Baricitinib for treating severe alopecia areata [ID 3979]

STA Report

Source of funding

This report was commissioned by the NIHR Evidence Synthesis Programme as project number 135648T.

Title:	Baricitinib for treating severe alopecia areata
Produced by:	BMJ Technology Assessment Group (BMJ TAG)
Authors:	Steve Edwards, Director of Health Technology Assessment, BMJ-TAG, London Ben Farrar, Senior Clinical Evidence Analyst, BMJ-TAG, London Tracey Jhita, Health Economist Manager, BMJ-TAG, London Victoria Wakefield, Principal Clinical Evidence Analyst, BMJ-TAG, London Archie Walters, Health Economist, BMJ-TAG, London
Correspondence to:	Steve Edwards, BMJ TAG, BMJ, BMA House, Tavistock Square, London, WC1H 9JR.
Date completed:	13/10/20211
Source of funding:	This report was commissioned by the NIHR Evidence Synthesis Programme as project number 135648T.
Declared competing interests of the authors	Ben Farrar was employed by Costello Medical between September 2020 to December 2020 and November 2021 to May 2022 but did not work on baricitinib or any of its comparators while employed there. BMJ Technology Assessment Group (BMJ-TAG) and the editorial team of The BMJ work independently to one another. The views and opinions expressed in this report are those of the BMJ-TAG.
Acknowledgments:	The EAG would like to thank Professor Andrew Messenger, Honorary Professor of Dermatology, Department of Infection, Immunity and Cardiovascular Disease, University of Sheffield and Dr Susan Holmes, Consultant Dermatologist, Alan Lyell Centre for Dermatology, Glasgow Royal Infirmary and Honorary Associate Clinical Professor, University of Glasgow for providing clinical advice throughout the project. The EAG would also like to thank Dr Leila Asfour, Consultant Dermatologist, Department of Dermatology, Salford Royal Foundation Trust, and Dr Nekma Meah, Consultant Dermatologist, St Helens and Knowsley NHS Trust and Honorary Clinical Lecturer, University of Manchester for providing feedback on the clinical sections of the report.
Rider on responsibility for report:	The views expressed in this report are those of the authors and not necessarily those of the NIHR Evidence Synthesis Programme. Any errors are the responsibility of the authors.
Report reference:	Edwards SJ, Farrar B, Jhita T, Wakefield V and Walters A. Baricitinib for treating severe alopecia areata : A Single Technology Appraisal. BMJ Technology Assessment Group, 2022.

Contribution of authors:

Steve Edwards	Critical appraisal of the company's submission; validated the statistical analyses; provided feedback on all versions of the report. Guarantor of the report.
Ben Farrar	Critical appraisal of the company's submission; critical appraisal of the clinical evidence; and drafted the summary, background and clinical results sections.
Victoria Wakefield	Critical appraisal of the company's submission; critical appraisal of the clinical evidence; cross checking of company's search strategies; and assisted with drafting the clinical results sections.
Tracey Jhita	Critical appraisal of the company's submission; critical appraisal of the economic model; cross checking of company's search strategies; critical appraisal of the economic evidence; carried out the economic analyses; and drafted the economic sections
Archie Walters	Critical appraisal of the company's submission; critical appraisal of the economic model; and drafted the economic sections

All authors read and commented on draft versions of the EAG report.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AA	Alopecia areata
AE	Adverse event
AESI	Adverse event of special interest
ANCOVA	Analysis of covariance
AU	Alopecia universalis
BAD	British Association of Dermatologists
BARI	Baricitinib
BID	Twice per day
BSC	Best supportive care
CENTRAL	Cochrane Central Register of Controlled Trials
CDSR	Cochrane Database of Systematic Reviews
CEAC	Cost-effectiveness acceptability curve
CI	Confidence interval
ClinRO	Clinician reported outcome
CMU	Confidential medicines unit
CRD	Centre for Reviews and Dissemination
CS	Company submission
CSR	Clinical study report
DLQI	Dermatology Life Quality Index
DPCP	Diphenylcyclopropenone
DSA	Deterministic sensitivity analysis
DSP	Disease Specific Programme
DSU	Decision Support Unit
EAG	Evidence assessment group
EB	Eyebrow
EL	Eyelash
EMA	European Medicines Agency
eMIT	Drugs and pharmaceutical electronic market information tool
EQ-5D	European Quality of Life-5 Dimensions
FAS	Ful analysis set
GP	General practitioner
HADS	Hospital Anxiety Depression Scale
HI	High intensity
HRQoL	Health-related quality of life
HSUV	Health state utility values
ICER	Incremental cost-effectiveness ratio
IL	Intralesional
IMT	Immunotherapy

ITC	Indirect treatment comparison
IWRS	Interactive web response system
JAK	Janus kinase
LSM	Least squares mean
LYG	Life-years gained
MACE	Major adverse cardiovascular event
MBCR	Mindfulness-based cognitive therapy
MHRA	Medicines and Healthcare products Regulatory Agency
mLOCF	Modified last observation carried forward
NHB	Net health benefit
NHS	National Health Service
NICE	National Institute for Health and Care Excellent
NMSC	Nonmelanoma skin cancer
NMA	Network meta-analysis
NMB	Net monetary benefit
NRI	Non-responder imputation
ONS	Office for National Statistics
OWSA	One-way sensitivity analysis
PAS	Patient Access Scheme
PBO	Placebo
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRO	Patient reported outcome
PSA	Probabilistic sensitivity analysis
PSSRU	Personal Social Services Research Unit
PWP	Psychological well-being practitioner
PTFU	Post-trial follow up
QALY	Quality-adjusted life year
QD	Once daily
RCT	Randomised controlled trial
SAE	Serious adverse event
SALT	Severity of Alopecia Tool
SD	Standard deviation
SE	Standard error
SF-36	Medical outcomes study 36-item short form health survey
SLR	Systematic literature review
STA	Single technology appraisal
STAT	Signal transducers and activators of transcription
TCS	Topical corticosteroids
TEAE	Treatment emergent adverse event
TSD	Technical Support Document

UK	United Kingdom
US	United States
VAS	Visual analogue scale
W	Week
WPAI	Work productivity and activity impairment
WTP	Willingness to pay

1 Executive summary

This summary provides a brief overview of the key issues identified by the Evidence Assessment Group (EAG) as being potentially important for decision making. It also includes the EAG's preferred assumptions and the resulting incremental cost-effectiveness ratios (ICERs).

Section 1.1 provides an overview of the key issues. Section 1.2 provides an overview of key model outcomes and the modelling assumptions that have the greatest effect on the ICER. Sections 1.3 to 1.6 explain the key issues in more detail. Background information on the condition, technology and evidence and information on non-key issues are in the main EAG report.

All issues identified represent the EAG's view, not the opinion of NICE.

1.1 Overview of the EAG's key issues

Table 1 presents a summary of the EAG's key issues on the evidence submitted on the clinical and cost effectiveness baricitinib for treating adults with severe alopecia areata (AA).

Table 1. Summary of key issues

ID xxx	Summary of issue	Report sections
1	Definition of the comparator	2.2.1, 2.3.3, 3.4, 4.2.3
2	Definition of treatment response at Week 36	4.2.5
2	Source of utilities in the model	4.2.8
4	Disease monitoring costs for best supportive care	4.2.9

The key differences between the company's preferred assumptions and the EAG's preferred assumptions are around the definition of the comparator and treatment response at Week 36, the source of utilities used in the model, and the assumptions of costs incurred in the best supportive care (BSC) health state. However, other secondary differences in the preferred assumptions between the company and EAG's approach include how long-term all-cause treatment discontinuation for baricitinib is calculated, inclusion of adverse events (AEs), removal of non-pharmacological psychological support costs and wig resource use in the induction phase of the model.

It should be noted that for AEs, the EAG was unable to verify the inputs used in the company's AE scenario provided in their clarification response and was unable to produce an alternative scenario due to a paucity of time. Nonetheless, the EAG requests that during technical engagement, the company provides a more thorough description and justification of their approach to the inclusion of AEs and assumed unit costs to treat each AE and update the scenario if necessary.

1.2 Overview of key model outcomes

NICE technology appraisals compare how much a new technology improves length (overall survival) and quality of life in a quality-adjusted life year (QALY). An ICER is the ratio of the extra cost for every QALY gained.

Overall, the technology is modelled to affect QALYs by:

- Improving and maintaining scalp hair regrowth.

Overall, the technology is modelled to affect costs by:

- Its higher unit price than established clinical management.

The modelling assumptions that have the greatest effect on the ICER are:

- Changing the definition of the comparator to 'discharged from care' and removing all monitoring costs in the induction phase and Maintenance health state as a result.
- Using utilities in the model sourced from the key trials of baricitinib, BRAVE-AA1 and BRAVE-AA2.
- Removing the costs associated with disease management in the BSC health state.

1.3 The decision problem: summary of the EAG's key issues

Table 2. Issue 1: Definition of the comparator

Report section	2.2.1, 2.3.3, 3.4, 4.2.3
Description of issue and why the EAG has identified it as important	The company's comparator is "Watch and wait", which is defined as no active treatment but frequent monitoring. In contrast, the EAG's clinical experts advised that, although no active treatment is a common management strategy for adults with severe AA, the company's definition of "Watch and wait" did not capture this adequately. The EAG considered three alternative comparators: • Treatment with DPCP, the most effective treatment currently used to treat severe AA in adults, which is the only active treatment recommended by the British Association of Dermatologists Guidelines for treating severe AA; • Treatment from a basket of "low-effectiveness" non-DPCP therapies sometimes used to treat severe AA in adults, primarily systemic immunosuppressants and systemic corticosteroids; • No active treatment and discharge from care. The EAG concluded that: • DPCP is not a reasonable comparator for the appraisal as DPCP is only available to a minority of patients with inequitable access and is associated with severe adverse events and a high rate of relapse; • No active treatment and discharge from care is the most commonly used approach for the prevalent population of adults with severe AA who would be eligible to receive baricitinib at the point of approval in the UK; • While systemic immunosuppressants and systemic steroids could be considered appropriate comparators for newly diagnosed cases of severe AA, their use is too heterogenous and their effectiveness too limited to be considered an established standard of care for severe AA. In lieu of robust treatment pattern data or comparative effectiveness data with baricitinib, the EAG considers no active treatment and discharge from care to be an acceptable comparator for this population.
What alternative approach has the EAG suggested?	The EAG recommends no active treatment and discharge from care as the most appropriate comparator for this appraisal. As such, in the economic model, the EAG considers that monitoring costs included for the induction phase and maintenance health state should be removed.
What is the expected effect on the cost-effectiveness estimates?	The impact on the company's ICER post clarification when monitoring costs are removed in the induction phase and maintenance health state for 'Watch and wait' (which redefines the comparator to 'discharged from care'), increases from £18,072 to £20,887.
What additional evidence or analyses might help to resolve this key issue?	Comprehensive treatment pattern data for AA and severe AA from a range of care settings in the UK would help to resolve some of the uncertainty in the treatment pathway of AA, especially for newly diagnosed severe AA patients. The EAG's clinical experts highlighted that such data do not exist to their knowledge.

Abbreviations: AA: alopecia areata: DPCP: diphencyprone EAG: evidence assessment group; ICER; incremental cost-effectiveness ratio; ITC: indirect treatment comparison NMA: network meta-analysis

1.4 The clinical and cost-effectiveness evidence: summary of the EAG's key issues

Table 3. Issue 2: Definition of treatment response at Week 36

Report section	4.2.5
Description of issue and why the EAG has identified it as important	In the company's base case, the primary outcome in the model was $SALT_{50}$ (defined as at least a 50% improvement from baseline SALT score). In addition to the outcome of $SALT_{50}$, the company also included the outcome of $SALT_{75}$ (defined as at least a 75% improvement from baseline SALT score), as a way of capturing additional quality of life benefit associated with achieving an increased relative improvement in scalp hair growth. In the key trials of BRAVE-AA1 and BRAVE-AA2, the primary endpoint was the proportion of patients achieving SALT≤20 at Week 36. A response of SALT≤20 indicated scalp hair loss of less than 20% (or ≥80% scalp coverage with hair). The EAG considers using SALT≤20 to be a more clinically meaningful benefit for patients. The EAG's clinical experts noted that a relative benefit of $SALT_{50}$ or $SALT_{75}$, is unlikely to be meaningful to patients unless it results in a similar increase in coverage to SALT≤20.
What alternative approach has the EAG suggested?	The EAG's preferred approach is to use SALT≤20 as the definition of treatment response at Week 36.
What is the expected effect on the cost-effectiveness estimates?	By using the outcome of SALT≤20 at Week 36, the company's ICER post clarification reduced from £18,072 to £17,071 for the overall population. For the severe and very severe subgroups, the company's ICER post clarification changed from £25,154 (severe) and £12,685 (very severe) to £18,773 (severe) and £16,929 (very severe), respectively.
What additional evidence or analyses might help to resolve this key issue?	No additional evidence required as the scenario resolves the issue.

Abbreviations: EAG, Evidence Assessment Group; ICER, incremental cost-effectiveness ratio; SALT, Severity of Alopecia Tool

Table 4. Issue 3: Source of utilities in the model

Report section	4.2.8
Description of issue and why the EAG has identified it as important	The BRAVE-AA1 and BRAVE-AA2 trials collected EQ-5D data up to Week 36 directly from patients but the company stated that the values obtained from the trials were insensitive to changes in the severity of AA and lacked content validity as baseline values were almost the same as UK age- and sex-adjusted general population values. Additionally, the company stated that [redacted] of participants in the BRAVE-AA1 and BRAVE-AA2 trials reported a score of perfect health at baseline (score of 11111) and as such an improvement in HRQoL would not be obtained at Week 36 for these patients. Thus, the utility values informing the economic model were derived from a company sponsored Adelphi DSP study. The company explained that in the Adelphi DSP study, the ceiling effect was also observed, but not to the same extent. However, the company did not provide the overall proportion of patients reporting a score of perfect health from the Adelphi DSP study.

	The EAG considers the company's justification for not using pooled EQ-5D data from the BRAVE-AA1 and BRAVE-AA2 trials is a criticism of the EQ-5D tool and not the methods to obtain the data used in the trial and thus extends to the EQ-5D data obtained from the company sponsored Adelphi DSP study. Furthermore, the company hasn't supplied sufficient evidence to validate the lack of content validity with the EQ-5D nor has it demonstrated why patients should have a substantial change in their QoL. As recommended in the NICE methods guide, the reference case should report the measurement of changes in health-related quality of life directly from patients. As such, the EAG considers the pooled EQ-5D data from the BRAVE-AA1 and BRAVE-AA2 trials represents a more robust source of utility data that matches the NICE reference case and should be used in the cost-effectiveness analysis for the base case.
What alternative approach has the EAG suggested?	During the clarification stage, the EAG requested, and the company provided, change from baseline at Week 36 for patients achieving SALT≤20 based on pooled EQ-5D data from the BRAVE-AA1 and BRAVE-AA2 trials. Thus, when using the outcome of SALT≤20 for treatment response at Week 36, the baseline utility and change from baseline associated with achieving SALT≤20 should be used in the model.
What is the expected effect on the cost-effectiveness estimates?	When implementing the baseline utility and change from baseline utility associated with achieving SALT≤20 in combination with using SALT≤20 at Week 36, the company's ICER post clarification increases from £18,072 to £118,494 for the overall population. For the severe and very severe subgroups, the ICERs are £130,303 and £117,510, respectively.
What additional evidence or analyses might help to resolve this key issue?	The EAG's clinical experts advised that for most patients' HRQoL may only be mildly affected and thus may not be that different to the general population but equally HRQoL is severely affected for a few patients (primarily driven by adverse mental health). Additionally, the EAG's clinical experts advised that overtime, some patients may come to terms with their hair loss, while a few may remain distressed about their condition. Thus, the EAG acknowledges that there is a small, but heterogenous, patient population that is more adversely affected in terms of HRQoL but that the demographics of this population are difficult to identify clinically and consistently, and it is beyond the scope of this assessment to identify that group. Nonetheless, the EAG has estimated the QALY gain needed to reach the £20,000 and £30,000 cost-effectiveness thresholds and advises the committee to consider if the estimated QALY gain needed for baricitinib 4 mg is plausible for the condition under consideration.

Abbreviations: AA, alopecia areata; DSP, disease specific programme; EAG, Evidence Assessment Group; HRQoL, health-related quality of life; ICER, incremental cost-effectiveness ratio; QALY, quality-adjusted life year; SALT, Severity of Alopecia Tool

Table 5. Issue 4: Disease monitoring costs for best supportive care

Report section	4.2.9.3 and 4.2.9.5
Description of issue and why the EAG has identified it as important	The EAG's clinical experts considered that it is likely that if response to treatment is not achieved, patients will not engage with further treatment and will not be followed up (effectively patients are discharged from care). The EAG considers that lack of engagement with treatment and being discharged from care has implications for the costs assumed in the BSC health state for both arms of the model, as patients transition to this health state upon loss of treatment response or treatment discontinuation for any other reason.
What alternative approach has the EAG suggested?	The EAG considers that disease management costs in the BSC health state should be excluded for both arms of the model.
What is the expected effect on the cost-effectiveness estimates?	Removal of disease monitoring costs in the BSC health state for both arms of the model increased the company's ICER post clarification from £18,072 to £63,941. However, when combined with a change to the treatment response definition (SALT≤20) and source of utilities, the ICER increases to £419,926.
What additional evidence or analyses might help to resolve this key issue?	No additional evidence required as the scenario resolves the issue.

Abbreviations: BSC, best supportive care; EAG, Evidence Assessment Group; HRQoL, health-related quality of life; ICER, incremental cost-effectiveness ratio; QALY, quality-adjusted life year; SALT, Severity of Alopecia Tool

1.5 Other key issues: summary of the EAG's view

- The EAG considers the BRAVE-AA trial populations to be narrower than the population of the final scope issued by NICE. Specifically, patients with current AA episodes >8 years and who had showed no sign of previous regrowth and patients >60 years (males) and >70 years (females) were excluded from the BRAVE-AA trials. These patients would be eligible to receive baricitinib as per the [redacted] marketing authorisation, but likely have a lower probability of response than the trial populations.
- The BRAVE-AA trial populations, having relatively long disease and episode durations at baseline and being treatment-experienced, are more similar to the prevalent population in the UK than to newly diagnosed patients severe AA. This may cause the trials to underestimate treatment effectiveness in newly diagnosed severe AA patients, as shorter current AA episodes are associated with favourable treatment response, and treatment inexperience may also be associated with favourable treatment response. The EAG notes, however, that the magnitude of any effect of treatment experience on response to baricitinib is uncertain because the mode of action of baricitinib is different to current therapies used to treat severe AA.
- Current AA episode duration and baseline SALT score are clinically meaningful variables that predict treatment response and vary substantially in the trial. Shorter AA episodes and lower

baseline SALT scores are associated with a higher probability of treatment response. Categorising AA episode duration and baseline SALT score could form clinically meaningful subgroups, however any categorisation of these continuous variables would be arbitrary.

1.6 Summary of EAG's preferred assumptions and resulting ICER

Table 6 presents the EAG preferred assumptions as well as the EAG deterministic and probabilistic base case ICER. Table 7 presents scenarios around the EAG base case.

Table 6. EAG preferred assumptions and base case ICER

Scenario	Incremental costs	Incremental QALYs	ICER (change from company base case
Company base case post clarification	[redacted]	[redacted]	18,072
SALT≤20 at Week 36	[redacted]	[redacted]	17,071
SALT≤20 at Week 36 + baseline and CFB utility from BRAVE trials	[redacted]	[redacted]	118,494
Long-term all-cause discontinuation based on Week 36-52 data for baricitinib 4 mg ([redacted])	[redacted]	[redacted]	107,217
No monitoring costs in the induction phase and Maintenance health state for 'Watch and wait' (comparator defined as 'discharged from care')	[redacted]	[redacted]	126,309
Removal of disease monitoring costs in the BSC health state for both arms of the model	[redacted]	[redacted]	419,926
Removal of non-pharmacological psychological support costs	[redacted]	[redacted]	423,809
One wig assumed in the induction phase for both arms of the model	[redacted]	[redacted]	423,775
EAG's preferred deterministic base case - combination of all scenarios	[redacted]	[redacted]	**423,775**
EAG's preferred probabilistic base case - combination of all scenarios	[redacted]	[redacted]	**379,030**

Abbreviations: BSC, best supportive care; CFB, change from baseline; EAG, Evidence Assessment Group; ICER, incremental cost-effectiveness ratio; QALY, quality-adjusted life year; SALT, Severity of Alopecia Tool

*It should be noted that QALY gain in the probabilistic analysis is [redacted] higher than the deterministic analysis. However, given that the incremental costs and QALYs are relatively small, the ICERs are sensitive to very small changes.

Table 7. Deterministic scenarios around the EAG base case

	Results per patient	Baricitinib 4 mg	'Discharged from care'	Incremental value
0	**EAG base case**			
	Total costs (£)	[redacted]	[redacted]	[redacted]
	QALYs	[redacted]	[redacted]	[redacted]
	ICER (£/QALY)			423,775
1	**Severe subgroup - baseline SALT 50-95**			

	Total costs (£)	[redacted]	[redacted]	[redacted]
	QALYs	[redacted]	[redacted]	[redacted]
	ICER (£/QALY)			407,212
2	**Very severe subgroup - baseline SALT 95-100**			
	Total costs (£)	[redacted]	[redacted]	[redacted]
	QALYs	[redacted]	[redacted]	[redacted]
	ICER (£/QALY)			456,573

Abbreviations: EAG, Evidence Assessment Group; ICER, incremental cost-effectiveness ratio; QALY, quality adjusted life year; SALT, Severity of Alopecia Tool.

Note: the same baseline utility ([redacted]), change from baseline ([redacted]) and treatment discontinuation rate ([redacted]) have been used for the subgroups as for the base case as the relevant data were not available by severity.

For further details of the exploratory and sensitivity analyses done by the EAG, see Section 6.3.

2 Introduction and background

2.1 Introduction

This document contains the Evidence Assessment Group's (EAG's) critique of the clinical and cost-effectiveness evidence submitted for the Single Technology Appraisal (STA) of baricitinib (brand name Olumiant®; Eli Lilly and Company) in the treatment of severe alopecia areata (AA) in adults.

2.2 Background

Section B.1 of the company submission (CS) provides information on:

- AA, including its aetiology, burden of disease and current pathway of clinical care in the NHS, and;
- Baricitinib, including its mechanism of action, details of its pending marketing authorisation, its costs and its method of administration and dosage.

The EAG's clinical experts agreed that Section B.1.3 of the CS provides a reasonable overview of AA, its aetiology and burden of disease. AA is an autoimmune disease that leads to non-scarring loss of hair on a person's scalp, face or body. In 2018, 0.58% of UK adults who were registered in electronic primary care records had an active or historic diagnosis of AA.[1] While the exact aetiology of AA is unknown, a suite of genetic risk factors[2] and environmental stressors[3, 4] exist that heighten the risk of AA.

AA can vary in severity, which can be measured using The Severity of Alopecia Tool (SALT). The SALT score ranges from 0-100 and measures the severity of scalp hair loss, with 0 corresponding to complete loss of scalp hair and 100 corresponding to a full head of hair. In the CS, severe AA is defined as SALT 50–94 and very severe AA is defined as SALT ≥95. Severity is associated with prognosis: those who are missing more scalp hair are less likely to regrow hair, either spontaneously or through treatment.[4, 5] Similarly, the length of an AA episode is related to prognosis: the longer a patient has had an AA episode for, the less likely the patient is to regrow hair.[5]

AA is caused by the loss of immune privilege of hair follicles. This occurs due to the production of pro-inflammatory cytokines, such as interferon-gamma, causing the stimulation of natural killer cell receptors and subsequent activation of the janus kinase (JAK) signal transducer and activator of transcription (STAT) signalling pathway (JAK/STAT). The inflammation associated with JAK/STAT activation causes the early termination of the anagen phase in hair follicles, preventing hair growth.[6, 7] Drugs that inhibit JAK therefore have the potential to prevent and reverse autoimmune hair loss in

AA.[8] Baricitinib is one such JAK inhibitor that selectively and reversibly inhibits JAK1 and JAK2, and is expected to receive marketing authorisation from the Medicines and Healthcare products Regulatory Agency (MHRA) for treating severe AA in adults in [redacted]. In this STA, baricitinib is being evaluated as an oral JAK inhibitor taken as a 4 mg dose once daily, with a lower dose of 2 mg once daily potentially being more suitable for some subgroups, such as those >75 years or those with a history of chronic or recurrent infections.

2.2.1 Treatment pathway for severe AA

There are no approved treatments for severe AA, defined as SALT ≥50, in England and Wales, and there is a clear unmet need for these patients. Several JAK inhibitors are in development for AA, but baricitinib is the first to undergo an appraisal by NICE for this indication. In the CS, Baricitinib is positioned as: i) a first-line treatment for severe AA, and ii) a later-line treatment to treat patients with severe AA who do not respond to other treatment strategies. The EAG's clinical experts thought this positioning is an accurate reflection of where baricitinib would be used in the treatment pathway for severe AA. The first wave of eligible patients, the prevalent population, would likely be later-line patients who have failed on or were intolerant to pre-existing therapies, and after this, baricitinib would become a preferred first-line therapy for newly diagnosed severe AA.

In Section B.1.3.3, the company outline their interpretation of the current treatment pathway for severe AA: patients may initially be left untreated under a "Watch and wait" approach similar to that used in mild AA, or patients can be treated from a range of often off-label therapies that have limited effectiveness in severe AA. These treatments include topical, intralesional (IL) or oral corticosteroids, topical immunotherapy, immunosuppressives such as methotrexate, and minoxidil and calcineurin inhibitors. In the economic analysis, the company defines established clinical management as "Watch and wait" followed by best supportive care—which comprised of the range of off-label therapies and psychological support.

The EAG's clinical experts believed it reasonable that patients with severe AA may be untreated up to around six months, however noted that this reflects the wait period to see a dermatologist rather than necessarily a decision to "Watch and wait". The EAG's clinical experts believed that most patients would have used a potent topical steroid during milder disease, or IL steroids if a patient visited a dermatologist. For severe AA, topical immunotherapy, systemic steroids and systemic immunosuppressants may be offered, and a small number of patients may have received these with milder disease, too. The EAG's clinical experts highlighted that there is no clear single standard management for severe AA and highlighted how only topical immunotherapy and wigs are

recommended by the British Association for Dermatologists (BAD) guidelines for treating severe AA.[9] Treatment for severe AA varies based on setting (primary care, specialist dermatologist, specialist dermatologist with an interest in AA), treatment availability and patient preference. Not all dermatologists will offer or have access to the more effective best supportive care therapies and not all patients will opt to take them, instead managing the condition with wigs or head shaving, if required. While the 2012 BAD guidelines for the management of alopecia areata recommend topical immunotherapy, e.g., DPCP, for extensive patchy hair loss and alopecia totalis/universalis,[9] DPCP is not widely available across the NHS, can lead to potent allergic reactions in patients and staff, and has a high rate hair-loss recurrence. For example, a meta-analysis reported a recurrence rate of 38% in patients receiving maintenance treatment and treatment-emergent severe eczema in 31% of DPCP treated patients, although the EAG's clinical experts noted this may be an overestimation of the rate of severe eczema.[10]

While the EAG's clinical experts did not recognise "Watch and wait" as a standard option for treating severe AA, they did recognise a similar no active treatment, and ultimately discharge from care, as a common management strategy opted for by severe AA patients. This was especially the case for the long-term care for patients who do not respond to treatment. Such patients may be prescribed wigs, or shave their heads, but this would not require intensive follow-up. This absence of intensive follow-up is the key difference between the no active treatment the EAG considers a common management strategy in clinical practice and the company's definition of "Watch and wait", which involves intensive surveillance and support. The EAG's clinical experts advised that most of the prevalent population will have opted for no active treatment and discharge from care, however they highlighted how there is a lack of comprehensive treatment pattern data for AA patients in the UK.

Hence, The EAG therefore considers there to be no established or highly- effective standard clinical management of severe AA, with no active treatment and discharge from care being a common endpoint. The EAG's clinical experts further highlighted the near absence of high-quality randomised controlled trials for the treatment of severe AA, excluding recent trials on JAK inhibitors.

2.3 Critique of the company's definition of the decision problem

In Table 1 of the CS, the company outlines: i) the final scope issued by NICE, ii) the decision problem addressed in the CS and, iii) the company's justification for differences between them. The EAG considers the decision problem addressed by the company to largely match the final scope issued by NICE.[11] However, the EAG notes that the two trials informing the clinical effectiveness data analysis in the submission, BRAVE-AA1 and BRAVE-AA2, have patient populations that differ in several

regards to the patient population that would be eligible to receive baricitinib in England and Wales, and that specified in the NICE final scope. Overall, however, the EAG considers the BRAVE-AA trial data to be suitable to inform decision making. An overview of the EAG's critique of the company's definition of the decision problem and the relevance of the BRAVE-AA trial populations can be found in Table 8.

Table 8. EAG critique of the decision problem

	Final scope issued by NICE	Decision problem addressed in the submission	Rationale if different from the scope	EAG comment
Population	Adults with severe alopecia areata	Adults with severe alopecia areata	NA	The decision problem matches that of the final scope issued by NICE: adults with severe alopecia areata. However, the BRAVE-AA trials: • excluded patients with baseline AA episodes >8 years; • excluded males >60 years and females >70 years. Hence, the BRAVE-AA trials provide data on a narrower population than those who could receive baricitinib in clinical practice, and has excluded some of the patients least likely to respond to treatment.
Intervention	Baricitinib	Baricitinib	NA	The intervention described in the CS, baricitinib, matches the intervention described in the final scope. Baricitinib is an oral JAK inhibitor that is expected to receive marketing authorisation for treating severe AA in adults in [redacted].
Comparator(s)	Established clinical management without baricitinib	Established clinical management without baricitinib, which may include supportive care	NA	As per the NICE final scope, the company has outlined what it believes to be established clinical management without baricitinib, informed by real-world dataset and three clinical experts. The EAG's clinical experts outline how there was no clear established clinical management for severe AA, with a large degree of variation between centres. The EAG notes that:

				• There is no clear standard clinical management for severe AA; • No active treatment or follow-up is currently a realistic end prospect for severe AA patients; • The company likely overestimates the amount of psychological support patients receive in the NHS. The EAG's clinical experts highlighted how the availability of psychological care is far below what is needed for severe AA patients.
Outcomes	The outcome measures to be considered include: • Disease severity e.g. Severity of Alopecia Tool (SALT); • Improvement in hair loss e.g. Scalp Hair Assessment Score, Measure for Eyebrow Hair Loss, Measure for Eyelash Hair Loss; • Adverse effects of treatment; • Health-related quality of life.	The outcome measures to be considered include: • Measures of disease severity and improvement in hair loss (including SALT, ClinRO for eyebrow hair loss and eyelash hair loss, PRO measures for scalp hair assessment, PRO measures for eyelashes and eyebrows); • Adverse effects of treatment (including AEs, SAEs, AESIs); • Health-related quality of life (including EQ-5D, Skindex-16 AA, HADS and SF-36).	NA	The outcomes in the company's submission match the outcomes described in the final scope.
Economic analysis	The reference case stipulates that the cost-effectiveness of treatments should be expressed in terms of	As per NICE final scope	NA	NA, as per NICE final scope

	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.			
Subgroups to be considered	Due to an assumed typographical error, the NICE final scope was ambiguous about which subgroups were to be considered, stating that: "If the evidence allows, the following subgroups based on severity and type of alopecia areata will be considered", but without specifying any subgroups.	The company provided subgroup analyses and a scenario analysis based on the baseline severity of alopecia areata (severe disease, SALT 50-95 and very severe disease, SALT 95-100) and current duration of AA at baseline. No scenario analyses were presented based on the type of alopecia areata.	NA	NA
Special considerations, including issues related to equity or equality	None identified.	None identified.	NA	The EAG's clinical experts highlighted for some cultures loss of beard hair can be an important issue.

Abbreviations: AA: alopecia areata; AE: adverse event; AESI: adverse event of special interest; CS: company submission; EAG: evidence assessment group; EQ-5D: the European Quality of Life-5 Dimensions; HADS: Hospital Anxiety Depression Scale; JAK: janus kinase; NICE: National Institute for Health and Care Excellence; PRO: patient reported outcome; SAE: serious adverse event; SALT: Severity of Alopecia Tool; SF-36: Short Form 36 Health Survey Questionnaire.

2.3.1 Population

Two Phase III trials of baricitinib in adults with severe AA, BRAVE-AA1 and BRAVE-AA2,[12] inform the clinical effectiveness evidence in the submission. Despite only including adults with severe AA, the populations of BRAVE-AA1 and BRAVE-AA2 are from a narrower population than that of the NICE final scope.[11] Specifically, the following patients were excluded from both BRAVE-AA trials:

- Patients with current AA episodes >8 years and who had showed no sign of previous regrowth;
- Male patients >60 years and female patients >70 years.

Such patients would be eligible to receive baricitinib per the [redacted] marketing authorisation but may be less likely to achieve hair regrowth. Patients with longer episodes of AA have a lower probability of hair regrowth and treatment response (Section 3.3.2.2), and a less effective 2 mg dose may be used in patients >75 years, leading to a lower probability of treatment response. In addition, patients with co-existing hair loss conditions, such as male patients >60 years who have male pattern baldness, the amount of scalp hair regrowth possible could be limited and difficult to determine. The EAG's clinical experts also noted that it is plausible that patients with long AA episodes who have disengaged will reengage with care to receive baricitinib, should it become available.

Any overestimation of the effectiveness of baricitinib because of the trial exclusion criteria may be balanced by the fact that around [redacted] of participants in the BRAVE-AA trials had received, and likely failed on, prior therapies, including [redacted] of participants having prior experience with therapies usually only given to patients with severe AA. The EAG's clinical experts highlighted that the level of treatment experience in the BRAVE-AA trials likely exceeds that seen in UK clinical practice for severe AA, both in terms of the percentage of patients receiving therapies such as contact immunotherapy and systemic immunosuppressants, but also that patients in the BRAVE-AA trials received therapies not currently used in the NHS, such as phototherapy, cryotherapy and platelet-rich-plasma injections. This may mean that the BRAVE-AA trials may underestimate of the effectiveness of baricitinib in the first-line setting, as patients who had succeeded on prior therapies would not have entered the trial. However, the EAG notes that the mode of action of baricitinib is different to the current therapies used to treat (severe) AA, and as such the magnitude of any effect of treatment experience is unknown.

Underestimation of the treatment effectiveness of baricitinib in the first-line setting in UK clinical practice is also likely due to the BRAVE-AA trial participants having varying, and often long, lengths of current AA episode at baseline, with mean durations ≥3.5 years for all arms. In contrast, patients presenting with newly diagnosed severe AAA in the first-line setting are likely to have shorter durations of AA episodes, and therefore a larger probability of treatment response.

Regarding the company's positioning of baricitinib both in the first-line setting and later-line settings for patients who have failed on previous treatment, the EAG considers that:

- The BRAVE-AA trial population is most similar to the prevalent population in clinical practice who would be eligible to receive baricitinib at the point of approval, i.e. a later-line treatment experienced population. The treatment effectiveness of baricitinib in this later-line population might be overestimated by the BRAVE-AA trial data due to the exclusion of patients with current AA episodes >8 years and who had showed no sign of previous regrowth and male patients >60 years and female patients >70 years;
- The BRAVE-AA trial data may underestimate the effectiveness of baricitinib in the first-line population, because of high rate of prior treatment with agents usually only given to patients with severe AA, and the presence of patients with relatively long baseline AA episode durations in the BRAVE-AA trial populations. In the company's Adelphi DSP study, [redacted] of severe/very severe AA patients were treatment experienced[13, 14], including both therapies given at milder stages of disease (e.g. topical corticosteroids), but also those primarily given to patients with severe disease (topical immunotherapy, systemic immunosuppressants and systemic steroids).

2.3.2 *Intervention*

The intervention under consideration is oral baricitinib 4 mg and matches the final scope issued by NICE.[11] Baricitinib is an oral JAK inhibitor that is expected to receive marketing authorisation for treating severe AA in adults in [redacted]. Further details of baricitinib for AA, including the method and administration and dosing can be found in Section B.1.2 of the CS. Baricitinib has previously been recommended in certain populations for treating moderate to severe atopic dermatitis in TA681,[15] and for treating severe rheumatoid arthritis in TA466.[16]

2.3.3 Comparators

The comparator in the final scope issued by NICE was, "established clinical management without baricitinib". To this, the company added, "which may include supportive care", which the EAG agrees is in-line with the final scope issued by NICE.[11] The EAG has discussed the current pathway of care for severe AA in Section 2.2.1 and notes that:

- There is no clear standard clinical management for severe AA and no single most suitable comparator therapy;
- Many severe AA patients may opt for or will end up receiving no active treatment, with only occasional, if any, follow-up, especially after disappointing results from available therapies.

The EAG is concerned with how the company defines of "Watch and wait" as established clinical management for adults with severe AA, which the company outline as involving continued monitoring. The EAG agrees with the company that the no active treatment component of "Watch and wait" is a common management strategy used for adults with severe AA, however the EAG's clinical experts highlighted that this would not require intensive follow-up as they would be discharged from care, or receive only occasional follow-up. In addition, the EAG's clinical experts highlighted how access to psychological support, while needed, is in practice minimal due to resource constraints.

Of the best supportive care options, the EAG considers there to be three candidates to provide comparative cost-effectiveness data against baricitinib in this submission:

- A comparison with DPCP, which the EAG's clinical experts highlighted might be the closest in effectiveness to baricitinib and the only active treatment recommended by the BAD Guidelines for treating severe AA in adults.[9] However, the EAG considers there to be no valid means of performing valid comparison between DPCP and baricitinib 4 mg in adult severe AA patients (see Section 3.4), and further notes that: i) DPCP is only available to a minority of patients with no equitable access and, ii) many patients discontinue treatment, and some will suffer strong adverse reactions to DPCP, iii) some patients will have already received or been eligible for DPCP for milder disease;
- A comparison with the systemic corticosteroids or systemic immunosuppressants currently used to treat severe AA, each at relatively low frequency. Again, the EAG does not consider there to be appropriate data available to perform a valid comparison between all or any of

these therapies and baricitinib 4 mg in adult severe AA patients (see Section 3.4), nor does the EAG consider any to be established standard of care. The EAG also notes that some of these treatments may have been given to patients when they had mild or moderate disease, and, because of the limited effectiveness of these treatments, the placebo group from the BRAVE-AA trials (i.e. no active treatment) may provide a reasonable approximation for the treatment effect;

- A comparison with no active treatment with discharge from care. The EAG considers this to be both the most appropriate comparison for treatment-experienced patients and for those newly diagnosed with severe AA who opt not to receive further treatment. Moreover, it is the only comparison for which robust comparative data are available with baricitinib 4 mg through the BRAVE-AA trials.

The EAG considers the comparison with no active treatment to be the most relevant for the current submission, as it reflects a viable treatment option across the prevalent and incident populations of adults with severe AA and is the only comparison for which high-quality data comparative effectiveness data are available. As outlined in Section 4.2.5.2, the EAG considers the placebo arm of the BRAVE-AA trials to provide a reasonable estimate of the treatment response a patient with no active treatment would receive.

The EAG considers there to be a distinction between the prevalent and incident populations of adults with severe AA when considering the most appropriate comparator for the appraisal, given the treatment experience of these patients may differ:

- Patients in the prevalent population are likely to have explored all treatment options available to them. Most of these patients will have opted for no active treatment and may manage their severe AA with head shaving or wigs. For this population — the population that would be treated at the point of approval in UK clinical practice — the EAG considers no active treatment and discharge from care to be the appropriate comparator;
- Patients in incident population will be less treatment-experienced than those in the prevalent population, and only a minority will have prior experience with topical immunotherapy (e.g. DPCP), systemic immunosuppressants or systemic corticosteroids. Upon progression to severe AA, they may opt to trial one or more of these therapies, assuming their dermatologist offers it to them. The EAG does not consider any specific systemic immunosuppressants or systemic corticosteroids to be an established standard of

care for these patients, and considers no active treatment to be a realistic endpoint for these patients. The EAG notes the absence of data available to permit meaningful comparisons of these aforementioned treatments with baricitinib 4 mg (see Section 3.4). The EAG agrees with the company that these therapies have very limited effectiveness for severe disease. The EAG highlights a lack of a: i) clear standard of care, ii) robust data on treatment patterns, and iii) robust comparative effectiveness of active treatments with baricitinib. Hence, the EAG considers a direct comparison between baricitinib and systemic immunosuppressants and/or systemic corticosteroids unlikely to adequately capture established clinical management in the UK until more data are available to demonstrate: i) that these therapies are frequently used by, and accessible to, most patients and, ii) to provide robust comparative effectiveness data with baricitinib. In lieu of such data, the EAG considers no active treatment and discharge from care is an acceptable comparator for adults newly diagnosed with severe AA. The EAG notes that at the point of approval this population will be small, but that baricitinib would become a preferred first-line therapy for newly diagnosed severe AA, if approved.

2.3.4 Subgroups

No subgroups were clearly identifiable from the NICE final scope, which stated, “If the evidence allows, the following subgroups based on severity and type of alopecia areata will be considered”. The company provided subgroup analyses and a scenario analysis based on the baseline severity of alopecia areata (severe disease, SALT 50-95 and very severe disease, SALT 95-100). No scenario analyses were presented based on the type of alopecia areata, however, the EAG notes that all alopecia totalis/universalis patients would be included in the SALT 95-100 subgroup.

Subgroup efficacy analyses based on baseline length of current AA episode were available in the clinical study reports (CSRs), and, in response to clarification question B2, the company provided an economic scenario analysis for subgroups of ≤4 years and >4 years (Section 5.1.2.2). In response to clarification question A14, the company also provided a subgroup analysis based on atopic background status, a factor highlighted by the EAG’s clinical expert as a potentially meaningful subgroup.

2.3.5 *Special conditions*

No special conditions were identified in the NICE final scope or by the company. The EAG's clinical experts highlighted how health related quality of life deficits may vary for different reasons across cultures – in line with the issues raised in the equality impact assessment.[17] For example, in some cultures loss of beard hair can be an important issue. However, the EAG's clinical experts highlighted that the negative consequences of AA can and do extend to patients of all demographics.

3 Clinical effectiveness

3.1 Critique of the methods review

The company conducted a systematic literature review (SLR) to identify randomised controlled trials (RCTs) and observational studies providing clinical efficacy and safety data for baricitinib for the treatment of alopecia areata (AA) comparators and supportive care therapies. The Evidence Assessment Group (EAG) notes that the original SLR was conducted in July 2021 and that it was subsequently updated in February 2022.

A total of 45 studies from 47 records were included from the SLR, including 12 RCTs. An overview of the methods used by the company for the SLR, together with the EAG's critique of the appropriateness of these methods, is presented in Table 9. In summary, the EAG considers the methods applied by the company to be adequate and likely to have identified most of the clinical evidence of relevance to the decision problem. One study of baricitinib was included from the SLR, King *et al.* 2021,[18] that reported on the Phase II portion of BRAVE-AA1. The company provided additional data on the Phase III portion of BRAVE-AA1 and from BRAVE-AA2 in the submission, as the primary publication for BRAVE-AA1 and BRAVE-AA2 was published after the SLR update search date.[12]

Table 9. Summary of EAG's critique of the methods implemented by the company to identify evidence relevant to the decision problem

Systematic review step	Section of CS in which methods are reported	EAG assessment of robustness of methods
Data sources	Appendix D1.1.	**The EAG considers the sources and dates searched to be appropriate.** Databases searched: Embase, MEDLINE In-Process and the Cochrane Library (CENTRAL and CDSR). Additional sources: Hand-searching of conference proceedings (published in 2019 to 2021) and clinical trial registers. Latest search update: 4 February 2022.
Search strategies	Appendix D1.1	**The EAG is satisfied that the searches have identified all evidence relevant to the decision problem.** Search strategies for the literature review combined comprehensive terms for the population, interventions and study designs, using free-text and medical subject headings.
Inclusion criteria	Appendix D1.2	**The EAG considers it likely that no relevant evidence was excluded, although the EAG notes that as young adolescent patients are often treated as adults in AA, many studies that**

		included a small number of patients <16 years were excluded. The inclusion criteria of the SLR were in line with the NICE final scope, except for severity and age. For age, studies reporting on patients ≥16 years were included. For severity, studies containing patients with moderate AA were included throughout the SLR, meaning the studies identified contain a wider population than specified in the NICE final scope.[11] Full reference details are available in the CS Appendix for included studies and excluded studies at full text review.
Screening and data extraction	Appendix D.1.2 and D.1.3	**The EAG considers the methods for screening and data extraction to be robust.** Two reviewers independently screened titles and abstracts, and subsequently studies selected for full text appraisal, against predefined criteria, with a third reviewer consulted when consensus could not be reached. Results of the literature screening processes were summarised in PRISMA diagrams. Conference proceedings and clinical trial registries were searched by a single reviewer and checked by a second reviewer. Data extraction was carried out by one reviewer, with a second researcher independently quality checking the extracted data.
Tool for quality assessment of included study or studies	B.2.5 & Appendix D.1.4 and D.3	**The EAG agrees with the company's choice of quality assessment tool for assessing BRAVE-AA1 and BRAVE-AA2.** The company used the using the Appraisal of RCT checklist by Cochrane[19] for the quality assessment of the RCTs included in the SLR. The quality of the included observational studies was assessed using the quality assessment tool developed by the York University CRD.[20]

Abbreviations: AA: alopecia areata; CENTRAL: Cochrane Central Register of Controlled Trials; CDSR: Cochrane Database of Systematic Reviews; CS: company submission; EAG: Evidence Assessment Group; NICE: National Institute for Health and Care Excellence; PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses; RCT: randomised controlled trial; SLR: systematic literature review.

3.1.1 SLR reporting quality

The EAG notes there were numerous reporting errors in the conduct of the SLR that raised concerns about the overall quality of the SLR, but that the company was able to clarify these adequately (clarification questions A16, A17, A21 and A22).

3.1.2 Age eligibility criterion

The EAG notes that studies reporting on patients ≥16 years were included in the SLR (Clarification question A22), a wider inclusion criterion than the BRAVE-AA trial inclusion criterion and the provisional marketing authorisation of baricitinib. The EAG considers including such studies to be

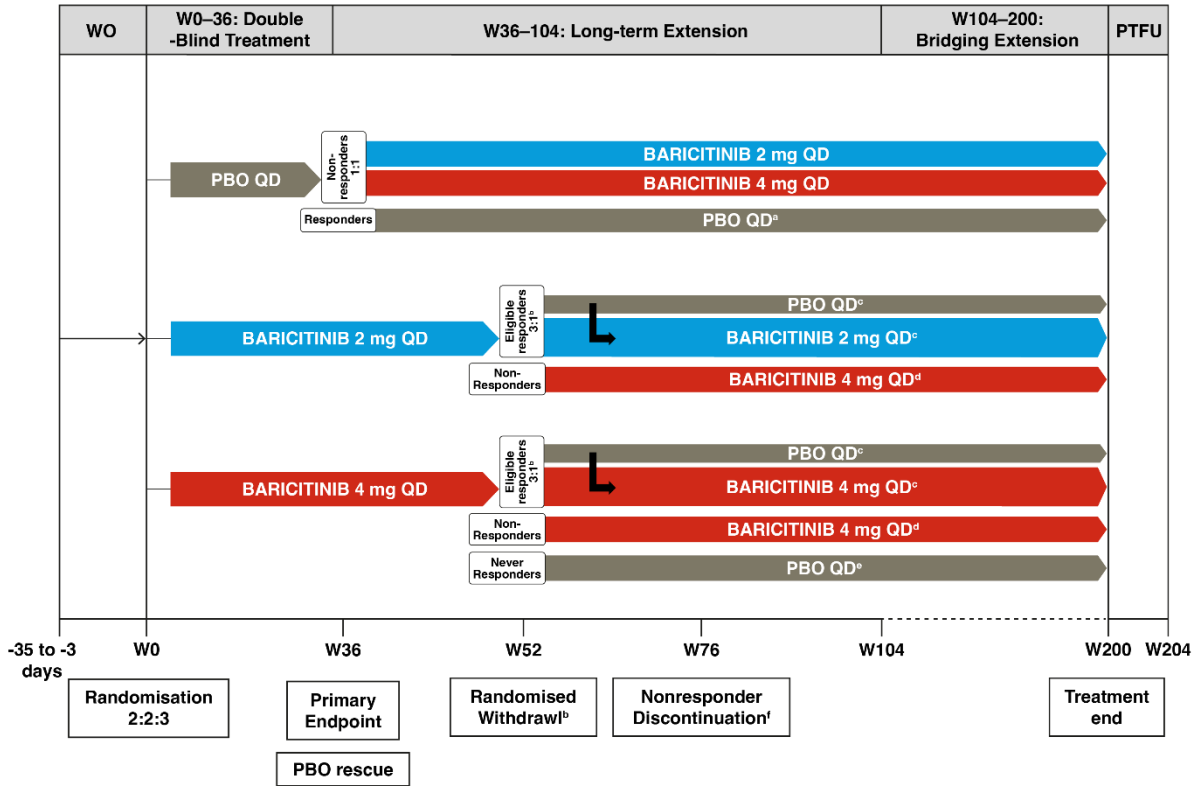
reasonable, as the EAG's clinical experts noted it was not uncommon for young adolescent AA patients be treated as adults, and that many studies identified in the SLR were single-centre reviews of all patients in a centre. Given the large number of records excluded from the SLR at the full text stage due to containing some patients <16 years (190 records vs 47 ultimately extracted in the SLR), the EAG believes that, if anything, the age edibility criterion of the SLR was too restrictive, and some relevant studies containing a small number of patients <16 years may have been excluded. Nevertheless, the EAG's clinical experts noted the absence of high-quality placebo controlled RCT data for treating severe AA, something also noted by other systematic reviews of the field.[21] Hence, the EAG considers it unlikely that any key studies have been excluded from the SLR due to containing some paediatric patients.

3.2 Critique of trials of the technology of interest

In this section, the EAG critiques the BRAVE-AA1 and BRAVE-AA2 trials that provide the key clinical effectiveness data used in the cost-effectiveness analysis of the CS.[12] Both BRAVE-AA1 and BRAVE-AA2 are international, double-blind, randomised, placebo-controlled trials comparing baricitinib 2 mg daily and baricitinib 4 mg daily with placebo. While BRAVE-AA1 is an adaptive Phase 2/3 trial, only the Phase 3 data are used in the submission and considered hereafter. As baricitinib 4 mg is the dosage under consideration in this STA, the data from the baricitinib 4 mg arms and placebo arms of BRAVE-AA1 and BRAVE-AA2 will be focused on in this critique. Supporting information will be cited from the baricitinib 2 mg data where appropriate, and the EAG notes that the results from the baricitinib 2 mg arms were consistent with the results of the baricitinib 4 mg throughout the results of the trial, albeit with a lower magnitude of benefit over placebo throughout.

The primary outcome of the BRAVE-AA trials was achieving an absolute SALT ≤20 at Week 36, i.e., at the end of the double-blind treatment stage. However, additional data up to Week 72 were provided for a randomised withdrawal sub study (BRAVE-AA1) and a randomised down-titration sub study (BRAVE-AA2). The EAG notes that because placebo non-responders at Week 36 were eligible for rescue therapy, robust comparative data between baricitinib and placebo are only available up to Week 36. The design of BRAVE-AA1 and BRAVE-AA2 are reproduced in Figure 1 and Figure 2.

Figure 1. Study design of BRAVE-AA1 (Reproduced from Figure 1, Clarification Response)

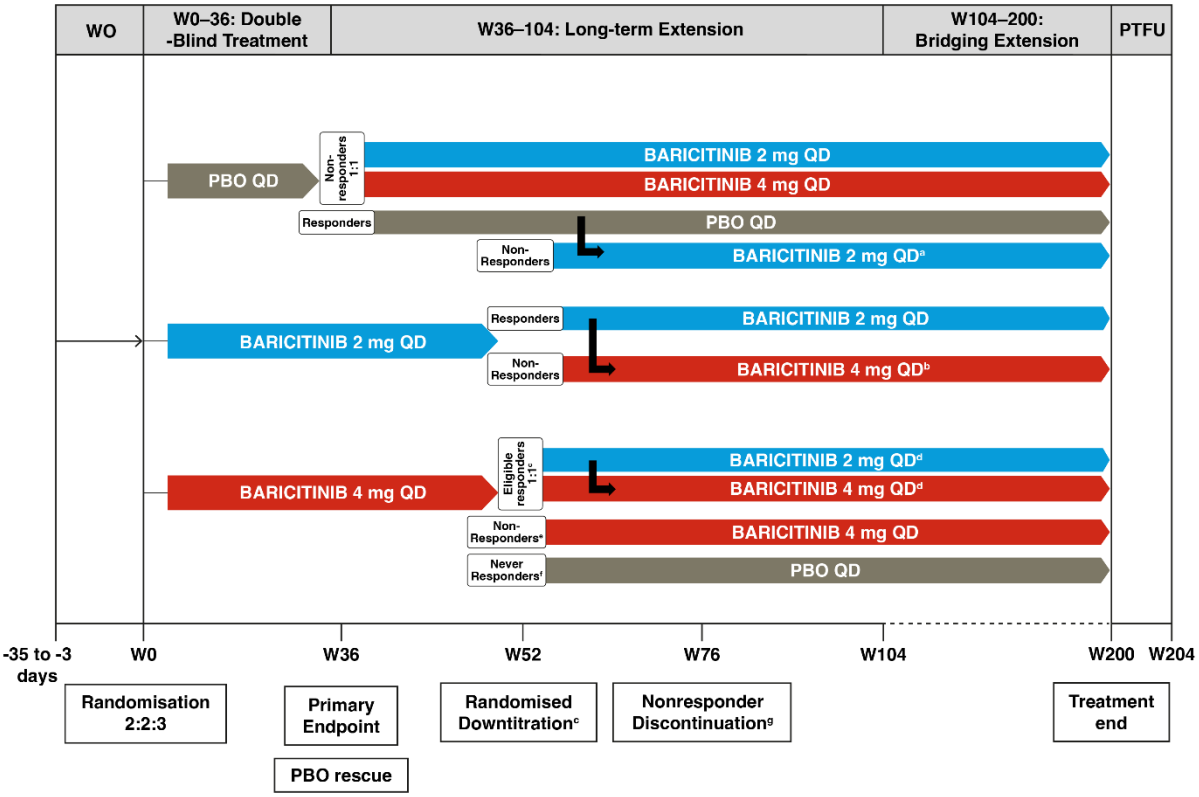


Footnotes: [a] Placebo responders stayed on placebo for remainder of the trial, even if relapse was observed later. [b] Patients with SALT ≤20 who stayed on the same dose of baricitinib from week 0 were randomised to stay on current baricitinib dose, or transitioned to placebo.[c] Responders participating in randomised withdrawal who experienced >20-point absolute worsening in total SALT score after week 52 were retreated with baricitinib dose to which they were originally randomised if they were randomised to placebo at week 52, OR continued to receive same dose of baricitinib if they were randomised to remain on baricitinib at week 52. [d] Non-responders at week 52 were rescued to baricitinib 4 mg if receiving baricitinib 2 mg from baseline, OR remained on baricitinib 4 mg if they were in the 4-mg group and achieved SALT ≤20 before week 52. [e] Never responders (never achieved SALT ≤20 by week 52 despite being in the baricitinib 4-mg group from baseline and had not experienced a ≥2-point improvement from baseline in ClinRO measures for EB or EL hair loss) were automatically transitioned to placebo.[f] Non-responders at week 52 AND week 76 were automatically discontinued at week 76 unless they had a ≥2-point improvement from baseline in ClinRO measures for EB or EL hair loss.

Abbreviations: EB: eyebrow; EL: eyelash; PTFU: post-trial follow up; QD: once daily; W: week.

Source: BRAVE-AA1 Clinical Study Report.

Figure 2. Study design of BRAVE-AA2 (Reproduced from Figure 2, Clarification Response)



Footnotes: [a] Placebo-treated patients not eligible for rescue to baricitinib at week 36 (due to spontaneous remission) were rescued to baricitinib if they were non-responders at week 52, OR if they experienced loss of treatment benefit after week 52. [b] Patients randomised to baricitinib 2 mg at week 0 were rescued to the 4-mg dose if they were non-responders at week 52, OR were responders at week 52 but experienced a >20-point worsening in SALT score after week 52. [c] Responders in the baricitinib 4-mg group (SALT ≤20 who stayed on 4 mg from week 0) were randomised to either stay on 4 mg OR transition to 2 mg.[d] Responders participating in the randomised down-titration who experienced a loss of treatment benefit after week 52 were re-treated with baricitinib 4 mg if they were randomised to the 2-mg dose at week 52, OR continued to receive baricitinib 4 mg if they randomised to remain on the 4-mg dose at week 52. [e] At week 52, non-responders (SALT >20) in the baricitinib 4-mg group since baseline who achieved SALT ≤20 before week 52 remained on 4 mg. [f] Never responders (never achieved SALT ≤20 by week 52 despite being in the baricitinib 4-mg group from baseline and had not experienced a ≥2-point improvement from baseline in ClinRO measures for EB or EL hair loss) were automatically transitioned to placebo. [g] Non-responders at week 52 AND week 76 were automatically discontinued at week 76 unless they had a ≥2-point improvement from baseline in ClinRO measures for EB or EL hair loss.

Abbreviations: PTFU: post-trial follow up; QD: once daily; W: week.

Source: BRAVE-AA2 Clinical Study Report.

The EAG's clinical experts agreed with the company that the design of BRAVE-AA1 and BRAVE-AA2 was sufficiently similar to justify pooling the data, and the EAG considers this to be appropriate. The EAG's quality assessment of BRAVE-AA1 (Phase III portion) and BRAVE-AA2 is provided in Table 10. Overall, the EAG considers BRAVE-AA1 and BRAVE-AA2 to be high quality RCTs, in-line with the quality assessment provided by the company in Table 14 of the CS.

Table 10. The EAG's quality assessment of BRAVE-AA1 and BRAVE-AA2

Aspect of trial design or conduct	Section of CS or supporting information where details are reported	EAG's critique	
		BRAVE-AA1	BRAVE-AA2
Randomisation	CS B.2.3.1	**Appropriate** Participants were randomised 2:2:3 to placebo, 2 mg baricitinib or 4 mg baricitinib using an interactive web response system (IWRS).	
Concealment of treatment allocation	CS B.2.3.1 BRAVE-AA1 Protocol Section 5.1 BRAVE-AA2 CSR Table 4.1	**Appropriate** Treatment allocation was concealed by use of an IWRS	
Eligibility criteria	CS Table 6	**Some concerns** The eligibility criteria of the trials excluded certain adults with severe alopecia areata who may be less likely to respond to treatment, namely: • Those with current AA episodes at baseline >8 years with no episodes of regrowth and; • Males >60 years and females >70 years.	
Blinding	CS B.2.3.1	**Appropriate** Up to Week 36 the trials were double blind, with the patients, investigators and study team blinded to baricitinib or placebo assignment.	
Baseline characteristics	CS B.2.3.3	**Appropriate** Key baseline characteristics were balanced between the baricitinib 4 mg and placebo arms in both BRAVE-AA1 and BRAVE AA-2 (Table 11, Appendix Table 44). While UK centres were not included in either trial, the EAG's clinical experts did not expect any difference in demographic variables between the included centres and UK centres to be potential treatment effect modifiers.	
Statistical analysis			
Sample size and power	BRAVE-AA1 Protocol Section 10.1 BRAVE-AA2 CSR Table 4.1	**Appropriate** Sample sizes of approximately 625 patients (BRAVE-AA1) and 476 patients (BRAVE-AA2) were targeted to provide over (BRAVE-AA1) or approximately (BRAVE-AA2) 90% power to differences between baricitinib and placebo using the graphical testing procedure and the	

		following assumed response rates: 30% for baricitinib 4-mg, 20% for baricitinib 2-mg, and 5% for placebo. The actual sample sizes in the full analysis sets exceed the target sizes: 654 (BRAVE-AA1) and 546 (BRAVE-AA2). While the response rates in the power calculations were not clinically justified, these sample sizes are likely to detect most meaningful benefits of baricitinib 4 mg over placebo.
Handling of missing data	CS B.2.4.3	**Some concerns** Missing SALT data for [redacted] of patients at Week 36. Missing data were imputed using: i) non-responder imputation for categorical endpoints and, ii) modified last observation carried forward for continuous endpoints. In comparison to multiple imputation, these methods provide lower power to detect treatment effects. The EAG does not consider the analyses in the CS to bias results in favour of baricitinib because the analysis in the primary publication,[12] that used multiple imputation, estimated a larger treatment effect for baricitinib 4 mg than in the CS. As such, the company's handling of missing data was conservative for SALT analyses.
Outcome assessment	BRAVE-AA1 Protocol Sections 9.1.3.3. and 9.1.5 BRAVE-AA1 Protocol Sections 9.1.3.3. and 9.1.5	**Appropriate** Measurement of SALT score, i.e., the proportion of the scalp without hair coverage, was conducted by blinded investigators and is relatively objective. EQ-5D data were collected by self-report from blinded participants.
Analysis for estimate of effect	CS B.2.6.1	**Appropriate** The company's primary analyses and data used in the economic model are responder analyses (SALT ≤20, $SALT_{50}$ and $SALT_{75}$). While these may have lower power than analyses using continuous outcome variables, the EAG's clinical experts considered SALT ≤20 and $SALT_{75}$ to be clinically meaningful outcomes, and the results from the responder analyses were consistent with the continuous change from baseline analyses.

Abbreviations: AA; alopecia areata; CS: company submission; CSR: clinical study report; EAG: evidence assessment group; EQ-5D: EuroQol-5 Dimension; IWRS: interactive web response system; SALT: severity of alopecia tool

3.2.1 Randomisation

Participants in BRAVE-AA1 and BRAVE-AA2 were randomised 2:2:3 to placebo, 2 mg baricitinib or 4 mg baricitinib at Visit 2. Randomisation was stratified based on geographic region (North America and Japan), and duration of current AA episode at baseline (<4 years versus ≥4 years). Errors in data

entry led to [redacted] ([redacted]) patients in BRAVE-AA1 and [redacted] ([redacted]) patients in BRAVE-AA2 patients) having incorrect duration of current AA episode data at baseline. This led to errors in stratified randomisation. However, given the overall balance in baseline characteristics between the baricitinib 4 mg arm and the placebo arm (see Section 3.2.3 and Table 11), the EAG did not determine these errors likely to bias the results of the trials.

3.2.2 Eligibility criteria

The EAG has outlined in Section 2.3.1 how it considers the eligibility criteria of BRAVE-AA1 and BRAVEE-AA2 to most closely reflect the later-line population of the positioning of baricitinib. Due to the exclusion of older patients and those with current AA episodes >8 years and who had showed no sign of previous regrowth, the EAG suggests that the BRAVE-AA trials may slightly overestimate the treatment effectiveness of baricitinib in UK clinical practice for the later line population. In contrast, because trial participants were not excluded based on prior therapies, including contact immunotherapy, the EAG considers the BRAVE-AA trial data to likely underestimate treatment effectiveness in the first-line population.

The EAG also notes that participants in BRAVE-AA1 and BRAVE-AA2 had very few permitted concomitant medicines, and only [redacted] of patients had any that may target AA. While this is a reasonable for the patients receiving baricitinib, the EAG's clinical experts considered that many severe AA patients not receiving baricitinib would be treated after 6 months. While the basket of currently used therapies for severe AA have low likelihoods of success, some may still be more effective than placebo, which the EAG discusses this further in Section 3.4.

3.2.3 Participant characteristics

The baseline characteristics of patients in the placebo and baricitinib 4 mg arms of BRAVE-AA1 and BRAVE-AA2 are presented in Section B.2.3.3 in the CS. The EAG's clinical experts stated that the duration of current AA episode, presence of an atopic background and baseline SALT score might predict the likelihood of hair regrowth, and these data are presented in Table 11 alongside baseline health-related quality of life (HRQoL) data. These characteristics were largely balanced between the placebo and baricitinib 4 mg arms of BRAVE-AA1 and BRAVE-AA2. However, the BRAVE-AA2 placebo arm had a higher mean duration of current AA episode at baseline (4.68 years) than the baricitinib 4 mg arm (3.94 years), which might equate to a lower chance of response in the BRAVE-AA2 placebo arm than all other arms.

Table 11. Baseline characteristics of patients in the placebo and baricitinib 4 mg arms of BRAVE-AA1 and BRAVE-AA2 (adapted from Tables 9 and 10 of the CS)

Characteristic	BRAVE–AA1		BRAVE–AA2	
	Placebo (N=189)	baricitinib 4 mg (N=281)	Placebo (N=156)	baricitinib 4 mg (N=234)
Baseline characteristic highlighted by EAG clinical experts				
Mean (SD) duration of current AA episode, years	3.53 (3.65)	3.46 (3.37)	4.68 (5.490)	3.94 (3.353)
Atopic background, n (%)	73 (38.6)	97 (34.5)	67 (42.9)	87 (37.2)
Duration of current AA episode, n (%)				
<4 years	134 (70.9)	189 (67.3)	94 (60.3)	140 (59.8)
≥4 years	55 (29.1)	92 (32.7)	62 (39.7)	94 (40.2)
Mean (SD) SALT score	84.7 (17.82)	85.3 (18.18)	85.0 (17.79)	84.8 (18.08)
SALT category, n (%)				
Severe (SALT 50–94)	92 (48.7)	133 (47.3)	74 (47.7)	115 (49.1)
Very severe (SALT 95–100)	97 (51.3)	148 (52.7)	81 (52.3)	119 (50.9)
HRQoL baseline characteristics				
Mean (SD) Skindex–16 AA baseline domain scores				
Emotions	[redacted]	[redacted]	[redacted]	[redacted]
Functioning	[redacted]	[redacted]	[redacted]	[redacted]
Symptoms	[redacted]	[redacted]	[redacted]	[redacted]
Mean (SD) HADS total score				
HADS-Anxiety	6.7 (3.92)	6.1 (3.80)	5.9 (4.01)	6.4 (3.95)
HADS-Depression	4.0 (3.15)	4.0 (3.39)	3.7 (3.46)	3.8 (3.49)
EQ-5D-5L health state index	[redacted]	[redacted]	[redacted]	[redacted]
EQ-5D-5L VAS score	[redacted]	[redacted]	[redacted]	[redacted]

Source: CS Table 9, Table 10, Table 26, and Table 27

Abbreviations: AA: alopecia areata; CS: company submission; HADS: hospital anxiety and depression score; PRO: patient reported outcome; SALT: severity of alopecia tool; SD: standard deviation

The EAG's clinical experts noted that several of the baseline characteristics of patients in the BRAVE AA trials indicate that the trial participants had particularly severe and difficult to treat alopecia areata: a mean disease duration from the first onset of AA diagnosis of 12.2 years; a mean episode duration of 3.9 years; >50% had SALT 95-100 and around [redacted] had alopecia universalis. Such a

severity of patients may mean the BRAVE-AA trials slightly underestimate treatment effectiveness relative to UK practice, although the absence of high-quality demographic data around severe AA in the UK makes this uncertain.

BRAVE-AA1 and BRAVE-AA2 were international trials with no UK centres. BRAVE-AA1 recruited most patients from the USA (54.7% of patients) and South Korea (37.8% of patients), and BRAVE-AA2 recruited most patients from the USA (34.8% of patients) and from Asian sites (26.9% of patients). Despite certain demographics, such as race, differing systematically from UK practice, the EAG's clinical experts did not believe treatment efficacy would differ substantially between geographic region or across races, and agreed that other baseline characteristics were broadly similar to those that would be seen in UK practice, which are presented in Appendix Table 44.

As highlighted in Section 2.3, around [redacted] of participants in the BRAVE-AA trials had received prior therapies for AA, and over [redacted] had received systemic immunosuppressants/immunomodulators that the EAG's clinical experts noted would only be given for severe AA in UK practice. These data are presented in Table 12. The EAG's clinical experts stated that while some of these prior treatments would be common in NHS patients, such as topical corticosteroids from their GPs, others are not widely available or used in the UK, such as cryotherapy and phototherapy, and some, such as cyclosporin and topical immunotherapy are used in the UK but at a lower prevalence than was reported in the BRAVE-AA trials.

Table 12. Prior therapies received by patients in the placebo and baricitinib 4 mg arms of BRAVE-AA1 and BRAVE-AA2 (Adapted from CS Table 11 and CS Table 12)

	BRAVE-AA1		BRAVE-AA2	
	Placebo (N=189)	Baricitinib 4 mg (N= 281)	Placebo (N=156)	Baricitinib 4 mg (N= 234)
Prior therapy, n (%)	173 (91.5)	247 (87.9)	149 (95.5)	211 (90.2)
Topical therapy, n (%)	108 (57.1)	173 (61.6)	98 (62.8)	148 (63.2)
Topical IMT, n (%)	45 (23.8)	84 (29.9)	41 (26.3)	63 (26.9)
Intralesional therapy, n (%)	101 (53.4)	152 (54.1)	88 (56.4)	104 (44.4)
Systemic agents, n (%)				
Immunosuppressant/ immunomodulator	101 (53.4)	138 (49.1)	97 (62.2)	124 (53.0)
Corticosteroids	68 (36.0)	103 (36.7)	77 (49.4)	102 (43.6)
JAK inhibitor[a]	12 (6.3)	15 (5.3)	9 (5.8)	10 (4.3)
Others	57 (30.2)	88 (31.3)	54 (34.6)	52 (22.2)

Cyclosporin	46 (24.3)	69 (24.6)	27 (17.3)	27 (11.5)
Methotrexate	15 (7.9)	28 (10.0)	27 (17.3)	31 (13.2)
Other systemic (non-immunosuppressant), n (%)	17 (9.0)	28 (10.0)	15 (9.6)	18 (7.7)
Phototherapy, n (%)	23 (12.2)	54 (19.2)	28 (17.9)	37 (15.8)
Procedures, n (%)	30 (15.9)	65 (23.1)	35 (22.4)	47 (20.1)

Source: CS Table 11 and Table 12

Abbreviations: AA: alopecia areata; CI: confidence interval; IMT: immunotherapy; JAK: janus kinase

[a]Patients with prior inadequate response to JAK inhibitors were excluded from the trial, although no patients failed screening for this reason (inclusion/exclusion criteria #9, BRAVE-AA1 CSR and BRAVE-AA2 CSR)

3.2.4 *Outcome assessment*

The key clinical effectiveness outcome, SALT score, was assessed by investigators blinded to the treatment a patient was receiving. SALT score measurement is a relatively objective procedure in which the assessor compares each quarter of the scalp with a chart detailing how much hair is missing. The EAG's clinical experts noted how SALT measurement can be quite imprecise, and as such cautioned against using strict absolute thresholds, such as SALT ≤20 or $SALT_{75}$ to determine whether treatment should be continued, i.e., they might be unlikely to recommend a patient achieving a SALT score 21 to discontinue treatment.

The key HRQoL data were collected by self-report from blinded participants and the EAG does not have concerns about the validity of this data collection. Section 3.3.5 and Section 4.2.8.1 contains a critique of the company's claim of a lack of content validity for the EQ-5D-5L measure for severe AA.

3.3 Critique of the clinical effectiveness analysis

The SALT score was the primary focus of the company's clinical effectiveness analysis, and the absolute measure, SALT ≤20 was the primary outcome of the BRAVE-AA trials at Week 36. In the base case analysis, the company focuses on relative rather than absolute measures of hair regrowth. The definitions of the SALT outcomes used throughout the CS are presented in Table 13.

Table 13. Example definition of SALT outcomes used in the CS.

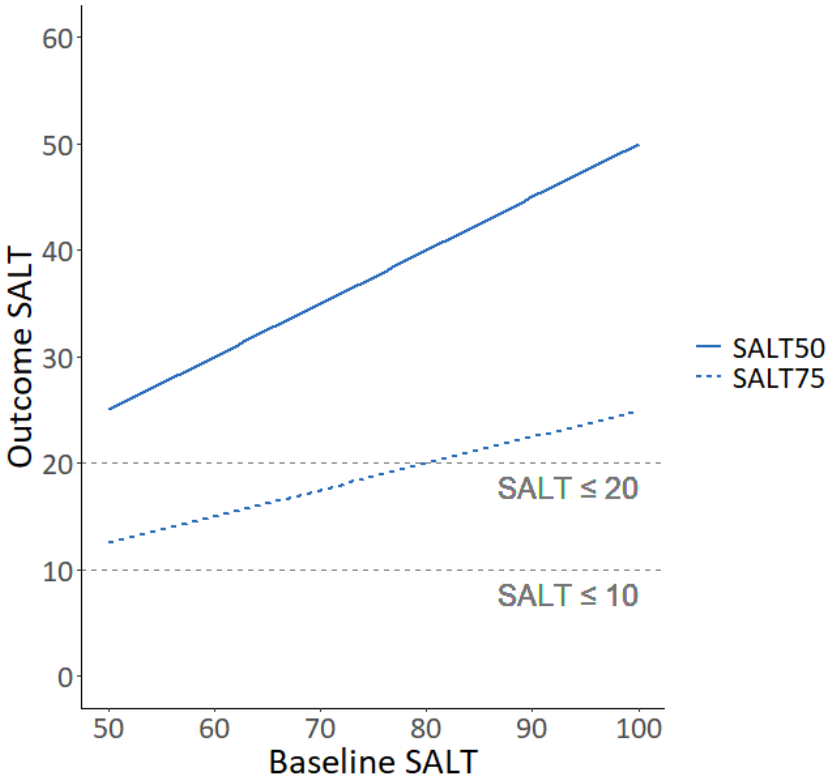
SALT Outcome	Definition
Absolute measures	
SALT ≤10, SALT ≤20	A SALT score of less than or equal to 10 (SALT ≤10) or 20 (SALT ≤20) at the timepoint.
Relative measures	

$SALT_{75}$, $SALT_{50}$	A 75% ($SALT_{75}$) or a 50% ($SALT_{50}$) reduction from baseline in an individual's SALT score at the timepoint.

Abbreviations: CS: company submission; SALT: severity of alopecia tool

Figure 3 displays the relationship between SALT ≤10, SALT ≤20, $SALT_{75}$ and $SALT_{50}$. SALT ≤10 is the most stringent criteria that the EAG's clinical experts agreed would be a strong clinically meaningful outcome for nearly all patients, alleviating the need for wig use. The EAG's clinical experts also agreed that SALT ≤20 and $SALT_{75}$ would be clinically meaningful outcomes and are near equivalent for severe AA patients. $SALT_{50}$, however, is a much less stringent criterion and the EAG's clinical experts doubted whether this would be a meaningful outcome for many patients, who would still have a large degree of hair loss and would likely still opt for wigs and/or head shaving.

Figure 3. The relationship between absolute and relative SALT thresholds used to define responders in the CS.



In the following section, the EAG critiques the clinical effectiveness analysis of the company in the CS. In general, the EAG considers the statistical comparisons between baricitinib 4 mg and placebo to be robust and to present a clear benefit of baricitinib 4 mg over placebo. The results of the additional analyses requested by the EAG in clarification question A11 were consistent with the

analyses presented in the CS, suggesting the company results are robust to different analytical approaches.

Throughout the CS, five different analysis sets were reported on, which are detailed in Table 13 of the CS. The clinical effectiveness used in the CS analyses were conducted in the full analysis sets (FAS) of BRAVE-AA1 and BRAVE-AA2, which comprised all patients randomised at baseline, or the pooled Week 36 efficacy population, which combines the BRAVE-AA1 FAS and the BRAVE-AA2 FAS. The EAG considers these the appropriate analysis sets to use for the clinical efficacy analysis. The safety analysis set comprised all patients randomised who receive at least one dose of study intervention and who did not discontinue from the study for the reason 'Lost to Follow-up' at the first post-baseline visit.

3.3.1 SALT responder outcomes

The proportion of patients achieving a SALT ≤20, SALT ≤10, $SALT_{50}$ and $SALT_{75}$ response in BRAVE-AA1 and BRAVE-AA2 at Week 36 are presented in Table 14. Pooled across BRAVE-AA1 and BRAVE-AA2 baricitinib 4 mg arms, the $SALT_{50}$ response rate was [redacted] and $SALT_{75}$ response rate was [redacted]. The SALT ≤20 response rate was similar to $SALT_{75}$ at [redacted], and the SALT ≤10 response rate was [redacted]. For all statistical comparisons, baricitinib 4 mg had a significantly higher Week 36 response rate than placebo, with all p-values [redacted], and all odds ratios [redacted] (lowest 95% CI: [redacted]). Detailed tables of statistical results for each outcome measure are provided in Table 45, Table 46 and Table 47 of the Appendix.

Table 14. SALT response rates at Week 36 in BRAVE-AA1 and BRAVE-AA2 (FAS)

Week 36	BRAVE-AA1		BRAVE-AA2	
	Placebo (N=189)	Baricitinib 4 mg (N=281)	Placebo (N=189)	Baricitinib 4 mg (N=281)
SALT ≤20, % (95% CI)	5.3 (2.9 to 9.5)	35.2 (29.9 to 41.0)	2.6 (1.0 to 6.4)	32.5 (26.8 to 38.7)
SALT ≤10, % (95% CI)	[redacted]	[redacted]	[redacted]	[redacted]
$SALT_{50}$, % (95% CI)	[redacted]	[redacted]	[redacted]	[redacted]
$SALT_{75}$, % (95% CI)	[redacted]	[redacted]	[redacted]	[redacted]

Source: CS Tables 15, 18, 19 and 22

Abbreviations: AA: alopecia areata; CI: confidence interval; FAS: full analysis set; SALT: severity of alopecia tool

The EAG notes that while response rates were significantly higher for baricitinib 4 mg than placebo for all outcomes, only around [redacted] of patients achieved SALT ≤20 or $SALT_{75}$ by Week 36. However, the EAG also notes that over [redacted] of SALT ≤20 responders were also SALT ≤10 responders at Week 36, suggesting that many responders at the SALT ≤20 or $SALT_{75}$ thresholds had a large and clinically meaningful response. The EAG further notes that response rate results were replicated successfully between two large, multi-site, high-quality international trials. In addition, the results of the baricitinib 2 mg arms consistent with the baricitinib 4 mg arms, with the expected lower magnitude for the lower dose. Overall, the EAG considers the trials to have strong internal validity and that the SALT responder results are likely robust within the inclusion and exclusion criteria of the BRAVE-AA trials.

In addition to the dichotomous responder-based analysis, the company provided some data on the mean change from baseline in SALT score at Week 36 in Table 16 and Table 17 of the CS. The mean (SE) change from baseline in SALT score for baricitinib 4 mg was -45.79 (2.66) in BRAVE-AA1 and -47.45 (2.23) in BRAVE-AA2, compared to -8.13 (3.10) and -2.96 (2.72) in the BRAVE-AA1 and BRAVE-AA2 placebo arms, respectively. The EAG considers the result from the change from baseline analyses to be consistent with the responder-based analysis.

3.3.1.1 Week 52 and Week 76 data

The EAG's clinical experts agreed that around Week 36 is a reasonable time to assess the effect of a JAK inhibitor on patients. At Week 36, the double-blind treatment phase of BRAVE-AA1 and BRAVE-AA2 ended and non-responders in the placebo arm were randomised to baricitinib 2 mg or 4 mg rescue treatment. In contrast, patients who started on one of the baricitinib arms continued on this arm until at least Week 52. The pooled SALT ≤20 response data for these patients by visit until Week 52 are presented in Figure 4. At Week 52, [redacted] of patients in the baricitinib 4 mg had achieved SALT ≤20, an increase of [redacted] from Week 36 ([redacted] had achieved SALT ≤20 at Week 36).

Figure 4. Proportion of patients achieving SALT≤20 through Week 52 in the BRAVE-AA studies (pooled Week 52 efficacy population; primary censoring [NRI]). Reproduction of CS Figure 20.

Abbreviations: BARI: baricitinib; NRI: non-responder imputation; SALT: Severity of Alopecia Tool.

Some Week 76 data were also presented in Section B.2.8.2 of the CS. These data come from the randomised withdrawal sub-study of BRAVE-AA1 and the down titration sub-study of BRAVE-AA2. Of the [redacted] responders at Week 52 who were re-randomised to stay on baricitinib 4 mg in BRAVE-AA1, [redacted] maintained their SALT ≤20 response at Week 76. In BRAVE-AA2, of the [redacted] responders at Week 52 re-randomised to stay on baricitinib 4 mg in BRAVE-AA1, [redacted] maintained their SALT ≤20 response at Week 76.

Overall, the EAG considers that the Week 52 and Week 76 data presented in the CS may indicate that the Week 36 SALT ≤20 response rates underestimate the long-term efficacy of baricitinib 4 mg treatment. However, because of the uncertainty in these data and absence of placebo data from Week 36, the EAG consider the Week 36 data to be most appropriate to use in the economic analyses.

3.3.2 Subgroup analyses

The EAG's clinical experts outlined three variables that might be associated with the probability of hair regrowth: baseline SALT score or disease severity, length of current AA at baseline and, presence of an atopic background. In the company's prespecified subgroup analyses (CS Section

B.2.7), no significant subgroup-by-treatment interaction terms were observed in BRAVE-AA1 or BRAVE-AA2 and subgroup effects were not considered further. The EAG considers these analyses likely underpowered to detect meaningful subgroup-by-treatment interaction terms because:

- Dichotomised outcome and predictor variables were used when continuous data were available;
- There was a floor effect in the placebo response rate;
- The trial sample size was chosen only to provide appropriate power to detect a main effect of treatment in the primary efficacy analysis.

While the EAG recognises interaction modelling is usually the appropriate approach to detecting subgroup effects, the EAG considers assessing the magnitude of main effect of subgroup within the baricitinib 4 mg arm to be an appropriate measure of subgroup effects in the current appraisal. The company provide this analysis for baseline SALT score in Appendix E of the CS, and for length of current episode at baseline in the BRAVE-AA clinical study reports. These are presented below, and a subgroup analysis based on atopic background status requested in clarification question A14 is presented in Appendix Section 8.3.

3.3.2.1 Baseline severity

Patients in the baricitinib 4 mg who had severe AA at baseline, i.e., SALT 50–94 at baseline, had a higher probability of achieving SALT ≤20 at Week 36 than patients with very severe AA at baseline, i.e., SALT 95–100 at baseline (BRAVE-AA1: severe responders, [redacted]; very severe responders, [redacted]. BRAVE-AA2: severe responders, [redacted]; very severe responders, [redacted]), a relationship that was replicated at the lower 2 mg dose. The EAG notes that a relationship in this direction is expected, as severe patients with SALT 50–94 at baseline are already closer to the SALT ≤20 threshold than very severe patients. Nevertheless, the magnitude of the difference is notable, with severe patients having over twice the SALT ≤20 response rate as very severe patients. These data are presented in Figure 5 and Figure 6.

Figure 5. Proportion of patients with SALT≤20 at Week 36 by baseline AA severity in BRAVE–AA1 (FAS population; primary censoring rule). Reproduction of Figure 3 from CS Appendix E.

Footnotes: **p<0.05 vs placebo; **p<0.01 vs placebo; ***p<0.001 vs placebo.*

Abbreviations: AA, alopecia areata; BARI, baricitinib; FAS, full analysis set; NRI, non–responder imputation; PBO, placebo; SALT, Severity of Alopecia Tool.

Source: CS Appendix E

Figure 6. Proportion of patients with SALT≤20 at Week 36 by baseline AA severity in BRAVE–AA2 (FAS population; primary censoring rule). Reproduction of Figure 4 from CS Appendix E.

Footnotes: **p<0.05 vs placebo; ***p<0.001 vs placebo.*

Abbreviations: AA, alopecia areata; BARI, baricitinib; FAS, full analysis set; NRI, non–responder imputation; PBO, placebo; SALT, Severity of Alopecia Tool.

Source: CS Appendix E

3.3.2.2 *Length of current AA episode at baseline*

The length of current AA episode was highlighted by the EAG's clinical experts as a meaningful variable likely to predict hair regrowth. Table 15 provides SALT ≤20 response data for the baricitinib 4 mg and placebo arms of BRAVE-AA1 and BRAVE-AA2 by the duration of current AA episode at baseline (<4 years and ≥4 years categories). For the baricitinib 4 mg arm, a larger proportion of patients in the <4 years category achieved SALT ≤20 at Week 36 (BRAVE-AA1, [redacted]; BRAVE-AA2, [redacted]) than those in the ≥4 years category (BRAVE-AA1, [redacted] BRAVE-AA2, [redacted]), a pattern also apparent in the placebo arm data (<4 years category: BRAVE-AA1, [redacted]; BRAVE-AA2, [redacted]; ≥4 years category; BRAVE-AA1, [redacted]; BRAVE-AA2, [redacted]). In response to clarification question A6, the company provided the correlation between baseline SALT score and baseline length of current AA episode at baseline. No meaningful correlation was observed (Pearson's r = 0.087 in the pooled placebo arms and Pearson's r = 0.039 in the pooled baricitinib 4 mg arms), suggesting that duration of current AA episode and baseline severity are relatively independent and meaningful subgroups.

Table 15. SALT ≤20 response of BRAVE-AA1 and BRAVE-AA2 by the duration of current AA episode at baseline (<4 years and ≥4 years categories)

	Duration of current AA episode at baseline	Week 36 SALT ≤20 response rate	
		BRAVE-AA1	BRAVE-AA2
Baricitinib 4 mg	<4 years	[redacted]	[redacted]
	≥4 years	[redacted]	[redacted]
Placebo	<4 years	[redacted]	[redacted]
	≥4 years	[redacted]	[redacted]

Source CS: Table 30, CS Table 31, BRAVE-AA1 CSR page 310, BRAVE-AA2 CSR page 324
Abbreviations: AA: alopecia areata; SALT: severity of alopecia tool

The company provided scenario analyses based on current duration of AA episode at baseline and results are given in Section 5.1.2.2.

3.3.3 *Withdrawal, down-titration, and relapse*

Data from the randomised withdrawal sub-study (BRAVE-AA1) and the down-titration sub-study (BRAVE-AA2) are presented in CS Section B.2.8.2. [redacted] SALT ≤20 responders in the BRAVE-AA1 4 mg baricitinib arm were re-randomised to placebo at Week 52. By Week 76, only [redacted] of these

patients ([redacted]) had maintained their SALT ≤20 response. [redacted] SALT ≤20 responders in the BRAVE-AA2 mg baricitinib arm were re-randomised to 2 mg baricitinib at Week 52. By Week 76, [redacted] of these patients ([redacted]) had maintained their SALT ≤20 response. The EAG notes that these data indicate that baricitinib is only therefore viable as a continued long-term treatment, and once patients have their treatment withdrawn, hair loss is common.

The company stated that data on trial-defined relapse were not yet available (clarification question A10). Relapse was measured in SALT ≤20 responders in the trial after Week 52 and was defined as a >20-point absolute worsening in total SALT score.

3.3.4 Non-SALT measures of hair regrowth

The company presents responder-based results of two non-SALT based hair loss-measures, the PRO Scalp Hair Assessment and the ClinRO measure for eyelash and eyebrow regrowth, in Table 16 and Table 17 of the CS. Similar to the SALT ≤20 responder analysis, approximately one third of patients in the baricitinib 4 mg arms achieved PRO responses and ClinRO responses at Week 36, compared to only around 5% of placebo patients for all measures. The EAG considers these results to be consistent with the SALT ≤20 responder analysis and assures that treatment with baricitinib 4mg leads to hair regrowth beyond the scalp.

3.3.5 Health-related quality of life

3.3.5.1 EQ-5D-5L

No meaningful differences were observed in EQ-5D-5L health state index or visual analogue score (VAS) between baseline and Week 36 for any arm in BRAVE-AA1 or BRAVE-AA2, with no more than a [redacted] mean increase in EQ-5D score across either the placebo or baricitinib 4 mg arm (Table 16).

Table 16. EQ-5D data from BRAVE-AA1 and BRAVE-AA2 at baseline and at Week 36.

EQ-5D	BRAVE-AA1		BRAVE-AA2	
	Placebo	Baricitinib 4 mg	Placebo	Baricitinib 4 mg
Health state index UK, mean (SD)				
Baseline	[redacted]	[redacted]	[redacted]	[redacted]
Week 36	[redacted]	[redacted]	[redacted]	[redacted]
VAS, mean (SD)				
Baseline	[redacted]	[redacted]	[redacted]	[redacted]
Week 36	[redacted]	[redacted]	[redacted]	[redacted]

Source: BRAVE-AA1 CSR pages 267, 272, 274 and 279; BRAVE-AA2 CSR pages 269, 274, 276 and 281

Abbreviations: EQ-5D: EuroQol-5 Dimension; SD: standard deviation; VAS; visual analogue scale

3.3.5.2 HADS

Mean change from baseline in the Hospital Anxiety Depression Scale (HADS) scores at Week 36 were presented in CS Table 25, and are reproduced for the placebo and baricitinib 4 mg arms of BRAVE-AA1 and BRAVE-AA2 in Table 17. Compared to baseline, HADS Anxiety decreased by a statistically significantly larger amount in the baricitinib 4 mg arms (BRAVE-AA1 mean change [SE]: [redacted]; BRAVE-AA2: [redacted]) than in the placebo arms (BRAVE-AA1 [SE]: [redacted]; BRAVE-AA2: [redacted]). However. these changes were lower than the most common definitions of the minimal clinically important difference (MCID) of around 1.7 to 2 for HADS scales,[22, 23] although this has not been validated in dermatology[24] or AA specifically. HADS Depression decreased in the baricitinib 4 mg arms of BRAVE-AA1 ([redacted]) and BRAVE-AA2 ([redacted]) but increased slightly from baseline in the placebo arms (BRAVE-AA1 [SE]: [redacted]; BRAVE-AA2: [redacted]). Only the difference between baricitinib 4 mg and placebo in BRAVE-AA2 for HADS Depression was statistically significant at $p < 0.05$.

Table 17. Mean change from baseline in HADS-Anxiety and HADS-Depression scores at Week 36 for the placebo and baricitinib 4 mg arms of BRAVE-AA1 and BRAVE-AA2

Week 36	BRAVE-AA1		BRAVE-AA2	
	Placebo (N = 189)	Baricitinib 4 mg (N = 281)	Placebo (N = 156)	Baricitinib 4 mg (N = 234)
HADS Anxiety				
Mean (SD) baseline score	[redacted]	[redacted]	[redacted]	[redacted]
LSM (SE)	[redacted]	[redacted]	[redacted]	[redacted]
p-value vs placebo	[redacted]	[redacted]	[redacted]	[redacted]
HADS Depression				
Mean (SD) baseline score	[redacted]	[redacted]	[redacted]	[redacted]
LSM (SE)	[redacted]	[redacted]	[redacted]	[redacted]
p-value vs placebo	[redacted]	[redacted]	[redacted]	[redacted]

Source: CS Table 25

Abbreviations: HADS: Hospital anxiety depression scale; LSM: least squares mean; SD: standard deviation; SE: standard error

3.3.5.3 SF-36 and Skindex-16

In addition to EQ-5D and HADS data, the company presented HRQoL data using the SF-36 and Skindex-16 measures at Week 36. In the SF-36 measure:

- There were no significant differences in change from baseline SF-36 physical component score between placebo and baricitinib 4 mg arms in either BRAVE-AA1 or BRAVE-AA2 (CS Table 28);
- There was a statistically significantly greater increase for the baricitinib 4 mg arm versus placebo in BRAVE-AA2 in the SF-36 mental component score, but not, however, in BRAVE-AA1 (CS Table 29). Moreover, the increase in baseline for the baricitinib 4 mg arm, [redacted] (95% CI difference from placebo: [redacted]), is below the most common definitions of the MCID for SF-36, around 3 to 5 points,[25, 26] although data in dermatology is scarce.[27-29] In the baricitinib 4 mg arm of BRAVE-AA1, the SF-36 mental component score numerically worsened from baseline (mean change from baseline: baricitinib 4 mg [SE], [redacted] [redacted]; placebo: [redacted].

In contrast to the SF-36 measure, a large benefit of baricitinib 4 mg over placebo was observed in the Skindex-16 measure, adapted for AA. Skindex-16 is a patient reported questionnaire designed to measure the effects of skin disease on quality of life, which comprises three domains: symptoms, emotions and functioning. Patients in the baricitinib 4 mg arms of BRAVE-AA1 and BRAVE-AA2 had a statistically significant improvement in mean change from baseline to Week 36 in the emotions and functioning domain, that was also significantly greater than the change from baseline in placebo (CS Table 23). While there was also a greater reduction in the symptom component of the Skindex-16 in the baricitinib 4 mg arms of BRAVE-AA1 and BRAVE-AA2, this reduction was only statistically significantly greater than placebo in BRAVE-AA2 (CS Table 23). Moreover, no specific MCID has been validated to date for the Skindex-16 or the AA-adapted version which make the clinical meaning of these results difficult to interpret.[30]

3.3.5.4 EAG critique of BRAVE-AA1 and BRAVE-AA2 trial HRQoL data

In the BRAVE-AA trials, no improvement in HRQoL was observed in EQ-5D for baricitinib 4 mg over placebo by Week 36 and only modest improvements were observed in the HADS and SF-36 measures. A larger improvement was observed in the Skindex-16 scale; a skin disease specific scale designed to be sensitive to quality-of-life changes caused by skin disease.

The company provided three arguments for why the EQ-5D data observed in BRAVE-AA1 and BRAVE-AA2 may be unsuitable for use in the economic models. The EAG does not find these arguments convincing and provides comments on them in Table 18, although the EAG's clinical experts did note reservations about the suitability of each of the EQ-5D scales for measuring QoL in AA, highlighting

how the majority of benefits will be most visible in psychosocial functioning, which the EQ-5D only captures partially.

Table 18. The EAG's critique of the company's discussion of the limitations and representative of the EQ-5D data collected din BRAVE-AA1 and BRAVE-AA2

Company Argument	EAG comment
Patients' baseline HRQoL data were near ceiling in BRAVE-AA1 and BRAVE-AA2, limiting the scope for patients' HRQoL to be improved by treatment. At baseline, patients median EQ-5D was [redacted] or higher in all BRAVE-AA arms, i.e., around the UK population norm for EQ-5D of 0.91 for males and females aged 35-44.[31]	BRAVE-AA1 and BRAVE-AA2 were judged to be high quality international clinical trials with large sample sizes, and the EQ-5D data were replicated between the trials. The EAG considers it likely the trials have appropriately measured a high baseline EQ-5D at the population level.
The AA patients who could have gained most utility benefit through hair regrowth may have been excluded in the BRAVE-AA studies. Specifically, patients with the presence of significant uncontrolled neuropsychiatric disorder, or who were clinically judged by the investigator to be at risk for suicide were excluded from the trials.	In response to clarification question A8 the company confirmed [redacted] of significant uncontrolled neuropsychiatric disorders, but noted that these patients may not have chosen or been invited to screening for the trial—something the EAG's clinical experts agreed with, and pointed to a systematic review and meta-analysis showing a negative relationship between HRQoL and anxiety and depression in AA.[32] However, the EAG does not consider the company to have provided sufficient evidence that either: i) the trial populations were missing a large cohort of patients who would have had low baseline EQ-5D scores nor, ii) provided evidence of the size of this cohort in UK clinical practice.
Scales such as the EQ-5D may lack content validity for indications like AA, which may not lead to issues with patients' mobility, cause pain or impede usual activities (three of the five domains of the EQ-5D).	The company have not provided relevant psychometric data demonstrating a lack of content validity of the EQ-5D in severe AA, and nevertheless continued to use the EQ-5D measure from the Adelphi DSP study in their base case analysis. While it is plausible that not all domains of the EQ-5D may be directly affected by AA—something noted by the EAG's clinical experts—it is possible that AA may have indirect effects on these domains.

Abbreviations: AA: alopecia areata; EAG: evidence assessment group; EQ-5D: EuroQol-5 Dimension; HADS: Hospital anxiety depression scale; HRQoL: health related quality of life.

Hence, the EAG finds it plausible that the BRAVE-AA trials have adequately estimated only a small gain in utility following baricitinib 4 mg treatment at the population level and considers the EQ-5D data collected in the BRAVE-AA1 and BRAVE-AA2 trials to be suitable to inform decision making. Nevertheless, the EAG recognises that severe AA can and does have large negative impacts on quality of life for some patients, something highlighted by the EAG's clinical experts. The EAG

believes that this may not equate to large changes in EQ-5D score at the population level, however, because:

- In the large sample, high-quality BRAVE-AA trials, many severe AA patients had a genuinely near-ceiling EQ-5D;
- Only [redacted] of baricitinib 4 mg patients achieved the clinically meaningful SALT ≤20 response rate at Week 36, meaning that any treatment effect at the population level on HRQoL would be diluted by the [redacted] non-responders.

Hence, because only a minority of patients may have EQ-5D deficits at baseline, and only a minority of these patients will likely respond to treatment, any EQ-5D improvements at the population level are likely to be small. The EAG's clinical experts also noted that:

- HRQoL benefits may lag behind a treatment response in severe AA, as a patient adjusts to the changes in their appearance;
- Baricitinib is not a curative treatment, and patients may continue to suffer anxiety because baricitinib needs to be taken continuously over a long period to maintain hair regrowth, with missed doses potentially resulting in hair loss.

3.3.6 Safety data

The company present the adverse events (AEs) observed in BRAVE-AA1 and BRAVE-AA2 in Section B.2.10 of the CS. A slightly higher proportion of patients treated with baricitinib 4 mg in BRAVE-AA1 ([redacted]) and BRAVE-AA2 ([redacted]) had at least one treatment emergent adverse event (TEAE) than in the placebo arms (BRAVE-AA1 placebo: [redacted]; BRAVE-AA2 placebo: [redacted]). Similarly, a slightly higher proportion of patients treated with baricitinib 4 mg in BRAVE-AA1 ([redacted]) and BRAVE-AA2 ([redacted]) had at least one serious adverse event (SAE) than in the placebo arms (BRAVE-AA1 placebo: [redacted]; BRAVE-AA2 placebo: [redacted]). These data are presented in Table 19, alongside pooled data from the baricitinib 4 mg arm extension phases up to August 2021 (providing approximately 6 months additional data following the first data cuts in February 2021). These data are in-line with the BRAVE-AA1 and BRAVE-AA2 data from Week 36

Table 19. Overview of adverse events in the BRAVE-AA studies up to Week 36, and from the extension phase up to August 2021.

	BRAVE-AA1	BRAVE-AA2	Pooled Extension Phase

	Placebo (N=189)	Baricitinib 4 mg (N=280)	Placebo (N=154)	Baricitinib 4 mg (N=233)	Baricitinib 4 mg (N=540)
Patients with ≥1 TEAE, n (%)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Deaths	0	0	0	0	[redacted]
SAEs, n (%)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
AEs leading to permanent discontinuation from study intervention, n (%)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
AEs leading to discontinuation from study, n (%)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Abbreviations: AE: adverse event; SAE: serious adverse event; TEAE: treatment emergent adverse event.

Source: CS Table 37 and Table 42

While the rates of AEs were slightly higher for baricitinib 4 mg over placebo, the EAG considers that baricitinib 4 mg had a relatively safe safety profile over the study period. Notably, adverse events of special interest due to the mechanism of baricitinib were not greatly elevated over placebo (Table 20).

Table 20. Adverse events of special interest across all treatment groups in the BRAVE-AA trials. Adapted from CS Table 41.

Adverse event	BRAVE-AA1		BRAVE-AA2	
	Placebo n (%)	Baricitinib 4 mg n (%)	Placebo n (%)	Baricitinib 4 mg n (%)
	(N = 189)	(N = 280)	(N = 154)	(N = 233)
Patients with ≥1 TE infection	[redacted]	[redacted]	[redacted]	[redacted]
TE herpes zoster	[redacted]	[redacted]	[redacted]	[redacted]
TE herpes simplex	[redacted]	[redacted]	[redacted]	[redacted]
Positively adjudicated MACE	[redacted]	[redacted]	[redacted]	[redacted]
Positively adjudicated VTE	[redacted]	[redacted]	[redacted]	[redacted]
Positively adjudicated ATE	[redacted]	[redacted]	[redacted]	[redacted]
Gastrointestinal perforation	[redacted]	[redacted]	[redacted]	[redacted]
Nonmelanoma skin cancer	[redacted]	[redacted]	[redacted]	[redacted]
Malignancies other than NMSC	[redacted]	[redacted]	[redacted]	[redacted]

Abbreviations: AE: adverse event; MACE: major adverse cardiovascular event; NMSC: nonmelanoma skin cancer; TE: treatment emergent.

Source: CS Table 41, CS page 104

The EAG notes that long-term safety data are not yet available for baricitinib in adults with severe AA. The EAG's clinical experts noted the importance of post marketing pharmacovigilance and highlighted the uncertainty they might feel in giving a patient such an immunomodulator for a long period of time, with serious infections, thromboembolic disease and malignancy being highlighted as long-term safety concerns. The EAG's clinical experts noted these are similar concerns that they have for atopic dermatitis patients who are already being prescribed baricitinib in the NHS, and that longer-term data with up to 4.6 years of follow-up for baricitinib in the treatment of rheumatoid arthritis are still somewhat uncertain, although the results were consistent with the short-term data from the primary publication of the clinical trials.[33]

3.3.6.1 Baricitinib 2mg dose

The EAG notes that the rate of AEs observed in the BRAVE-AA trials was lower for the baricitinib 2 mg. This dose may be used for patients:

- Aged ≥75 years;
- With a history of chronic or recurrent infections;
- Who have dose tapered.

The EAG notes, however, that no patients aged ≥75 years were included in the BRAVE-AA trials. As such the safety data from the trial are unlikely to be representative of this population.

3.4 Critique of the indirect comparison and/or multiple treatment comparison feasibility assessment

The company conducted a feasibility assessment for a network meta-analysis (NMA) comparing the drugs comprising best supportive care for severe AA with baricitinib (Section B.2.9.2 of the CS). The company concluded that no NMA or indirect treatment comparisons (ITCs) were feasible because only two placebo-controlled randomised controlled trials (RCTs) formed a connected evidence loop with BRAVE-AA1 and BRAVE-AA2. Neither of these RCTs reported similar outcomes to BRAVE-AA1 or BRAVE-AA2 at similar timepoints, nor reported sufficient or similar treatment effect modifying baseline characteristics to the BRAVE-AA trials.

Twenty-one RCTs or observational studies did not form a connected network with BRAVE-AA1 and BRAVE-AA2. For these studies:

- Key baseline characteristics were poorly reported and often differed to the BRAVE-AA studies;
- Most studies (17 out of 21) included patients with mild or moderate AA;
- Most studies did not report similar outcomes to BRAVE-AA1 and BRAVE-AA2, and not at similar timepoints.

The EAG considers the company's feasibility assessment to be thorough, and the EAG agrees with the company's decision not to perform an NMA or ITCs. Any unanchored comparisons would be at very high risk of bias and the EAG does not believe it would be possible to appropriately adjust for treatment effect modifiers that differ between the studies.

In the absence of viable NMA or ITC data, the EAG considers there to be some unresolved uncertainty concerning the comparative effectiveness of baricitinib against some, but not all, of the current best supportive care therapies for severe AA. The EAG agrees with the company that most supportive care therapies have very limited effectiveness in treating severe AA. For these therapies, the placebo arm of BRAVE-AA1 and BRAVE-AA2 is an acceptable approximation for the treatment response. However, for DPCP there is evidence of a treatment effect above that of no treatment in some severe AA patients who can tolerate the treatment. Nevertheless, the EAG does not consider comparing baricitinib directly with DPCP to be relevant to this submission as:

- Only a minority of patients receive DPCP, and it causes strong allergic reactions in many of these patients (e.g., serve eczema was reported as treatment-emergent adverse event in 31% of patients in a large meta-analysis[10]);
- The magnitude of the effectiveness of DPCP may have been overestimated by various biases in the efficacy analyses of DPCP trials (see company's response to clarification question A9);
- Over 25% of patients in the BRAVE-AA1 and BRAVE-AA2 trials had been previously treated with topical immunotherapy (and hence are likely DPCP failures), and as such are likely a more severe population than those in the DPCP trials.

Hence, while the EAG maintains that there is some unresolved uncertainty around how much more effective baricitinib might be over DPCP, the EAG notes this is likely a sizeable benefit both in efficacy and safety. As outlined in Section 2.3.3, the EAG considers no active treatment, informed by the placebo arm of the BRAVE-AA trials, to be the most appropriate comparator for this appraisal.

3.5 Conclusions of the clinical effectiveness section

In the CS, the company has presented clinical effectiveness and safety evidence in support of baricitinib for treating adults with severe AA, an indication for which patients have a clear unmet need. The company's evidence comes primarily from the international, placebo-controlled BRAVE-AA1 and BRAVE-AA2 randomised controlled trials. The EAG assessed BRAVE-AA1 and BRAVE-AA2 to be high quality trials, and to provide strong evidence of a clinically meaningful benefit of baricitinib 4 mg over placebo. Pooled across the BRAVE-AA trials, [redacted] of participants in the baricitinib 4 mg achieved the primary SALT ≤20 outcome at Week 36, compared to only [redacted] in the placebo arms, although [redacted] is still a minority of patients.

The EAG considers the trial populations of BRAVE-AA1 and BRAVE-AA2 to be suitably similar to patients in the UK to inform decision making, despite neither of the trials including UK centres. The EAG noted that the exclusion of males >60 years and females >70 years, and the exclusion of patients with AA episode durations >8 years at baseline and no previous sign of regrowth may have biased the efficacy estimates in BRAVE-AA1 and BRAVE-AA2 in favour of baricitinib over placebo. However, the EAG considered any overestimation of the efficacy of baricitinib likely to be balanced by several features of the trial that might cause the efficacy to be underestimated, namely a baseline population with a high rate of treatment-experience, a relatively severe population at baseline and a conservative use of non-responder imputation. The EAG therefore concludes that the trials provide a reasonably unbiased estimate of the efficacy of baricitinib 4 mg in adults with severe AA.

The EAG considers there to be two clinically meaningful variables that predict a patient's response to baricitinib 4 mg treatment: baseline SALT score and baseline duration of current AA episode. Patients with a shorter duration of AA episode and patients with a lower SALT score at baseline are more likely to achieve a clinically meaningful response than patients with a longer duration of current AA episode and patients with higher SALT scores at baseline. The EAG notes, however, that AA episode duration and SALT score are continuous variables for which categorical subgroups cannot be clinically defined.

The EAG considers the BRAVE-AA trials to provide high-quality health-related quality of life data on adults with severe AA. There were no significant increases in baseline for EQ-5D or SF-36 in either of the baricitinib 4 mg arms of BRAVE AA-1 or BRAVE AA-2, and the EAG considers this to be likely due to: i) a genuine high baseline quality of life in many adults with severe AA and, ii) a low absolute response rate to baricitinib 4 mg treatment, i.e., [redacted]. Hence, while many patients may have a

greatly reduced quality of life because of their severe AA, treatment with baricitinib 4 mg may only lead to modest HRQoL benefits at the population level.

The EAG considers baricitinib 4 mg to have displayed a relatively strong short-term safety profile in the BRAVE-AA trials. Over the 36-week double-blind treatment phases of BRAVE-AA1 and BRAVE-AA2, there was only a slightly higher rate of treatment emergent adverse events (TEAEs) and serious adverse events (SAEs) in the baricitinib 4 mg arm versus the placebo arm. The absolute rate of SAEs was low, and the safety profile of baricitinib 4 mg in the BRAVE-AA trials was consistent with its safety profile from trials informing TA681[15] and TA466.[16] Nevertheless, the EAG considers there to be uncertainty concerning the long-term safety of baricitinib 4 mg in treating adults with severe AA, with the EAG's clinical experts noting reservations about providing long-term immunomodulators to otherwise healthy young adult patients with severe AA.

Similarly, the EAG considers there to be uncertainty concerning the long-term effectiveness of baricitinib 4 mg in treating adults with severe AA. The small amount of data available at Week 52 and Week 72 suggests that the [redacted] response rate at Week 36 may be an underestimate of the long-term effectiveness of baricitinib. However, in the absence of comparative data at these timepoints, the EAG considers the Week 36 response rates to be the most robust estimate of the relative efficacy of baricitinib at a clinically relevant timepoint. However, data from later timepoints in the withdrawal sub study of BRAVE-AA1 and down-titration study in BRAVE-AA2 suggest that hair loss is common as soon as treatment is stopped, and treatment efficacy reduced upon down-titration.

In general, the EAG considers the submitted evidence to suitably match the decision problem defined in the final scope issued by NICE.[11] However, the EAG notes that the comparator proposed by the company, "Watch and wait" with active monitoring, is not a common management strategy used in clinical practice for adults with severe AA. The EAG agrees with the company that the no active treatment component of "Watch and wait" is a common management strategy used by patients with severe AA, but notes that this is not usually associated with intensive monitoring. The EAG considered three plausible comparators for baricitinib 4 mg:

- A comparison with DPCP, the most effective treatment used to treat severe AA in clinical practice and only active treatment recommended by the 2012 BAD Guidelines.[9] The EAG notes, however, that there exist no data to perform a valid indirect comparison between

DPCP and baricitinib 4 mg in adult severe AA patients, that DPCP is only available to a minority of AA patients, leads to severe adverse reactions in many patients and has a high rate of relapse;

- A comparison with the "basket" of non-DPCP therapies currently used to treat severe AA, primarily systemic immunosuppressants and systemic corticosteroids. Again, the EAG notes the lack of suitable evidence to perform such indirect comparisons. Moreover, where evidence exists, it suggests these therapies are ineffective;
- A comparison with no active treatment and discharge from care.

The EAG, in agreement with the company, considers a comparison with no active treatment to be the most relevant for the current submission as it reflects a commonly used treatment option by adults with severe AA, and is the only comparison for which high-quality data comparative effectiveness data are available. However, the EAG considers no active treatment and discharge from care, rather than "Watch and wait" with active monitoring, to be the appropriate comparison. The EAG nevertheless notes a large degree of heterogeneity and a lack of data around the treatment pathway of AA and severe AA in UK clinical practice, and potential differences between the prevalent and incident populations of adults with severe AA in clinical practice. For treatment experienced severe AA patients, i.e., the prevalent population in UK practice that would receive baricitinib at the point of approval, the EAG considers no active treatment and discharged from care to be the most appropriate comparator. For newly diagnosed cases of severe AA, the EAG considers it likely that contact immunotherapy, systemic immunosuppressants or systemic steroids may be trialled, but their use is heterogenous and access inequitable, and only contact immunotherapy and wig use are recommended by the 2012 BAD Guidelines. The EAG does not consider there to be a widely accepted standard of care for these patients, and in lieu of robust treatment pattern or comparative effectiveness data, the EAG considers no active treatment and discharge from care to be an acceptable comparator for this population.

4 Cost effectiveness

Table 21 below presents the incremental cost-effectiveness results of the company's updated (i.e., post clarification) base case results.

Table 21. Company's base case results post clarification

Interventions	Total Costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER (£/QALY)
Deterministic results							
'Watch and wait'	[redacted]	22.60	[redacted]	-	-	-	-
Baricitinib	[redacted]	22.60	[redacted]	[redacted]	0.00	[redacted]	18,072
Probabilistic results							
'Watch and wait'	[redacted]	-	[redacted]	-	-	-	-
Baricitinib	[redacted]	-	[redacted]	[redacted]	-	[redacted]	17,942

Abbreviations: ICER, incremental cost effectiveness ratio, LYG, life year gained; QALY, quality adjusted life year.

4.1 EAG comment on the company's review of cost effectiveness evidence

The company carried out a systematic literature review (SLR), using a single search strategy, to identify existing:

- Cost-effectiveness studies for the treatment of adult patients with severe alopecia areata (AA);
- Health-state utility values (HSUVs) for patients with severe AA; and,
- Cost and resource use evidence for the treatment of adult patients with severe AA.

Searches were initially run in August 2021 and were last updated in January 2022. A summary of the Evidence Assessment Group's (EAG's) critique of the methods implemented by the company to identify relevant evidence is presented in Table 22. Due to time constraints, the EAG was unable to replicate the company's searches and appraisal of identified abstracts.

Table 22. EAG's critique of company's systematic literature review

Systematic review step	Section of CS in which methods are reported			EAG assessment of robustness of methods
	Cost effectiveness evidence	HRQoL evidence	Resource use and costs evidence	
Search strategy	Appendix G	Appendix G	Appendix G	Appropriate

Inclusion/ exclusion criteria	Appendix G	Appendix G	Appendix G	Appropriate
Screening	Appendix G	Appendix G	Appendix G	Appropriate
Data extraction	Appendix G	Appendix H	Appendix I	Appropriate
Quality assessment of included studies	Appendix G	Appendix G	Appendix G	Appropriate

Abbreviations: CS, company submission; EAG, evidence assessment group; HRQoL, health related quality of life.

The SLR identified a total of 597 records. The SLR did not identify any cost-effectiveness studies for any treatment for AA. A total of 30 publications related to health-related quality of life (HRQoL) and four costs studies were identified by the SLR.

Of the 30 extracted and HRQoL studies, one reported AA quality of life (AAQOL) index values, 17 reported Dermatology Life Quality Index (DLQI) scores, four reported Hospital Anxiety and Depression Scale (HADS) scores, eight reported Skindex values, three reported EQ-5D values directly and SF-36 values were reported in six studies. However, none of the extracted utility data were deemed suitable by the company and instead utility values used in the model are from a company sponsored study (the Adelphi disease-specific programme (DSP) study)[13, 14] with scenarios informed by utility data from the BRAVE-AA1 and BRAVE-AA2 trials (See Section 4.2.8).

The company considered none of the four cost papers to be useful to inform the economic analysis as the cost data were not UK specific. As with utilities informing the model, the company obtained UK specific resource use from the Adelphi DSP study.[13, 14] Please refer to Section 4.2.9 for further details on the cost and resource use data applied in the model.

4.2 Summary and critique of company's submitted economic evaluation by the EAG

4.2.1 NICE reference case checklist

Table 23 summarises the EAG's appraisal of the company's economic evaluation against the requirements set out in the NICE reference case checklist for the base-case analysis, with reference to the NICE final scope outlined in Section 2.

Table 23. NICE reference case checklist

Element of health technology assessment	Reference case	EAG comment on company's submission
Perspective on outcomes	All direct health effects, whether for patients or, when relevant, carers	All relevant health effects for adult patients with severe AA have been included

Perspective on costs	NHS and PSS	All relevant costs have been included and are based on the NHS and PSS perspective
Type of economic evaluation	Cost–utility analysis with fully incremental analysis	Cost-utility analysis has been provided by the company. Fully incremental analysis not required as there is only one relevant comparator in the analysis.
Time horizon	Long enough to reflect all important differences in costs or outcomes between the technologies being compared	Lifetime horizon (100 years of age)
Synthesis of evidence on health effects	Based on systematic review	The company performed an appropriate systematic review
Measuring and valuing health effects	Health effects should be expressed in QALYs. The EQ-5D is the preferred measure of health-related quality of life in adults.	QALYs based on EQ-5D from a company sponsored Adelphi DSP study.[13, 14]
Source of data for measurement of health-related quality of life	Reported directly by patients and/or carers	EQ-5D obtained from a company sponsored Adelphi DSP study[13, 14] which included AA patients with mild, moderate, severe and very severe disease. EQ-5D data were available directly from patients in the BRAVE-AA1 and BRAVE-AA2 trials, but the company only explored this in a scenario analysis.
Source of preference data for valuation of changes in health-related quality of life	Representative sample of the UK population	The EQ-5D data from the company sponsored Adelphi DSP study[13, 14] were based only on responses from UK patients.
Equity considerations	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	The economic evaluation matches the reference case.
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	Costs included in the analysis have been sourced using NHS reference costs, PSSRU and the NHS Drug tariff.[34-36]
Discounting	The same annual rate for both costs and health effects (currently 3.5%)	Discount rate of 3.5% has been used for both costs and health effects.

Abbreviations: AA, alopecia areata; DSP, disease specific programme; EAG, evidence review group; NHS, national health service; PSS, personal social services; QALY, quality adjusted life year

4.2.2 Population

The modelled population considered by the company for this Single Technology Appraisal (STA) are adults with severe AA, aligned with the [redacted] marketing authorisation for baricitinib. Severe AA was defined as patients with a Severity of Alopecia Tool (SALT) score higher or equal to 50 at baseline and is reflective of the inclusion criteria for the key trials, BRAVE-AA1 and BRAVE-AA2. Additionally, the company explored subgroups based on disease severity (severe defined as SALT score between 50-94 and very severe, defined as SALT score between 95-100).

Baseline characteristics of the modelled population are based on pooled data from BRAVE-AA1 and BRAVE-AA2. Baseline age and sex of the population included in the model are [redacted] years of age and [redacted] male, which the EAG's clinical experts considered were reflective of the patient population in the UK. Baseline characteristics for the severe subgroup analysis are presented in Table 47 of the company submission (CS).

Generally, the modelled population and subgroups are in line with NICE final scope. However, as mentioned in Section 2.3.1, the BRAVE-AA trial population is most similar to the prevalent population in clinical practice who would be eligible to receive baricitinib at the point of approval, i.e. a later-line treatment experienced population. Furthermore, the BRAVE-AA trials provide data on a narrower population than those who could receive baricitinib in clinical practice, as patients with baseline AA episodes >8 years and males >60 years and females >70 years were excluded. Such patients would be eligible to receive baricitinib per the [redacted] marketing authorisation but may be less likely to achieve hair regrowth. Although, the EAG considers that impact on treatment response based on excluded patients is balanced out by the fact that around [redacted] of participants in the BRAVE-AA trials had received, and likely failed on, prior therapies, including over [redacted] receiving therapies used specifically for severe AA in UK practice. Please see Section 2.3.3 for further details.

4.2.3 Interventions and comparators

The intervention considered in the economic analysis is baricitinib 4 mg, once daily. Additionally, the SmPC states that, "*a dose of 2 mg once daily may be appropriate for patients such as those aged ≥ 75 years and for patients with a history of chronic or recurrent infections. A dose of 2 mg once daily may also be considered for patients who have achieved sustained control of disease activity with 4 mg once daily and are eligible for dose tapering*".[37] However, the company has not considered dose

tapering or subgroup analysis by age to account for the 2 mg dose in the CS. However, it should be noted that in the model, less than [redacted] of patients aged over 75 years remain on treatment.

The comparator in the analysis is 'Watch and wait', which the company assumes is akin to the placebo arm in the BRAVE-AA1 and BRAVE-AA2 trials. As such, the 'Watch and wait' arm in the economic model is associated with no drug acquisition costs but includes costs associated with regular monitoring. The NICE final scope lists the main comparator as established clinical management without baricitinib.[11]

4.2.3.1 EAG critique

As mentioned in Sections 2.2.1 and Section 2.3.3, the EAG agrees with the company that the no active treatment component of "Watch and wait" is a common management strategy used for adults with severe AA. However, the EAG's clinical experts considered that patients would not be regularly monitored if they are not on treatment. Additionally, the EAG's clinical experts advised that there is no standard treatment pathway. Patients with severe disease are most likely to have had systemic corticosteroids or systemic immunosuppressants but response to treatment is limited. In particular, the EAG's clinical experts advised that for the prevalent population, patients are likely to have explored all available treatment options and that a significant proportion of patients may not take up treatment. As such, for the prevalent population, patients are likely to manage their condition using wigs or complete hair removal. For the incident (or newly diagnosed) population, the EAG considers (based on advice from its clinical experts) that patients are likely to be less treatment-experienced than the prevalent population but does not consider any specific systemic immunosuppressants or systemic corticosteroids to be an established standard of care for these patients, and considers no active treatment to be a realistic endpoint for these patients.

As such, the EAG considers that the relevant comparator for the patient population is 'discharged from care' and thus the assumption of active monitoring of patients not on any treatment is not reflective of UK clinical practice. As such, for the EAG preferred assumptions, the comparator is defined as 'discharged from care' and the removal of associated costs of monitoring (discussed further in Section 4.2.9) are excluded.

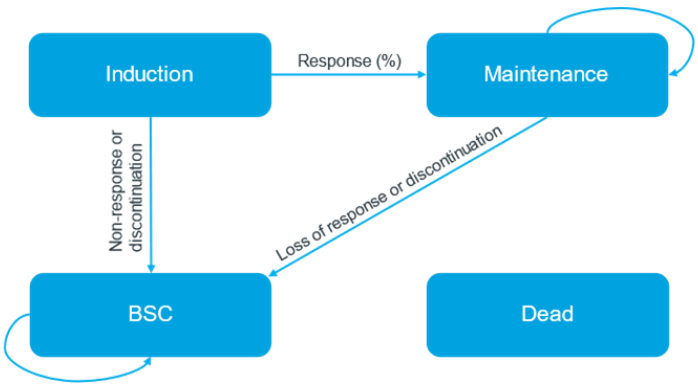
During the clarification stage, the EAG requested the company to explore dose tapering scenarios in line with the guidance in the Summary of Product Characteristics (SmPC). In their clarification response, the company stated that the cost for 4 mg and 2 mg is the same and thus does not affect

costs. Furthermore, the company explained that patients with a sustained response who are down titrated to 2 mg and do not maintain their response will resume the 4 mg dose to restore their previous response. As such the company did not consider a dose tapering scenario to be informative as remedial measures will be employed and thus in the long-term, the cost-effectiveness of baricitinib is unlikely to be affected. However, there is no direct evidence to suggest a "loss and regain" effect when remedial measures are used for patients who have lost response based on dose tapering. Nonetheless, the EAG considers the company's justification for not exploring dose tapering to be reasonable.

4.2.4 *Modelling approach and model structure*

A single *de novo* Markov model was developed in Microsoft Excel© to assess the cost-effectiveness of baricitinib 4 mg compared with a 'Watch and wait' approach for the treatment of adults with severe AA. The company structured the model using previous economic models for other dermatological disorders, such as psoriasis and atopic dermatitis due to a lack of AA models identified in the literature. The aim of the model developed by the company was to estimate the treatment pathway for patients beginning treatment for severe AA (first-line treatment). To capture all costs and benefits associated with treatment until death, the health states within the model include induction, maintenance, best supportive care (BSC) and death. Figure 7 presents the schematic of the Markov model.

Figure 7. Model structure (Figure 26 in company submission)



Abbreviations: BSC, best supportive care.

All patients enter the model via the induction state and either start treatment on baricitinib 4 mg or are regularly monitored ('Watch and wait'). The duration of the induction phase is 36 weeks, and patients transition through nine tunnel states, each lasting four weeks in duration. At any point

during the induction phase, patients can transition to the BSC health state due to all-cause treatment discontinuation (excluding discontinuation due to lack of efficacy).

Following the end of the 36-week induction phase, patients in the baricitinib 4 mg and 'Watch and wait' treatment groups are assessed on their response to treatment. Responders to treatment at Week 36 (defined as achieving $SALT_{50}$) transition to the Maintenance health state where they remain until loss of response or treatment discontinuation due to other causes (all cause discontinuation). Baricitinib patients that transition to the Maintenance health state at Week 36 continue to remain on a 4 mg dose as treated in the induction phase. Patients on 'Watch and wait' who enter the maintenance phase continue with regular monitoring. Patients transitioning to the Maintenance health state are stratified into $SALT_{50}$ and $SALT_{75}$ subgroups depending on relative hair regrowth to allow for differences in utility to be captured. In the model, after the 36-week treatment response assessment, patients remained either $SALT_{50}$ or $SALT_{75}$ unless they discontinued treatment for any reason and thus transition to the BSC health state. Please see Section 4.2.5 for further details on the definition of treatment response and treatment discontinuation applied in the model.

Non-responders were classified as those who fail to achieve $SALT_{50}$ at the end of the induction phase and transition to the BSC state alongside those who discontinued treatment during the induction phase. Patients in the BSC state remain there until the end of the model time horizon or death. At any point in the model time horizon, patients can transition to death from all health states and no patients can experience remission after the 36-week treatment response assessment (that is, transition from being a non-responder to a responder). Transition probabilities to death reflect the UK general population mortality rates (see Section 4.2.7 for further details).

The model was designed to capture responses to treatment over a lifetime horizon (until a patient reaches 100 years of age) and model cycle length was 4 weeks. No half-cycle correction was included in the model due to the short cycle length. The perspective of the analysis was based on the UK National Health Service (NHS) with an annual discount rate of 3.5% being applied for both costs and quality-adjusted life-years (QALYs) captured by the model as per the NICE reference case.

4.2.4.1 EAG critique

The EAG considers the company's model structure to be appropriate and allows important differences in costs and QALYs to be captured. Additionally, the model structure is similar to previous analyses of similar dermatological diseases, such as atopic dermatitis.[15, 38] However, the

EAG has key issues with the underlying assumptions included for each the health states, which are explored throughout the rest of Section 4. In particular, the EAG considers the company's definition of treatment response and the distinction between a patient achieving $SALT_{50}$ and $SALT_{75}$ to allow for addition utility gain may not be reasonable and this is further explored in Sections 4.2.5 and 4.2.8.

4.2.5 Treatment response

In the model, the primary treatment response measure was the achievement of $SALT_{50}$ at Week 36 based on pooled data from the BRAVE-AA1 and BRAVE-AA2 trials. The treatment response of $SALT_{50}$ is a relative measure of response and is defined by the company as at least a 50% improvement from baseline SALT score. In addition to the outcome of $SALT_{50}$, the company also included the outcome of $SALT_{75}$ (defined as at least a 75% improvement from baseline SALT score), as a way of capturing additional quality of life benefit associated with achieving an increased relative improvement in hair growth. The company also explored treatment response by severity as additional scenarios.

Treatment response data included in the company's base case is presented in Table 24.

Table 24. Pooled treatment response at Week 36 from BRAVE-AA1 and BRAVE-AA2

Intervention	$SALT_{50}$ (SE)	$SALT_{75}$ (SE)
Baseline SALT 50-100 patients (FAS population)		
Baricitinib 4 mg	[redacted]	[redacted]
'Watch and wait'	[redacted]	[redacted]
Baseline SALT 50-94 patients (severe population)		
Baricitinib 4 mg	[redacted]	[redacted]
'Watch and wait'	[redacted]	[redacted]
Baseline SALT 95-100 patients (very severe population)		
Baricitinib 4 mg	[redacted]	[redacted]
'Watch and wait'	[redacted]	[redacted]

Abbreviations: FAS, full analysis set; SALT, Severity of Alopecia Tool; SE, standard error.

At any point during the induction phase (prior to the Week 36 treatment response assessment point), patients can transition to the BSC health state due to all cause discontinuations excluding lack of efficacy. Table 25 presents the treatment discontinuation data used during the induction phase of the model. The 36-week data were adjusted to reflect the 4-week cycles included in the model.

Table 25. Pooled 36-week treatment discontinuation (excluding lack of efficacy) from BRAVE-AA1 and BRAVE-AA2

Population (Baseline SALT scores)	Baricitinib 4 mg	'Watch and wait'
SALT 50-100 patients (FAS population)	[redacted]	[redacted]
SALT 50-94 patients (severe population)	[redacted]	[redacted]
SALT 95-100 patients (very severe population)	[redacted]	[redacted]

Abbreviations: FAS, full analysis set; SALT, Severity of Alopecia Tool; SE, standard error.

4.2.5.1 Long-term treatment discontinuation

In the model, at Week 36, patients will either move into the maintenance health state if they have achieved a treatment response at Week 36 or move to BSC if they do not achieve a response. Patients in the maintenance health remain on treatment (baricitinib 4 mg or 'Watch and wait') and only transition to the BSC health state due to all-cause discontinuation. All-cause discontinuation was defined as discontinuation from treatment for all causes including lack of efficacy.

For the 'Watch and wait' arm of the model, the pooled Week 0-36 all-cause discontinuation rate from the placebo arms of the BRAVE-AA1 and BRAVE-AA2 trials was used, due to a lack of data beyond 36 weeks. The 36-week all-cause discontinuation data for 'Watch and wait' was then converted into an annual rate to be used for the model.

For baricitinib 4 mg, pooled Week 0-52 all cause discontinuation data from the BRAVE-AA1 and BRAVE-AA2 was used. Table 26 presented the long-term discontinuation data used in the model. The annual all-cause discontinuation data were adjusted to reflect the 4-week cycles included in the model.

Table 26. Annual pooled all-cause discontinuation from BRAVE-AA1 and BRAVE-AA2

Population (Baseline SALT scores)	Baricitinib 4 mg	'Watch and wait'
SALT 50-100 patients (FAS population)	[redacted]	[redacted]
SALT 50-94 patients (severe population)	[redacted]	[redacted]
SALT 95-100 patients (very severe population)	[redacted]	[redacted]

Abbreviations: FAS, full analysis set; SALT, Severity of Alopecia Tool; SE, standard error.

4.2.5.2 EAG critique

The EAG primary concern with the company's approach to treatment response included in the model was the definition of response employed at Week 36. In the BRAVE-AA1 and BRAVE-AA2 trials, the primary endpoint was the proportion of patients achieving SALT≤20 at Week 36. A response of SALT≤20 indicated scalp hair loss of less than 20% (or ≥80% scalp coverage with hair). The EAG considers that SALT≤20 represents an absolute measure of response, which its clinical experts considered was a more clinically meaningful outcome for patients as, definitively, they will have at least 80% hair regrowth on the scalp and thus may stop wearing wigs or shaving their head. The company's base case approach of using $SALT_{50}$ is a relative improvement from baseline in hair regrowth on the scalp and thus may still be patchy and require the use of wigs or hair removal.

During the clarification stage, the EAG requested the company to provide a scenario exploring the outcome of SALT≤20 in the model. Additionally, based on the advice from the EAG's clinical experts, a scenario exploring SALT≤10 (defined as scalp hair loss of less than 10% or ≥90% scalp coverage with hair) was also requested. The company provided pooled Week 36 treatment response data using SALT≤20 and SALT≤10 (Table 27) and ran these data in a scenario (presented in Section 5.1.2.2).

Table 27. Pooled treatment response at Week 36 from BRAVE-AA1 and BRAVE-AA2

Intervention	SALT≤20 (95% CI)	SALT≤10 (95% CI)
Baseline SALT 50-100 patients (FAS population)		
Baricitinib 4 mg	[redacted]	[redacted]
'Watch and wait'	[redacted]	[redacted]
Baseline SALT 50-94 patients (severe population)		
Baricitinib 4 mg	[redacted]	[redacted]
'Watch and wait'	[redacted]	[redacted]
Baseline SALT 95-100 patients (very severe population)		
Baricitinib 4 mg	[redacted]	[redacted]
'Watch and wait'	[redacted]	[redacted]

Abbreviations: CI, confidence interval; FAS, full analysis set; SALT, Severity of Alopecia Tool.

The EAG considers that SALT≤20 is the most appropriate definition of response at Week 36 for use in the model as it is the primary endpoint in the key BRAVE trials and based on the EAG's clinical experts, is a more clinically meaningful outcome for patients. As such, the EAG has included SALT≤20 at Week 36 in its preferred assumptions, presented in Section 6.4.

The EAG notes that in a randomised double-blind trial, where placebo is the comparator, it can be argued that any observed placebo response is due to: increased medical attention as a result of being in an RCT, an unconscious expectation by the patient and the investigator that the patient will improve, as well as a patient's profound desire to get better. This is particularly the case when considering outcomes that have a subjective component like an assessment of SALT score. In clinical practice, where patients will know what treatment they are receiving or, as in this particular case, know when they have been discharged from clinical care, the EAG considers it unlikely that an observed placebo response from a clinical trial would occur.

However, it could also be argued that the factors influencing a perceived placebo response are also present in the active treatment group, in this case baricitinib. To negate the potential additional benefit present in the outcomes for baricitinib, the EAG considers it reasonable to use the placebo group from the trial as a surrogate for 'discharged from care'. With the rationale being that any "placebo effect" in both arms will "cancel out" and the incremental results from the cost-effectiveness analysis (and so the resulting ICER) will be based solely on the "true" treatment effect of baricitinib. With regards to treatment discontinuation for the comparator arm in the model, the EAG considers that this should be viewed as the placebo effect waning, as patients are not on active treatment.

The EAG considered that for long-term all-cause discontinuation, using Week 0-52 data for baricitinib 4 mg for the maintenance phase may not be representative of discontinuation of patients with a sustained response and during the clarification stage, requested the company to explore all cause discontinuation based on data for Week 36-52. The company supplied the requested data (adjusted to an annual rate), which estimated all-cause discontinuation for baricitinib 4 mg to be [redacted], based on Week 36-52 data, which is lower than the company's base case estimate of [redacted]. Results of the scenario are presented in Section 5.1.2.2 and this has been included in the EAG preferred assumptions.

4.2.6 Adverse events

The company did not include the impact of adverse events (AEs) in the model as they considered observed AEs from BRAVE-AA1 and BRAVE-AA2 to be mild and would not have a significant impact on HRQoL or costs.

4.2.6.1 EAG critique

The company state that their approach of not including AEs in the economic model is aligned with TA681 and TA534, but the EAG considers this is not accurate.[15, 38] In both TA681 and TA534, the impact of AEs was included in the economic models.[15, 38] As such, during the clarification stage, the EAG requested the company to include a scenario which considers the impact of AEs in terms of costs. The EAG focussed only on costs as it considered that the impact of AEs would be captured in the utility estimates derived from BRAVE-AA1 and BRAVE-AA2, which is the preferred source of utility data (see Section 4.2.8). The company provided this scenario using pooled AE rates from BRAVE-AA1 and BRAVE-AA2 (presented in Table 28). However, the company provided no justification for the inclusion of specific AEs nor the definition used (such as treatment-emergent or serious AEs). Additionally, no sources were provided for the costs used for the scenario.

The impact on the incremental cost-effectiveness ratio (ICER) from the inclusion of costs for AEs was minimal (see Section 5.1.2.2), but the EAG considers it is good practice to include costs associated with AEs in the model, especially when treatment is long-term if response is achieved. As the EAG was unable to verify the inputs used in the company's AE scenario and was unable to produce an alternative scenario due to a paucity of time, costs of AEs have not been included in the EAG base case. Nonetheless, the EAG requests that during technical engagement, the company provides a more thorough description and justification of their approach to the inclusion of AEs and assumed unit costs to treat each AE and update the scenario if necessary.

Table 28. AEs and costs included in the company scenario analysis

Adverse event	Unit cost*	Baricitinib 4 mg		'Watch and wait'	
		Induction (36-weeks)	Maintenance (annual)	Induction (36-weeks)	Maintenance (annual)
Upper respiratory tract infection	£39.00	[redacted]	[redacted]	[redacted]	[redacted]
Nasopharyngitis	£39.00	[redacted]	[redacted]	[redacted]	[redacted]
Headache	£206.34	[redacted]	[redacted]	[redacted]	[redacted]
Acne	£171.53	[redacted]	[redacted]	[redacted]	[redacted]
Total cost	-	**£19.48**	**£27.79**	**£15.86**	**£22.70**

Abbreviations: AE, adverse event

*Source or justification of cost not provided by the company

4.2.7 Mortality

Treatment with baricitinib is assumed not to impact mortality. As such, the company included background mortality such that the transition probability to the death state per model cycle was equal for both arms. All-cause mortality was based on Office for National Statistics (ONS) UK lifetables.[39]

4.2.8 Health-related quality of life

In the base case analysis, QALYs accrued by the patient cohort in each model cycle are dependent on the utility attributable to each model health state and an age-related reduction in quality of life.

The BRAVE-AA1 and BRAVE-AA2 trials collected EQ-5D data, as well as data from the SKINDEX-16 AA, SF-36 and HADS questionnaires, up to Week 36 directly from patients but the company stated that the values obtained from the trials were insensitive to changes in the severity of AA and lacked content validity as baseline values were almost the same as UK age- and sex-adjusted general population values. As such, the utility values informing the economic model were derived from a company sponsored Adelphi DSP study.[13, 14] Utilities based on EQ-5D and HADS data from the BRAVE-AA1 and BRAVE-AA2 trials were explored in a scenario analysis.

The Adelphi DSP study collected EQ-5D-5L data from patients with AA in Europe (including the UK). The study was initiated in October 2021. Details of the Adelphi DSP study were limited in the CS and as such, the EAG requested further information during the clarification stage. In their response to clarification, the company provided the questionnaire used for patient reported disease burden, a data file of the responses from the study as well as the overall objective (used as proxy for the study protocol) for the utility aspect of the Adelphi DSP study (the other aspect was to obtain resource use from treating physicians of patients with AA).

The objective of the utility aspect of the Adelphi DSP study was to characterise the patient reported disease burden of AA based on physician-rated current severity by:

- Describing SKINDEX-16 AA, HADS, EQ-5D-5L and Work Productivity and Activity Impairment (WPAI);
- Reporting on concordance in patient and physician ratings of severity; and
- Assessing predictors of patient reported burden measures (Skindex-16 AA, HADS, EQ-5D and WPAI).

However, there were sections in the questionnaire (after the patient filled out their responses to the utility instruments) that were focused on the effects of AA on work and daily life as well as the patients' feelings about their condition.

Overall, there were [redacted] responses to the EQ-5D questionnaire. Responses were stratified by physician-reported current severity of a patient's AA episode and only responses for severe and very severe were considered for the analysis ([redacted] responses). The crosswalk algorithm by Hernandez et al.[40] was used to convert the EQ-5D-5L values to the EQ-5D-3L. The health state utility values (HSUVs) from the final analysis of the Adelphi DSP study (provided during the clarification stage) informing the model are presented in Table 29. It should be noted that the same utility values were used for the severity subgroup analysis.

Table 29. Utility values informing the model from the Adelphi DSP study

Health state	Utility value (SE)	Comments
Induction (up to Week 36)	[redacted]	Baseline score for the severe and very severe subgroup.
Maintenance - $SALT_{50}$	[redacted]	Utility value for the moderate severity subgroup. In the model, the company implemented the utility value as a change from baseline utility gain, calculated as the difference between baseline scores for moderate and severe/very severe subgroup ([redacted])
Maintenance - $SALT_{75}$	[redacted]	Utility value for the mild severity subgroup. In the model, the company implemented the utility value as a change from baseline utility gain, calculated as the difference between baseline scores for mild and severe/very severe subgroup ([redacted])
BSC	[redacted]	Baseline score for the severe and very severe subgroup.

Abbreviations: BSC, best supportive care; DSP, disease-specific programme; SALT, Severity of Alopecia Tool; SE, standard error.

Utilities in the model were adjusted for age, as per the NICE methods guide.[41] The multiplicative approach was used as recommended by the Decision Support Unit (DSU) Technical Support Document (TSD) 12.[42] General population utility values adjusted for age and sex were obtained from the HSE 2014 dataset, as recommended by the DSU.[40]

4.2.8.1 EAG critique

The EAG considers the company's justification for not using pooled EQ-5D data from the BRAVE-AA1 and BRAVE-AA2 trials (lack of sensitivity and content validity) is a criticism of the EQ-5D tool and not the methods to obtain the data used in the trial. As such, the EAG considers the company's criticism

of the trial EQ-5D data extends to the EQ-5D data obtained from the company sponsored Adelphi DSP study. The EAG asked the company to explain why EQ-5D data from the Adelphi DSP study were more appropriate than the EQ-5D data from the BRAVE-AA1 and BRAVE-AA2 trials. The company stated that there was a substantial ceiling effect present in the trial data. The company estimated that around [redacted] of participants in the BRAVE-AA1 and BRAVE-AA2 trials reported a score of perfect health at baseline (score of 11111) and as such an improvement in HRQoL would not be obtained at Week 36 for these patients. The company explained that in the Adelphi study, the ceiling effect was observed, but not to the same extent. However, the company did not provide the overall proportion of patients reporting a score of perfect health from the Adelphi DSP study, rather they presented data by each domain of the EQ-5D (Table 26 of the company clarification response).

Additionally, the company only reported perfect health score data for the severe and very severe subgroup but did not supply any information for the mild and moderate subgroup which inform the utility gain in the model for patients who achieve $SALT_{50}$ and $SALT_{75}$ outcomes used in the company base case analysis. The EAG considers that mild and moderate severity patients in the Adelphi DSP study are more likely to report scores of perfect health as their disease is, by definition, less severe, and unlikely to have a more significant impact on HRQoL compared with severe and very severe patients. As such, the utility gain in the economic model may be biased if there was a high proportion of mild and moderate severity patients reporting a score of perfect health in the Adelphi DSP study.

Additionally, based on the physician survey used for the Adelphi DSP study,[13, 14] definition of severity was not based on SALT score, but categories of severity (mild, moderate, severe, and very severe). The EAG considers the severity categories from the Adelphi DSP study represent absolute scalp hair coverage and the difference between the severe/very severe and moderate/mild severity reflects an assumed change in absolute hair regrowth rather than the relative change from baseline assumed in the economic model.

Conversely, the pooled EQ-5D data from the BRAVE-AA1 and BRAVE-AA2 trials is based only on patients with severe and very severe disease at baseline (defined as SALT 50-100) and the change from baseline at Week 36 estimated by the company is an observed change from baseline score for patients achieving $SALT_{50}$ and $SALT_{75}$ outcomes used in the company base case analysis, rather than baseline scores for patients with mild and moderate severity AA.

In BRAVE-AA1 and BRAVE-AA2, EQ-5D-5L was measured at Weeks 0, 12, 24 and 36. In the long-term extension phase of BRAVE-AA1 and BRAVE-AA2, EQ-5D-5L was measured at Weeks 52, 64 and 76. Patients were followed up for 200 weeks. Overall, there were 341 responses in the placebo arm and 514 responses in the baricitinib 4 mg arm but pooled data were used to inform the HSUVs. The following linear model was used to analyse the EQ-5D data:

$$cEQ5D = a0 + a1\ EQ5Dbl + a2\ SALTcat + a3\ AGE$$

Abbreviations: cEQ5D, Change in EQ5D; EQ5Dbl, Baseline EQ5D; SALTcat, SALT improvement categories at Week 36 (<50%, ≥50% to <75%, ≥75%); AGE, Age in years.

As with the utility values from the Adelphi DSP study, the company used the crosswalk algorithm by Hernandez et al.[40] to convert EQ-5D-5L values to the EQ-5D-3L. Table 30 presents the pooled EQ-5D health state data from BRAVE-AA1 and BRAVE-AA2 trials. As mentioned previously, the company ran a scenario using HSUVs from the BRAVE-AA1 and BRAVE-AA2 trials and results of the scenario are presented in Section 5.1.2.2. It should be noted that age-matched general population utility value is 0.91. The utility values for patients that achieve $SALT_{50}$ or $SALT_{75}$ using data from BRAVE-AA1 and BRAVE-AA2 are only just below the general population value. Based on feedback from the EAG's clinical experts, this may not be unreasonable as for the majority of patients with severe AA, there is not a significant impact on HRQoL.

Table 30. Health state utility data - pooled EQ-5D data from BRAVE-AA1 and BRAVE-AA2

Health state	Utility value (SE)	Comment
Induction (up to Week 36)	[redacted]	Baseline score for SALT 50-100 FAS population
Maintenance - $SALT_{50}$	[redacted]	Change from baseline for patients achieving $SALT_{50}$
Maintenance - $SALT_{75}$	[redacted]	Change from baseline for patients achieving $SALT_{75}$
BSC	[redacted]	Baseline score for SALT 50-100 FAS population

Abbreviations: BSC, best supportive care; FAS, full analysis set; SALT, Severity of Alopecia Tool; SE, standard error.

The source of utility values in the model is a primary driver of cost-effectiveness. As recommended in the NICE methods guide,[41] the reference case should report the measurement of changes in health-related quality of life directly from patients. As such, the EAG considers the pooled EQ-5D data from the BRAVE-AA1 and BRAVE-AA2 trials represents a more robust source of utility data that matches the NICE reference case and should be used in the cost-effectiveness analysis for the base case.

As mentioned in Section 4.2.5.2, the EAG considers that the primary trial outcome of SALT≤20 is a more appropriate measure of treatment effectiveness. As such, the EAG requested, and the company provided, change from baseline at Week 36 for patients achieving SALT≤20 based on pooled EQ-5D data from the BRAVE-AA1 and BRAVE-AA2 trials ([redacted]). Upon request, the company also supplied the change from baseline at Week 36 for patients achieving SALT≤10 based on pooled EQ-5D data from the BRAVE-AA1 and BRAVE-AA2 trials ([redacted]). The EAG ran a scenario using the SALT≤20 change from baseline utility gain in combination with the outcome of SALT≤20 for treatment effectiveness in the model, as well as the same scenario using data for SALT≤10 and results are presented in Section 6.3. The EAG considers the SALT≤20 combined scenario is a more appropriate approach to the cost-effectiveness analysis and has included it in the EAG preferred assumptions, presented in Section 6.4.

The EAG notes that the experience of severe AA can vary between patients. The EAG's clinical experts advised that for most patients, HRQoL may only be mildly affected and thus may not be that different to the general population but equally HRQoL is severely affected for a few patients (primarily driven by adverse mental health). Additionally, the EAG's clinical experts advised that overtime, patients may come to terms with their hair loss, while a few may remain distressed about their condition. Thus, the EAG acknowledges that there is a small, but heterogenous, patient population that is more adversely affected in terms of HRQoL but that the demographics of this population are difficult to identify clinically and consistently, and it is beyond the scope of assessment to identify that group. Nonetheless, the EAG ran two scenarios around the EAG base case to identify the QALY gain needed for the ICER to reach the £20,000 and £30,000 cost-effectiveness threshold and these are presented in Section 6.4.

4.2.9 Resource use and costs

The costs included in the economic model consist of drug acquisition costs, monitoring resource use and costs, costs associated with BSC, and costs associated with the management of the psychological burden of AA. The details of each are given in the following subsections. Unit costs used in the model were based on 2020/21 price years.

Many of the company's resource use assumptions were informed by the company sponsored Adelphi DSP study.[13, 14] One objective of the Adelphi DSP study was to describe treatment patterns associated with AA based on physician rated severity and this feedback was used to inform the

resource use in the model (described in the following subsections). Only data obtained from physicians treating severe and very severe patients in the UK ([redacted]) were used for the model.

4.2.9.1 Drug acquisition costs

Baricitinib is given as a fixed-dose 4 mg tablet taken once daily and is also available as a 2 mg dose. The list price of a 28-tablet pack of 2 mg or 4 mg is £805.56. There is currently a patient access scheme (PAS) in place for baricitinib such that the fixed price pack is [redacted]. The daily cost of baricitinib 4 mg is [redacted].

No drug acquisition costs were applied to the 'Watch and wait' arm of the model. Instead, patients who are allocated to 'Watch and wait' are actively monitored and these costs are described in Section 4.2.9.2.

4.2.9.2 Monitoring resource use and costs

During the induction and maintenance phases of the economic model, patients on baricitinib and 'Watch and wait' are actively monitored. The company used feedback from clinical experts to inform the assumptions around monitoring during the induction and maintenance phases of the economic model, presented in Table 31. Unit costs for monitoring resource are presented in Table 48 of the company's clarification response and were sourced from NHS reference costs 2020/21 and PSSRU.[34, 35]

The total cost of monitoring in the induction phase (36 weeks) baricitinib 4 mg and 'Watch and wait' was £1,022.75 and £1,011.86, respectively. The annual cost of monitoring in the maintenance health state for baricitinib and 'Watch and wait' was £371.71 and £357.19, respectively. The main difference in costs in the induction and maintenance health state between baricitinib 4 mg and 'Watch and wait' was the inclusion of blood monitoring for baricitinib patients.

Table 31. Monitoring resource use and costs

Resource use	Proportion of patients	Baricitinib 4 mg		'Watch and wait'	
		Induction (36 weeks)	Maintenance (annual)	Induction (36 weeks)	Maintenance (annual)
Dermatologist outpatient consultation	100%	4.0	2.0	4.0	2.0
Dermatologist nurse visit	100%	1.0	0.5	1.0	0.5
Full blood count	100%	3.0*	4.0	0.0	0.0
Wig use (modacrylic wig)	80%	2.0	0.0	2.0	0.0

Orthotics	80%	1.0	0.0	1.0	0.0
*Updated in the company's clarification response.					

4.2.9.3 BSC health state costs

In the BSC health state, costs included drug acquisition and monitoring costs as well as disease management costs. Costs associated with the management of the psychological burden of AA were also included in the BSC health state and are described further in Section 4.2.9.4.

Treatments and the proportion of patients on each treatment included in the BSC health state are presented in Table 32 and were informed by the company's Adelphi DSP study. Treatment dosage was based on each treatment's SmPC. Unit costs were obtained from the NHS drug tariff[36] and are presented in Table 56 of the CS. Confidential medicines unit (CMU) and Drugs and pharmaceutical electronic market information tool (eMIT) prices are available for medicines included in BSC, as such the EAG has produced a confidential appendix to the EAG report. Please refer to Appendix 8.4. for the source of the confidential prices used in the confidential appendix.

The company also included costs of monitoring patients while on treatment in the BSC (Table 33) and assumptions were based on clinical expert opinion obtained by the company. Unit costs for monitoring in the BSC health state were based on NHS reference costs 2020/21.[35] The total annual cost of drug acquisition and monitoring in the BSC health state was estimated to be £3,683.10.

Table 32. Drug acquisition costs in the BSC health state

Treatment	Dose and frequency	Number of doses per year	Proportion of patients*	Annual cost
Ciclosporin	4 mg/kg QD	108,114 (4mg * 74kg * 365.25 days)	13.72%	£355.70
Methotrexate	20 mg per week	1,040 (20mg * 52 weeks)	12.86%	£3.25
Azathioprine	2 mg/kg body weight QD, for 1 year	54,057 (2mg * 74kg * 365.25 days)	2.57%	£3.10
Intralesional steroids (triamcinolone acetonide)	5 mg repeated every other week	130 (5mg * 26 weeks)	9.43%	£0.46
DPCP (contact immunotherapy) treatment	Weekly treatment for 9 months	36 (4 times per month for 9 months)	21.63%	£890.79
Prednisolone*	0.4 mg/kg QD	10,811 (0.4mg * 74kg * 365.25 days)	17.15%	£10.46

TCS: Mometasone scalp lotion*	2ml QD	730.5 (2 ml * 365.25 days)	24.77%	£13.15
Minoxidil 5% foam (topical)*	1 g BID (men) or 1 g QD (women) - discontinue if no improvement after 16 weeks (men) or 24 weeks (women)	202 [39.3%§ *(2g * 7 days * 16 weeks) + 60.7%§*(1g * 7 days * 24 weeks)]	5.72%	£3.02
Minoxidil tablets	20 mg QD	7,305 (20mg * 365.25 days)	0.00%	£0.00
Mycophenolate Mofetil	1 g BID, for 1 year	730,500 (2,000mg * 365.25 days)	2.86%	£4.59
Anthralin 0.1% cream	1.5 g QD	242 (1.5g * 7 days * 23 weeks)	5.72%	£1.04
Patients not currently on treatment	4 mg/kg QD	108,114 (4mg * 74kg * 365.25 days)	12.00%	£0.00
Total cost		-	-	**£1,285,56**

Abbreviations: BID, twice per day; DPCP, diphenylcyclopropenone; QD, once per day; TCS, topical corticosteroids

§Sex distribution based on SALT 50-100 pooled FAS population from BRAVE-AA1 and BRAVE-AA2.

*Updated as part of the company's clarification response

Table 33. Drug monitoring in the BSC health state

Treatment	Description of monitoring[35]	Unit cost[35]	Frequency of visits per year	Proportion of patients*	Annual cost
Ciclosporin	Weighted average of WF01A-D and WF02A-D - Dermatology	£171.53	9	13.72%	£211.81
Methotrexate	Weighted average of WF01A-D and WF02A-D - Dermatology	£171.53	9	12.86%	£198.57
Azathioprine	Weighted average of WF01A-D and WF02A-D - Dermatology	£171.53	9	2.57%	£39.71
Intralesional steroids (triamcinolone acetonide)	JC43C – OPROC – Minor Skin Procedures, 19 years and over – Dermatology	£250.70	18	9.43%	£425.65
DPCP (contact immunotherapy) treatment	JC43C – OPROC – Minor Skin Procedures, 19 years and over – Dermatology	£250.70	36	21.63%	£1,952.11
Prednisolone	Weighted average of WF01A-D and WF02A-D - Dermatology	£171.53	13	17.15%	£382.43
TCS: Mometasone ointment	Weighted average of WF01A-D and WF02A-D - Dermatology	£171.53	4	24.77%	£169.97

Minoxidil 5% foam (topical)	Weighted average of WF01A-D and WF02A-D - Dermatology	£171.53	4	5.72%	£39.22
Minoxidil tablets	Weighted average of WF01A-D and WF02A-D - Dermatology	£171.53	4	0.00%	£0.00
Mycophenolate Mofetil	Weighted average of WF01A-D and WF02A-D - Dermatology	£171.53	9	2.86%	£44.13
Anthralin 0.1% cream	Weighted average of WF01A-D and WF02A-D - Dermatology	£171.53	4	5.72%	£39.22
Patients not currently on treatment	-	-	-	12.00%	-
Total cost	-	-	-	-	**£3,502.82**

Abbreviations: DPCP, diphenylcyclopropenone; TCS, topical corticosteroids

*Based on company clinical expert opinion and updated as part of the company's clarification response.

In addition to drug acquisition and monitoring costs, the company included additional disease monitoring costs, based on feedback obtained from clinical experts (i.e. tests, wig use and orthotics) and the company's Adelphi DSP study (dermatologist visits). Table 34 presents the disease management costs applied in the BSC health state. Unit costs were sourced from NHS reference costs 2020/21 and PSSRU.[34, 35] The total annual cost of disease management in the BSC health state was estimated to be £468.47.

Table 34. Disease monitoring costs in the BSC health state

Resource use	Description of monitoring[34, 35]	Unit cost[34, 35]	Frequency per year	Proportion of patients*	Annual cost
Dermatologist outpatient consultation	Weighted average of WF01A-D and WF02A-D - Dermatology	£171.53	2.00	13%	£41.17
Dermatologist nurse visit	PSSRU, 15 minutes of hospital nurse Band 6 patient related time	£28.25	0.50	13%	£1.70
Thyroid function	DAPS05 - Haematology	£3.63	4.00	100%	£14.52
Vitamin D	DAPS05 - Haematology	£3.63	4.00	100%	£14.52
Ferritin	DAPS05 - Haematology	£3.63	4.00	100%	£14.52
Full blood count	DAPS05 - Haematology	£3.63	4.00	100%	£14.52
Liver function	DAPS04 - Clinical biochemistry	£1.85	4.00	100%	£7.40
Renal function	DAPS04 - Clinical biochemistry	£1.85	4.00	100%	£7.40

Tuberculosis	DAPS07 - Microbiology	£10.18	4.00	100%	£40.72
Lipids	DAPS05 - Haematology	£3.63	4.00	100%	£14.52
Wig use (modacrylic wig)	Wigs and fabric supports on the NHS	£75.70	2.00	80%	£121.12
Orthotics	Service Code 658 - Total Outpatient Attendances	£220.46	1.00	80%	£176.37
Total	-	-	-	-	**£468.47**

Abbreviations: BSC, best supportive care.

4.2.9.4 Costs of Psychological management of AA

The company assumed that patients with severe AA will incur costs to manage the psychological burden of AA. The costs of psychological management of severe AA were split between pharmacological (Table 36) and non-pharmacological costs (Table 35) and were assumed to occur in the induction phase and the BSC health state of the economic model. Costs were sourced from PSSRU and CG90.[34, 43] It should be noted that CG90 was replaced by NG222 after the company produced the CS, but costs were not dissimilar between the two sources.[43, 44]

Table 35. Non-pharmacological support costs included in the model

Resource use & description	Unit cost (PSSRU)[34]	Proportion of patients*	Resource use in induction*	Resource use in BSC*
Psychiatrist visit - NICE NG222** - band 7 HI therapist (with MBCT qualification).	£112.00	5.00%	3.00	4.00
Psychologist visit - NICE NG222** - One-hour direct contact (band 5 PWP).	£50.00	10.00%	3.00	4.00
Self-help with support - 1 GP session.	£39.23	12.38%	0.75	1.00
Group exercise & one GP referral visit - 30 sessions x 1 hour each; 1 therapist (band 5 PWP) and 8 participants per group = 30 therapist.	£186 + £39.23	0.75%	0.75	1.00
Interpersonal psychotherapy & one GP referral visit - 8 sessions x 1 hour each = 8 therapist hours per service user (band 7 HI therapist).	£873 + £39.23	0.75%	0.75	1.00
Counselling & one GP referral visit - 12 sessions x 1 hour each = 12 therapist hours per service user (band 7 HI therapist).	£873 + £39.23	1.13%	0.75	1.00
Total costs	-	-	**£49.54**	**£66.05**

Abbreviations: BSC, best supportive care; HI, high intensity; MBCT, mindfulness-based cognitive therapy; PWP, psychological well-being practitioner.

*Based on company clinical expert opinion

**CG90 was replaced by NG222 after the company submission and as such, the relevant tables in NG222 are Table 85-86 of Evidence Review B.[43, 44]

Table 36. Cost of pharmacological treatment for the psychological treatment of severe AA

Treatment	Proportion of patients*	Total cost (incl. 4 GP visits)**	Induction		BSC	
			Resource use	Cost	Resource use	Cost
Sertraline	16.50%	£161.42	0.75	£19.98	1.00	£26.63
Escitalopram	16.50%	£160.26	0.75	£19.83	1.00	£26.44
Duloxetine	5.00%	£164.59	0.75	£6.16	1.00	£8.23
Total costs	-	-	-	**£45.98**	-	**£61.31**

Abbreviations: AA, alopecia areata; BSC, best supportive care; GP, general practitioner; HI, high intensity; MBCT, mindfulness-based cognitive therapy; PWP, psychological well-being practitioner.

*Based on company clinical expert opinion

**Obtained from Table 86, Evidence Review B of CG90. CG90 was replaced by NG222 after the company submission and as such, the relevant table in NG222 is Table 83 of Evidence Review B.[43, 44] However, the costs between the original guidance and the update are not dissimilar.

† Cost of GP visit = £39.23[34]

4.2.9.5 EAG critique

The EAG identified several issues with the company's assumptions around resource use and costs that were deemed by the EAG's clinical experts not to align with UK clinical practice. Primarily, the EAG considers the costs in the model for 'Watch and wait' and the BSC health state to be overestimated based on feedback from the EAG's clinical experts. Overestimation of costs in the BSC health state is a key issue as patients in both arms of the model spend a substantial amount of time in the BSC health state accruing costs with no associated benefit (utility for this health state is set to baseline, see Section 4.2.8).

As mentioned in Section 4.2.3, the EAG's clinical experts considered that 'Watch and wait' for patients with severe AA does not happen in the NHS and patients would not be regularly monitored if they were not receiving treatment. Additionally, the EAG's clinical experts considered that a range of treatments may be given to patients but that these are not very effective. In the company's own Adelphi DSP study, it was estimated that the majority of severe/very severe patients were treatment experienced ([redacted]).[13, 14] Thus, it is likely that if response to treatment is not achieved, patients will not engage with further treatment or will not be followed up (effectively patients are discharged from care). Furthermore, a significant proportion of patients may not take up treatment and instead opt for using wigs to manage their hair loss. The EAG considers that lack of engagement with

treatment or being discharged from care has implications for the costs assumed in the BSC health state for both arms of the model, as patients transition to this health state upon loss of treatment response or treatment discontinuation for any other reason. Thus, the EAG considers that monitoring costs for patients on 'Watch and wait' (what the EAG refers to as 'discharged from care') and disease management costs in the BSC health state should be excluded.

During the clarification stage, the EAG requested, and the company supplied, a scenario where monitoring costs for 'Watch and wait' in both the induction phase and Maintenance health state were removed from the model (see Section 5.1.2.2). Additionally, the EAG requested a scenario where disease monitoring costs in the BSC health state are removed from both arms of the model. However, the company only provided a scenario where disease monitoring costs in the BSC were excluded for the baricitinib 4 mg arm only. As such, the EAG ran a scenario where disease monitoring costs are excluded from the BSC health state for both arms of the model and results are presented in Section 6.3.

A key assumption made by the company which affects both arms of the model in the induction phase and the BSC health state, is the inclusion of costs associated with psychological support for severe AA patients (non-pharmacological interventions). The EAG's clinical experts advised that the company's assumptions of psychological care support were optimistic, and that provision of support is extremely limited given the current pressures faced by the NHS. The EAG's clinical experts did consider the company's assumptions around pharmacological treatment for the management of the psychological burden of severe AA to be reasonable. During the clarification stage, the EAG requested the company to provide a scenario where psychological support costs were removed from the model. The company supplied the requested scenario, but upon further investigation, the EAG found that the company's scenario excluded pharmacological treatment costs in addition to the costs of psychological support. As such, the EAG ran a scenario excluding only the non-pharmacological psychological support costs and the results are presented in Section 6.3.

For patients in the induction phase of the model and in the BSC health state, provision of wigs and orthotics has been assumed to occur for 80% of patients, with 2 wigs and 1 orthotic supplied for both induction and annually in the BSC health state. The EAG's clinical experts advised that wigs and orthotics are predominantly used by female patients, of which at baseline in the model, 60.7% are female. Furthermore, the induction phase of the model is only 36 weeks, yet the assumptions made for wigs and orthotics resource use are the same as the BSC health state, which represents annual

usage. The EAG requested scenarios exploring wigs and orthotics for only female patients and only one wig for the induction phase. The company supplied the requested scenarios, and this demonstrated that changes to the assumptions around wigs and orthotics had minimal impact on the ICER (see Section 5.1.2.2).

Overall, the EAG considers the following assumptions to be a more accurate reflection of costs incurred by severe AA patients in the model and has included these in its preferred assumptions, presented in Section 6.4:

- Exclusion of monitoring costs for 'Watch and wait' in the induction phase of the model (comparator is assumed to be 'discharged from care').
- Exclusion of disease management costs in the BSC health state for both arms of the model.
- Exclusions of psychological support costs (non-pharmacological intervention) in the induction phase and BSC health state for both arms of the model.
- Only one wig assumed in the induction phase for both arms of the model. The EAG decided not to include the assumption of wigs and orthotics use only for female patients only as this may be a strong assumption and the impact on the ICER was minimal.

5 Cost effectiveness results

5.1.1 *Company's cost effectiveness results*

Table 37 presents the cost-effectiveness results of the company's updated (i.e., post clarification) base case deterministic and probabilistic analyses. The company performed probabilistic sensitivity analysis (PSA) to assess the joint parameter uncertainty around base case results. Incremental results from the company's PSA, arising from 1,000 simulations.

In the base case probabilistic analysis, an incremental quality-adjusted life-year (QALY) gain of [redacted] over 'Watch and wait' along with additional costs of [redacted] for the baricitinib 4 mg, generates an incremental cost-effectiveness ratio (ICER) of £17,942 per QALY. The net monetary benefit (NMB) using the £30,000 threshold is [redacted] and the net health benefit (NHB) is [redacted]. A positive NHB implies that overall population health would be increased as a result of the new intervention

A proposed confidential patient access scheme (PAS) discount for baricitinib is applied in the company's base case and is therefore reflected in the results presented in this report. Confidential medicines unit (CMU) and Drugs and pharmaceutical electronic market information tool (eMIT) prices are available for medicines included in best supportive care (BSC) and as such the Evidence Assessment Group (EAG) has produced a confidential appendix to the EAG report. Analyses included in the confidential appendix include the company base case results, and sensitivity and scenario analyses.

Table 37. Company's base case results post clarification

Interventions	Total Costs (£)	Total LY	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER (£/QALY)
Deterministic results							
'Watch and wait'	[redacted]	22.60	[redacted]	-	-	-	-
Baricitinib	[redacted]	22.60	[redacted]	[redacted]	0.00	[redacted]	18,072
Probabilistic results							
'Watch and wait'	[redacted]	-	[redacted]	-	-	-	-
Baricitinib	[redacted]	-	[redacted]	[redacted]	-	[redacted]	17,942

Abbreviations: ICER, incremental cost effectiveness ratio, LYG, life year gained; QALY, quality adjusted life year.

A PSA scatterplot is presented in Figure 8 and a cost-effectiveness acceptability curve (CEAC) is presented in Figure 9. Based on these analyses, the probability that baricitinib is cost effective versus

'Watch and wait' is [redacted] at a willingness to pay (WTP) threshold of £20,000 and [redacted] at a WTP threshold of £30,000.

The EAG considers the parameters and respective distributions chosen for PSA to be generally sound. The EAG also considers the probabilistic results to be comparable to the deterministic results.

Figure 8. Cost-effectiveness plane - PSA scatterplot: baricitinib 4 mg vs 'Watch and wait' (company's clarification response appendix, Figure 18)

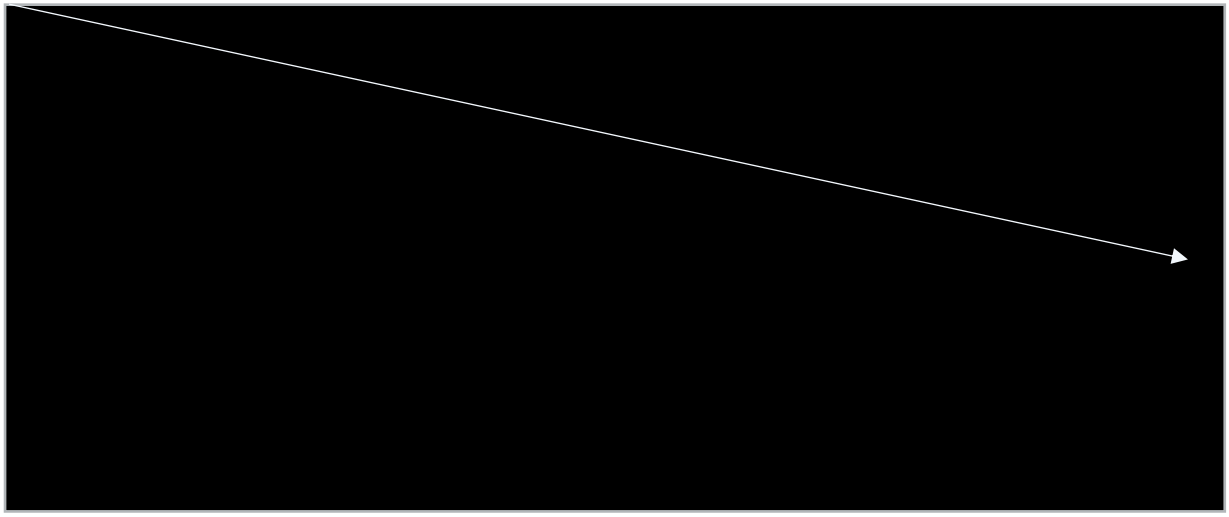


Figure 9. Cost-effectiveness acceptability curve: baricitinib 4 mg vs 'Watch and wait' (company's clarification response appendix, Figure 19)

5.1.2 Company's sensitivity analyses

5.1.2.1 One way sensitivity analysis

The company conducted one-way sensitivity analyses (OWSAs) to assess the impact, on the ICER, of varying specific parameters in isolation and to identify the main model drivers. The results are illustrated using the tornado diagram in Figure 10. The ICER was most sensitive to the frequency and monitoring resource use for diphenylcyclopropenone (DPCP) treatment included in the BSC health state, followed by the Severity of Alopecia Tool 50 ($SALT_{50}$) 36-week response rate for baricitinib 4 mg.

Figure 10. Tornado plot (company's clarification response appendix, Figure 20)

Abbreviations: BSC, best supportive care; DPCP, diphenylcyclopropenone; HSUV, health state utility value; SALT, severity of alopecia tool.

5.1.2.2 Scenario analysis

The company undertook a series of scenario analyses to assess the impact of applying alternative assumptions to key model parameters. In addition, the company conducted several additional scenario analyses requested by the EAG. Results of all the scenario analyses conducted by the company are presented in Table 38. Several requested scenarios were not provided by the company, as such the EAG have conducted these additional scenario analyses and provided the results in Section 6.3.

Table 38. Company scenario analyses - deterministic

Results per patient	Baricitinib 4 mg (1)	'Watch and wait' (2)	Incremental value (1-2)
Company updated base case - post clarification			
Total costs (£)	[redacted]	[redacted]	[redacted]
QALYs	[redacted]	[redacted]	[redacted]
ICER (£/QALY)			18,072
Starting population with SALT 50-94 (severe subgroup)			
Total costs (£)	[redacted]	[redacted]	[redacted]
QALYs	[redacted]	[redacted]	[redacted]
ICER (£/QALY)			25,154
Starting population with SALT 95-100 (very severe subgroup)			
Total costs (£)	[redacted]	[redacted]	[redacted]
QALYs	[redacted]	[redacted]	[redacted]
ICER (£/QALY)			12,685
Response based on $SALT_{75}$			
Total costs (£)	[redacted]	[redacted]	[redacted]
QALYs	[redacted]	[redacted]	[redacted]
ICER (£/QALY)			16,490
Utilities based on pooled EQ-5D data from BRAVE-AA1 and BRAVE-AA2			
Total costs (£)	[redacted]	[redacted]	[redacted]
QALYs	[redacted]	[redacted]	[redacted]
ICER (£/QALY)			174,446
Utilities based on pooled HADS data from BRAVE-AA1 and BRAVE-AA2			
Total costs (£)	[redacted]	[redacted]	[redacted]
QALYs	[redacted]	[redacted]	[redacted]
ICER (£/QALY)			55,483
Proportion of patients on BSC drugs based on clinical expert opinion			
Total costs (£)	[redacted]	[redacted]	[redacted]
QALYs	[redacted]	[redacted]	[redacted]
ICER (£/QALY)			Dominant
EAG requested scenarios			
Response based on SALT≤20			
Total costs (£)	[redacted]	[redacted]	[redacted]
QALYs	[redacted]	[redacted]	[redacted]
ICER (£/QALY)			17,071
Response based on SALT≤20 - Starting population with SALT 50-94 (severe subgroup)			
Total costs (£)	[redacted]	[redacted]	[redacted]
QALYs	[redacted]	[redacted]	[redacted]

ICER (£/QALY)			18,773
Response based on SALT ≤20 - Starting population with SALT 95-100 (very severe subgroup)			
Total costs (£)	[redacted]	[redacted]	[redacted]
QALYs	[redacted]	[redacted]	[redacted]
ICER (£/QALY)			16,929
Response based on SALT≤10			
Total costs (£)	[redacted]	[redacted]	[redacted]
QALYs	[redacted]	[redacted]	[redacted]
ICER (£/QALY)			20,782
Response based on SALT≤20 and duration of AA episode <4 years			
Total costs (£)	[redacted]	[redacted]	[redacted]
QALYs	[redacted]	[redacted]	[redacted]
ICER (£/QALY)			16,154
Response based on SALT≤20 and duration of AA episode >4 years			
Total costs (£)	[redacted]	[redacted]	[redacted]
QALYs	[redacted]	[redacted]	[redacted]
ICER (£/QALY)			18,982
Long-term all-cause discontinuation based on Week 36-52 data for baricitinib 4 mg ([redacted])*			
Total costs (£)	[redacted]	[redacted]	[redacted]
QALYs	[redacted]	[redacted]	[redacted]
ICER (£/QALY)			16,293
Inclusion of costs for AEs			
Total costs (£)	[redacted]	[redacted]	[redacted]
QALYs	[redacted]	[redacted]	[redacted]
ICER (£/QALY)			18,348
Removal of monitoring costs for 'Watch and wait' in the induction phase and Maintenance health state			
Total costs (£)	[redacted]	[redacted]	[redacted]
QALYs	[redacted]	[redacted]	[redacted]
ICER (£/QALY)			20,887
Wig costs weighted by proportion of females (60.67%)			
Total costs (£)	[redacted]	[redacted]	[redacted]
QALYs	[redacted]	[redacted]	[redacted]
ICER (£/QALY)			18,732
Inclusion of only one wig on the induction phase			
Total costs (£)	[redacted]	[redacted]	[redacted]
QALYs	[redacted]	[redacted]	[redacted]
ICER (£/QALY)			18,068

Abbreviations: AA, alopecia areata; BSC, best supportive care; EAG, Evidence Review Group; HADS, Hospital Anxiety and Depression Scale; ICER, incremental cost effectiveness ratio; QALY, quality adjusted life year; SALT, Severity of Alopecia Tool

*The EAG has presented results for this scenario from the model, as it could not verify the company's ICER presented in the clarification response (B4bi)

5.1.3 Model validation and face validity check

For the model validation, the company stated that quality control checks were performed by an analyst not involved in the development of the economic model. Additionally, the company provided the model quality assurance checklist used for the validation, which the EAG considers provided a thorough and appropriate check of the model.[45] Consequently, the EAG did not identify any model errors.

6 Additional economic analysis undertaken by the EAG

6.1 Model corrections

The Evidence Assessment Group (EAG) did not identify any model corrections.

6.2 Exploratory and sensitivity analyses undertaken by the EAG

In Section 4 of this report, the EAG has described several scenarios that warrant further exploration in addition to the company's own sensitivity and scenario analyses to ascertain the impact of these changes on the incremental cost effectiveness ratio (ICER). The deterministic scenarios that the EAG has performed are as follows and results are presented in Section 6.3:

- Treatment response at Week 36 defined as achieving Severity of Alopecia Tool (SALT) score of less than or equal to 20 (SALT≤20) in combination with utility values for baseline and change from baseline associated with achieving SALT≤20 from the BRAVE trials (Section 4.2.5.2 and 4.2.8.1):
    - Full analysis set (FAS) - baseline SALT values of 50-100.
    - Severe subgroup - baseline SALT values of 50-94.
    - Very severe subgroup - baseline SALT values of 95-100.
- SALT≤10 and in combination with utility values for baseline and change from baseline associated with achieving SALT≤10 from the BRAVE trials - FAS only (Section 4.2.5.2 and 4.2.8.1).
- No placebo response - all patients in the comparator arm move to the best supportive care (BSC) health state at Week 36 (Section 4.2.5.2)
- Removal of disease monitoring costs in the BSC health state for both arms of the model (Section 4.2.9.5).
- Removal of non-pharmacological support costs (Section 4.2.9.5).

6.3 EAG scenario analysis

Table 39 presents the deterministic results of the EAG exploratory analyses described in Section 6.2. Results reported include the company's proposed patient access scheme (PAS); a fixed pack price of [redacted].

Table 39. Results of the EAG's scenario analyses

	Results per patient	Baricitinib 4 mg	'Watch and wait'	Incremental value
0	**Company base case**			
	Total costs (£)	[redacted]	[redacted]	[redacted]
	QALYs	[redacted]	[redacted]	[redacted]
	ICER (£/QALY)			18,072
1	**SALT≤20 at Week 36 + baseline and CFB utility from BRAVE trials - FAS population**			
	Total costs (£)	[redacted]	[redacted]	[redacted]
	QALYs	[redacted]	[redacted]	[redacted]
	ICER (£/QALY)			118,494
2	**SALT≤20 at Week 36 + baseline and CFB utility from BRAVE trials - severe population**			
	Total costs (£)	[redacted]	[redacted]	[redacted]
	QALYs	[redacted]	[redacted]	[redacted]
	ICER (£/QALY)			130,303
3	**SALT≤20 at Week 36 + baseline and CFB utility from BRAVE trials - very severe population**			
	Total costs (£)	[redacted]	[redacted]	[redacted]
	QALYs	[redacted]	[redacted]	[redacted]
	ICER (£/QALY)			117,510
4	**SALT≤10 at Week 36 + baseline and CFB utility from BRAVE trials - FAS population**			
	Total costs (£)	[redacted]	[redacted]	[redacted]
	QALYs	[redacted]	[redacted]	[redacted]
	ICER (£/QALY)			129,068
5	**No placebo response (SALT≤20 at Week 36 + CFB utility from BRAVE trials - FAS population)**			
	Total costs (£)	[redacted]	[redacted]	[redacted]
	QALYs	[redacted]	[redacted]	[redacted]
	ICER (£/QALY)			86,343
6	**Removal of disease monitoring costs in the BSC health state for both arms of the model**			
	Total costs (£)	[redacted]	[redacted]	[redacted]
	QALYs	[redacted]	[redacted]	[redacted]
	ICER (£/QALY)			63,941
7	**Removal of non-pharmacological support costs**			
	Total costs (£)	[redacted]	[redacted]	[redacted]
	QALYs	[redacted]	[redacted]	[redacted]
	ICER (£/QALY)			18,679

Abbreviations: BSC, best supportive care; CFB, change from baseline; EAG, Evidence Assessment Group; FAS, full analysis set; ICER, incremental cost-effectiveness ratio; QALY, quality adjusted life year; SALT, Severity of Alopecia Tool.

6.4 EAG preferred assumptions

In this section, the EAG presents its base case ICER for baricitinib 4 mg for treating severe alopecia areata (AA). As discussed in Section 4, the EAG considers that the definition of the comparator arm should be 'discharged from care' based on advice obtained from its clinical experts. As a reminder, the EAG's clinical experts did not recognise the company's comparator of 'Watch and wait' as reflective of what happens in the NHS for patients with severe AA. According to the EAG's clinical experts, patients would not be regularly monitored if they are not on treatment. Additionally, the EAG's clinical experts considered that a significant proportion of patients may not take up treatment (effectively discharged from care) and instead opt for using wigs to manage their hair loss. As such, the comparator for the EAG's base case is 'discharged from care'.

The following assumptions were incorporated into the EAG's base case:

- Treatment response at Week 36 defined as achieving SALT≤20.
    - utility values for baseline and change from baseline associated with achieving SALT≤20Long-term all-cause discontinuation based on Week 36-52 data for baricitinib 4 mg.
- No monitoring costs in the induction phase and Maintenance health state for 'Watch and wait' (comparator defined as 'discharged from care').
- Removal of disease monitoring costs in the BSC health state for both arms of the model.
- Removal of non-pharmacological psychological support costs.
- One wig assumed in the induction phase for both arms of the model.

The EAG considers that costs of adverse events (AEs) should be included in the EAG's preferred base case. However, as mentioned in Section 4.2.6.1, the EAG was unable to verify the inputs used in the company's AE scenario and was unable to produce an alternative scenario due to a paucity of time. Nonetheless, the EAG requests that during technical engagement, the company provides a more thorough description and justification of their approach to the inclusion of AEs and assumed unit costs to treat each AE and update the scenario if necessary.

Table 40 presents the impact of each assumption on the ICER and Table 41 presents the EAG's deterministic and probabilistic base case results. Table 42 presents the severity subgroup analysis around the EAG base case but it should be noted that probabilistic subgroup results could not be obtained due to a problem with the probabilistic sensitivity analysis (PSA) function in the model.

In the EAG base case probabilistic analysis, an incremental quality-adjusted life-year (QALY) gain of [redacted] over 'discharged from care' along with additional costs of [redacted] for the baricitinib 4 mg, generates an ICER of £423,775 per QALY. The net monetary benefit (NMB) using the £30,000 threshold is [redacted] and the net health benefit (NHB) is [redacted]. The EAG considers that the ICERs are highly sensitive due to the small incremental costs and quality-adjusted life-year (QALY) gain, such that small changes cause a substantial impact.

Additionally, as mentioned in Section 4.2.8.1, the EAG acknowledges that there is a small, but heterogenous, patient population that is more adversely affected in terms of Health-related quality of life (HRQoL) but that the demographics of this population are difficult to identify clinically and consistently, and it is beyond the scope of assessment to identify that group. Nonetheless, the EAG ran two scenarios around the EAG base case and severity subgroup analysis to identify the QALY gain needed for the ICER to reach the £20,000 and £30,000 cost-effectiveness threshold and these are presented in Table 43. The results of the threshold analysis demonstrate that for the overall population, a QALY gain of [redacted] to [redacted] is needed for the ICER to be within the £20,000 to £30,000 threshold. Thus, the EAG advises the committee to consider if the estimated QALY gain needed for baricitinib 4 mg to be cost-effective is plausible for the condition under consideration.

Table 40. EAG's preferred model assumptions - FAS population

Preferred assumption	Section in EAG report	Deterministic ICER (£/QALY)	Cumulative ICER (£/QALY)
Company base case post clarification	**-**	**18,072**	**18,072**
SALT≤20 at Week 36	4.2.5.2	17,071	17,071
SALT≤20 at Week 36 + baseline and CFB utility from BRAVE trials	4.2.5.2 and 4.2.8.1	118,494	118,494
Long-term all-cause discontinuation based on Week 36-52 data for baricitinib 4 mg ([redacted])	4.2.5.2	16,293	107,217
No monitoring costs in the induction phase and Maintenance health state for 'Watch and wait' (comparator defined as 'discharged from care')	4.2.9.5	20,887	126,309
Removal of disease monitoring costs in the BSC health state for both arms of the model	4.2.9.5	63,941	419,926
Removal of non-pharmacological psychological support costs	4.2.9.5	18,679	423,809
One wig assumed in the induction phase for both arms of the model	4.2.9.5	18,068	423,775
EAG preferred base case	**-**	-	**423,775**

Abbreviations: BSC, best supportive care; CFB, change from baseline; EAG, Evidence Assessment Group; ICER, incremental cost-effectiveness ratio; QALY, quality adjusted life year; SALT, Severity of Alopecia Tool.

Table 41. EAG's base case

Interventions	Total Costs (£)	Total LY	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER (£/QALY)
Deterministic results							
'Discharged from care'	[redacted]	22.60	[redacted]	-	-	-	-
Baricitinib 4 mg	[redacted]	22.60	[redacted]	[redacted]	0.00	[redacted]	423,775
Probabilistic results							
'Discharged from care'	[redacted]	22.60	[redacted]	-	-	-	-
Baricitinib 4 mg	[redacted]	22.60	[redacted]	[redacted]	-	[redacted]	379,030

Abbreviations: ICER, incremental cost effectiveness ratio, LYG, life year gained; QALY, quality adjusted life year.

Table 42. Deterministic scenarios around the EAG base case

	Results per patient	Baricitinib 4 mg	'Discharged from care'	Incremental value
0	**EAG base case**			
	Total costs (£)	[redacted]	[redacted]	[redacted]
	QALYs	[redacted]	[redacted]	[redacted]
	ICER (£/QALY)			423,775
1	**Severe subgroup - baseline SALT 50-95**			
	Total costs (£)	[redacted]	[redacted]	[redacted]
	QALYs	[redacted]	[redacted]	[redacted]
	ICER (£/QALY)			407,212
2	**Very severe subgroup - baseline SALT 95-100**			
	Total costs (£)	[redacted]	[redacted]	[redacted]
	QALYs	[redacted]	[redacted]	[redacted]
	ICER (£/QALY)			456,573

Abbreviations: ICER, incremental cost-effectiveness ratio; QALY, quality adjusted life year; SALT, Severity of Alopecia Tool.

Note: the same baseline utility ([redacted]), change from baseline ([redacted]) and treatment discontinuation rate ([redacted]) have been used for the subgroups as for the base case as the relevant data were not available by severity.

Table 43. Threshold analysis on QALY gain needed for £20,000 to £30,000 cost-effectiveness threshold

Population	QALY gain - £20,000 threshold	QALY gain - £30,000 threshold
Full analysis set - baseline SALT 50-100	[redacted]	[redacted]
Severe subgroup - baseline SALT 50-94	[redacted]	[redacted]
Very severe subgroup - baseline SALT 95-100	[redacted]	[redacted]

Abbreviations: EAG, evidence review group; ICER, incremental cost-effectiveness ratio; QALY, quality adjusted life year; SALT, Severity of Alopecia Tool.

6.5 Conclusions of the cost effectiveness sections

Generally, the EAG considers the company's submitted cost-effectiveness analysis adheres to the decision problem defined in the NICE final scope. However, for the comparator (which is listed in the final scope as established clinical management without baricitinib 4 mg), current treatment of severe alopecia areata is variable across the NHS and clinician dependent as many treatments are ineffective and there is use of off label medicines. Additionally, the EAG's clinical experts considered that a significant proportion of patients may not take up treatment (effectively discharged from care) and instead opt for using wigs to manage their hair loss. The EAG agrees with the company that the no active treatment component of "Watch and wait" is a common management strategy used for adults with severe AA. However, the EAG's clinical experts considered that patients would not be regularly monitored if they are not on treatment. As such, the EAG considers that 'discharged from care' is the relevant comparator for the analysis. The EAG's preferred definition for the comparator has implications for costs included in the model as monitoring costs in the induction phase and Maintenance health states are no longer relevant. Additionally, the EAG's clinical experts considered that if patients do achieve a sufficient treatment response on any treatment (including Janus Kinase [JAK] inhibitors), they are unlikely to engage with further care. As such, much of the costs included in the BSC health state is likely to not be incurred by patients.

As such, the EAG considers that a true reflection of the cost-effectiveness of baricitinib 4 mg is a comparison where, in the absence of baricitinib 4 mg, patients manage their hair loss with the use of wigs and orthotics and for a small proportion of patients, antidepressants are required to manage the psychological burden of severe AA.

Additionally, the EAG considers (based on advice from its clinical experts) that patients value an absolute change in scalp hair regrowth and a relative change from their baseline hair loss (as measure by SALT scores) may still require hair removal or use of wigs due to patchy regrowth. As such, the use of $SALT_{50}$ as the measure of treatment response at Week 36 is not considered clinically meaningful and instead SALT≤20 is the EAG's preferred definition of treatment response and aligns with the primary endpoint of the key baricitinib trials, BRAVE-AA1 and BRAVE-AA2.

The utilities used are a key driver in the model as based on the EQ-5D data from the BRAVE-AA1 and BRAVE-AA2 trials, patients have a relatively high baseline utility and there is not a substantial increase in health-related quality of life (HRQoL) from achieving a response to treatment (whether that is the company's base case definition of $SALT_{50}$ or the primary endpoint of SALT≤20 in the trials).

The company argue that the EQ-5D is insensitive to changes in the severity of AA and lacked content validity as baseline values were almost the same as UK age- and sex-adjusted general population values. However, the company hasn't supplied sufficient evidence to validate the lack of content validity with the EQ-5D nor has it demonstrated why patients should have a substantial change in their QoL.

The EAG's clinical experts advised that for most patients' HRQoL may only be mildly affected and thus may not be that different to the general population but equally HRQoL is severely affected for a few patients (primarily driven by adverse mental health). Additionally, the EAG's clinical experts advised that overtime, patients may come to terms with their hair loss, while a few may remain distressed about their condition. Thus, the EAG acknowledges that there is a small, but heterogenous, patient population that is more adversely affected in terms of HRQoL but that the demographics of this population are difficult to identify clinically and consistently, and it is beyond the scope of assessment to identify that group. However, the EAG has estimated the QALY gain needed to reach the £20,000 and £30,000 cost-effectiveness thresholds and advises the committee to consider if the estimated QALY gain needed for baricitinib 4 mg to be cost-effective is plausible for the condition under consideration.

7 References

1. Harries M, Macbeth AE, Holmes S, Chiu WS, Gallardo WR, Nijher M, et al. The epidemiology of alopecia areata: a population-based cohort study in UK primary care*. *British Journal of Dermatology* 2022; 186: 257-65.
2. Petukhova L, Cabral RM, Mackay-Wiggan J, Clynes R, Christiano AM. The genetics of alopecia areata: What's new and how will it help our patients? *Dermatol Ther* 2011; 24: 326-36.
3. Darwin E, Hirt PA, Fertig R, Doliner B, Delcanto G, Jimenez JJ. Alopecia Areata: Review of Epidemiology, Clinical Features, Pathogenesis, and New Treatment Options. *Int J Trichology* 2018; 10: 51-60.
4. Pratt CH, King LE, Jr., Messenger AG, Christiano AM, Sundberg JP. Alopecia areata. *Nat Rev Dis Primers* 2017; 3: 17011.
5. Trueb RM, Dias M. Alopecia Areata: a Comprehensive Review of Pathogenesis and Management. *Clin Rev Allergy Immunol* 2018; 54: 68-87.
6. Xing L, Dai Z, Jabbari A, Cerise JE, Higgins CA, Gong W, et al. Alopecia areata is driven by cytotoxic T lymphocytes and is reversed by JAK inhibition. *Nature medicine* 2014; 20: 1043-9.
7. Fukuyama M, Ito T, Ohyama M. Alopecia areata: Current understanding of the pathophysiology and update on therapeutic approaches, featuring the Japanese Dermatological Association guidelines. *J Dermatol* 2022; 49: 19-36.
8. Triyangkulsri K, Suchonwanit P. Role of janus kinase inhibitors in the treatment of alopecia areata. *Drug design, development and therapy* 2018; 12: 2323-35.
9. Messenger AG, McKillop J, Farrant P, McDonagh AJ, Sladden M. British Association of Dermatologists' guidelines for the management of alopecia areata 2012. *British Journal of Dermatology* 2012; 166: 916-26.
10. Lee S, Kim BJ, Lee YB, Lee W-S. Hair Regrowth Outcomes of Contact Immunotherapy for Patients With Alopecia Areata: A Systematic Review and Meta-analysis. *JAMA Dermatology* 2018; 154: 1145-51.
11. National Institute for Health and Care Excellence. Baricitinib for treating severe alopecia areata - Final scope [10941], 2022. Available from: https://www.nice.org.uk/guidance/gid-ta10941/documents/final-scope. Date accessed: September 2022.
12. King B, Ohyama M, Kwon O, Zlotogorski A, Ko J, Mesinkovska NA, et al. Two Phase 3 Trials of Baricitinib for Alopecia Areata. *The New England journal of medicine* 2022.
13. Adelphi Real World. Alopecia Areata Disease Specific Programme Physician Survey - Survey Materials. 2020.
14. Eli Lilly and Company. Disease Burden, Disease Management, Treatment Patters and Resouce Use in Patients with Alopecia Areata - Protocol. 2022.
15. National Institute for Health and Care Excellence (NICE). Baricitinib for moderate to severe atopic dermatitis: Technology appraisal guidance [TA681]. 2021. Available from: https://www.nice.org.uk/guidance/ta681
16. National Institute for Health and Care Excellence (NICE). Baricitinib for moderate to severe rheumatoid arthritis: Technology appraisal guidance [TA466]. 2017. Available from: https://www.nice.org.uk/guidance/ta466
17. National Institute for Health and Care Excellence. Equality impact assessment – Scoping: Baricitinib for treating severe alopecia areata [ID3979] 2022. Available from: https://www.nice.org.uk/guidance/indevelopment/gid-ta10941/documents. Date accessed.
18. King B, Ko J, Forman S, Ohyama M, Mesinkovska N, Yu G, et al. Efficacy and safety of the oral Janus kinase inhibitor baricitinib in the treatment of adults with alopecia areata: Phase 2 results from a randomized controlled study. *Journal of the American Academy of Dermatology* 2021; 85: 847-53.

19. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011; 343: d5928.
20. Centre for Reviews and Dissemination. *Systematic Reviews: CRD's guidance for undertaking reviews in health care*. York: Centre for Reviews and Dissemination, University of York; 2008.
21. Gupta AK, Carviel JL, Foley KA, Shear NH, Piraccini BM, Piguet V, et al. Monotherapy for Alopecia Areata: A Systematic Review and Network Meta-Analysis. *Skin Appendage Disord* 2019; 5: 331-7.
22. Lemay KR, Tulloch HE, Pipe AL, Reed JL. Establishing the Minimal Clinically Important Difference for the Hospital Anxiety and Depression Scale in Patients With Cardiovascular Disease. *J Cardiopulm Rehabil Prev* 2019; 39: E6-E11.
23. Wynne SC, Patel S, Barker RE, Jones SE, Walsh JA, Kon SS, et al. Anxiety and depression in bronchiectasis: Response to pulmonary rehabilitation and minimal clinically important difference of the Hospital Anxiety and Depression Scale. *Chron Respir Dis* 2020; 17: 1479973120933292.
24. Canadian Agency for Drugs and Technologies in Health. Clinical Review Report. Dupilmab (Dupixent), 2020. Available from: https://www.cadth.ca/dupilumab-0.
25. Frendl DM, Ware JE, Jr. Patient-reported functional health and well-being outcomes with drug therapy: a systematic review of randomized trials using the SF-36 health survey. *Med Care* 2014; 52: 439-45.
26. Ogura K, Bartelstein MK, Yakoub MA, Nikolic Z, Boland PJ, Healey JH. Minimal clinically important differences in SF-36 global score: Current value in orthopedic oncology. *J Orthop Res* 2021; 39: 2116-23.
27. Both H, Essink-Bot ML, Busschbach J, Nijsten T. Critical review of generic and dermatology-specific health-related quality of life instruments. *J Invest Dermatol* 2007; 127: 2726-39.
28. Shikiar R, Willian MK, Okun MM, Thompson CS, Revicki DA. The validity and responsiveness of three quality of life measures in the assessment of psoriasis patients: results of a phase II study. *Health Qual Life Outcomes* 2006; 4: 71.
29. Jayadevappa R, Cook R, Chhatre S. Minimal important difference to infer changes in health-related quality of life-a systematic review. *Journal of clinical epidemiology* 2017; 89: 188-98.
30. Secrest AM, Hopkins ZH, Frost ZE, Taliercio VL, Edwards LD, Biber JE, et al. Quality of Life Assessed Using Skindex-16 Scores Among Patients With Acne Receiving Isotretinoin Treatment. *JAMA Dermatol* 2020; 156: 1098-106.
31. Kind P, Hardman G, Macan S. UK population norms for EQ-5D. Discussion paper 172. *University of York* 1999.
32. Rencz F, Gulacsi L, Pentek M, Wikonkal N, Baji P, Brodszky V. Alopecia areata and health-related quality of life: a systematic review and meta-analysis. *Br J Dermatol* 2016; 175: 561-71.
33. Taylor PC, Takeuchi T, Burmester GR, Durez P, Smolen JS, Deberdt W, et al. Safety of baricitinib for the treatment of rheumatoid arthritis over a median of 4.6 and up to 9.3 years of treatment: final results from long-term extension study and integrated database. *Ann Rheum Dis* 2022; 81: 335-43.
34. Jones K, Burns A. PSSRU: Unit Costs of Health and Social Care (2021). 2021. Available from: https://www.pssru.ac.uk/project-pages/unit-costs/unit-costs-of-health-and-social-care-2021/.
35. National Health Service (NHS). 2020/21 National Cost Collection data. 2020. Available from: https://www.england.nhs.uk/national-cost-collection/. Date accessed.
36. Service NH. National Health Service (NHS) Business Services Authority. Drug Tariff (April 2022). 2022.
37. European Medicinces Agency (EMA). Olumiant (baricitinib) SmPC. Available from: https://www.ema.europa.eu/en/documents/product-information/olumiant-epar-product-information_en.pdf. . Date accessed: September 2022.

38. National Institute for Health and Care Excellence. Dupilumab for treating moderate to severe atopic dermatitis [TA534]. 2018. Available from: https://www.nice.org.uk/guidance/ta534.
39. Statistics OfN. Population Estimates for the UK, England and Wales, Scotland and Northern Ireland: mid-2020. June 2021. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/mid2020. Date accessed: September 2022.
40. Hernández Alava M, Pudney S, Wailoo A. Estimating EQ-5D by age and sex for the UK: Report by the Decision Support Unit., 2022. Available from: https://www.sheffield.ac.uk/sites/default/files/2022-02/DSU%20Age%20based%20utility%20-%20Final%20for%20website.pdf. Date accessed: October 2022.
41. National Institute for Health and Care Excellence (NICE). NICE health technology evaluations: the manual, 2022. Available from: https://www.nice.org.uk/process/pmg36/resources/nice-health-technology-evaluations-the-manual-pdf-72286779244741. Date accessed: September 2022.
42. Ara R, Wailoo AJ. NICE DSU Technical Support Document 12: The use of health state utility values in decision models., 2011. Available from: http://nicedsu.org.uk/wp-content/uploads/2016/03/TSD12-Utilities-in-modelling-FINAL.pdf. Date accessed.
43. National Institute for Health and Care Excellence. Depression in adults: recognition and management [CG90], 2009. Available from: https://www.nice.org.uk/guidance/cg90.
44. National Institue for Health and Care Excellence. Depression in adults: treatment and management [NG222], 2022. Available from: https://www.nice.org.uk/guidance/ng222. Date accessed: September 2022.
45. Lilly E. NICE Cost-Effectiveness Model for Baricitinib in AA: Model Sanity Checklist. 2022.

8 Appendices

8.1 Additional baseline characteristics

Baseline characteristics reported in the CS but not presented in the main body of the EAG report are reproduced in Table 44.

Table 44. Baseline characteristics reported for BRAVE-AA1 and BRAVE-AA2 not included in Table 11.

Characteristic	BRAVE–AA1		BRAVE–AA2	
	Placebo (N=189)	baricitinib 4 mg (N=281)	Placebo (N=156)	baricitinib 4 mg (N=234)
Mean (SD) age, years	37 (12.91)	36 (13.27)	37 (12.35)	38 (12.65)
Female, n (%)	109 (57.7)	165 (58.7)	98 (62.8)	144 (61.5)
Geographic region, n (%)				
North America	103 (54.5)	153 (54.4)	54 (34.6)	82 (35.0)
Asia	70 (37.0)	107 (38.1)	42 (26.9)	63 (26.9)
Rest of the world	16 (8.5)	21 (7.5)	60 (38.5)	89 (38.0)
Race, n (%)				
White	83 (44.1)	123 (43.9)	85 (54.5)	144 (61.5)
Asian	78 (41.5)	114 (40.7)	51 (32.7)	67 (28.6)
Black or African American	17 (9.0)	28 (10.0)	16 (10.3)	18 (7.7)
American Indian or Alaska Native	8 (4.3)	8 (2.9)	0 (0.0)	0 (0.0)
Native Hawaiian or Other Pacific Islander	1 (0.5)	1 (0.4)	0 (0.0)	0 (0.0)
Mean (SD) BMI, kg/m2	[redacted]	[redacted]	[redacted]	[redacted]
Mean (SD) duration since onset of AA, years	12.6 (11.2)	11.8 (11.1)	11.79 (10.190)	11.89 (11.122)
Age of AA onset, n (%)				
<18 years	[redacted]	[redacted]	57 (36.5)	74 (31.6)
≥18 years	[redacted]	[redacted]	99 (63.5)	160 (68.4)
Patients with AU, n (%)	74 (39.2)	127 (45.2)	66 (42.3)	111 (47.4)
ClinRO for eyebrow hair loss, n (%)				
2	53 (28.3)	73 (26.3)	46 (30.1)	49 (21.0)
3	71 (38.0)	115 (41.4)	66 (43.1)	112 (48.1)
ClinRO for eyelash hair loss, n (%)				
2	38 (20.3)	74 (26.6)	31 (20.3)	43 (18.5)
3	58 (31.0)	93 (33.5)	59 (38.6)	97 (41.6)

PRO for Scalp Hair Assessment				
3 (50–94% hair loss)	72 (38.1)	102 (36.4)	60 (38.5)	78 (33.3)
4 (95–100% hair loss)	109 (57.7)	173 (61.8)	91 (58.3)	137 (58.5)

Abbreviations: AA: alopecia areata; AU: alopecia universalis; ClinRO; Clinician reported outcome; PRO: patient reported outcome SD: standard deviation

8.2 SALT responder statistical analyses

Table 45 provides the detailed statistical analysis, including number of responders and differences, odds ratios and p-values versus placebo for baricitinib 4 mg in BRAVE-AA1 and BRAVE-AA2 at Week 36 for SALT ≤20. Table 46 provides these data for SALT ≤10, and Table 47 for the relative treatment response outcomes, $SALT_{50}$ and $SALT_{75}$.

Table 45. SALT ≤20 response at Week 36 for BRAVE-AA1 and BRAVE-AA2 (FAS)

SALT ≤20, Week 36	BRAVE-AA1		BRAVE-AA2	
	Placebo (N=189)	4 mg baricitinib (N=281)	Placebo (N=189)	4 mg baricitinib (N=281)
Response, n (%) (95% CI)	10 (5.3) (2.9, 9.5)	99 (35.2) (29.9, 41.0)	4 (2.6) (1.0, 6.4)	76 (32.5) (26.8, 38.7)
Difference (95% CI) vs placebo	N/A	29.9 (23.2, 36.2)	NA	29.9 (23.1, 36.3)
Odds ratio (95% CI) vs placebo	[redacted]	[redacted]	[redacted]	[redacted]
p-value vs placebo	N/A	<0.001	NA	<0.001

Source: CS Table 15

Abbreviations: AA: alopecia areata; CI: confidence interval; FAS: full analysis set; SALT: severity of alopecia tool

Table 46. SALT ≤10 response at Week 36 for BRAVE-AA1 and BRAVE-AA2 (FAS)

SALT ≤10, Week 36	BRAVE-AA1		BRAVE-AA2	
	Placebo (N=189)	4 mg baricitinib (N=281)	Placebo (N=189)	4 mg baricitinib (N=281)
Response, n (%) (95% CI)	[redacted]	[redacted]	[redacted]	[redacted]
Difference (95% CI) vs placebo	[redacted]	[redacted]	[redacted]	[redacted]
Odds ratio (95% CI) vs placebo	[redacted]	[redacted]	[redacted]	[redacted]
p-value vs placebo	[redacted]	[redacted]	[redacted]	[redacted]

Source: CS Table 22

Abbreviations: AA: alopecia areata; CI: confidence interval; FAS: full analysis set; SALT: severity of alopecia tool

Table 47. $SALT_{50}$ and $SALT_{75}$ response at Week 36 for BRAVE-AA1 and BRAVE-AA2 (FAS)

Week 36	BRAVE-AA1		BRAVE-AA2	
	Placebo (N=189)	4 mg baricitinib (N=281)	Placebo (N=156)	4 mg baricitinib (N=234)
$SALT_{50}$				
Response, n (%) (95% CI)	[redacted]	[redacted]	[redacted]	[redacted]
Difference (95% CI) vs PBO		[redacted]		[redacted]
Odds ratio (95% CI) vs PBO		[redacted]		[redacted]
p-value vs placebo		[redacted]		[redacted]
$SALT_{75}$				
Response, n (%) (95% CI)	[redacted]	[redacted]	[redacted]	[redacted]
Difference (95% CI) vs PBO		[redacted]		[redacted]
Odds ratio (95% CI) vs PBO		[redacted]		[redacted]
p-value vs placebo		[redacted]		[redacted]

Abbreviations: AA: alopecia areata; CI: confidence interval; FAS: full analysis set; SALT: severity of alopecia tool

8.3 Atopic background subgroup analysis

Table 48 provides SALT ≤20 response data for the baricitinib 4 mg and placebo arms of BRAVE-AA1 and BRAVE-AA2 by atopic background status (no atopic background and atopic background categories). For the baricitinib 4 mg arm, a larger proportion of patients in the atopic background category achieved SALT ≤20 at Week 36 (BRAVE-AA1, [redacted]; BRAVE-AA2, [redacted]) than those in the no atopic background category (BRAVE-AA1, [redacted]; BRAVE-AA2, [redacted]), a pattern also apparent in the placebo arm data (no atopic background category: BRAVE-AA1, [redacted]; BRAVE-AA2, [redacted]; atopic background category; BRAVE-AA1, [redacted]; BRAVE-AA2, [redacted]).

Table 48. SALT ≤20 response of BRAVE-AA1 and BRAVE-AA2 by atopic background status at baseline

	Atopic background	Week 36 SALT ≤20 response rate	
		BRAVE-AA1	BRAVE-AA2
Baricitinib 4 mg	No atopic background	[redacted]	[redacted]
	Atopic background	[redacted]	[redacted]

Placebo	No atopic background	[redacted]	[redacted]
	Atopic background	[redacted]	[redacted]

Source: Clarification questions Table 6 and Table 7

Abbreviations: AA: alopecia areata; SALT: severity of alopecia tool

8.4 Source of the confidential prices used in the confidential appendix

Table 49. Source of the confidential prices used in the confidential appendix

Treatment	Source of price/type of commercial arrangement
Ciclosporin	CMU
Methotrexate	List price
Azathioprine	CMU
Intralesional steroids (triamcinolone acetonide)	List price
DPCP treatment	NHSE
Prednisolone	eMIT
Prednisolone	eMIT
TCS: Mometasone ointment	eMIT
Minoxidil 5% foam (topical)	List price
Minoxidil tablets	List price
Mycophenolate Mofetil	List price
Anthralin / dithranol 0.1% cream	List price
Sertraline	List price
Escitalopram	eMIT
Duloxetine	eMIT

Abbreviations: CMU, confidential medicines unit; DPCP, diphenylcyclopropenone; contact immunotherapy; eMIT, Drugs and pharmaceutical electronic market information tool; NHSE, National Health Service England.