NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

SINGLE TECHNOLOGY APPRAISAL

APPEAL HEARING

# Baricitinib for treating severe alopecia areata [ID3979]

# Decision of the panel

## Introduction

1. An appeal panel was convened on 12 September 2023 to consider an appeal against NICE’s final draft guidance (FDG) to the NHS, on baricitinib for treating severe alopecia areata (AA).
2. The appeal panel consisted of:
* Professor Jonathan Cohen (Chair),
* Jackie Fielding (Non-executive director of NICE),
* Dr Biba Stanton (Health service representative),
* Adrian Griffin (Industry representative); and
* Rosemary Harris (Lay representative).
1. None of the members of the appeal panel had any competing interest to declare.
2. The panel considered the appeal submitted by Alopecia UK.
3. Alopecia UK was represented by:
* Sue Schilling (CEO of Alopecia UK)
* Lynn Wilks (Trustee and volunteer at Alopecia UK)
* Emily Doogan (Volunteer at Alopecia UK)
* Dr Leila Asfour (Consultant dermatologist, NHS)
* Dr Susan Holmes (Consultant dermatologist, NHS)
1. In addition, the following individuals involved in the appraisal were present and available to answer questions from the Appeal Panel:
* Dr Radha Todd (Chair of Technology Appraisal committee A, NICE)
* Jacoline Bouvy (Programme Director, NICE)
* Janet Robertson (Associate Director, NICE)
* Becky Pennington (committee member, Technology Appraisal Committee A, NICE)
* Sharlene Ting (Technical Analyst, NICE)
1. The appeal panel’s legal adviser, Alistair Robertson (DAC Beachcroft LLP), was also present.
2. Kawitha Helme (Industry Representative), a member of the appeal panel for technology appraisals and highly specialised technologies was present as a silent observer throughout the hearing and panel discussions.
3. The panel is grateful to Emily Doogan who gave patient testimony, describing her own experience of the psychosocial impacts of severe AA and of privately funded treatment with baricitinib.
4. Under NICE’s appeal procedures, members of the public are admitted to observe appeal hearings and several members of the public and NICE staff observed the proceedings which were held via Zoom.
5. There are two grounds under which an appeal can be lodged:

Ground One: In making the assessment that preceded the recommendation, NICE has:

(a) Failed to act fairly; and/or

(b) Exceeded its powers.

Ground Two: The recommendation is unreasonable in light of the evidence submitted to NICE.

1. Dr Mark Chakravarty, NICE Lead non-executive director for appeals, in preliminary correspondence had confirmed that Alopecia UK had valid grounds for appeal under Ground Two.
2. The evaluation that is the subject of the current appeal provided advice to the NHS on the use of baricitinib for treating severe alopecia areata.
3. Before the appeal panel inquired into the detailed complaints the following made preliminary statements: Sue Schilling on behalf of Alopecia UK and Dr Radha Todd on behalf of the appraisal committee (the committee).
4. The numbering of appeal points in this letter reflects those that were used during the hearing. The text of this document does not represent a verbatim account of the proceedings nor a documentation of the order of events that took place, but rather provides a brief summary of the submission from both Alopecia UK and the committee for the points that were discussed relevant to the decisions of the panel.

## Appeal by Alopecia UK

## Appeal Ground 1a: In making the assessment that preceded the recommendation, NICE has failed to act fairly

1. There was no valid appeal under this ground.

## Appeal Ground 1b: In making the assessment that preceded the recommendation, NICE has exceeded its powers

1. There was no appeal under this ground.

## Appeal Ground 2: The recommendation is unreasonable in the light of the evidence submitted to NICE.

### Appeal Ground 2.2: The committee's reliance on the health-related quality of life (HRQoL) tool rather than a more disease specific HRQoL tool was unreasonable.

1. Lynn Wilks, for Alopecia UK, introduced Alopecia UK's challenge to the reasonableness of the committee's use of EQ-5D-5L as a HRQoL tool. Lynn Wilks noted that of the five domains measured in EQ-5D-5L, only one (anxiety and depression) was relevant for those with severe alopecia areata (AA) and that the use of this tool by the committee did not allow for an accurate reflection of the psychosocial impacts of severe AA.
2. The main evidence considered for baricitinib in treating severe AA (from the BRAVE-AA1 and BRAVE-AA2 (BRAVE) trials) excluded from enrolment those who had severe depression. In Alopecia UK's view, this resulted in HRQoL data that failed to reflect the experiences of those who suffered the greatest psychosocial impact of severe AA. Lynn Wilks noted that Alopecia UK was surprised at the high number of trial participants who were measured as having normal quality of life from the outset. In Alopecia UK's view, this suggested that the tool itself was inadequate.
3. Lynn Wilks concluded by noting that evidence provided to the committee by clinical experts and clinicians ought to have demonstrated that the EQ-5D-5L tool was inappropriate to capture the full breadth of HRQoL impacts.
4. Dr Susan Holmes, for Alopecia UK, elaborated on Lynn Wilks' submissions. She asserted that the clinical experts who had spoken at the committee stage had explained to the committee that EQ-5D-5L is not an appropriate tool for measuring the HRQoL impacts of severe AA, and that the committee had itself acknowledged this fact at pages 9 and 11 of the FDG on baricitinib for treating severe AA. The EQ-5D-5L tool used showed significant differences between the patients and the control groups in only one domain (the anxiety and depression domain) due to the nature of the condition.
5. Dr Susan Holmes explained that the Skindex-16 Alopecia Areata Scale (Skindex) provided data showing significant improvement in the treatment group compared to the placebo group, and that there were improvements shown in the Hospital Anxiety and Depression Scale (HADS) scoring, albeit below what is considered minimal clinical significance.
6. Dr Leila Asfour, for Alopecia UK, compared the use of the EQ-5D-5L model in the current appeal with the single technology appraisal (STA) convened for baricitinib for treating atopic dermatitis (AD). She referred to the FDG for baricitinib in treating AD, where it is noted that the committee acknowledged EQ-5D-5L's inability to fully capture quality of life effects for those with dermatological conditions and a different Quality of Life (QoL) tool was used.
7. Dr Radha Todd, for NICE, responded by explaining that the generic nature of EQ-5D-5L is appropriate in most instances to facilitate consistent decision making. Referring to the NICE Health Technology Evaluations: The Manual (31 January 2022) (the Manual), Dr Todd explained that although there is provision to use alternative disease specific quality of life measures, no alternative was provided by Eli Lilly & Company (the Company) and as a result the committee continued to use the recommended tool. Janet Robertson, also for NICE, supported this view. She explained that the Manual does allow for alternatives if EQ-5D-5L is shown not to be suitable, but those other measures must be capable of producing a utility value that NICE can use.
8. Becky Pennington, for NICE, explained in response to Dr Susan Holmes’ challenge relating to Skindex, that although Skindex was considered in the committee’s decision making it cannot generate utility values that can be used for economic modelling. Both BRAVE and Adelphi results fed into the utility data used in the economic model. She noted the absence of empirical evidence from the literature reviewed by the committee that EQ-5D-5L was inappropriate. Jacoline Bouvy, for NICE, added that the committee require a preference-based measure in order to calculate health utility that can be used in a cost effectiveness model. This is not possible using Skindex. Sharlene Ting, for NICE, explained that the committee had asked the Company whether there was a mapping algorithm that could be used to calculate health utility values from Skindex, and the Company confirmed that there was not.
9. The panel asked Alopecia UK which tool would, in Alopecia UK's view, be more appropriate as a HRQoL measure. Lynn Wilks answered confirming that there is no tool that can fully capture the HRQoL impacts of severe AA – and that in its absence the committee ought to have taken all of the submitted evidence (in particular that of clinical experts) into account.
10. Dr Susan Holmes explained that it is not settled which tool best captures the HRQoL, but that the Global Registry of Alopecia Areata Disease Severity and Treatment Safety UK (GRASS) provides a list of possible tools. She noted that although she understands the reasons why the committee cannot use Skindex alone in the STA discussions, the disease specific data it provides is important and should be considered.
11. The panel asked the committee to address how they took account in their decision-making of the circumstances in which all participants appear to acknowledge the limitations of EQ-5D-5L in measuring the impact of AA on HRQoL.
12. Jacoline Bouvy responded that there are two points at play; whether EQ-5D-5L is sensitive enough to measure changes in HRQoL for those with AA; and whether the utility values in the BRAVE trial gave a fair representation of the impact on the HRQoL for those living with AA. In respect of the latter, she explained that those taking part in the BRAVE trial had a high HRQoL as a baseline, which might mean that the population in the trial were not a good representation of those with severe AA.
13. Jacoline Bouvy noted that this was not the case in the Adelphi study, which also used EQ-5D-5L. The Adelphi study participants reported more severe health state utility values than those in the BRAVE trial. Jacoline Bouvy stated that this suggests that the problem was not solely EQ-5D-5L being insensitive to measure HRQoL of people with severe AA, but that the different characteristics of those in Adelphi compared with those in BRAVE contributed to the different results. The committee concluded that the true answer lay between the results reported by BRAVE and Adelphi. Jacoline Bouvy explained that the committee recognised the limitations of the BRAVE trial and that is why it was not completely relied upon by the committee in reaching its conclusions.
14. The panel noted that the committee did receive substantial information from other sources pointing out the inadequacy of EQ-5D-5L and asked how that was taken into account. Dr Radha Todd explained that that was why the committee moved from looking only at BRAVE, why the committee's recommendation was to use a utility value figure between the values of both BRAVE and Adelphi trials, and why the committee recommended research to try to identify an appropriate disease-specific measure of quality of life.
15. The panel questioned the utility changes shown in both the BRAVE trial (showing a 2% increase in utility measure) and the Adelphi study (showing a 12% increase in utility measure). The panel asked the committee how and whether it used input from clinicians and patients to verify that the true benefit was somewhere in between both sets of results.
16. Becky Pennington explained that the committee was faced with a 'gold-standard' trial – BRAVE, which had clear limitations as discussed above. The committee also had the Adelphi trial which was not as robust because it used a smaller sample size, different definitions of severity, and relied on cross-sectional data rather than the preferred longitudinal data. The Adelphi study did not show how QoL would change when treated with baricitinib. The committee noted that the Company's base case relied on Adelphi and that this would align more with the clinical view. However, the committee had also to note the limitations of that sort of evidence. She explained further that patients and experts did not give evidence in terms of "2%" or "12%" difference to QoL, and they would not have expected to. The committee did hear testimony from patients and clinical experts in the meetings, and that testimony drove the committee's concerns about BRAVE and is why the committee also considered the Adelphi results.
17. The appeal panel concluded as follows. There was agreement from everyone at the hearing that AA can have a profound impact on quality of life. The committee had acknowledged that the very high health related quality of life (QoL) at baseline in the BRAVE trial seemed to lack face validity when compared with both patient and expert testimony and other sources of data. They had noted that this could suggest that the EQ-5D is not sensitive enough to measure QoL in alopecia area, and/or that the population in the trial were not representative of all patients with the condition (perhaps because those with more severe anxiety and depression were excluded). The committee had noted that in the Adelphi study, the EQ-5D was able to capture some impact of AA on health-related quality of life, but nevertheless acknowledged in the FDG (paragraph 3.18) that the EQ-5D may not capture all the benefits of treatment for this condition. The committee specifically addressed this in making a recommendation for further research at paragraph 4.2 of the FDG.
18. The panel had in mind both the need for consistency in NICE evaluations, which is why generic tools for measuring QoL are generally preferred, and also the fact that this might not be the best approach for all conditions. The panel noted that the Manual sets out the circumstances in which the EQ-5D may not be the most appropriate measure for cost-utility analysis at paragraph 4.3.10. This states “To make a case that the EQ-5D is inappropriate, provide qualitative empirical evidence on the lack of content validity for the EQ-5D, showing that key dimensions of health are missing. This should be supported by evidence that shows that EQ-5D performs poorly on tests of construct validity (that is, it does not perform as would be expected) and responsiveness in a particular patient population.” The panel agreed that the committee had not been provided with this evidence.
19. More importantly, the committee had not been provided with any disease-specific source of health-related quality of life data that could generate utilities for decision-making. The BRAVE trial did use a disease specific QoL measure, the Skindex, but this could not be used in the modelling because this instrument cannot generate utilities. The panel noted that the committee had specifically asked about whether there is a mapping algorithm to generate utilities from the Skindex-16 QoL data and the company had confirmed that there is not. Alopecia UK were not able to suggest a disease specific QoL measure that could have been used and acknowledged that further research in this area is needed. The panel therefore concluded that it was not unreasonable for the committee to base their decision-making on utilities from EQ-5D data.
20. The panel judged that the committee responded reasonably to the acknowledged limitations of the utility data from BRAVE by also considering utilities from the Adelphi study which seemed to have better face validity. They panel judged that it was reasonable for the committee to also note the limitations of this cross-sectional data, and to conclude that the real utilities were likely to lie somewhere between these two data sources.
21. The panel also considered Alopecia UK’s point that evaluations of baricitinib for other conditions had used disease-specific tools for deriving utilities. They noted that for these conditions, unlike in AA, there are disease specific QoL tools that can be used to generate utilities. Therefore, the circumstances were very different, and the committee was reasonable to take a different approach.
22. Overall, the panel fully recognised Alopecia UK’s concern that the impact of AA may not be fully captured by the EQ-5D but judged that the committee had been reasonable in its approach to managing this limitation and could only work with the data it was given.
23. The appeal panel therefore dismissed the appeal on this point.

### Appeal Ground 2.3: Not considering those people for whom severe alopecia areata can have a large negative impact on quality of life as a subgroup was unreasonable.

1. Lynn Wilks, for Alopecia UK, opened the discussion by reference to evidence from clinicians that approximately a third of patients living with severe AA experience anxiety and/or depression.
2. Emily Doogan, for Alopecia UK, spoke movingly about the psychological morbidities she had experienced as a person living with severe AA. She explained that she has now been privately treated with baricitinib for 10 months and describes her anxiety as mild and she is no longer depressed.
3. Dr Radha Todd, for NICE, explained the process by which a subgroup is identified – by recommendation of the Company and ideally at an early stage in the process alongside the initial submission. No subgroup was identified by the Company and the committee was not presented with any characteristics of a potential subgroup. The committee did, nevertheless, consider whether there might be an appropriate subgroup, and discussed this at some length. However, in the meeting the committee was unable to identify a set of characteristics for a subgroup of people with severe AA for whom baricitinib might be more effective. The committee did not therefore consider the impact on a subgroup.
4. The panel asked Dr Radha Todd whether the committee considered the subgroup identified by the evidence assessment group, described as "small but heterogenous", and whether the committee considered it important to pursue assessment of a subgroup given the previously acknowledged insensitivity of the HRQoL data.
5. Becky Pennington, for NICE, explained that although the committee did ask the Company for an identifiable subgroup, they did not provide one. Further, the Manual recommends that subgroups are not identified on the basis of utility data. Dr Radha Todd noted that the external assessment group (EAG) did not identify the subgroup because they did not have any data with which to do so.
6. Both Dr Susan Holmes and Dr Leila Asfour, for Alopecia UK, presented the panel with their clinical experience of significant psychological morbidities in those suffering severe AA. Dr Leila Asfour estimated that around 10% of severe AA patients that she sees in tertiary clinics report suicidal ideation. Dr Susan Holmes quoted the evidence assessment group's conclusion that there is a "small but heterogenous patient population that is more adversely affected in terms of HRQoL" (page 4 of of the EAG response to company draft guidance comments) and expressed that her clinical experience would dictate the opposite (that those adversely affected are the larger in number).
7. Lynn Wilks suggested that the subgroup ought to be formed of those who are affected by severe anxiety and/or depression and suicidal ideation as a psychological comorbidity of severe AA. Dr Susan Holmes suggested that as clinicians (in consultation with the British Dermatology Association and the British Hair and Nail Society) – they could identify a group of patients who would benefit most from being treated with baricitinib.
8. Janet Robertson, for NICE, explained that assessing the effect on a subgroup is important regardless of its size, but that the committee was not provided with the baseline utility for any particular subgroup – and therefore could not draw any conclusions as to HRQoL improvement of a subgroup compared to the broader severe AA population.
9. In response to a question from the panel, Becky Pennington agreed that in theory anyone with access to patient level data from the trials could look at patients with at least moderate anxiety. She confirmed that neither the committee nor the EAG had access to that level of data.
10. The appeal panel concluded as follows. At the hearing, there was consensus that there may be some patients who have greater benefit from treatment with baricitinib. The panel agreed that it would be desirable to define a subgroup of patients that might gain greater benefit from treatment. The Manual states at paragraph 4.9.3 that “the characteristics of patients in the subgroup should be clearly defined and should preferably be identified based on an expectation of differential clinical or cost effectiveness because of known, biologically plausible mechanisms, social characteristics or other clearly justified factors”. The panel accepted that the committee had considered whether such a subgroup could be identified. When the committee explored this with the Company at the committee meeting, the Company were not able to define one. There was agreement at the hearing that such a subgroup would be difficult to define. Alopecia UK were not able to clearly define a subgroup and acknowledged that further work on this would be needed. The panel considered whether patients with greater psychological impact from their condition could have been defined as a subgroup but noted that the committee had seen no evidence to show that this group would gain additional benefit from treatment. The panel concluded that no subgroup had been clearly defined, and the committee had not been presented with data to allow them to make a recommendation for a subgroup of patients.
11. The appeal panel therefore dismissed the appeal on this point.

### Appeal Ground 2.5: Having acknowledged unmet need (paragraph 3.2 of the Final Appraisal Document), it was unreasonable for the committee not to recommend baricitinib.

1. Dr Susan Holmes, for Alopecia UK, described how those with severe AA are often neglected in NHS treatment. She said there are no standard or managed pathways for treatment, there is no urgency in treatment, provision for wigs vary geographically and is often inadequate. Dr Susan Holmes also explained how in comparison to the appraisal of baricitinib for AD, as baricitinib is the first in class technology to treat AA there are no direct comparators to assist in assessing cost effectiveness. Dr Leila Asfour, for Alopecia UK, echoed this, by describing those with severe AA as being penalised in the evaluation process due to lack of authorised comparators (in comparison with relatively costly JAK inhibitors as comparators for treating psoriasis, for example).
2. Dr Radha Todd, for NICE, responded that although the committee did consider the lack of alternative treatment, the recommendation must be made through balancing both clinical and cost efficacy. Janet Robertson, for NICE, supported this view, noting that the committee often considers technologies which are the first in class – but where the technology is not cost-effective, the committee's powers are limited.
3. Jacoline Bouvy, for NICE, acknowledged Dr Susan Holmes and Dr Leila Asfour's statements above in respect of comparators, and noted that NICE frequently recommend technologies that are first in class but that on the basis of evidence and data presented of baricitinib's cost and clinical efficacy in treating severe AA the committee were unable to recommend its use despite the recognised unmet need.
4. The panel enquired how unmet need is weighed in the committee's decision making. Dr Radha Todd confirmed that it certainly forms part of the committee's discussion (as acknowledged in the FDG), but that the ultimate arbiter is the data and evidence presented to the committee. Jacoline Bouvy explained that where the committee is presented with incremental cost-effectiveness ratios (ICERs) above £20,000 per Quality Adjusted Life Year (QALY) gained, the additional considerations (including unmet need) are especially important. She noted that committees can and do recommend technologies with ICERs above £20,000 following consideration of unmet need. She explained that on this occasion, the committee was not able to make a positive recommendation, even after this was taken into account.
5. Lynn Wilks, for Alopecia UK, concluded by suggesting that the uncertainties reflected in the ICER are down to lack of a current patient pathway – and expressed a desire to work with NICE to meet the unmet need and resolve the uncertainties moving forward. Janet Robertson responded by explaining that there is a process outlined in section 8 of the Manual for monitoring guidance to make sure it is up to date and decide what action to take it is no longer valid or accurate, and for reviewing guidance if significant new evidence comes to light which could lead to a change in recommendation.
6. The appeal panel concluded as follows. The panel agreed with Alopecia UK that there is considerable unmet need for new treatments for alopecia areata. The panel noted the profound impact that alopecia areata can have on QoL, the lack of established care pathways for this condition, and the fact that baracitinib is the first licensed treatment for this indication. As acknowledged by the committee during the hearing, it can be harder for the first licensed treatment for a condition to show that it is cost-effective. However, the panel were aware that its remit was to consider whether the recommendation of the committee was reasonable in the light of the evidence submitted to NICE. The panel judged that the committee had been aware of the extent of unmet need in alopecia areata when they reached their decision. The panel accepted that the committee had considered this during their meetings, but concluded the ICERs were such that they could not recommend baricitinib even once this unmet need was taken into account. The panel did not judge that this was conclusion was unreasonable.
7. The appeal panel therefore dismissed the appeal on this point.

### Appeal Ground 2.7: Having acknowledged uncaptured treatment benefits and concluded that baricitinib is innovative (paragraph 3.18 of the final appraisal document), it was unreasonable for the committee not to consider baricitinib with managed access.

1. Lynn Wilks, for Alopecia UK, argued that it was unreasonable for the committee not to recommend baricitinib for severe AA with a managed access agreement or with restricted patient access.
2. Dr Radha Todd, for NICE explained that managed access will be appropriate in instances where there is a gap in data and there is a plausible ICER that could be improved by that data. Such a proposal is usually advanced by the Company. Janet Robertson, for NICE, agreed and elaborated that in order to recommend managed access the committee have to be confident that the data gathered from the scheme will resolve the uncertainty. To do that, the Company has to submit a managed access proposal, and the committee would assess the feasibility of that, including whether the proposal can be implemented without undue burden on the NHS and whether at the end of the managed access period specified uncertainties would be resolved. In this case, however, the ICERs presented did not suggest plausible cost effectiveness even following a period of managed access, and therefore even by resolving some uncertainty through managed access, there remained little plausible potential for cost effectiveness. Further, managed access can only be considered following receipt of a proposal from the Company and the Company did not submit any proposal for managed access in this case.
3. In response to a question from the panel, Lynn Wilks confirmed that her understanding of 'restricted patient access' was broadly the same as recommending the technology for a defined subgroup (referred to as an “optimised recommendation” in paragraph 6.4.4 of the Manual). The panel concluded that this option had been considered in detail as part of the discussion of subgroups under appeal point 2.3, and therefore refers to its conclusions above in respect of this point.
4. With regard to managed access, the appeal panel concluded as follows. The panel noted that the Manual contains detailed guidance on when a recommendation with managed access can be made at paragraph 6.4.6 – 6.4.11. The committee set out its considerations with regard to this guidance in detail during the hearing and at paragraph 3.16 of the FDG. The panel agreed that the committee was not in a position to recommend managed access when no proposal was submitted by the company (because they could not reach a conclusion on whether relevant data that might resolve uncertainty could feasibly be collected). The panel considered whether the committee should have encouraged the company to submit a managed access proposal. However, the panel accepted that the committee’s conclusion that there was no ICER that has plausible potential to be cost effective was reasonable. The panel therefore judged that it was reasonable not to seek out a managed access proposal.
5. The appeal panel therefore dismissed the appeal on this point.

##  Conclusion and effect of the appeal panel’s decision

1. The appeal panel dismissed the appeal against this appraisal on all grounds.
2. There is no possibility of further appeal against this decision of the appeal panel. However, this decision and NICE’s decision to issue the final guidance may be challenged by applying to the High Court for permission to apply for a judicial review. Any such application must be made within three months of NICE publishing the final guidance.