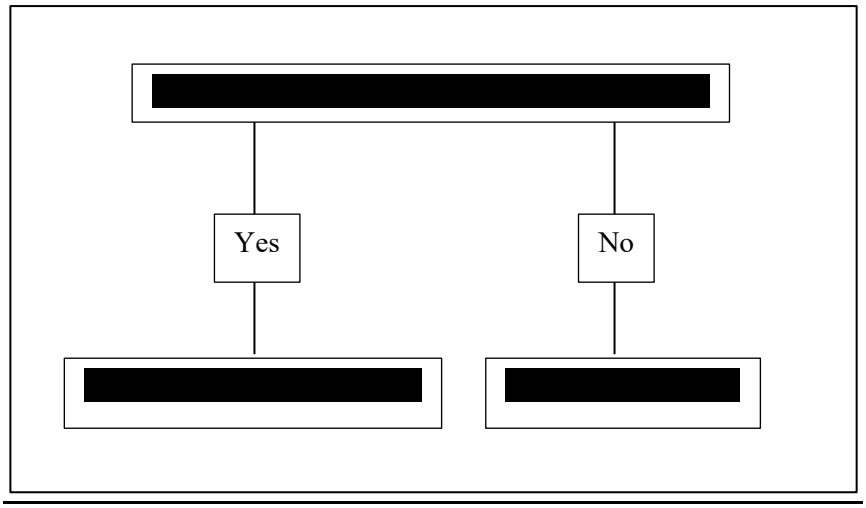


Figure 1: Process for deciding which modelling approach to use


Yes
No

[redacted]

[redacted]

3 OTHER CHANGES TO THE COMPANY MODEL

The company made several other changes to their model in addition to the changes made to the company model in response to the six key issues already described in section 2.

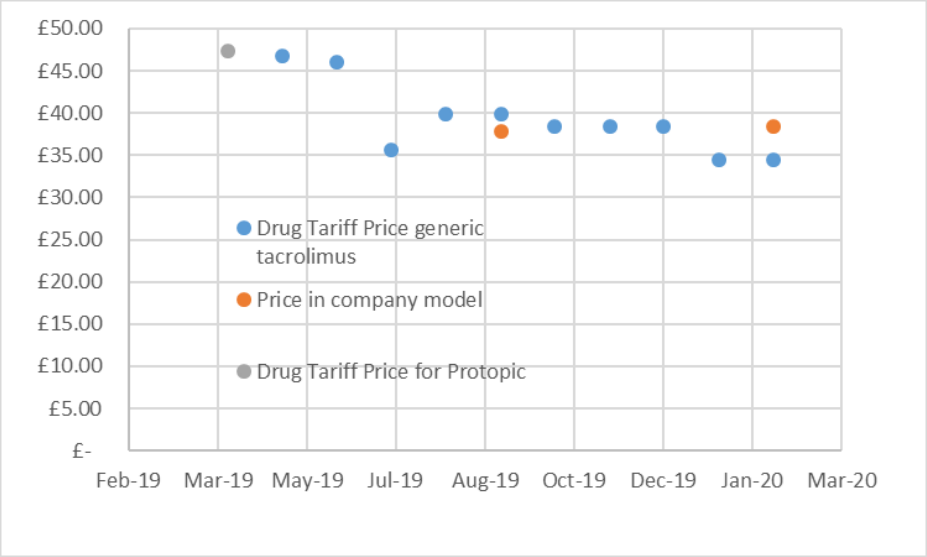
3.1 Adjusting clinician visits

Patients are now assumed to receive one consultant visits per on-treatment cycle in subsequent therapy (phototherapy or systemic therapy) instead of 6 per year, although these costs are now only applied to those who have uncontrolled disease as per the ERG's preferred assumption. The ERG notes that this single change approximately halves the incremental cost savings in the company's analyses (see Table 5 to 8 of the CRTE). This further highlights the importance of correctly estimating downstream costs in order to accurately quantify the size of any cost savings or QALY gains from avoiding patients progressing to subsequent therapies which are managed mostly in secondary care.

3.2 Updating the cost of tacrolimus

The company's updated model uses an acquisition cost for generic 0.1% tacrolimus of £38.46. Although this matches the drug tariff price reported in the current BNF online (accessed 17th Feb 2020), the ERG notes that there is lag between updates to the drug tariff and updates to the BNF online and the drug tariff price for February 2020 is £34.52. The ERG notes that the price of generic 0.1% tacrolimus has varied several times over the course of this appraisal and the costs may continue to fluctuate in future. Figure 2 shows how the price of tacrolimus has varied since a price was listed for the generic version in May 2019. The ERG have applied the latest price from the February 2020 for generic 0.1% tacrolimus (£34.52) in their analyses presented in Section 5.

Figure 2: Drug tariff prices for 0.1% tacrolimus over time and prices applied in the company's models submitted in Sept 2019 and Feb 2020



3.3 Phototherapy costs

Phototherapy costs are now applied once when patients initiate phototherapy. The cost applied is £93 which is based on the reference cost for phototherapy. Therefore, the company appears to be assuming that £93 is the cost of a whole course of phototherapy. The ERG notes that in the appraisal of dupilumab, the reference cost for phototherapy was £86.85 per session (2016/17 prices), and the number of session over 3 months was assumed to be 22 giving a cost per course of £1,910.70 (TA534: Sanofi response to ACD, Appendix C). Therefore, the ERG does not believe that the company has properly estimated the costs of phototherapy. The impact of this is that the cost savings from avoiding patients progressing to subsequent therapy will have been underestimated. This will obviously favour crisaborole in the company's base-case but would favour TCIs when using the ERG's preferred simple random effects NMA. This failure by the company to properly estimate the cost of phototherapy adds further to the uncertainty regarding the cost savings and QALY gains attributable in the model to avoiding treatment with subsequent therapies.

4 COST COMPARISONS ASSUMING EQUIVALENT EFFICACY

In the technical report conclusions on issue 6, the NICE technical team requested that the company provide some analyses assuming that crisaborole has the same efficacy as comparator treatments. The company has provided results tables for these comparison but no spreadsheet model has been provided therefore the ERG had to determine what assumptions had been used in these analyses by trial and error.

The ERG were able to replicate the results for patients with mild AD (i.e. Table 9 for children with mild AD and Table 12 for adults for mild AD) by setting the response probability for crisaborole equal to that for pimecrolimus. No change was made to the duration of treatment for pimecrolimus which remained twice daily for four weeks.

The ERG were able to replicate the results in Table 11 for patients with children moderate AD when assuming 3 weeks of twice daily treatment followed by 3 weeks of once daily treatment for tacrolimus 0.03% and setting the efficacy of crisaborole equal to that of pimecrolimus, but again no change was made to the duration of treatment for pimecrolimus.

The ERG were unable to replicate the results in Table 10 by then setting the efficacy of crisaborole equal to that of tacrolimus 0.03%. The results presented in Table 10 by the company lack face validity because the QALY gains are not equivalent between crisaborole and tacrolimus 0.03% suggesting that there is an error in these results. The ERG believe that the results in Table 10 of the CRTE were obtained in error by setting crisaborole to have equivalent efficacy to tacrolimus 0.1% instead of tacrolimus 0.03%. The ERG has produced corrected results for this scenario in Table 1 below.

Table 1: Children with moderate disease: Crisaborole efficacy equivalent to tacrolimus 0.03% and assuming 6 weeks of treatment tacrolimus 0.03%†

	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER vs. baseline	ICER incremental
Crisaborole	[redacted]	[redacted]	[redacted]	[redacted]	-	Dominant
Tacrolimus 0.03%	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominated
Pimecrolimus	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	Dominated

† ERG correction to Table 10 of the CRTE in which efficacy was mistakenly set equal to tacrolimus 0.1%

The ERG was able to replicate the results for adults in Tables 12 to 14 by setting the duration of treatment for tacrolimus 0.03% and tacrolimus 0.1% to 6 weeks and setting the efficacy of crisaborole to the appropriate TCI option.

The ERG notes that the results presented in Tables 9 to 14 of the CRTE do not apply a 6 week duration for pimecrolimus but they do assume a 6 week duration for both tacrolimus 0.03% and tacrolimus 0.1%. This is despite the company stating on page 12 that they assumed *"up to 6 weeks of therapy per treatment cycle for TCIs*" and not just tacrolimus. The ERG notes that results assuming 6 weeks of treatment with pimecrolimus are presented in the bottom row in Tables 15 and 16 of the CRTE and in these scenarios crisaborole dominates due to the additional costs incurred over these additional two weeks.

The ERG was able to validate the results in Tables 15 to 17 of the CRTE. It is noted that in Tables 17 and 18 which present results for moderate AD where there is more than one comparator treatment, the efficacy of treatment for all three options has been set to the same value. This is in contrast to Tables 10, 11, 13 and 14, where the efficacy of crisaborole was set to match one comparator but the efficacy of the other comparator was left at its original value. The cost assumptions applied for the three options in Tables 15 to 17 are summarised in Table 2. These are based on the description provided by the company on page 7 to 8 of the CRTE and the ERG's attempts to replicate the results in Tables 15 to 18. However, it should be noted that the ERG were unable to replicate the costs presented in the final row of Table 18 of the CRTE, although it believes that the costs in this row should match those in Table 14 so these results are still provided in the CRTE.

Despite the ERG being unable to replicate some results, it is clear that when assuming equivalent efficacy, and equivalent treatment duration, crisaborole is never cost saving. In contrast, when assuming that TCIs are used for 6 weeks instead of 4 weeks, the treatment with the lowest cost is always crisaborole. Therefore, the key decision uncertainty is whether the cost of TCIs should be assumed to apply for 6 weeks or 4 weeks.

Table 2 : Cost assumptions applied in the equivalence scenarios (Tables 15 to 18 of the CRTE)

Company's description of scenario	Crisaborole dosing	Pimecrolimus dosing	Tacrolimus 0.03% in children	Tacrolimus 0.03% in adults	Tacrolimus 0.1%
Equivalent efficacy	4 weeks of twice daily	4 weeks of twice daily	3 weeks of twice daily and 1 week of once daily	4 weeks of twice daily	4 weeks of twice daily
Equivalent efficacy and dosing	4 weeks of twice daily	4 weeks of twice daily	4 weeks of twice daily	4 weeks of twice daily	4 weeks of twice daily
Equivalent efficacy and 6 weeks of therapy for TCIs	4 weeks of twice daily	6 weeks of twice daily	3 weeks of twice daily and 3 weeks of once daily	6 weeks of twice daily*	6 weeks of twice daily*

*the results for adults with moderate disease for this options could not be replicated so the ERG cannot verify if this was the assumption applied in adults

The ERG wishes to point out that the company's assumption that patients receive 6 weeks of initial treatment because the summary of product characteristics (SmPCs) for TCIs specify that they can be used for up to 6 weeks, is at odds with their modelling of partial responders who would receive 6 weeks of treatment cost during the first model cycle and then 6 weeks of treatment costs in the second model cycle. Given that the company claims that TCI treatment should be given for up to 6 weeks based on the SmPCs for TCIs, it seems unreasonable for the model to include costs for up to 12 weeks for the proportion who have a partial response. In the company's original base-case model, patients received 4 weeks of treatment in the initial 4 week cycle, and partial responders received a second cycle of 4 weeks treatment allowing them to receive up to 8 weeks of treatment with TCIs, which already is longer than the 6 weeks that the company claims is the maximum duration of treatment for TCIs based on the SmPCs for TCIs. The ERG would also point out that the draft SmPC for crisaborole (as described in Table 2 of the company submission) states, *"Staquis can be used for up 4 weeks per treatment course. If any signs/and or symptoms persist, or new areas affected with atopic dermatitis appear, further treatment courses can be used. Staquis should be discontinued if signs and/or symptoms on treated areas persist after 3 consecutive treatment courses of 4 weeks each or if the signs and/or symptoms worsen during treatment."* Therefore, the draft SmPC for crisaborole suggests that it can be used for up to 12 weeks, but cost for 12 weeks of crisaborole are not explored in any of the company's scenario analyses.

The company's rationale for assuming 6 weeks of TCI treatment is that they had previously assumed that the costs for tacrolimus reflected 4 weeks usage, even though, the efficacy data for tacrolimus used in the NMA, captured outcomes for up to 6 weeks tacrolimus treatment. The ERG noted that whilst three of the TCI trials included in the efficacy NMA reported outcomes at 6 weeks,[2-4] two studies reported outcomes at four weeks [5, 6] and two studies reported outcomes at both 4 weeks and 6 weeks,[7, 8] but the 4 week data was included in the NMA as this was closer to the duration of the crisaborole studies. In addition, all of the 5 studies that had a duration longer than 4 weeks,[2-4, 7, 8] mentioned that patients could stop TCIs early if symptoms cleared. Therefore, whilst treatment was allowed to be continued up to 6 weeks in some of the TCI studies included in the NMA, it is not clear that all patients required 6 weeks of treatment to achieve the response rates incorporated in the NMA, and outcomes in the NMA were based on data from 4 weeks in 4 of the 7 studies.

The ERG also notes that all of the comparisons assuming equivalence of efficacy are heavily dependent on the fact that it assumed that the same amount of treatment is needed per application for crisaborole and all TCIs and that there is no wastage due to a mismatch between the tube size and the total amount needed to treat a single flare. In practice, the costs could be quite different if one intervention needed just under 2 tubes of treatment and the other needed just over 2 tubes. The bottom line is that these interventions have [redacted]

5 IMPACT ON THE ICER OF ADDITIONAL ECONOMIC ANALYSES UNDERTAKEN BY THE ERG

The ERG ran all exploratory analyses deterministically using the latest cost for tacrolimus 0.1% (£34.52 per tube). A summary of the exploratory analyses undertaken by the ERG is presented in Table 3 for children with mild AD, Table 4 for children with moderate AD, Table 5 for adults with mild AD, and Table 6 for adults with moderate AD. It can be seen from these results that when using the simple random effects NMA, crisaborole is dominated by TCIs in all four populations. Conversely when using the MAIC, TCIs are dominated by crisaborole in all populations except adults with moderate AD. Therefore, the key area of decision uncertainty relates to whether the [redacted] as this determines whether the results from the simple random effects NMA or MAIC should be preferred.

There is only one scenario using the MAIC where TCIs are not dominated by crisaborole, and this is the comparison between tacrolimus 0.1% and crisaborole in adult patients with moderate AD. In this case crisaborole is associated with a small additional cost and a small QALY gain with an ICER of [redacted] per QALY gained for crisaborole versus tacrolimus 0.1%. For comparison, the ICER for crisaborole versus tacrolimus 0.1% in adults with moderate AD was [redacted] per QALY gained in the company's revised model when selecting the MAIC and using the latest cost for tacrolimus 0.1%.

It should be noted that only pair-wise cost-effectiveness analyses can be presented when using the MAIC as the rate of response for crisaborole is estimated based on the MAIC specific to the individual comparator. Therefore, the rate of response for crisaborole when comparing against tacrolimus 0.03% will be different to the rate of response for crisaborole when comparing against tacrolimus 0.1%. For this reason, incremental analysis cannot be conducted when using the MAIC if there are two or more comparators.

Additionally, the ERG highlights that the small decrement in tacrolimus 0.1% price (from £38.46 to £34.52 per tube) had a considerable impact on the analyses for the adult moderate population, and crisaborole does not dominate tacrolimus 0.1% anymore, even in the company's base case. This illustrates that the cost differences between comparators are minimal, and the conclusion that one or another treatment dominates would be sensitive to any price changes that could happen in the future.

Table 3: ERG exploratory model results for mild child patients

Analysis	Discounted costs		Discounted QALYS		ICER (crisaborole versus pimecrolimus)
	Crisaborole	Pimecrolimus	Crisaborole	Pimecrolimus	
Company base case	[redacted]	[redacted]	[redacted]	[redacted]	**Crisaborole dominates** [redacted]
1) Correcting acquisition costs of subsequent systemic therapy	[redacted]	[redacted]	[redacted]	[redacted]	Crisaborole dominates [redacted]
2) Assuming non responders receive 4 weeks of treatment	[redacted]	[redacted]	[redacted]	[redacted]	Crisaborole dominates [redacted]
ERG base case (scenarios 1 – 2) using the company's preferred NMA	[redacted]	[redacted]	[redacted]	[redacted]	**Crisaborole dominates** [redacted]
3) Adjusting costs of subsequent therapy to reflect the whole time on treatment†	[redacted]	[redacted]	[redacted]	[redacted]	Crisaborole dominates [redacted]
ERG base case (scenarios 1 – 2) using the company's MAIC results	[redacted]	[redacted]	[redacted]	[redacted]	**Crisaborole dominates** [redacted]
Company base case using the ERG's simple random effects NMA results	[redacted]	[redacted]	[redacted]	[redacted]	**Pimecrolimus dominates** [redacted]
4) Correcting acquisition costs of subsequent systemic therapy	[redacted]	[redacted]	[redacted]	[redacted]	Pimecrolimus dominates [redacted]
5) Assuming non responders receive 4 weeks of treatment	[redacted]	[redacted]	[redacted]	[redacted]	Pimecrolimus dominates [redacted]
ERG base case (scenarios 4 – 5) using the ERG's simple random effects NMA results	[redacted]	[redacted]	[redacted]	[redacted]	**Pimecrolimus dominates** [redacted]
6) Adjusting costs of subsequent therapy to reflect the whole time on treatment†	[redacted]	[redacted]	[redacted]	[redacted]	Pimecrolimus dominates [redacted]

†In conjunction with the ERG base case mentioned above
ΔC, difference in costs, ΔQ, difference in QALYs; ICER, incremental cost-effectiveness ratio; MAIC, matching adjusted indirect comparison; QALY, quality adjusted life year

Table 4: ERG exploratory model results for moderate child patients

Analysis	Discounted costs			Discounted QALYS			ICER (crisaborole versus tacrolimus 0.03%[†])
	Crisaborole	Tacrolimus 0.03%	Pimecrolimus	Crisaborole	Tacrolimus 0.03%	Pimecrolimus	
Company base case	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	**Crisaborole dominates tacrolimus 0.03%[†]** [redacted]
1) Correcting acquisition costs of subsequent systemic therapy	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Crisaborole dominates tacrolimus 0.03%[†] [redacted]
2) Assuming non responders receive 4 weeks of treatment	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Crisaborole dominates tacrolimus 0.03%[†] [redacted]
ERG base case (scenarios 1 – 2) using the company's preferred NMA	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	**Crisaborole dominates tacrolimus 0.03%[†]** [redacted]
3) Adjusting costs of subsequent therapy to reflect the whole time on treatment[†]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Crisaborole dominates tacrolimus 0.03%[†] [redacted]
ERG base case (scenarios 1 – 2) using the company's MAIC results (Crisaborole vs tacrolimus 0.03%)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	**Crisaborole dominates tacrolimus 0.03%[†]** [redacted]
ERG base case (scenarios 1 – 2) using the company's MAIC results (Crisaborole vs	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	**Crisaborole dominates pimecrolimus**

Analysis	Discounted costs			Discounted QALYS			ICER (crisaborole versus tacrolimus 0.03%†)
	Crisaborole	Tacrolimus 0.03%	Pimecrolimus	Crisaborole	Tacrolimus 0.03%	Pimecrolimus	
pimecrolimus)							[redacted]
Company base case using the ERG's simple random effects NMA results	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	**Tacrolimus 0.03%† dominates crisaborole** [redacted]
4) Correcting acquisition costs of subsequent systemic therapy	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Tacrolimus 0.03%† dominates crisaborole [redacted]
5) Assuming non responders receive 4 weeks of treatment	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Tacrolimus 0.03%† dominates crisaborole [redacted]
ERG base case (scenarios 4 – 5) using the ERG's simple random effects NMA results	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	**Tacrolimus 0.03%† dominates crisaborole** [redacted]
6) Adjusting costs of subsequent therapy to reflect the whole time on treatment†	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Tacrolimus 0.03%† dominates crisaborole [redacted]

†Tacrolimus 0.03% always dominates pimecrolimus in all scenarios
†In conjunction with the ERG base case mentioned above
ΔC, difference in costs, ΔQ, difference in QALYs; ICER, incremental cost-effectiveness ratio; MAIC, matching adjusted indirect comparison; QALY, quality adjusted life year
NR = not reportable – the analyses based on the MAIC can only be used to conduct pairwise comparisons as the MAIC estimates different response rates for crisaborole when different comparators are selected

Table 5: ERG exploratory model results for mild adult patients

Analysis	Discounted costs		Discounted QALYS		ICER (crisaborole versus pimecrolimus)
	Crisaborole	Pimecrolimus	Crisaborole	Pimecrolimus	
Company base case	[redacted]	[redacted]	[redacted]	[redacted]	**Crisaborole dominates** [redacted]
1) Correcting acquisition costs of subsequent systemic therapy	[redacted]	[redacted]	[redacted]	[redacted]	Crisaborole dominates [redacted]
2) Assuming non responders receive 4 weeks of treatment	[redacted]	[redacted]	[redacted]	[redacted]	Crisaborole dominates [redacted]
ERG base case (scenarios 1 – 2) using the company's preferred NMA	[redacted]	[redacted]	[redacted]	[redacted]	**Crisaborole dominates** [redacted]
3) Adjusting costs of subsequent therapy to reflect the whole time on treatment†	[redacted]	[redacted]	[redacted]	[redacted]	Crisaborole dominates [redacted]
ERG base case (scenarios 1 – 2) using the company's MAIC results	[redacted]	[redacted]	[redacted]	[redacted]	**Crisaborole dominates** [redacted]
Company base case using the ERG's simple random effects NMA results	[redacted]	[redacted]	[redacted]	[redacted]	**Pimecrolimus dominates** [redacted]
4) Correcting acquisition costs of subsequent systemic therapy	[redacted]	[redacted]	[redacted]	[redacted]	Pimecrolimus dominates [redacted]
5) Assuming non responders receive 4 weeks of treatment	[redacted]	[redacted]	[redacted]	[redacted]	Pimecrolimus dominates [redacted]
ERG base case (scenarios 4 – 5) using the ERG's	[redacted]	[redacted]	[redacted]	[redacted]	**Pimecrolimus dominates** [redacted]

Analysis	Discounted costs		Discounted QALYS		ICER (crisaborole versus pimecrolimus)
	Crisaborole	Pimecrolimus	Crisaborole	Pimecrolimus	
simple random effects NMA results					
6) Adjusting costs of subsequent therapy to reflect the whole time on treatment†	[redacted]	[redacted]	[redacted]	[redacted]	Pimecrolimus dominates [redacted]

†In conjunction with the ERG base case mentioned above
ΔC, difference in costs, ΔQ, difference in QALYs; ICER, incremental cost-effectiveness ratio; MAIC, matching adjusted indirect comparison; QALY, quality adjusted life year

Table 6: ERG exploratory model results for moderate adult patients (using £34.52 as the cost per tube of tacrolimus 0.1%)

Analysis	Discounted costs			Discounted QALYS			ICER (crisaborole versus tacrolimus 0.1%†)
	Crisaborole	Tacrolimus 0.1%	Tacrolimus 0.03%	Crisaborole	Tacrolimus 0.1%	Tacrolimus 0.03%	
Company base case	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted] **per QALY versus tacrolimus 0.1%†** [redacted]
1) Correcting acquisition costs of subsequent systemic therapy	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	Crisaborole dominates tacrolimus 0.1%† [redacted]
2) Assuming non responders receive 4 weeks of treatment	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted] per QALY tacrolimus 0.1%† [redacted]
ERG base case (scenarios 1 – 2) using the company's preferred NMA	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted] **per QALY versus tacrolimus 0.1%†** [redacted]
3) Adjusting costs of subsequent therapy to reflect	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted] per QALY tacrolimus 0.1%†

Analysis	Discounted costs			Discounted QALYS			ICER (crisaborole versus tacrolimus 0.1%†)
	Crisaborole	Tacrolimus 0.1%	Tacrolimus 0.03%	Crisaborole	Tacrolimus 0.1%	Tacrolimus 0.03%	
the whole time on treatment†							
ERG base case (scenarios 1 – 2) using the company's MAIC results (Crisaborole vs tacrolimus 0.1%)							**per QALY versus tacrolimus 0.1%**
ERG base case (scenarios 1 – 2) using the company's MAIC results (Crisaborole vs tacrolimus 0.03%)							**Crisaborole dominates tacrolimus 0.03%**
Company base case using the ERG's simple random effects NMA results							**Tacrolimus 0.1% dominates crisaborole**
4) Correcting acquisition costs of subsequent systemic therapy							Tacrolimus 0.1% dominates crisaborole
5) Assuming non responders receive 4 weeks of treatment							Tacrolimus 0.1% dominates crisaborole
ERG base case (scenarios 4 – 5) using the ERG's simple random effects NMA results							**Tacrolimus 0.1% dominates crisaborole**
6) Adjusting costs of subsequent therapy to reflect the whole time on treatment†							Tacrolimus 0.1% dominates crisaborole

†Tacrolimus 0.1% always dominates Tacrolimus 0.03% in all scenarios
†In conjunction with the ERG base case mentioned above
ΔC, difference in costs, ΔQ, difference in QALYs; ICER, incremental cost-effectiveness ratio; MAIC, matching adjusted indirect comparison; QALY, quality adjusted life year

Analysis	Discounted costs			Discounted QALYS			ICER (crisaborole versus tacrolimus 0.1%[†])
	Crisaborole	Tacrolimus 0.1%	Tacrolimus 0.03%	Crisaborole	Tacrolimus 0.1%	Tacrolimus 0.03%	

NR = not reportable – the analyses based on the MAIC can only be used to conduct pairwise comparisons as the MAIC estimates different response rates for crisaborole when different comparators are selected

6 OVERALL CONCLUSIONS

The company has not provided any updated analyses using either the MAIC or the simple random effects NMA. However, the ERG's exploratory analyses using these show that they provide very different results with crisaborole being dominated by TCIs when using the simple random effects NMA and crisaborole dominating TCIs when using the MAIC in all except one population (adults with moderate AD). Therefore, the key area of decision uncertainty relates to whether the [redacted] as this determines whether the results from the simple random effects NMA or MAIC should be preferred. The ERG notes that double blind placebo controlled trials comparing crisaborole head-to-head against TCIs would be the best way to determine the relative effectiveness of crisaborole and TCIs and details of ongoing studies including some that compare against TCIs are provided in section 4.2.9 of the ERG report.

In the CRTE it is argued that crisaborole should be compared to [redacted]. However, the ERG would argue that a full incremental analysis should always be conducted to determine whether crisaborole is cost-effective compared to the comparator which reflects the most cost-effective use of NHS resources, [redacted]
The ERG does not accept that crisaborole is cost-saving relative to TCIs on the basis that it will be used for 4 weeks instead of 6 weeks as the draft SmPC for crisaborole suggests that it may be used [redacted] and the company has not presented any comparison assuming longer than 4 weeks treatment with crisaborole. The ERG would argue that head-to-head studies of crisaborole versus TCIs would be needed to determine whether one treatment or the other required a longer duration to achieve an adequate response.

The company repeatedly claims in their response to technical engagement that each of the changes made to the model indicate that the issues raised in the technical engagement report are not significant areas of decision uncertainty. The ERG would agree with this, in so much that the key area of decision uncertainty remains whether crisaborole is more, less or equally as effective as TCIs in achieving a response in mild to moderate AD. The other areas of uncertainty are only relevant in determining the likely size of cost savings or QALY gains for the more effective therapy, but it is still important to determine these accurately. In addition, the ERG believes that there remains considerable uncertainty regarding whether having any additional treatment option for managing a mild to moderate AD flare is likely to result in fewer patients receiving subsequent treatments such as phototherapy or systemic therapies further down the treatment pathway given that these subsequent treatments are usually reserved for patients with severe AD.

7 REFERENCES

1. Ridd MJ, Roberts A, Grindlay D, Williams HC. Which emollients are effective and acceptable for eczema in children? *BMJ* 2019;367:l5882.
2. Chapman MS, Schachner LA, Breneman D, Boguniewicz M, Gold MH, Shull T*, et al.* Tacrolimus ointment 0.03% shows efficacy and safety in pediatric and adult patients with mild to moderate atopic dermatitis. 2005;53:S177-S85.
3. Abramovits W, Fleischer AB, Jr., Jaracz E, Breneman D. Adult patients with moderate atopic dermatitis: tacrolimus ointment versus pimecrolimus cream. *J Drugs Dermatol* 2008;7:1153-8.
4. Paller AS, Lebwohl M, Fleischer AB, Jr., Antaya R, Langley RG, Kirsner RS*, et al.* Tacrolimus ointment is more effective than pimecrolimus cream with a similar safety profile in the treatment of atopic dermatitis: results from 3 randomized, comparative studies. *J Am Acad Dermatol* 2005;52:810-22.
5. Eichenfield LF, Lucky AW, Boguniewicz M, Langley RG, Cherill R, Marshall K*, et al.* Safety and efficacy of pimecrolimus (ASM 981) cream 1% in the treatment of mild and moderate atopic dermatitis in children and adolescents. *J Am Acad Dermatol* 2002;46:495-504.
6. Levy AL SM, Roberts R. Tacrolimus Cream 0.03% Is Safe and Effective in the Treatment of Mild and Moderate Atopic Dermatitis in Adults. *J Allergy Clin Immunol* 2005;115:S103.
7. Schachner LA, Lamerson C, Sheehan MP, Boguniewicz M, Mosser J, Raimer S*, et al.* Tacrolimus ointment 0.03% is safe and effective for the treatment of mild to moderate atopic dermatitis in pediatric patients: results from a randomized, double-blind, vehicle-controlled study. *Pediatrics* 2005;116:e334-42.
8. Kempers S, Boguniewicz M, Carter E, Jarratt M, Pariser D, Stewart D*, et al.* A randomized investigator-blinded study comparing pimecrolimus cream 1% with tacrolimus ointment 0.03% in the treatment of pediatric patients with moderate atopic dermatitis. *J Am Acad Dermatol* 2004;51:515-25.