CONFIDENTIAL UNTIL PUBLISHED

External Assessment Group Report commissioned by the NIHR Systematic Reviews Programme on behalf of NICE

Mirikizumab for treating moderately to severely active ulcerative colitis

Produced by	**Southampton Health Technology Assessments Centre (SHTAC)**
Authors	Karen Pickett, Senior Research Fellow, SHTAC Neelam Kalita, Senior Research Fellow, SHTAC Emma Maund, Research Fellow, SHTAC Jaime Peters, Senior Research Fellow, University of Exeter Marcia Takahashi, Research Fellow, SHTAC Joanna Picot, Senior Research Fellow, SHTAC
Correspondence to	Dr Karen Pickett Southampton Health Technology Assessments Centre (SHTAC) School of Healthcare Enterprise and Innovation University of Southampton Alpha House Enterprise Road, University of Southampton Science Park Southampton SO16 7NS www.southampton.ac.uk/shtac
Date completed	20/03/2023

Source of Funding: This report was commissioned by the National Institute for Health and Care Research (NIHR) Evidence Synthesis Programme as project number NIHR135730.

Acknowledgements

We thank Professor Alan Lobo, Consultant Gastroenterologist, Inflammatory Bowel Disease Centre, Royal Hallamshire Hospital, Sheffield Teaching Hospitals NHS Foundation Trust, for providing clinical advice to the project team through commenting on a draft version of this report.

We also thank Lorna Hazell, Senior Research Assistant, Southampton Health Technology Assessments Centre (SHTAC), for providing a quality assurance review of the draft report. We additionally thank Lois Woods, Senior Research Assistant, SHTAC, for critically appraising the systematic literature review search strategies.

Declared competing interests of the authors and advisors

The authors report none. Prof Alan Lobo reports the following financial relationships with a company associated with this appraisal in the previous 12 months: receipt of consulting fees from Takeda UK and being a virtual advisory board Chair for Takeda UK in relation to vedolizumab for Crohn's disease and in relation to aspects of the management of Crohn's disease. Prof Lobo also reports contributing to a non-promotional virtual policy summit and a subsequent report, organised and funded by Takeda UK, on care for people with inflammatory bowel disease.

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.

This report should be referenced as follows:

Pickett, K; Kalita, N; Maund, E; Peters, J; Takahashi, M; Picot, J. Mirikizumab for treating moderately to severely active ulcerative colitis*: A Single Technology Appraisal.* Southampton Health Technology Assessments Centre, 2023.

Contributions of authors

Karen Pickett critically appraised the company's decision problem and network meta-analysis, drafted the report, project managed the review, and is the project guarantor; Neelam Kalita critically appraised the company's economic evaluation, and drafted the report; Emma Maund critically appraised the clinical efficacy evidence from the company's trials, and the network meta-analysis, and drafted the report; Jaime Peters critically appraised the company's network meta-analysis and drafted the report; Marcia Takahashi critically appraised the company's economic evaluation, and drafted the report; Joanna Picot critically appraised the company's background information, the decision problem, the clinical efficacy evidence from the company's trials, and the network meta-analysis, and drafted the report.

Table of Contents

LIST OF TABLES

LIST of FIGURES

LIST of APPENDICES

LIST OF ABBREVIATIONS

AE	Adverse event
BNF	British National Formulary
CRD	Centre for Reviews and Dissemination
CrIs	Credible intervals
CS	Company submission
CSR	Clinical study report
DIC	Deviance information criteria
EAG	External Assessment Group
eCOA	Electronic clinical outcome assessment
EMA	European Medicines Agency
EQ-5D-5L	European Quality of Life Working Group Health Status Measure 5 Dimensions, 5 Levels
FDA	Food and Drug Administration
IBDQ	Inflammatory Bowel Disease Questionnaire
ITT	Intention-to-treat
IV	Intravenous
JAK	Janus kinas
MIMS	Monthly Index of Medical Specialities
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NMA	Network meta-analysis
ONS	Office for National Statistics
OR	Odds ratio
PATT	Proportionate approach to technology appraisals
PAS	Patient access scheme
PSA	Probabilistic sensitivity analysis
PSS	Personal Social Services
RCT	Randomised controlled trial
SAP	Statistical analysis plan
SC	Subcutaneous
SLR	Systematic literature review
SP	Sphingosine 1-phospate
TA	Technology appraisal

TNFi	Tumour necrosis factor alpha inhibitor
TSD	Technical Support Document
UC	Ulcerative colitis

1 Executive summary

The company (Eli Lilly) submitted evidence to NICE for mirikizumab, in the treatment of people with moderately to severely active ulcerative colitis (UC), to be considered under NICE's proportionate approach to technology appraisals (PATT) streamlined cost-comparison process. This report is the external assessment group's (EAG's) critique of the company's submission (CS). It identifies the strengths and weaknesses of the CS. This summary provides a brief overview of the issues identified by the EAG as being potentially important for decision making. All issues identified represent the EAG's view, not the opinion of the National Institute for Health and Care Excellence (NICE).

During the EAG's evaluation of the CS, the company submitted an addendum to the CS to NICE, in which the company amended input errors identified in the network meta-analysis (NMA) presented in their original CS. We refer to this document as the 'CS addendum' in this report. The company also submitted a revised cost-comparison model as part of the addendum.

The company is using the PATT streamlined cost-comparison process for this appraisal as they argue a case in the company submission (CS) that mirikizumab has similar or better clinical efficacy for treating moderately to severely active UC than the company's two chosen comparators, ustekinumab and vedolizumab, in the induction and maintenance phases of UC treatment. The EAG is overall satisfied that the company's argument is supported by the evidence in the CS.

1.1 Overview of the EAG's key issues

The EAG has identified no critical issues with the evidence included in the CS that, in our opinion, would prevent a cost-comparison approach proceeding. Below, however, we detail uncertainties we identified with an aspect of the company's decision problem and with the evidence base they present.

1.2 The decision problem: summary of the EAG's critique

The company's decision problem overall appears appropriate and the EAG suggests, based on advice from our clinical expert and based on NICE committee discussions in previous appraisals, that the company's selection of vedolizumab and ustekinumab as comparators for the cost-comparison is reasonable.

The only uncertainty we have identified with the decision problem is that, from the information supplied in the CS, it is not fully clear what the company mean when they state they are partly positioning mirikizumab for managing moderately to severely active UC in biologic-naïve patients (that is, people for whom conventional treatment cannot be tolerated or is not working well enough) in whom "*other biologic treatment is not suitable*" (CS section B.1.1).

1.3 The clinical effectiveness evidence: summary of the EAG's critique

The company conducted a network meta-analysis (NMA) to provide support for their claim that mirikizumab has similar clinical efficacy to ustekinumab and vedolizumab. We judged that the methodology of the NMA was overall appropriate, but we had some concerns about the NMA. These included that:

- the searches for the systematic literature review that informed the NMA were performed over six months ago, meaning that there is a risk that there may have been relevant studies published recently that will not have been included in the NMA;
- the study eligibility criteria of the systematic literature review that informed the NMA focused on a broad population of "*adult patients (≥18 years) with moderate to severe UC*" (CS Appendix D, section D.1.3, Table 19); eligibility was not limited to studies of only adults with moderately to severely active UC who were intolerant of, or whose disease has had an inadequate response, or loss of response to previous biologic therapy or conventional therapy, as per the population of interest specified in the NICE scope. As a consequence of this, the biologic-naïve subgroup analyses in the NMA (of people "*who had not received any prior biologic, including a JAKi [Janus kinas inhibitor]*", CS section B.3.9.3.1) do not fully reflect the population of interest in the NICE scope, as the participants included in these analyses were not necessarily intolerant of, or had had an inadequate response to or loss of response to conventional therapy. The NMA biologic-naïve subgroup also does not fully reflect the biologic-naïve population in whom the company is partly positioning mirikizumab (that is, those in whom "*Conventional treatment cannot be tolerated or is not working well enough and other biologic treatment is not suitable ("biologic-naïve")*", CS section B.1.1);
- the company did not model baseline effect using representative UK-specific data as is recommended in Technical Support Document (TSD) 5[1] and the impact of this on the results is unclear;
- the company's NMA network was broad, including a range of approved targeted therapies and emerging therapies for UC. There was considerable statistical and

clinical heterogeneity in the analysis. We suggest this may have been reduced through using a narrower network, with fewer comparators included (i.e. by limiting the NMA to the treatments of interest in the cost-comparison: mirikizumab, ustekinumab, vedolizumab and placebo). Reduced heterogeneity would provide more confidence in the potential clinical efficacy equivalence of the drugs (through providing more precise credible intervals).

We also note that the similarity of the treatment effects and safety of mirikizumab versus ustekinumab and vedolizumab is based on findings of statistical significance in the NMA. Non-inferiority and equivalence have not been statistically assessed in the available evidence in the CS (e.g. through an equivalence or non-inferiority trial).

The concerns we detail above, however, are not, in our opinion, critical issues affecting the robustness of the NMA efficacy and safety results.

1.4 The cost-effectiveness evidence: summary of the EAG's critique

The company conducted a cost-comparison analysis of mirikizumab versus ustekinumab and vedoluzimab. The EAG conclusions are as follows:

- The company's cost comparison analyses considered two patient cohorts: biologic-naïve and biologic failed. The patients' characteristics, based on the pivotal mirikizumab LUCENT trials' intention-to-treat (ITT) populations, are consistent with a previous NICE appraisal (TA633;[2] ustekinumab for treating moderately to severely active UC).
- The comparators included in the analysis are appropriate and consistent with the NICE scope.
- The company's model structure and assumptions are appropriate and consistent with a previous NICE appraisal (TA633). Overall, the model was well-implemented, although we identified two errors in the company's scenario analyses.
- The model assumes equal clinical efficacy for mirikizumab, ustekinumab and vedolizumab based on the NMA results. While there are uncertainties with the NMA, none of these are critical. Hence, we view it is reasonable to assume equal clinical efficacy for all three drugs.
- At the list price, mirikizumab is [redacted] – ustekinumab and vedolizumab. This applies for the company's base case analysis and for all the company and EAG scenario analyses.

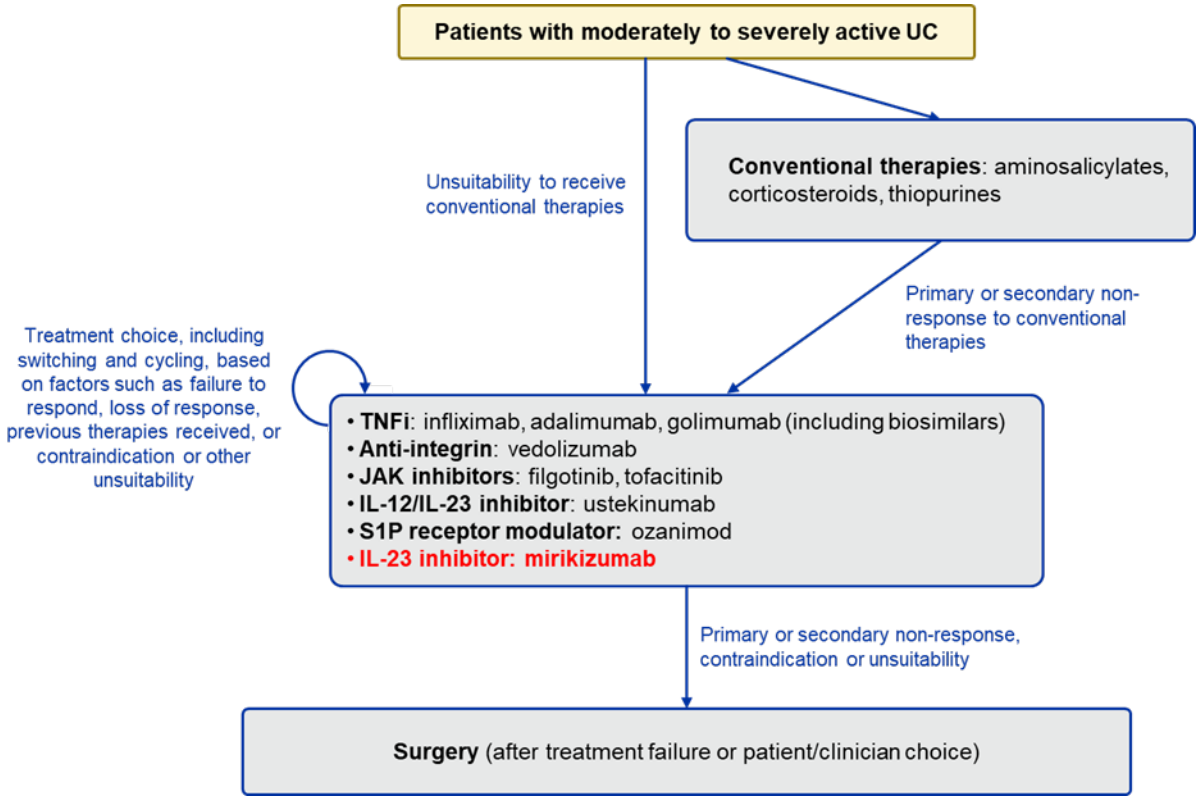
- The cost difference between mirikizumab and the two comparators is most sensitive to assumptions about re-induction rates and delayed response assessment.

2 Background

Mirikizumab for treating moderately to severely active ulcerative colitis (UC) is being considered using cost-comparison methodology as part of the recently introduced proportional approach to technology appraisals (PATT) process. This is because:

- at the time the final scope was produced NICE had already released technology appraisal (TA) guidance for similar medicines used for the same indication: TA329[3] (the TNF inhibitors infliximab, adalimumab and golimumab), TA342[4] (vedolizumab), TA547[5] (tofacitinib) and TA633 (ustekinumab).[2] TA828[6] on ozanimod for treating moderately to severely active UC, TA792[7] on filgotinib for treating moderately to severely active UC and TA856[8] on upadacitinib for treating moderately to severely active UC were released after the final scope for this appraisal.
- The CS states that mirikizumab has similar or better efficacy for treating moderately to severely active UC than the company's two chosen comparators, ustekinumab and vedolizumab, in the induction and maintenance phases of treatment (CS section B.1.1). Relative efficacy was estimated by an indirect treatment comparison that compared mirikizumab to the full range of comparators specified in the final scope.

The company provides a succinct and accurate description of the disease area in CS section B.1.3.1 covering the primary and secondary symptoms of UC, epidemiology and diagnosis, disease staging (severity and extent). Burden of disease is summarised in CS section B.1.3.2. The clinical pathway of care, focussing on patients with moderately to severely active UC, is provided in CS section B.1.3.3 and summarised in CS Figure 2 which is reproduced below as Figure 1. As Figure 1 shows, first-line treatment for suitable patients is conventional therapies (e.g. aminosalicylates, corticosteroids, thiopurines). If conventional therapies are not suitable for a patient, or when a patient has an inadequate response to or loses response to conventional therapies a variety of biological therapies form the next (second-line) treatment options. The company show the intended positioning of mirikizumab is at the same step of the pathway as the biological therapies, Janus kinas (JAK) inhibitors and sphingosine 1-phospate (SP) receptor modulator. As stated in the figure, the biologic ustekinumab (a comparator in this appraisal) is restricted for use only where a tumour necrosis factor alpha inhibitor (TNFi) has failed or cannot be tolerated. The final treatment option available either for patients unable to receive biological therapies, or for patients who experience inadequate disease control despite receipt of a biological therapy, is surgery to remove the colon.



Patients with a response or in remission remain on the same therapy with a 12-month review. In the biologic-naïve setting, ustekinumab is restricted for use only where a TNFi has failed (that is, the disease has responded inadequately or has lost response to treatment) or cannot be tolerated, and ozanimod is for use where conventional treatment cannot be tolerated or is not working well enough and infliximab is not suitable.

Figure 1 Current treatment pathway for moderately to severely active UC in UK clinical practice and the anticipated positioning of mirikizumab within it.

Source: reproduction of CS Figure 2
IL, interleukin; JAK, Janus kinase; S1P, sphingosine-1-phosphate; TNFi, tumour necrosis factor alpha inhibitors, UC: ulcerative colitis.

Mirikizumab's mechanism of action is shown in CS Figure 1 (this is within CS Table 2). Mirikizumab is a recombinant humanised IgG4 monoclonal antibody that selectively binds to the p19 subunit of the IL-23 cytokine. When mirikizumab is bound to the p19 subunit, the interaction of the IL-23 cytokine with the IL-23 receptor is inhibited, thereby reducing the inflammatory processes driven via IL-23 that contribute to the inflammatory processes underlying UC. Mirikizumab is administered by intravenous (IV) infusion during induction and thereafter by subcutaneous injection for maintenance. Mirikizumab does not yet hold a license in the UK.

As summarised in Appendix 1 mirikizumab's mechanism of action is most similar to that of ustekinumab, one of the company's chosen comparators. Ustekinumab also inhibits the inflammatory cascade underlying UC via inhibition of the IL-23 cytokine but because

ustekinumab binds to the p40 subunit, it also inhibits the IL-12 cytokine which shares this subunit (whereas mirikizumab targets the p19 subunit of the IL-23 cytokine). Additionally, mirikizumab and ustekinumab share a similar method of administration (initially IV infusion for induction, followed by subcutaneous injection for maintenance). The other eight therapies recommended by NICE and listed as potential comparators in the final scope (including the company's other chosen comparator, vedolizumab) have different mechanisms of action. Two can also be administered by IV infusion for induction, followed by subcutaneous injection for maintenance (infliximab and vedolizumab, infliximab can be administered solely by IV infusion), two are administered subcutaneously (adalimumab, golimumab) and four orally (tofacitinib, filgotinib, ozanimod, upadacitinib).

3 Critique of the decision problem in the company's submission

CS Table 1 outlines the decision problem addressed by the company in the CS in relation to the final scope issued by NICE. The table shows deviations from the scope, as highlighted by the company. Here we provide a critique of the company's deviations from the NICE scope and the company's stated reasons for these.

3.1 Population

The population addressed in the company's decision problem is "*Adults with moderately to severely active ulcerative colitis for whom conventional treatment cannot be tolerated or is not working well enough and other biologic treatment is not suitable, or biological treatment cannot be tolerated or is not working well enough*" (CS Table 1). The population specified by the company (see CS Table 1) broadly matches that specified in the NICE scope, but differs in that among people who cannot tolerate conventional treatment or in whom conventional treatment has not worked well enough, the company is positioning mirikizumab treatment only in the subgroup for whom other biologic treatments are not suitable. This population is referred to by the company as "biologic-naïve"; see CS section B.1.1. The company state this is a sub-population of the proposed marketing authorisation (see CS Table 1). In the CS, the population in whom biological treatment cannot be tolerated or is not working well enough is also addressed in the company's decision problem and is referred to by the company as "biologic-failed".

From the information supplied in the CS, the EAG is not fully clear about what the company mean when they state mirikizumab is partly positioned for managing UC in biologic-naïve patients in whom "*other biologic treatment is not suitable*" (CS Table 1). We note that none of the comparator drugs specified in the NICE scope, for which NICE recommendations have been published,[2-7] have the same restriction as proposed by the company for mirikizumab. Ustekinumab (TA 547) is more specifically recommended as an option when conventional treatment or a biologic cannot be tolerated, or the disease has not responded adequately or lost response to treatment, only if a TNFi has failed, cannot be tolerated or is unsuitable.[2] Ozanimod (TA 828) is more specifically recommended as an option when conventional treatment cannot be tolerated or is not working well and infliximab is unsuitable (as well as being recommended for as a treatment option when a biologic cannot be tolerated or is not working well enough).[6]

The clinical expert advising the EAG stated that they thought clinicians would want to have the option of using mirikizumab to treat patients who cannot tolerate either conventional or existing available biologic treatments. The expert also estimated that the proportion of patients for whom other biologic treatment would be unsuitable would be low – around 10% to 15% of patients. They commented that the criteria clinicians would use to judge unsuitability would be subjective and not clearly defined (the judgement might be based on, for example, cancer risk or patient preference).

3.2 Intervention

The intervention specified by the company in their decision problem (mirikizumab) matches the NICE scope.

3.3 Comparators

In a cost-comparison NICE appraisal, companies are not expected to provide a comparison of the intervention against all the comparators specified in the NICE scope.[9] Only one of the scoped comparators needs to be selected, which should represent NICE recommended treatments as a whole in terms of costs and effects, and which has a significant market share. In the company's decision problem for this appraisal, they have selected ustekinumab and vedolizumab as comparators, for the reasons outlined in CS Table 1 and in CS section B.1.1, which include that the company state that their NMA shows that mirikizumab has a similar or possibly greater efficacy than ustekinumab and vedolizumab. The company state ustekinumab and vedolizumab are the relevant comparators for the biologic-failed subgroup (CS section B.1.1; see section 3.1 above for how this subgroup is defined). The company does not explicitly state the relevant comparator(s) for biologic-naïve population in whom other biologic treatments are not suitable (see section 3.1 above for how this subgroup is defined).

The company does not provide an estimate in the CS of the market share for either ustekinumab or vedolizumab in treating people with moderately to severely active UC who are intolerant of, or have failed treatment with, prior biologic therapy. Clinical expert advice to the EAG is that ustekinumab and vedolizumab are used extensively in these patients. The expert notes that the treatment landscape is currently changing and would also include tofacitinib, filogotinib and (if recommended by NICE) upadacitinib. The EAG's expert estimated that the market share of vedolizumab is 40%, tofacitinib 35%, ustekinumab 20% and surgery or other treatments 5%.

We consider the company's selection of ustekinumab and vedolizumab as comparators for mirikizumab in this cost-comparison appraisal is reasonable based on Committee meeting discussions in previous NICE appraisals of treatments for moderately to severely active UC, the NICE recommended indications for these drugs in moderately to severely active UC,[4,10] and based on clinical expert advice to the EAG for this appraisal. The EAG's clinical expert noted that treatment options are changing rapidly for moderately to severely active UC. They noted that vedolizumab and ustekinumab are reasonable comparators to choose, but that tofacitinib, filgotinib and ozanimod would also be treatment options. The expert noted that tofacitinib is quite frequently used, but that its use is variable due to differing familiarity with it and some concern about side effects. The EAG's expert commented that there is considerable uncertainty about how the various treatments for moderately to severely active UC should be positioned and sequenced.

Regarding the use of TNF-alpha-inhibitors in treating moderately to severely active UC, we note that in the NICE appraisals of ustekinumab, filgotinib and ozanimod (TA633, TA792 and TA828, respectively), clinical experts informed the NICE Committees that, in practice, TNF-alpha inhibitors are typically offered as a first biologic treatment after failure on or due to intolerance of conventional therapy,[2,6,7] with infliximab commonly used at this stage.[2,6] The clinical experts advising the Committee on the ustekinumab appraisal, for which guidance was published 17 June 2020, stated that if a patient produces antibodies to a TNF-alpha inhibitor and loses response, another TNF-alpha inhibitor may be tried.[2] If the patient has produced no antibodies and the condition has not responded adequately or lost response to the first TNF-alpha inhibitor, the patient may be offered vedolizumab or tofacitinib.[2] The expert advising us in this appraisal agreed with this depiction of the use of TNF-alphas in clinical practice.

3.4 Outcomes

The company has included all the outcomes specified in the NICE scope in the CS, except for rates of and duration of relapse. The company, however, models loss of response in the cost comparison model (CS section B.4.2.1.5). The expert advising the EAG confirmed that loss of response is clinically the same as relapse. The company provide a definition of loss of response in CS Table 12. The EAG's expert was of the opinion that the definition is appropriate.

Mortality is not reported as an efficacy outcome in the CS, but is reported as an adverse effect.

The outcomes of clinical response and clinical remission were measured in the comparator vedolizumab and ustekinumab pivotal trials (GEMINI I[11] and UNIFI,[12] respectively) and were outcomes used in the cost-effectiveness economic models that informed the NICE appraisals of these drugs.[2,4] We note that the definitions of these outcomes used in the mirikizumab pivotal trials (LUCENT-1 and LUCENT-2) differ to those used in the previous appraisals. This is discussed further in section 4.5.4.3, where we note that the expert advising the EAG confirmed that the way these outcomes had been defined in the pivotal mirikizumab trials was appropriate.

3.5 Economic analysis

The company has submitted a cost comparison analysis for the reasons outlined in section 2. The company's base case analysis uses a 10-year time horizon (CS section B.4.2.2). The expert advising the EAG was of the opinion that this time horizon would be sufficient for capturing any differences in costs between mirikizumab and ustekinumab and vedolizumab. The CS details that a patient access scheme (PAS) discount has been submitted to the Patient Access Schemes Liaison Unit and provides details of the proposed discount (CS Table 2). The company provides base case and scenario analyses results using both the list and PAS prices (CS sections B.4.3 and B.4.4.2, and updated in sections 3.2 and 3.3 of the CS addendum). We note confidential commercial arrangements are in place for ustekinumab and vedolizumab.

3.6 Subgroups to be considered

Two patient subgroups are specified in the company's decision problem: 'biologic-failed' and 'biologic-naïve' (these subgroups are defined in CS Table 1). The company's definitions of these groups broadly align with those of the subgroups specified in the NICE scope, except that the company includes tofacitinib (which is a small molecule JAK inhibitor) in addition to biologics. The clinical expert advising the EAG, confirmed it is reasonable to group tofacitinib with biologics, as, collectively, these therapies are now sometimes described as 'advanced therapies'. The expert additionally noted that while grouping tofacitinib with the biologics was reasonable, there is sparse information available about whether people who fail on tofacitinib differ in an important way to those who fail on a TNF-alpha inhibitor. This is partly because tofacitinib is not often used as a first-line treatment. The expert notes that general clinical experience is that there are higher response rates in biologic naïve patients than those who have been biologic exposed, but it is unclear if the same pattern of response would be observed in people who have received tofacitinib but who have not been exposed to a biologic.

4 Summary of the EAG's critique of clinical effectiveness evidence submitted

4.1 Overview of the clinical effectiveness evidence submitted by the company

The company identified the submitted clinical effectiveness evidence by conducting a systematic literature review (SLR) and by including data on their own pivotal Phase III trials (LUCENT-1 and LUCENT 2) which were not published when searches for the SLR were conducted. The final evidence included comprises:

- LUCENT-1.[13] The company's phase III randomised controlled trial (RCT) of mirikizumab versus placebo designed to evaluate the safety and efficacy of mirikizumab over a 12-week induction period.
- LUCENT-2.[14] The company's phase III RCT of mirikizumab versus placebo designed to evaluate the safety and efficacy of mirikizumab in maintaining a treatment response to Week 40, with the primary study population comprising of LUCENT-1 participants who were randomised to mirikizumab and who achieved a clinical response at week 12.
- 35 additional studies included in the company's NMAs that compare mirikizumab with a broader range of comparators than that listed in the NICE scope for this appraisal.

The company's two pivotal studies of mirikizumab are described and critiqued in sections 4.2 to 4.4 of this report and the company's NMAs in section 4.5 below.

4.2 Description of pivotal studies of mirikizumab

CS sections B.3.2 and B.3.3 provide details of the design and methodology of the company's two pivotal mirikizumab studies, LUCENT-1 and LUCENT-2. Patients who completed the 12-week induction period of LUCENT-1 were eligible to enrol in the LUCENT-2 study, which was a 40-week maintenance study. Treatment received in LUCENT-2 was based on the patient's randomised treatment arm and clinical response in LUCENT-1 and whether they experienced loss of response in LUCENT-2. These studies are discussed individually in sections 4.2.1 and 4.2.2 below.

4.2.1 LUCENT-1

LUCENT-1 was a multi-national, phase III, randomised, double-blind placebo-controlled study evaluating the superiority of mirikizumab versus placebo in inducing clinical remission at 12 weeks in patients with moderately to severely active UC whose prior treatment with either conventional therapy or with biologic therapy had failed.

- Moderately to severely active UC was defined as a modified Mayo score of 4 to 9 out of a possible total score of 9 (i.e. a score based on three of four total Mayo subscores (Stool frequency subscore (0–3), Rectal bleeding subscore (0–3), and Endoscopic subscore (0–3) but excluding the Physician's global assessment subscore (0-3) (CS Table 12)),[15] and an endoscopic subscore of ≥2. The EAG agree with the company that the modified Mayo score has been shown to highly correlated with the full Mayo score and the exclusion of the Physician's Global Assessment subscore is in line with guidance published by the Food and Drug Administration (FDA).[16]
- Conventional-failed ("biologic-naïve") patients were defined as having had an inadequate response to, loss of response to, or intolerance to corticosteroids or immunomodulators and having never failed nor demonstrated an intolerance to a biologic medication (TNFis, anti-integrins) indicated for the treatment of UC. [redacted]
- Biologic-failed patients were defined as having had an inadequate response to, loss of response to, or intolerance to biologic (TNFis, anti-integrins) or JAK inhibitors (e.g. tofacitinib). Further details of medication failure criteria are in CS Appendix J. [redacted]

The EAG notes that the LUCENT-1 trial definition of the conventional-failed subgroup encompasses people who are biologic-naïve who have not failed on or are intolerant to a biologic. It is not clear if these people were not suitable for treatment with a biologic. This LUCENT trial subgroup therefore does not fully reflect biologic-naïve subgroup stated to be of interest in the company's decision problem ("*adult patients with moderately to severely active ulcerative colitis for whom: Conventional treatment cannot be tolerated or is not working well enough and other biologic treatment is not suitable ("biologic-naïve)"*; CS section B.1.1 and Table 1).

The study design is shown in Figure 2. The study had a screening period of up to 28 days followed by double-blind treatment for 12 weeks. Patients who completed 12 weeks of treatment were eligible to enrol in LUCENT-2. Patients who discontinued LUCENT-1 before week 12 or completed LUCENT-1 but did not enrol in LUCENT-2, completed a post-treatment follow-up period for 16 weeks after their last visit.

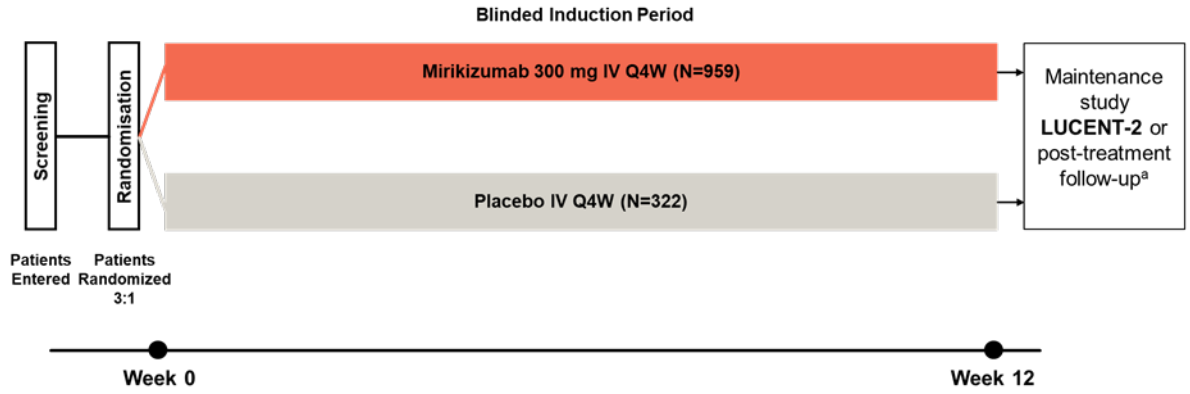


Figure 2 Trial design of LUCENT-1

[a] Patients who completed LUCENT-1 through Week 12 either completed post-treatment follow-up within the study or were eligible to participate in the maintenance study LUCENT-2.
IV: intravenous; Q4W: every 4 weeks.
Source: reproduced from CS Figure 3

In the LUCENT-1 trial, 1281 patients were randomised 3:1 to IV mirikizumab 300 mg every 4 weeks or IV placebo every 4 weeks stratified by biologic-failed status, baseline corticosteroid use, baseline disease activity (modified Mayo score of 4–6 or 7–9) and region. Patients received visually identical IV treatment by blinded personnel at weeks 0, 4 and 8 and were allowed to continue ongoing therapy with stable doses of protocol specified non-biologic treatments (CS Table 6, LUCENT-1 Trial Protocol point 9).

Eligibility criteria for LUCENT-1 are shown in CS Table 6 and CS Appendix J, with baseline characteristics shown in CS Tables 8 and 9. The company states that baseline characteristics were well-balanced across treatment groups (CS section B.3.3.2.1); the EAG agrees with this.

The primary outcome of LUCENT-1 was the proportion of patients in clinical remission at week 12 defined using the modified Mayo score (i.e. Stool frequency subscore = 0 or 1, with ≥1-point decrease from baseline, Rectal bleeding subscore = 0, Endoscopic subscore = 0 or 1 (excluding friability), CS Table 12). Major secondary outcomes are listed in CS Table 6, defined in CS Table 12, adverse reactions in CS Appendix F and additional secondary outcomes in Appendix M and the clinical study report (CSR).

Electronic clinical outcome assessment (eCOA) devices were used to record patient reported outcomes, including the Stool frequency and Rectal bleeding subscore components of the modified Mayo score. During the trial, errors in the Turkish and Polish wording of these two components on the eCOA devices were discovered (LUCENT-1 statistical analysis plan (SAP) section 4.3). One hundred and seventeen patients from Turkey and

Poland therefore had baseline data collected using incorrect questions (CSR section 3.1.2.2). As a result, and in agreement with the FDA, the primary efficacy analysis for all endpoints was based on a modified intention to treat population (LUCENT-1 SAP section 5.4). This population (n=1162, 90.7% of randomised patients) included all randomised patients who received any amount of study treatment, regardless of whether they received the correct treatment, or otherwise did not follow the protocol, but excluded those 117 patients impacted by the eCOA wording errors in Turkey and Poland (CS Table 13, LUCENT-1 SAP section 5.4). Sensitivity analyses that included impacted patients from Turkey and Poland by using methods of imputation were performed (LUCENT-1 SAP section 5.3.4); results were presented in the CSR only. In contrast, the primary analysis of adverse events was based on the safety population (n=1279) which included impacted patients from Turkey and Poland. Descriptions of trial populations used in the analysis of LUCENT-1 outcomes are presented in CS Table 13 and a summary of the statistical analyses undertaken for LUCENT-1 is provided in CS Table 15. The EAG note that to account for multiple testing a two-sided alpha of 0.00125 was used for all primary and major secondary endpoints. For all other endpoints, a significance level of 0.05 was used (LUCENT-1 SAP section 5.1.4).

4.2.2 LUCENT-2

LUCENT-2 was a multi-national, phase III, 40 week-long maintenance study comprising five treatment arms (n=1177, LUCENT 2 CSR Table 8.1). Patients in LUCENT-1 who received at least one dose of study drug and completed assessments at week 12 were eligible to enrol in LUCENT-2; eligibility criteria are detailed in CS Table 7 and CS Appendix K. The study design is shown in Figure 3.

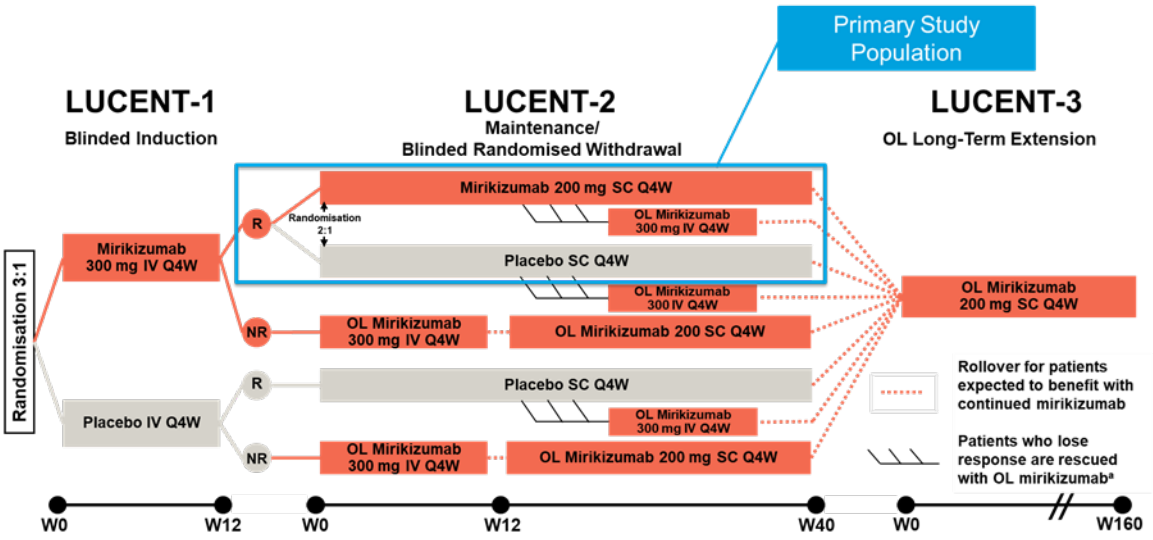


Figure 3 The trial design of LUCENT-2

[a] Patients for whom re-induction ("rescue therapy") with open-label mirikizumab was not deemed to demonstrate clinical benefit discontinued treatment and were not eligible to enter the open-label extension.
IV: intravenous; NR: non-responder OL: open-label; Q4W: every 4 weeks; R: responder; SC: subcutaneous; W: week.
Source: reproduced from CS Figure 4

The primary study population of LUCENT-2 (the two study arms within the blue box in Figure 3) were patients randomised to mirikizumab in LUCENT-1 and who achieved clinical response at week 12 of LUCENT-1, i.e. mirikizumab responders. In LUCENT-2, these patients (n= 581, LUCENT-2 CSR Table AMAG.8.1) were re-randomised (stratified by biologic-failed status, induction remission status, baseline corticosteroid use, and region) 2:1 to blinded subcutaneous mirikizumab 200mg maintenance treatment or blinded subcutaneous placebo every 4 weeks (weeks 0, 4, 8, 12, 16, 20, 24, 28, 32, and 36). Patients were allowed to continue ongoing therapy with stable doses of protocol specified non-biologic treatments (CS Table 7). If patients experienced loss of response to either mirikizumab or placebo at or after week 12 of LUCENT-2, they received open-label IV mirikizumab 300mg treatment every 4 weeks for three doses and no subcutaneous injections. Loss of response was defined as:

- ≥2-point increase in the combined stool frequency and rectal bleeding subscores (relative to LUCENT-1 baseline)
- ≥4 points combined stool frequency and rectal bleeding subscores on 2 consecutive visits
- Confirmation of negative Clostridium difficile testing (from week 8)

And

- Confirmed by a centrally read endoscopic subscore of 2 or 3 from week 12 and no later than week 28 (CS Table 12).

Patients who, in the investigator's opinion, received clinical benefit (not further defined in the CS or CSR) from IV mirikizumab were considered for a longer-term extension study (LUCENT-3) but were discontinued from LUCENT-2. Patients who did not receive clinical benefit from IV mirikizumab discontinued study treatment and went into post-treatment follow up.

The three remaining treatment arms in LUCENT-2 were not assigned by randomisation. These were:

- Patients randomised to placebo in LUCENT-1 who achieved clinical response at week 12 of LUCENT-1, i.e. placebo responders. These patients received subcutaneous blinded placebo every 4 weeks in LUCENT-2. Loss of response and subsequent procedures were the same as those defined for those patients in the primary study population.
- Patients randomised to mirikizumab in LUCENT-1 who did not achieve clinical response at week 12 of LUCENT 1, i.e. mirikizumab non-responders. These patients received open-label extended induction therapy, i.e. IV mirikizumab 300 mg every at Weeks 0, 4 and 8 of LUCENT-2. At Week 12, these patients were assessed for clinical response, i.e. delayed clinical response. Patients who achieved delayed clinical response, as compared with LUCENT-1 baseline, received open-label subcutaneous mirikizumab 200 mg every four weeks from Week 12. Patients who did not achieve delayed clinical response discontinued the study.
- Patients randomised to placebo in LUCENT-1 who did not achieve clinical response at week 12 of LUCENT-1, i.e. placebo non-responders. These followed the same procedures in LUCENT-2 as for mirikizumab non-responders, described above.

The primary outcome of LUCENT-2 was the proportion of patients in the primary study population who achieved clinical remission at week 40, using the modified Mayo score. Major secondary outcomes are listed in CS Table 7 with additional secondary outcomes detailed in CS Appendix N and in the CSR.

Inferential statistics were only carried out for the primary study population (CS section B.3.4.1). As in LUCENT-1, due to the issue with eCOA devices described in section 4.2.1, primary efficacy analyses were based on the modified intention-to-treat population and

included patients who were deemed as mirikizumab induction responders (n=544). Safety analyses were performed for this “mirikizumab induction responders” subset of the overall safety population (n=581).” Baseline characteristics, shown in CS Tables 10 and 11, were balanced between the two arms of the primary study population. A summary of the statistical analyses undertaken for LUCENT-2 is provided in CS Table 15. A statistical significance level of 0.05 was used for all primary and major secondary endpoints.

4.3 Key results from pivotal studies of mirikizumab

Key results for LUCENT-1 and LUCENT-2 are presented individually in sections 4.3.1 and 4.3.2 below. Caution is required in the interpretation of subgroup results given that neither trial was powered to demonstrate statistically significant treatment differences according to subgroups (LUCENT-1 CSR section 5.1.2 and LUCENT-2 CSR section 5.1.2). Although we note that the LUCENT-2 trial protocol states an expected 80% power to assess clinical remission among biologic-failed participants who were induction remitters (LUCENT-2 protocol, page 15).

4.3.1 LUCENT-1 trial results

4.3.1.1 Primary outcome - Proportion of patients in clinical remission at week 12

A statistically significant greater percentage of patients achieved clinical remission at week 12 (defined using the modified Mayo score), in the mirikizumab group compared to the placebo group (24.2% versus 13.3%, p=0.00006). A statistically significant difference in favour of mirikizumab versus placebo was also seen in the biologic-naïve subgroup (30.9% versus 15.8%, p= <0.001) but not in the biologic-failed subgroup (15.2% versus 8.5%, p=0.065; CS.B.3.6.1.1).

4.3.1.2 Key secondary outcomes

Results using the alternative definition of clinical remission at week 12 were consistent with those of the primary outcome (CS B.3.6.1.2).

For the following efficacy outcomes there was [redacted] for the whole trial population, and for both the biologic-naïve and biologic-failed subgroups:

- Clinical response at week 12 (CS B.3.6.1.3)
- Endoscopic remission at week 12 (CS B.3.6.1.4)
- Symptomatic remission at week 12 (CS B.3.6.1.5)
- Bowel urgency numeric rating scale improvement at week 12 (CS B.3.6.1.6)

- Histologic-endoscopic mucosal improvement at week 12 (CS B.3.6.1.7)

The health-related quality of life outcomes of the European Quality of Life Working Group Health Status Measure 5 Dimensions, 5 Levels (EQ-5D-5L) total score at week 12 (CS Appendix M.1) and the Inflammatory Bowel Disease Questionnaire (IBDQ) total score change from baseline at week 12 (CS Appendix M.2) were reported for the whole trial population only. Both were statistically significantly in favour of mirikizumab versus placebo.

Data on adverse events in LUCENT-1 were presented in CS Appendix F and in the CSR. Overall, the safety and tolerability of mirikizumab appeared similar to or better than placebo:

- Treatment-emergent adverse events were similar between the two treatment groups (44.5% in the mirikizumab group versus 46.1% in the placebo group). However, the proportion of severe adverse events was approximately three times greater in the placebo group compared to mirikizumab (7.2% versus 2.2%)
- There were no deaths in the 12-week induction period of LUCENT-1. However, two patients randomised to mirikizumab died during the 16 week follow up period. Both deaths (sudden cardiac death and disseminated intravascular coagulation) were considered unrelated to study drug or protocol procedures (LUCENT-1 CSR section 5.2.3).
- Serious adverse events in the placebo group were nearly double that of the mirikizumab group (5.3% versus 2.8%). Ulcerative colitis, pneumonia and cytomegalovirus colitis were the only serious adverse events to occur in more than one patient.
- The proportion of patients discontinuing due to adverse events was over four times greater in the placebo group compared to the mirikizumab group (7.2% versus 1.6%). The most common adverse event leading to discontinuation in both groups was ulcerative colitis (5.9% in the placebo group versus 0.5% in the mirikizumab group), the second most common adverse event leading to discontinuation was infusion-related hypersensitivity reaction in the mirikizumab group (0.3% versus none in the placebo group).

4.3.2 LUCENT-2 trial results

4.3.2.1 Primary population study

The following results relate to the primary study population only of LUCENT-2; that is, patients who were mirikizumab responders at week 12 of LUCENT-1 and were subsequently re-randomised to mirikizumab or placebo in LUCENT-2.

4.3.2.1.1 Primary outcome – clinical remission

A statistically significant greater percentage of patients achieved clinical remission at week 40 (defined using the modified Mayo score), in the mirikizumab group compared to the placebo group (49.9% versus 25.1% of patients, p<0.001). [redacted] was also seen in both the biologic-naïve subgroup (51.5% versus 30.7% of patients, p<0.001) and in the biologic-failed subgroup (46.1% versus 15.6% of patients, p<0.001; CS B.6.2.1).

4.3.2.1.2 Key secondary outcomes

Results using the alternative definition of clinical remission at week 12 were consistent with the primary outcome (CS B.3.6.2.2). [redacted] in both the primary study population and in the biologic-failed subgroup only (CS B.3.6.2.3).

For the following efficacy outcomes there was [redacted] for the primary study population, and for both the biologic-naïve and biologic-failed subgroups:

- Endoscopic remission at Week 40 (CS B.3.6.2.4)
- Corticosteroid-free remission without surgery at Week 40 (CS B.3.6.2.5)
- Histologic-endoscopic mucosal remission rates at Week 40 (CS B.3.6.2.6)
- Bowel urgency numeric rating scale improvement at Week 40 (CS B.3.6.2.7)
- Bowel urgency remission at Week 40 among clinical responders with urgency numeric rating scale ≥3 at induction baseline (CS B.3.6.2.8)

Data for symptomatic remission were reported for the primary study population only (CS Appendix N.3). There were statistically significant differences in favour of mirikizumab versus placebo for symptomatic remission rates at week 40 and stable maintenance of symptomatic remission at Week 40.

During the 40-week randomised phase of LUCENT-2 (CS Appendix N.6):

- [redacted] patients in the placebo group and [redacted] patients in the mirikizumab group had UC-related hospitalisation.
- [redacted] underwent UC-related surgery

The health-related quality of life outcomes of EQ-5D-5L total score at week 40 (CS Appendix N.1) and IBDQ total score change from baseline at week 40 (CS Appendix N.2) were reported for the primary study population only. Both were statistically significantly in favour of mirikizumab versus placebo.

Data on adverse events in LUCENT-2 were presented in CS section B.3.10 and in the CSR. Overall, the safety and tolerability of mirikizumab appeared similar or better than placebo:

- The proportion of patients who experienced a treatment-emergent adverse event (TEAE) was similar between the two treatment groups as was the proportion who experienced severe adverse events (CS Table 31). Nasopharyngitis was the most frequently reported TEAE in the mirikizumab group (7.2% compared with 5.7% in the placebo group), while ulcerative colitis was the most frequent event in the placebo group (20.8% versus 6.7% in the mirikizumab group).
- There was one death, in the placebo group, during LUCENT-2.
- Serious adverse events in the placebo group were more than double that of the mirikizumab group (7.8% versus 3.3%), with ulcerative colitis the most frequent event in the placebo group (3.1% versus 0% in the mirikizumab group). (CS Table 33)
- The proportion of patients discontinuing due to adverse events was over five times greater in the placebo group compared to the mirikizumab group (8.3% versus 1.5%, respectively), with ulcerative colitis the most frequent event in both groups.

However, the EAG note that in the mirikizumab group four patients experienced depression and one patient experienced “depression suicidal”, which were adverse events of special interest. No patients in the placebo arm experienced such events (CS Tables 33 and 35). Our clinical expert noted that depression is more frequent in people with IBD and is probably associated with disease activity. They were unaware of depression as an adverse event of other treatments of UC, therefore the occurrence of these events in the mirikizumab arm only of LUCENT-2 were of potential concern.

4.3.2.2 Placebo or mirikizumab non-responders in LUCENT-1

In patients who were placebo or mirikizumab non-responders in LUCENT-1 and subsequently received three initial doses of 300 mg, open-label IV mirikizumab therapy in LUCENT-2 (CS B.3.6.2.10):

- [redacted] of patients previously treated with placebo in LUCENT-1 achieved clinical remission versus [redacted] previously treated with mirikizumab
- [redacted] of patients treated with placebo in LUCENT-1 achieved a clinical response versus [redacted] of patients previously treated with mirikizumab

- [redacted]% of patients treated with placebo in LUCENT-1 achieved endoscopic remission versus [redacted]% of patients previously treated with mirikizumab.

4.4 Critique of the company's risk of bias assessment of the pivotal studies of mirikizumab

The company assessed the LUCENT studies for risk of bias with results reported in Appendix D.3.5 of the CS. The EAG agree with the company's assessment and is not concerned with the risk of bias of either study. (The EAG's full risk of bias assessment is available in Appendix 2.)

4.5 Critique of the network meta-analyses (NMAs) submitted by the company

The company carried out NMAs to compare the efficacy and safety of mirikizumab with a wide range of approved targeted therapies for UC, including ustekinumab and vedolizumab, as well as emerging therapies (see CS Appendix D, Table 19). They carried out the NMA due to an absence of RCTs directly comparing mirikizumab with comparators (CS section B.3.9). The company stated they conducted a wide NMA, comparing mirikizumab with comparators other than just ustekinumab and vedolizumab, for "completeness" (CS section B.3.9); the EAG has found no other justification in the CS for the wide network. The EAG suggests that such a broad network may introduce greater heterogeneity.

The outcomes of main interest in the NMA were clinical response and remission (both in the induction and maintenance phases of treatment), for the reasons described in CS section B.3.9.2.4. The NMA additionally focused on mucosal healing (also both during the induction and maintenance periods) for the reasons outlined in CS section B.3.9.2.4. The safety outcomes of all cause discontinuation during induction and serious adverse events during the induction phase only were also analysed; see CS section B.3.9.2.4 for the company's reason for only analysing AEs in the induction period.

Separate clinical efficacy analyses were conducted in the NMA for the biologic-naïve and biologic-failed subgroups. In the NMA, the biologic-naïve group was defined as "*patients who had not received any prior biologic, including a JAKi*" (CS section B.3.9.3.1). The biologic-failed group was defined as "*patients who had failed previous biologic therapy, including with a JAKi*" (CS section B.3.9.3.1). This is in line with definition of the biologic-failed subgroup used in the LUCENT trials and in line with where the company is partly positioning mirikizumab treatment in their decision problem (see CS section B.1.1). The NMA subgroup

definitions also broadly match the subgroups specified to be of interest in the NICE scope. Safety analyses were conducted for the total trial populations.

4.5.1 How the NMA results are used in the company's cost-comparison model

The company used the following efficacy parameters derived from the NMA results in their cost-comparison model (see CS sections B.4.2.1.4 and sections B.4.2.1.5 and section 5.3.1 of this report):

- the distributions of the response status (response, including remission) at the end of the induction period, and,
- loss of response estimates, calculated from the NMA maintenance phase clinical response results, to model treatment discontinuation during maintenance treatment.

4.5.2 Identification and selection of studies included in the NMA

A systematic literature review was carried out to identify relevant RCTs to include in the NMA (CS B.3.9.1). The methodology of the review is detailed in CS Appendix D. Reflecting the broad scope of the review, the study eligibility criteria were wide (CS Appendix D, Table 19) and included a range of approved targeted therapies (including all eight comparators listed in the NICE scope) and emerging therapies for UC, which could be either the intervention or comparator drugs in the screened studies. These drugs could either be used alone or in combination with conventional drugs (as shown in the company's inclusion criteria in: CS Appendix D, Table 19; Table 3 in the NMA report appendices accompanying the CS;[17] and, CS Addendum, Appendix 1.3, Table 3). Clinical expert advice to the EAG is that the use of concomitant medications in clinical practice depends on the drug. Patients might receive a steroid alongside vedolizumab and ustekinumab until a drug effect is observed. Adalimumab and infliximab are usually used in combination with thiopurine/methotrexate. Tofacinib, ozanimod and filgotinib tend to be used alone.

The stated population in the study eligibility criteria for the NMA was "*adult patients (≥18 years) with moderate to severe UC*" (CS Appendix D, Table 19). The population was not limited to those who were intolerant of, or whose disease has had an inadequate response, or loss of response to previous biologic or conventional therapy, as specified by the NICE scope. Therefore, the NMA population does not fully reflect the population of interest in this appraisal (the implications of this are discussed in our summary of our critique of the company's NMA presented in section 4.5.6 below). The company do not explain why the inclusion criteria population differs to the population specified in the NICE scope. The company state that separate clinical efficacy analyses were conducted in the NMA for the

biologic-naïve and biologic-failed subgroups (CS section B.3.9; also see section 4.5.4.2 below).

Overall, the EAG has no other concerns with how the systematic literature review was carried out, but we note that the review searches were last updated in June 2022 (CS section B.3.9.1). This means that there is a risk that there may be recently published, relevant studies available that have not been included.

4.5.3 Studies included in the NMAs and the company's feasibility assessment of the studies

A total of 71 RCTs were included in the company's systematic literature review, including the mirikizumab phase III LUCENT trials (see CS section B.3.91 and CS Appendix D, section D.1.4.1 for details). The company included the 71 RCTs in an NMA feasibility assessment before the NMAs were conducted, to assess if any important heterogeneity in the study populations, interventions, outcomes and methodology was present (CS section B.3.9.2). At this stage, the company included only studies that used European Medicines Agency (EMA) and FDA approved dosing regimens in the NMA (CS section B.3.9.2.3) (thus effectively applying another inclusion criteria to the review by excluding studies that did not use the approved regimens). We note that EMA and FDA approved doses for ustekinumab and vedolizumab are the same for induction treatment, but there are some differences in the recommended dosing regimens for these two drugs in the maintenance treatment period, as shown in CS Appendix D, Table 27, and as highlighted in bold in Table 1 here. We note, however, that the FDA approved maintenance doses match part of the maintenance doses clinicians can opt to use as outlined by the EMA (see Table 1), so it appears to be appropriate to include data from studies using the FDA approved doses. Different dosing regimens of the same drug were used as separate comparators in the NMA (CS section B.3.9.2.3 and CS Appendix D, Table 33) and this also appears appropriate.

Table 1 Comparison of EMA and FDA approved dosing regimens for ustekinumab and vedolizumab

Drug	EMA approved dose and regimen		FDA approved dose and regimen	
	Induction	Maintenance	Induction	Maintenance
Ustekinumab	Approx. 6mg/ kg (260 mg (IV) or 390 mg (IV) or 520 mg (IV) based on weight, single dose	**90 mg (SC) Q12W** (or Q8W if needed)	260 mg (IV) or 390 mg (IV) or 520 mg (IV) based on weight single dose	90 mg (SC) Q8W from week 8

Vedolizumab	300 mg (IV) week 0, 2 and 6	300 mg (IV) Q8W **(or Q4W if needed)** **108 mg (SC) Q2W**	300 mg (IV) week 0, 2 and 6	300 mg (IV) Q8W

Source: this is a shorted, reproduced version of CS Appendix D, section D.1.6.1, Table 27. Bold text shows where the EMA approved maintenance dosing regimen differs to that specified by the FDA. EMA, European Medicines Agency; FDA, Food and Drug Administration; IV, intravenous; Q2W, once every two weeks; Q4W, once every four weeks; Q8W, once every eight weeks; Q12W, once every 12 weeks; SC, subcutaneous.

After the feasibility assessment, 34 studies were excluded from the NMA (CS section B.9.2.5) and the exclusions appear appropriate based on the reasons supplied by the company (CS Appendix D, Table 26). Of the 71 originally identified studies, 28 assessed an EMA or FDA approved UC treatment in the induction period and 21 assessed an EMA or FDA approved treatment in the maintenance period (CS section B.3.9.2.2). When the LUCENT trials were added to these numbers (mirikizumab is currently undergoing regulatory consideration; CS Table 2), along with the included PUSUIT SC study being split into two separate studies, there were 30 induction and 22 maintenance studies considered for the NMAs.

4.5.4 Clinical heterogeneity assessment

As with the NMA conducted for the ustekinumab NICE appraisal (TA633),[18] we and the company have identified a number of sources of potential heterogeneity across the studies included in the NMA, as we detail below (sections 4.5.4.1 to 4.5.4.4). The company discusses heterogeneity in CS section B.3.9.2 and CS Appendix D, section D.1.6.1.

4.5.4.1 Treat-through and re-randomised responder trials

As detailed in CS Appendix D, section D.1.6.1, the studies included in the maintenance treatment phase NMAs were of either a ‘treat-through’ or ‘re-randomised responder’ design. The differences between these two types of trial designs are described in CS Appendix D, section D.1.6.1, and so are not repeated here for brevity. Nine of the maintenance studies were of a treat-through design, while 13 were re-randomised studies. As pointed out in CS Appendix D, section D.1.6.1, participants entering the maintenance phases therefore differ from each other in each of these trial designs in terms of their exposure to the study drug. Those who have received active treatment during induction who are re-randomised to placebo may show a better response during maintenance than those who have remained on placebo in the treat-through trials. To account for this source of heterogeneity (e.g. in patients’ potential level of response to treatment), statistical adjustments were carried out to make the populations more comparable (CS Appendix D, section D.1.6.1) – see section

4.5.5.1 below for the EAG's explanation and critique of this. The company carried out a sensitivity analysis of clinical response and remission in the maintenance phase in which studies with a treat-through design were excluded (CS Appendix D.1.6.3).

4.5.4.2 Subgroup definitions

There was some heterogeneity between studies in how the groups of patients from which the company used data to inform their 'biologic-naïve' and 'biologic-failed' subgroup analyses in the NMA were defined (see CS Addendum, Appendix 1.5, Table 8). The EAG, however, has no concerns about this.

4.5.4.3 Outcome definitions

There was heterogeneity across the studies included in the NMA in how clinical response and remission were defined, as we outline below. It should be noted that the outcome of response encompasses patients in clinical remission.

4.5.4.3.1 Clinical response in the induction and maintenance phases

Five different definitions of clinical response in the induction and maintenance phases were used across the studies included in the NMA, where definitions were reported (see CS Appendix D, section D.1.6.1, Tables 28 and 29). We note that 22 studies in the NMA used the same definition as used in the GEMINI I and UNIFI pivotal trials of vedolizumab and ustekinumab, respectively (see Table 2 below for definitions) in the induction phase and 10 studies used this definition in the maintenance phase. The definition in the LUCENT trials in the maintenance and induction phases differs to this, as is also shown in Table 2 and as is detailed in CS Appendix D, section D.1.6.1. The LUCENT-1 trial is the only study included in the NMA that uses this definition in the induction phase NMA and the LUCENT-2 trial is one of only two studies that uses this definition in the maintenance phase NMA (as assumed by the EAG from information provided in CS Appendix D, section D.1.6.1, Tables 28 and 29). The clinical expert advising the EAG confirmed the definition of clinical response used in the LUCENT trials is not used in clinical practice *per se*, but is appropriate and reflects FDA guidance. The expert also felt the differences in the definitions used by the GEMINI I and UNIFI trials (and thus the majority of the other studies in the NMA) and the LUCENT trials were unlikely to be important, as the differing elements would make little difference to whether or not a patient would be classed as having responded or not.

4.5.4.3.2 Clinical remission in the induction and maintenance phases

Similarly to the discussion above about the definition of clinical response, the majority of the studies included in the NMA used the same definition of clinical remission in the induction (n = 17) and maintenance (n = 15) phases as used in the ustekinumab and vedolizumab pivotal trials (see Table 2 below for the definition used in these studies, and see CS Appendix D, section D.1.6.1, Tables 30 and 31, for the definitions used in the studies included in the NMA). The LUCENT trials, however, used a different definition, and so did the remaining NMA studies (where the definition was reported). Again, the clinical expert advising the EAG confirmed the definition used in the LUCENT trials does not reflect clinical practice as such, but is appropriate and in line with FDA guidance, and that missing elements from the definition would not impact on whether or not patients would be classed as being in clinical remission.

Table 2 Definitions of clinical response and clinical remission used in the mirikizumab, vedolizumab and ustekinumab pivotal clinical trials

Trials (intervention)	Definition of clinical response	Definition of clinical remission
LUCENT-1 and -2 trials (mirikizumab)	• ≥2-point and ≥30% decrease in the modified Mayo score from baseline • Rectal bleeding subscore = 0 or 1, or ≥1 point decrease from baseline	Definition 1: • Stool frequency subscore = 0 or 1, with ≥1-point decrease from baseline • Rectal bleeding subscore = 0 • Endoscopic subscore = 0 or 1 (excluding friability) Definition 2: RBS of 0, stool frequency score ≤1 and decrease from baseline ≥1, and endoscopy subscore ≤1 (excluding friability)
GEMINI I (vedolizumab) and UNIFI (ustekinumab) trials	Reduction in complete Mayo score of ≥ 3 points and ≥ 30% from baseline (Week 0) with an accompanying decrease in rectal bleeding subscore of ≥ 1 point or absolute rectal bleeding subscore of ≤ 1 point	Complete Mayo score of ≤ 2 points and no individual subscore > 1 point.

Source: the LUCENT-1 and -2 trials' outcome definitions are reproduced from CS Table 12. The GEMINI and UNIFI trials' outcome definitions were sourced from the company submissions to NICE in the associated NICE appraisals.[2,4]

4.5.4.4 Other sources of heterogeneity

In Table 3 below, we outline some of the other potential sources of heterogeneity in the company's NMAs. In addition to these, we note, as was highlighted by the EAG in the filgotinib NICE appraisal (TA792) and as discussed at the NICE Committee meeting for that appraisal,[7] that due to including trial designs in the NMA in which participants have been re-randomised, there is heterogeneity in the maintenance networks in the treatments patients in the common comparator placebo arms received during induction and their response to those treatments. For example, those who were re-randomised to placebo after responding to mirikizumab or other comparator drugs during the induction phase will be included in the placebo comparator of the NMA. There is therefore heterogeneity between the participants based on how they responded to treatment in the induction phase.

4.5.4.5 Risk of bias assessment for studies included in the NMA

The company assessed the risk of bias associated with studies included in the NMA using the Centre for Reviews and Dissemination (CRD)'s[19] quality assessment checklist for RCTs and presents their judgements on each of the CRD checklist domains in CS Appendix D, section D.3, Table 46. The company's critical appraisal of the LUCENT trials is available in CS section B.3.5, and the EAG and the company's assessments are summarised in section 4.4 of this report. The company does not provide an overall conclusion about the risk of bias associated with the NMA studies. Based on the company's judgements, the EAG notes the studies were generally rated to be of a low risk of bias across most of the risk of bias domains assessed, but with most of the studies having one or more unclear or high risk of bias judgements on some of the domains.

Table 3 Other potential sources of heterogeneity in the company's NMA

Study aspect	Heterogeneity across studies	EAG comments
Induction timepoint of assessment	Varied from 6 to 14 weeks (CS Appendix D, section D.1.5)	The EAG suggests that studies with a shorter assessment timepoint in the induction period may be at risk of not identifying patient clinical response or remission that may have occurred at later timepoints.
Maintenance phase assessment timepoint	Ranged from 30 to 60 weeks. To address this the company restricted inclusion of studies in the maintenance NMAs to those with assessment points between 52 and 60 weeks (CS Appendix D, section D.1.6.1).	The EAG considers this reasonable. The EAG report for the ustekinumab NICE appraisal[18] notes that inclusion of studies with a shorter maintenance assessment timepoint may bias results in favour of the treatment (e.g. there may be less loss of response than if the outcome had been measured at a later timepoint). [redacted]
Baseline risk adjustment (i.e. placebo response rate)	Please see discussion in CS Appendix D, section D.1.6.2 about this. The Company addressed potential heterogeneity through carrying out baseline risk adjustment NMAs, using an exploratory analysis utilising meta-regression to adjust for baseline risk. The results of the adjusted and unadjusted NMAs were compared and the adjusted or unadjusted results were chosen for use in the CS based on goodness-of-fit statistics and covariate coefficient statistics (CS Appendix D, section D.1.6.2).	As we critique further in section 4.5.5.2 below, the company has used placebo-arm data from all included RCTs and has not used representative UK-specific data as is recommended in the NICE Decision Support Unit (DSU) Technical Support Document 5.[1]

Source: The information in this table was synthesised from CS Appendix D, sections D.1.5 D.1.6.1 and D.1.6.2 by the EAG. CS, company submission; EAG, External Assessment Group; NMA, network meta-analysis; UC, ulcerative colitis.

4.5.5 Critique of the NMA modelling approach and statistical procedures

4.5.5.1 Data inputs to the NMA

The company report the data inputs from RCTs included in the NMA for each outcome analyses (CS Appendix sections D.1.10.1, D.1.10.2 and D.1.10.3). As with similar TAs in UC, relevant trials include treat-through RCTs and re-randomised RCTs (as described above in section 4.5.4.1). This difference in study design only impacts on the analysis of outcomes in the maintenance phase. To deal with these differences, the company have taken a similar approach to that reported in previous TAs, in particular TA792[7] (filgotinib). Raw data are calculated for the treat-through RCTs to reflect the results that would have been seen had these been re-randomised RCTs (CS Appendix D.1.7.7). The company assume that 1) the total number of responders at the end of the induction phase in the treat-through RCTs is a proxy for the total number of patients entering the maintenance phase, 2) the number of patients with a durable or sustained response at maintenance from the treat-through RCTs can be used to estimate the number of patients with a response at the end of the maintenance phase, and 3) the proportion of patients in remission at the end of the maintenance phase in the treat-through RCTs is a proxy for the number of those with a response in remission. Where such data are not reported in the relevant treat-through RCTs, the company make assumptions to enable estimation. As LUCENT-1 and LUCENT-2 are re-randomised RCTs, no adjustments are made to their results. The four RCTs affected are ACT1 for infliximab, Suzuki 2014 and ULTRA 2 for adalimumab, and VARSITY which compared adalimumab with vedolizumab (CS Appendix D Table 37 and Table 42). There are two points to note in the company's calculations. The first relates to the number of remitters in the placebo arm of ACT1 for the biologic-naïve population. The company use a weighted average of the percentage of responders who were remitters across all placebo arms of the re-randomised trial, which is appropriate in the circumstances. However, we could not replicate the weighted average in the original CS (57.7%; CS Appendix D, Table 37) unless we assumed that 100% of responders were remitters in TRUE NORTH and VISIBLE1, and then a weighted average of 57.8% was obtained. For TRUE NORTH and VISIBLE1, the number of responders without remission in the placebo arm is not reported, and the company do not state how they dealt with this when estimating the weighted average for ACT1, e.g. whether or not they assumed 100% were remitters or excluded these RCTs from the calculations. In the CS Addendum the number of responders without remission is reported for TRUE NORTH (CS Addendum Table 16) and the weighted average reported, and applied, is 78.7% (CS Addendum Table 17), yet the EAG could not replicate this figure regardless of whether it is assumed that all responders were remitters in VISIBLE1 or data from VISIBLE1 are excluded from calculations. The second point relates to

the placebo arm of ULTRA2. For the biologic-naïve subgroup there is a difference in the raw data calculated by the company compared to that reported by Lu et al 2022[20] (a publication based on TA792). However, the company have followed their own described approach, which the EAG agree with. As stated in section 4.5.4.1 above, for clinical response and remission in the maintenance phase, the company undertook sensitivity analyses which excluded all treat-through study designs.

4.5.5.2 Statistical methods for the NMA

The company used a Bayesian framework, implemented in Stan,[21] for their NMAs (CS Appendix section D.1.7.2). The statistical models chosen followed recommendations made in NICE DSU TSD 2:[22] a multinomial model with probit link function to estimate clinical response and remission (accounting for correlation between these outcomes); and a binomial model with logit link to estimate mucosal healing (CS Appendix section D.1.7.4). The company undertook fixed and random effects modelling, assessed the impact of assuming different prior distributions on the between-trial heterogeneity parameter in the random effects models (CS Appendix section D.1.7.3), and explored the use of meta-regression to adjust for different levels of baseline risk across studies, as recommended in TSD 3[23] (CS Appendix section D.1.7.4). The statistical models chosen for the different outcomes were appropriate, and addressed limitations noted in previous TAs on this topic.

To model the baseline effect, the company incorporated placebo-arm data from all included RCTs rather than using representative UK-specific data as is recommended in TSD 5.[1] In related TAs, reporting of the data used to inform the baseline effects does not appear to be stated explicitly (e.g. TA792 and TA633). This could suggest that the same approach was taken, as there is no statement of other UK-relevant data being used instead. In TA828[6] (ozanimod), the company used placebo-arm data from all RCTs, and the EAG conducted an additional analysis limiting the placebo-arm data to RCTs that were deemed to be more generalisable to the UK. The EAG reported that this led to lower response rates observed in the placebo arms, and in many of the active treatment arms. It is not clear how the results for mirikizumab would change had the baseline effects been more representative of the UK. We therefore highlight this as an additional source of uncertainty in the NMA results.

Methods reported by the company for assessing model convergence (CS Appendix section D.1.7.6) are appropriate. Homogeneity was assessed by noting the value of tau (as recommended in TSD3[23]), and where there were closed loops in the network, consistency was assessed and reported (as recommended in TSD 4[24]), CS Appendix section D.1.7.8.

The company summarise the posterior distributions from the NMA using the mean and 95% credible intervals (CS section B.3.9.4, CS Appendix sections D.1.10.1, D.1.10.2 and D.1.10.3). When the posterior distribution is asymmetric, reporting the median is preferred. It is unclear whether the posterior distributions from the company NMAs are asymmetric, so whether different estimates would be seen had the medians been reported instead of the means. Given that the credible intervals would remain the same, and treatment rankings, which are reported for the different outcomes and population subgroups, also contribute to an assessment of whether mirikizumab can be considered to have similar, or greater, effectiveness than ustekinumab and vedolizumab, it is unlikely that reporting of posterior medians would have led to different conclusions.

4.5.5.3 Choice between NMA models

The company conduct fixed effects and random effects models with and without adjustment for baseline risk (CS Appendix sections D.1.7.3 and D.1.7.4). To help choose between fixed or random effects models for each outcome and population subgroup (biologic-naïve or biologic-failure), the company report using goodness-of-fit statistics, in particular the deviance information criteria (DIC), and also refer to the magnitude of heterogeneity within the network. In deciding whether the base case model should include adjustment for baseline risk or not, the company consider goodness of fit and evidence on whether differences in baseline risk are observed. Thus, the base case models are not the same across each outcome and population subgroup.

There is some inconsistency in justification of whether a fixed effects or random effects model is the most appropriate. For instance, for induction of clinical response and remission in a biologic-naïve population and for serious adverse events in induction, the DIC is lowest for the fixed effects model (indicating a better fitting model), however a random effects model is preferred by the company due to the heterogeneity observed across the network. In other analyses (sensitivity analyses for maintenance of clinical response and remission in the biologic-naïve population and all cause discontinuation, although the DIC indicates the random effects model would be preferred, and there is evidence of a great deal of heterogeneity across the network, a fixed effects model is chosen by the company. The company justify the choice of fixed effects over random effects models for all cause discontinuation on the basis of "*parsimony and the uncertain estimates provided by the random effects model*" (CS Addendum Section 5.3.3.1). Although not explicitly stated by the company, it is assumed that their argument follows that these very wide credible intervals

lead to NMA results that have limited usefulness in determining the comparative effectiveness of treatments. Given limitations in available data when a network is sparse (as in these cases), use of vague prior distributions can lead to estimates of heterogeneity that are unrealistically high (TSD3[23]). The use of more informative prior distributions for the between-trial parameters has been recommended, however the EAG believes that use of the fixed effects model in the company's submission is reasonable, especially given the small difference in DIC values between models in the Company NMA (≤3); any difference in DIC values between models of <5 is not considered to be important (TSD 3[23]). The EAG note that fixed effects NMA models were deemed appropriate in similar analyses for ustekinumab (TA633[2]) and filgotinib (TA792[7]).

Where results of baseline risk adjusted models indicate evidence of differences in baseline risk across trials, the company have chosen to report results from these adjusted models. The EAG agrees with this approach. However, for mucosal healing in the maintenance period for the biologic-naïve population a baseline risk adjusted model is preferred by the company when the DIC suggests an unadjusted model is a better fit and there is no evidence from the meta-regression that this coefficient should be included. No appropriate justification is given by the company for this decision.

Comparison of results from the base case NMA models chosen by the company, with results from models with the lowest DIC tends to show a slightly more favourable finding from the company chosen models, in terms of the magnitude of the mean of the posterior distribution for mirikizumab. As expected, where a fixed effects model is chosen over a random effects model, the credible intervals are generally much narrower. However, the overall conclusions across the outcomes and populations do not change depending on the model selected, except for the outcome of all cause discontinuation: results from the fixed effects model (the company preferred model) are more favourable to mirikizumab compared to placebo (OR [redacted]) than results from the random effects model (OR [redacted]) due the narrower 95% credible intervals.

4.5.6 Summary of EAG critique of the NMA

- Overall, the EAG does not have any major concerns about the studies selected for inclusion in the NMA, but we note the following:
    - The range of treatments that studies could examine to be included in the NMA was broad. As with other appraisals of treatments for moderately to severely active UC,[18,25] many sources of heterogeneity across the included

studies were identified. As there is no justification for analysing such a broad network (other than for completeness), a smaller network may have resulted in less heterogeneity observed in the network. Reduced heterogeneity could provide more confidence in the NMA results through providing more precise credible intervals.

  - The NMA study eligibility criteria did not limit inclusion of studies to only people with moderately to severely active UC *who were intolerant of, or whose disease has had an inadequate response, or loss of response to previous biologic therapy or conventional therapy*, as per the population of interest specified in the NICE scope. This does not affect the interpretation of the results for the biologic-failed subgroup, as the studies that contributed data to these subgroup analyses included various populations of people who had had an inadequate response, loss of response, or intolerance to TNF antagonists, biologic therapy or specified biologics, or treatment failure on TNF or biologic. It does mean, however, that the biologic-naïve subgroup analyses do not fully reflect the population of interest in the NICE scope. This is because the studies contributing evidence to these analyses included people who had mainly just not previously received a TNF inhibitor therapy or biologic (CS Addendum, Appendix 1.5, Table 8) (i.e. they were not intolerant of, or had had an inadequate response to or loss of response to conventional therapy).
  - The searches for the systematic literature review informing the network meta-analysis were performed over six months ago and there is a risk that there may have been relevant studies published recently that will have been missed.

- Regarding how the NMA was conducted, the general approach to imputation of data used in NMA maintenance phase analyses from RCTs with a treat-through design was described as used in TA792[7] (filgotinib). However, there is an inconsistency in the weighted average applied to the placebo arm of ACT1 in the CS Addendum, and a difference in the raw data calculated from the company compared to that reported by Lu et 2022[20] (publication based on TA792[7]). The impact of these on the results for mirikizumab are likely to be minimal and the company conducted sensitivity analyses removing these 4 RCTs.
- The statistical models chosen for the different outcomes are appropriate and addressed limitations noted in previous TAs on this topic. Reporting of methods is generally clear.

- To model the baseline effect, the company incorporated placebo-arm data from all included RCTs rather than using representative UK-specific data as is recommended in TSD 5. The impact of this on the results is unclear.
- Justification for base case model choice was not consistent across the outcomes and subgroups. However, given limitations of a sparse network, the degree of heterogeneity observed and the small differences in estimates of model fit (DIC), the EAG believes the company's approach is reasonable

4.5.7 Results from the NMAs

Results of the NMA are presented below by treatment phase (induction phase, maintenance phase). Summaries of analyses, statistical models used and results, by subgroup (biologic-naïve, biologic-failed), are presented in Table 4 (induction phase) and Table 5 (maintenance phase), with statistically significant results highlighted in bold.

4.5.7.1 Induction phase

Results of the NMA for the induction phase are described below and presented in Table 4.

4.5.7.1.1 Clinical response – biologic-naïve and biologic-failed subgroups

For clinical response in the induction period, results of the biologic-naïve base case multinomial probit random effects model unadjusted for baseline risk show [redacted] (odds ratio (OR) [redacted]; Table 4; CS Addendum Figure 1). These findings are [redacted] (OR [redacted]; Table 4; CS Addendum Figure 7). For both the biologic-naïve subgroup and biologic-failed subgroup, [redacted] (Table 4; CS Addendum Figure 1 and 7). Exploratory analyses, which were adjusted for baseline risk, [redacted] (Table 4; CS Addendum, Appendix 3.1.1, Figure 114), but [redacted]; Table 4; CS Addendum, Appendix 3.1.2, Figure 124).

4.5.7.1.2 Clinical remission – biologic-naïve and biologic-failed subgroups

NMA results of clinical remission for the biologic-naïve subgroup and for the biologic-failed subgroup [redacted] respectively; Table 4; CS

Addendum Figure 2 and Figure 7), but [redacted]

(Table 4; CS Addendum Figure 2 and Figure 7). Results of exploratory analyses, i.e. with baseline risk adjustment, showed [redacted] (Table 4; CS Addendum, Appendix 3.1.1, Figure 115), but [redacted],Table 4; CS Addendum, Appendix 3.1.2, Figure 124).

4.5.7.1.3 Mucosal healing – biologic-naïve and biologic-failed subgroups

In both the biologic-naïve subgroup and biologic-failed subgroups, NMA results of mucosal healing show [redacted] respectively; Table 4; CS Addendum Figure 3 and Figure 8), but [redacted] (Table 4; CS Addendum Figure 3 and Figure 8).

4.5.7.1.4 All-cause discontinuation and serious adverse events – overall population

For the outcome of all cause discontinuation for the overall population (i.e. biologic-naïve and biologic-failed), [redacted]; Table 4; CS Addendum Figure 12), but [redacted]. However as we have previously noted (section 4.5.5.3 above) the company's results come from the fixed effect model whereas the model with the lowest DIC was the random effects model and this produced [redacted].

NMA results of serious adverse events for the overall population (i.e. biologic-naïve and biologic-failed), showed [redacted] (Table 4; CS Addendum Figure 13).

Table 4 Summary of NMA analyses and results for the induction phase

OUTCOME	ANALYSIS	STATISTICAL MODEL FEATURES		MIRI vs. PBO OR (95% CrI)	MIRI vs. VED OR (95% CrI)	MIRI vs. UST OR (95% CrI)
		FIXED/ RANDOM EFFECTS	BASELINE RISK ADJUSTMENT			
BIOLOGIC-NAÏVE INDUCTION PHASE						
Clinical response	Base case	Random[a]	No	[redacted]	[redacted]	[redacted]
	Exploratory	Random[a]	Yes	[redacted]	[redacted]	[redacted]
Clinical remission	Base case	Random[a]	No	[redacted]	[redacted]	[redacted]
	Exploratory	Random[a]	Yes	[redacted]	[redacted]	[redacted]
Mucosal healing	Base case	Random[b]	Yes	[redacted]	[redacted]	[redacted]
BIOLOGIC-FAILED INDUCTION PHASE						
Clinical response	Base case	Fixed[a]	No	[redacted]	[redacted]	[redacted]
	Exploratory	Fixed[a]	Yes	[redacted]	[redacted]	[redacted]
Clinical remission	Base case	Fixed[a]	No	[redacted]	[redacted]	[redacted]
	Exploratory	Fixed[a]	Yes	[redacted]	[redacted]	[redacted]
Mucosal healing	Base case	Fixed[b]	Yes	[redacted]	[redacted]	[redacted]
OVERALL/MIXED POPULATION INDUCTION PHASE						
All cause discontinuation	Base case	Fixed[b]		[redacted]	[redacted]	[redacted]
SAEs	Base case	Random[b]		[redacted]	[redacted]	[redacted]

[a] Multinomial model with ordered categories and probit link; [b] Binomial model with logit link
CrI: credible interval; MIRI: mirikizumab; OR: odds ratio; PBO: placebo; Q12W: every 12 weeks; Q2W: every 2 weeks; Q4W: every 4 weeks; Q8W: every 8 weeks; SAEs: serious adverse events; UST: ustekinumab; VED: vedolizumab.
For efficacy outcomes OR > 1 is in favour of mirikizumab. For safety outcomes OR <1 is in favour of mirikizumab.
Bold text: an OR and 95% CrIs which show a statistically significant result in favour of mirikizumab
Source: CS Appendices Table 32; CS Addendum Figures 1, 2, 3, 7, 8, 12 and 13; CS Addendum section 2.3.1; CS Addendum, Appendix 3.1.1, Figures 114 and 115; CS Addendum, Appendix 3.1.2, Figure 124

4.5.7.2 Maintenance phase

Results of the NMA for the maintenance phase are described below and presented in Table 5.

4.5.7.2.1 Clinical response – biologic-naïve and biologic-failed subgroups

For clinical response in the maintenance phase, results of the biologic-naïve base case multinomial probit fixed effect model adjusted for baseline risk show [redacted]; Table 5; CS Addendum Figure 4) and [redacted]. However, there was [redacted]. Exploratory analyses, unadjusted for baseline risk, also show [redacted]; Table 5; CS Addendum, Appendix 3.2.1, Figure 133). There was however [redacted] (Table 5; CS Addendum, Appendix 3.2.1, Figure 133).

Results of the biologic-failed base case multinomial probit fixed effect model unadjusted for baseline risk, also show [redacted]; Table 5; CS Addendum Figure 9). There was however [redacted] (Table 5; CS Addendum Figure 9). Exploratory analyses adjusted for baseline risk showed [redacted] Table 5; CS Addendum, Appendix 3.3.1, Figure 145).

The company note that results for maintenance phase clinical response (and clinical remission) should be interpreted with caution due to the imputation of data to account for the differing RCT designs. The EAG agree with the company. Results of sensitivity analyses (including re-randomised RCTs only) show that in the biological-naïve subgroup for clinical response for mirikizumab versus placebo [redacted] (Table 5; CS Addendum Figure 4 and CS Addendum, Appendix 2.1.2.2, Figure 42). However, this difference is also likely to be affected by the fact that these results are from a different base case NMA model (the sensitivity analysis results are from a model not adjusted for baseline

risk, while the analysis including the imputed data are from a model where baseline risk is included). The company have not reported results for the sensitivity analysis using a model with adjustment for base line risk. In the biologic-failed subgroup for maintenance phase clinical response, [redacted]. The same model (not including baseline risk adjustment) is used in both analyses.

4.5.7.2.2 Clinical remission – biologic-naïve and biologic-failed subgroups

For clinical remission in the maintenance phase, results of the biologic-naïve base case multinomial probit fixed effect model adjusted for baseline risk show [redacted]; Table 5; CS Addendum Figure 5) and [redacted]. However, there was [redacted]. Exploratory analyses, which were unadjusted for baseline risk, [redacted] Table 5; CS Addendum, Appendix 3.2.1, Figure 134). Furthermore, there was [redacted] (Table 5; CS Addendum, Appendix 3.2.1, Figure 134).

Results of the biologic-failed base case multinomial probit fixed effect model unadjusted for baseline risk, also show [redacted]; Table 5; CS Addendum Figure 10). [redacted] (Table 5; CS Addendum Figure 10). [redacted] (Table 5; CS Addendum, Appendix 3.3.1, Figure 145)

As stated earlier, the above results for clinical remission in the maintenance phase should be interpreted with caution due to the imputation of data to account for the differing RCT designs. Results of sensitivity analyses (including re-randomised RCTs only) for remission in the maintenance phase [redacted]:

- in the biologic-naïve subgroup there are [redacted]

[redacted] (Table 5; CS Addendum Figure 5 and CS Addendum, Appendix 2.1.2.2, Figure 43).

- in the biologic-failed subgroup, results of sensitivity analysis for clinical remission are [redacted] (Table 5; CS Addendum, Appendix 2.2.2.2, Figure 86).

4.5.7.2.3 Mucosal healing – biologic-naïve and biologic-failed subgroups

In the biologic-naïve subgroup, NMA results of mucosal healing in the maintenance phase showed [redacted]; Table 5; CS Addendum Figure 6), but [redacted] (Table 5; CS Addendum Figure 6). In the biologic-failed subgroup, NMA results of mucosal healing showed [redacted] Table 5; CS Addendum Figure 11).

4.6 Summary

In the absence of a trial directly comparing mirikizumab against vedolizumab and ustekinumab, the evidence for the comparability of mirikizumab with these drugs comes from the company's NMA results and is based on the statistical significance of the results only. There are no data available in the CS to directly show whether mirikizumab may be statistically equivalent to or non-inferior to ustekinumab and vedolizumab (i.e. there are no data from equivalence or non-inferiority trials). Acknowledging this limitation as an area of uncertainty, the EAG observes that based on the results reported in the NMA, mirikizumab appears to result in [redacted] (CS Addendum Figures 1, 2, 7 and 12). There is evidence from the base case NMA that mirikizumab results in [redacted] (CS Addendum Figures 4 and 5). [redacted] (CS Addendum Figures 4, 5, 9 and 10).

Table 5 Summary of NMA analyses and results for the maintenance phase

OUTCOME	ANALYSIS	STATISTICAL MODEL FEATURES		MIRI vs. PBO OR (95% CrI)	MIRI vs. VED OR (95% CrI) (108mg Q2W; 300mg Q4W; 300mg Q8W)	MIRI vs. UST OR (95% CrI) (90mg Q8W; 90mg Q12W)
		FIXED/ RANDOM EFFECTS	BASELINE RISK ADJUSTMENT			
BIOLOGIC-NAÏVE MAINTENANCE PHASE						
Clinical response	Base case	Fixed[a]	Yes	[redacted]	[redacted]	[redacted]
	Sensitivity[b]	Fixed[a]	No	[redacted]	[redacted]	[redacted]
	Exploratory	Fixed[a]	No	[redacted]	[redacted]	[redacted]
Clinical remission	Base case	Fixed[a]	Yes	[redacted]	[redacted]	[redacted]
	Sensitivity[b]	Fixed[a]	No	[redacted]	[redacted]	[redacted]
	Exploratory	Fixed[a]	No	[redacted]	[redacted]	[redacted]
Mucosal healing	Base case	Fixed[c]	Yes	[redacted]	[redacted]	[redacted]
BIOLOGIC-FAILED MAINTENANCE PHASE						
Clinical response	Base case	Fixed[a]	No	[redacted]	[redacted]	[redacted]
	Sensitivity[b]	Fixed[a]	No	[redacted]	[redacted]	[redacted]
	Exploratory	Fixed[a]	Yes	[redacted]	[redacted]	[redacted]
Clinical remission	Base case	Fixed[a]	No	[redacted]	[redacted]	[redacted]
	Sensitivity[b]	Fixed[a]	No	[redacted]	[redacted]	[redacted]

OUTCOME	ANALYSIS	STATISTICAL MODEL FEATURES		MIRI vs. PBO OR (95% CrI)	MIRI vs. VED OR (95% CrI) (108mg Q2W; 300mg Q4W; 300mg Q8W)	MIRI vs. UST OR (95% CrI) (90mg Q8W; 90mg Q12W)
		FIXED/ RANDOM EFFECTS	BASELINE RISK ADJUSTMENT			
	Exploratory	Fixed[a]	Yes	[redacted]	[redacted]	[redacted]
Mucosal healing	Base case	Fixed[c]	No	[redacted]	[redacted]	[redacted]

[a] Multinomial model with ordered categories; [b] Re-randomised studies only sensitivity analysis; [c] Binomial model with logit link
CrI: credible interval; MIRI: mirikizumab; OR: odds ratio; PBO: placebo; Q12W: every 12 weeks; Q2W: every 2 weeks; Q4W: every 4 weeks; Q8W: every 8 weeks; UST: ustekinumab; VED: vedolizumab.
For efficacy outcomes OR > 1 is in favour of mirikizumab. For safety outcomes OR <1 is in favour of mirikizumab.
Bold text: an OR and 95% CrIs which show a statistically significant result in favour of mirikizumab
Source: CS Appendices Table 32; CS Addendum Figures 4, 5, 6, 9, 10 and 11; CS Addendum sections 2.1.2, 2.2.2, 5.3.1.3 and 5.3.2.3; CS Addendum, Appendix 2.1.2.2, Figures 42 and 43; CS Addendum, Appendix 2.2.2.2, Figures 86 and 87; CS Addendum, Appendix 3.2.1, Figures 133 and 134; CS Addendum, Appendix 3.3.1, Figures 145 and 146

5 Summary of the EAG's critique of the cost comparison evidence submitted

5.1 Introduction

The following sections critique:

i. the company's cost comparison evidence submitted on 8th December 2022 for this appraisal (henceforth, referred to as the 'original CS' and the 'original economic model')

ii. the new evidence, received on 15th February 2023, submitted as an addendum to the original CS and a revised economic model (henceforth, referred to as 'addendum to the CS' and 'revised economic model', respectively).

The company produced the addendum and the corresponding revised economic model to correct errors in the NMA in the original CS (as discussed earlier in Section 4.5). This amendment was in response to the EAG's correspondence with NICE, seeking further clarifications on the NMA inputs that informed the company's original economic model.

5.2 Decision Problem for the cost comparison

5.2.1 Population

The EAG has determined that the characteristics of the population used by the company in the cost-comparison analysis adequately reflects the indications in the NICE recommendations for the comparator drugs. The company's cost analyses modelled two patient cohorts with mean age of [redacted] years for biologic-naïve and [redacted] years for biologic-failed patients respectively (CS Table 39). These patient characteristics, based on the modified ITT populations of the LUCENT trials, are similar to those of the modelled cohort for the comparator appraisal TA633 (ustekinumab).[2]

5.2.2 Comparators

The analysis compares mirikizumab with ustekinumab and vedolizumab. As stated in section 3.3, the EAG consider that these comparators are appropriate for the cost-comparison analysis.

5.2.3 Cost-comparison model

The company describe their cost-comparison model in CS section B.4.2.1. The model structure is illustrated in CS Figures 29 and 30. We outline the model features and structure below.

Model features:

- Markov model with four components:
    - an induction period of up to 26 weeks comprising two-week tunnel states,
    - an on-treatment maintenance state,
    - an off-treatment state, and
    - a death state.
- Efficacy parameters (response rates) in the induction and maintenance phases are informed by the NMA results (discussed earlier in Section 4.5 of this report).
- Patients incur no costs in the off-treatment state.
- Time horizon: 10 years
- No discounting
- Perspective: National Health Service (NHS)/Personal Social Services (PSS)
- Cycle length: 2 weeks (induction phase); 12 weeks (maintenance phase)

Model structure:

Induction phase:

- Variable and treatment-specific lengths of induction periods for the treatments, varying between 2-12 weeks depending on the drug, and up to 12 additional weeks for delayed responders. For the mirikizumab arm, the induction phase is 12 weeks for the base case and a scenario was conducted to include an extended induction phase for delayed responders up to 12 weeks.
- All non-responders at the end of the induction period either enter the no-treatment state or continue to be treated for an additional 8 weeks on ustekinumab (16 weeks total induction), an additional 4 weeks on vedolizumab (10 weeks total treatment) or an additional 12 weeks on mirikizumab (24 weeks total treatment) to assess delayed response. The timepoints for delayed response are based on the pivotal trials for the respective drugs.
- At the end of the induction period, patients are classified as responders or non-responders. The responders transition to the maintenance state and the non-responders to the no-treatment state.

Maintenance phase:

- Responders at the end of induction phase enter the maintenance phase, which includes:
    - on treatment,

    - off treatment, and
    - death
- Non-responders transition to ‘no treatment’ state.
- In their base case, the company included re-induction of mirikizumab in the maintenance phase (rather than dose-escalation as modelled in the comparator arms) as this is anticipated in their marketing authorisation. The re-induction dose is 300 mg IV mirikizumab at Week 12, 16 and 20. In the base case, [redacted] of patients receiving mirikizumab were modelled to undergo re-induction (equating to [redacted] per cycle), to reflect the proportion of patients who were re-inducted in the LUCENT-2 trial. A scenario was conducted with 30% of patients undergoing re-induction, to align with the comparator arms where 30% of the patients receive dose escalation in the maintenance phase.
- Given the assumption of equal efficacy for all treatments, dose escalation and reinduction were assumed to affect only costs, not efficacy.
- The cost of re-induction was applied only to the cycle in which the patient is re-induced.

EAG conclusions:

- The model structure is a reasonable simplification. We agree with the company’s approach to exclude other states (such as surgery/ post-surgery) due to similar downstream costs driven by similar efficacy.
- The company explored the impact of varying model features in their scenario analysis. These included: increasing the model time horizon, applying discount rates, extending the induction phase for delayed responders, assuming similar proportion of patients receiving re-induction as patients in the comparator arms receiving dose-escalation. Further details are in Section 6.
- Based on our clinical expert’s advice, we view it is reasonable to assume that dose escalation and re-induction are likely to impact only costs and not efficacy of the drugs, due to the assumption of equal efficacy for all the treatments.

5.3 Model parameters

5.3.1 Efficacy

As stated earlier in section 5.1, the company corrected their NMA inputs in the original CS and submitted a revised economic model. We present a detailed critique of the revised NMA in section 4.5 of this report. The efficacy parameters discussed in the following sub-sections are obtained from the revised NMA to populate the company’s revised economic model.

Induction phase

- The model assumed similar response rates across all treatments, although the rates differed between the two sub-groups: biologic-naïve and biologic-failed.
- For their base case, the response rates were obtained from the revised NMA inputs shown in Table 2, Figure 1 (biologic-naïve) and Figure 7 (biologic-experienced), respectively, of the addendum to the CS.
- For the scenario analysis (extended induction), the absolute probability of response was obtained from the previous NICE appraisal on ustekinumab TA633 (Table 3 of the addendum to the CS).

Maintenance phase

- All treatments were assumed to have the same risk of treatment discontinuation. The odds ratio obtained for response at the end of maintenance for mirikizumab relative to placebo from the revised NMA were converted to absolute probability. Further details on the NMA are in Section 4.5 above.
- Those patients who are off treatment remained in the state until the end of the model simulation or death.

In Table 6, we present a summary of the estimated probabilities obtained from the revised NMA response rates results (as presented in the addendum to the CS) that are used in inform the revised economic model.

Table 6 Probabilities (per cycle) used in the company model for the base case

Sub-group	Induction		Maintenance	
	Response	***Non-response (estimated as 1-response)***	***Response***	***Non-response (Estimated as 1-response)***
Biologic-naïve	[redacted]	[redacted]	[redacted]	[redacted]
Biologic-failed	[redacted]	[redacted]	[redacted]	[redacted]
Sources: Table 2 and section 3.1.1.2 of the addendum to the CS				

EAG conclusions:

- Overall, we agree with the company's assumptions which are reasonable simplifications.
- The company's methodological approach to obtain the probabilities from response rates is appropriate (further details are in company's response to EAG clarification Question 1). We did not have access to the CODA output to produce the mean

absolute probabilities of response, which was calculated for 20,000 NMA samples. Therefore, we are unable to verify the company's estimates for the probability of response.

- With respect to the efficacy inputs in the model, obtained from the revised NMA, the EAG has a few concerns including i) the broad NMA structure leading to clinical and statistical heterogeneity, ii) the lack of representative UK-specific data for modelling baseline effect; and iii) inconsistency in the population characteristics for the biologic-naïve subgroup included in the NMA and those stated in the NICE scope. For further details, see Section 4.5. However, none of these concerns are critical and we do not anticipate these to have any significant impact on the efficacy parameters.

5.3.2 Mortality

The Office for National Statistics (ONS) life tables, adjusted for age and gender, were used for mortality estimation. No increased mortality was assumed due to ulcerative colitis. This is consistent with previous NICE TAs (TA633, TA342, TA792 and TA547).

5.3.3 Costs

Acquisition costs

Details of the company's inputs and assumptions for acquisition costs of the intervention – mirikizumab – and the comparators ustekinumab and vedolizumab are summarised in CS Table 40. Drug acquisition costs, sourced from Monthly Index of Medical Specialities (MIMS) and the British National Formulary (BNF), are summarised in CS Table 41 (induction phase) and CS Table 42 (maintenance phase).

Mirikizumab patients who did not respond after initial induction therapy or who lost response in the maintenance phase received re-induction. Whereas patients in the two comparator arms who did not respond after initial induction therapy or who lost response in the maintenance phase received dose-escalation. Irrespective of patients re-induced or who received dose-escalation, drug acquisition costs took into account the proportion of patients on standard and escalated doses during the maintenance phase.

All other costs

- Drug administration costs are summarised in CS Tables 43 and 44.
- Disease management costs, costs for monitoring and tests during the induction phase, and adverse event costs are not modelled.

EAG conclusions: Overall, we agree with the company's costs estimates. Their approach for estimating acquisition costs is appropriate and that for administration costs is consistent with previous appraisals (TA633, TA547 and TA792). Based on our clinical expert's opinion, we view it is reasonable to exclude the costs associated with disease management, monitoring, and adverse events, provided the assumption that all the treatments have similar efficacy holds true. Furthermore, our expert indicated that the provision of mirikizumab is unlikely to incur any other additional costs that are not incurred in the provisions of ustekinumab and vedolizumab.

5.4 EAG model checks

The EAG conducted a range of checks on the company's original cost-comparison model submitted on 8th December 2022. These included:

- verification that all input parameters and model results matched the values cited in the CS and, where available, values in published sources.
- Inspection of formulae in the Markov trace and intermediate calculations ('white box' verification)
- checking that changes to input parameters had a plausible impact on results ('black box' verification).
- re-running all the company's sensitivity and scenario analyses. The probabilistic sensitivity analysis (PSA) was not implemented in the model, which is acceptable, as the PSA is not required for a cost-comparison model.

We conducted the following checks on the company's revised model received on 15th February 2023:

- re-produced the revised cost comparison results from the original company model (received on 8th December 2022) by applying the revised NMA estimates into the original model.
- verified no other changes have been made to the remaining model parameters including baseline characteristics, life tables, costs, and adverse events, in the revised model.
- re-ran all the company's sensitivity and scenario analyses.

We identified two inconsistencies in the company's scenario analyses:

- A minor inconsistency in the estimation of adverse event costs. The company applied adverse events costs of £4000; we estimated a slightly different AE cost of £3,898. This minor difference does not have any significant impact on the results.

- For the scenario of extended induction (when non-responders at the end of induction continue for an additional treatment phase), the company did not apply the correct treatment duration for mirikizumab which is 24 weeks in total (12 weeks of induction + 12 weeks of extended induction) (see CS Document B section B.4.2.1.1). We corrected this error (in cell K96 of Sheet!Model Settings of the company's revised mdel); the results, in Table 7 below, show that mirikizumab is [redacted] than the two comparators, in both the sub-groups.

Table 7 Corrected results from the company's scenario analysis of delayed response (extended induction) (list price)

Scenario	Incremental costs relative to mirikizumab (biologic-naive)			Incremental costs relative to mirikizumab (biologic-experienced)		
	Ustekinumab	Vedolizumab IV	Vedolizumab IV/SC	Ustekinumab	Vedolizumab IV	Vedolizumab IV/SC
Company's Base case	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Scenario with delayed response	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

EAG conclusions: Overall, the model is well-implemented, although we identified two errors in the company's scenario analyses, as discussed above.

6 Company and EAG cost comparison results

6.1 Company's cost comparison analysis results

The company revised base case cost comparison results are presented in Table 6 (for biologic-naïve) and Table 7 (for biologic-failed) of the addendum to the CS. These results are based on the list price and PAS price for mirikizumab, and list prices for the two comparators, respectively. We present the results of the company's analyses using the PAS prices for mirikizumab and vedolizumab and CMU price for ustekinumab in a confidential addendum.

Uncertainty over model assumptions was assessed with one-way sensitivity analyses (parameters described in Table 8 of the addendum to the CS) and scenario analyses (Table 9 of the addendum to the CS), respectively. The one-way sensitivity analysis was conducted using an outdated PAS price discount for mirikizumab and list prices for the two comparators. Hence, we have not commented on these results. We have, however, run these analyses using the list prices for all the three treatments, as discussed in the following Section 6.2. We also conducted the corresponding analyses using PAS prices for mirikizumab and vedolizumab and CMU price for ustekinumab in the confidential addendum.

Results from the company's scenario analyses using the list prices for all the three drugs (see Table 9 of the addendum to the CS) show that:

- for biologic-naïve population (Table 10 of the addendum to the CS), mirikizumab remained [redacted] than the comparators in most of the scenarios. Increasing the percentage of patients in treatment re-induction from [redacted] to [redacted] per cycle (scenario 5) had the highest impact for vedolizumab IV, [redacted] the incremental cost from [redacted] (revised base case) to [redacted]. Scenario 6 (with a delayed response assessment for mirikizumab and the comparators) had the biggest impact for ustekinumab and vedolizumab IV/SC, [redacted] the incremental cost from [redacted] (revised base case) to [redacted] for ustekinumab, and from [redacted] (revised base case) to [redacted] for vedolizumab IV/SC respectively.
- Regarding the biologic-failed population results (Table 11 of the addendum to the CS), the EAG observed similar effect as the biologic-naïve population, where mirikizumab remained [redacted] than the comparators in all scenarios, [redacted] the incremental cost from [redacted] (revised base case) to [redacted]. Scenario 6 (with a delayed response assessment for mirikizumab and the comparators) had the biggest impact for ustekinumab and vedolizumab IV/SC, [redacted] the incremental cost from

[redacted] (revised base case) to [redacted] for ustekinumab, and from [redacted] (revised base case) to [redacted] for vedolizumab IV/SC, respectively.

6.2 EAG analyses

6.2.1 Company's one-way sensitivity analysis using list prices

The EAG has run the company's one-way sensitivity analysis using the list prices for all three drugs (mirikizumab, ustekinumab and vedolizumab) for biologic-naïve and biologic-failed populations using the revised company model as the company conducted the one-way sensitivity analysis results using an outdated PAS price for mirikizumab. Tornado plots are presented in the Appendix 3 of this report (see Figure 4, Figure 6, and Figure 8 for the biologic-naïve population, and Figure 5, Figure 7, and Figure 9 for the biologic-failed population, respectively). For both the subgroups, the key model drivers are the response rates for the induction and the maintenance phases. Changing the proportion of patients for dose escalation also impacted the model results, but to a lesser extent.

6.2.2 Additional scenarios by EAG

We performed three additional analyses with the company's base case to complement the company's scenarios and analyse the impact of changing some of the model assumptions in the final cost-comparison results.

- Mirikizumab arm: change the re-induction rate from [redacted] to [redacted] and [redacted], and maintain dose escalation in 30% for the comparators
- Include AE costs (for completeness: £3898 EAG estimated vs company's estimate of £4000)
- Time horizon: 15 years.

Table 8 presents the results for biologic-naïve and Table 9 for biologic-failed populations. These analyses are conducted using the list prices for mirikizumab and the comparators-ustekinumab and vedolizumab. The EAG notes:

- For the biologic-naïve population, mirikizumab [redacted] than the comparators. Varying the re-induction rate to [redacted] the cost difference between mirikizumab and the comparators by [redacted], which increased to [redacted] at a [redacted] re-induction rate. Using a 15-year time horizon had a marginal impact on the cost difference between mirikizumab and the comparators. For example, the cost difference between mirikizumab and vedolizumab [redacted] by [redacted] compared to the revised base case result, by [redacted] between mirikizumab and vedolizumab (IV) and

[redacted] between mirikizumab and vedolizumab (IV/Sc), respectively. The scenario including revised adverse event costs [redacted] the costs [redacted].

- For the biologic-failed population, mirikizumab remained [redacted] than the comparators in all the scenarios. Varying the re-induction rate [redacted] the cost difference between mirikizumab and the comparators by [redacted] (10% re-induction rate) and [redacted] (15% re-induction rate), respectively compared to the company's revised base case results. The scenarios including adverse event costs and time horizon [redacted] the costs negligibly ([redacted]).

Table 8 EAG scenario analysis for mirikizumab for biologic naïve population – incremental cost mirikizumab versus comparators (list price for all drugs)

EAG scenario	Treatments	Total costs	Incremental costs for Mirikizumb vs comparators
Revised company base case	Mirikizumab	[redacted]	
	Ustekinumab	£23,310	[redacted]
	Vedolizumab IV	£35,732	[redacted]
	Vedolizumab SC/IV	£26,644	[redacted]
Re-induction rate per cycle to 10%	Mirikizumab	[redacted]	
	Ustekinumab	£23,310	[redacted]
	Vedolizumab IV	£35,732	[redacted]
	Vedolizumab SC/IV	£26,644	[redacted]
Re-induction rate per cycle to 15%	Mirikizumab	[redacted]	
	Ustekinumab	£23,310	[redacted]
	Vedolizumab IV	£35,732	[redacted]
	Vedolizumab SC/IV	£26,644	[redacted]
Include adverse event costs (£3,898)	Mirikizumab	[redacted]	
	Ustekinumab	£23,521	[redacted]
	Vedolizumab IV	£35,938	[redacted]
	Vedolizumab SC/IV	£26,850	[redacted]
Time horizon 15 years	Mirikizumab	[redacted]	
	Ustekinumab	£24,090	[redacted]
	Vedolizumab IV	£37,101	[redacted]
	Vedolizumab SC/IV	£27,615	[redacted]

Table 9 EAG scenario analysis for mirikizumab considering for biologic failed population – incremental cost mirikizumab versus comparators (list price for all drugs)

EAG scenario	Treatments	Total costs	Incremental costs for Mirikizumb vs comparators
Revised company base case	Mirikizumab	[redacted]	
	Ustekinumab	£10,542	[redacted]
	Vedolizumab IV	£12,952	[redacted]
	Vedolizumab SC/IV	£10,481	[redacted]
Re-induction rate per cycle to 10%	Mirikizumab	[redacted]	
	Ustekinumab	£10,542	[redacted]
	Vedolizumab IV	£12,952	[redacted]
	Vedolizumab SC/IV	£10,481	[redacted]
Re-induction rate per cycle to 15%	Mirikizumab	[redacted]	
	Ustekinumab	£10,542	[redacted]
	Vedolizumab IV	£12,952	[redacted]
	Vedolizumab SC/IV	£10,481	[redacted]
Include adverse event costs (£3,898)	Mirikizumab	[redacted]	
	Ustekinumab	£10,609	[redacted]
	Vedolizumab IV	£13,015	[redacted]
	Vedolizumab SC/IV	£10,544	[redacted]
Time horizon 15 years	Mirikizumab	[redacted]	
	Ustekinumab	£10,543	[redacted]
	Vedolizumab IV	£12,954	[redacted]
	Vedolizumab SC/IV	£10,482	[redacted]

7 Equalities and innovation

Mirikizumab is not a particularly innovative medicine in comparison to the comparators either in terms of mechanism of action (targeting the IL-23 cytokine pathway, which is similar to ustekinumab that targets the IL-23 and IL-12 cytokine pathways as summarised in section 0) or in terms of method of administration (initially IV infusion for induction then subcutaneous injection for maintenance treatment). No equality considerations have been raised during this appraisal.

8 EAG commentary on the robustness of evidence submitted by the company

The EAG overall does not have any major concerns about how the clinical efficacy and safety estimates for mirikizumab versus ustekinumab and vedolizumab have been derived from the company's NMA. We have not identified any critical issues, that, in our opinion, would prevent progression with a cost-comparison approach. We have identified some uncertainties associated with the evidence base, however. We note that:

- with regard to results presented in the CS for the biologic-naïve population from the NMA, there is an issue that the characteristics of this group in the NMA studies do not fully reflect the exact biologic-naïve population stated in the NICE scope and the biologic-naïve group in whom the company is partly positioning mirikizumab.
- the NMA methodology on the whole appears appropriate, but the company has not modelled baseline effect using representative UK-specific data as is recommended in TSD 5.[1] The impact of this on the results is unclear.
- there was considerable clinical and statistical heterogeneity in the broad NMA network; a narrower network may have resulted in more precise estimates of clinical efficacy (i.e. through providing narrower credible intervals, and thus providing more confidence in mirikizumab having [redacted] to comparators).
- there are no data available in the CS to show whether mirikizumab may be statistically non-inferior or equivalent to ustekinumab and vedolizumab (i.e. there are no data from equivalence or non-inferiority trials).

Based on the statistical significance of the NMA findings, mirikizumab appears to have [redacted], treatment effects (i.e. clinical response and remission in the induction and maintenance treatment phases) than, and a [redacted] safety profile to, vedolizumab and ustekinumab.

The EAG's conclusions on the company's cost-comparison analysis are:

- The model structure and key assumptions of the company's cost comparison model are appropriate, and consistent with the previous NICE ustekinumab appraisal TA633.
- The model assumes equal clinical efficacy for mirikizumkab, ustekinumab and vedolizumab based on the NMA results. While there are uncertainties with the NMA (discussed in Section 4 and reiterated above), none of these are critical. Therefore, we view that it is appropriate to assume equal clinical efficacy for all three drugs.

- With the list prices for mirikizumab, ustekinumab and vedolizumab, mirikizumab is [redacted] than the two comparators. This applies for the company's base case analysis and for all the company and EAG scenario analyses.
- The cost difference between mirikizumab and the two comparators is most sensitive to assumptions about re-induction rates and delayed response assessment.

9 References

1. Dias S, Welton NJ, Sutton AJ, et al. NICE DSU Technical Supoort Docuemnt 5: Evidence synthesis in the baseline natural history model - report by the Decision Support Unit. Sheffield: Decision Suport Unit, 2012.
2. National Institute for Health and Care Excellence (NICE). *Ustekinumab for treating moderately to severely active ulcerative colitis [TA633]*. https://www.nice.org.uk/guidance/ta633/evidence.
3. National Institute for Health and Care Excellence (NICE). *Infliximab, adalimumab and golimumab for treating moderately to severely active ulcerative colitis after the failure of conventional therapy [TA329]*. https://www.nice.org.uk/guidance/ta329.
4. National Institute for Health and Care Excellence (NICE). *Vedolizumab for treating moderately to severely active ulcerative colitis [TA342]*. https://www.nice.org.uk/guidance/ta342.
5. National Institute for Health and Care Excellence (NICE). *Tofacitinib for moderately to severely active ulcerative colitis. Technology appraisal guidance [TA547]*. https://www.nice.org.uk/guidance/ta547.
6. National Institute for Health and Care Excellence (NICE). *Ozanimod for treating moderately to severely active ulcerative colitis [TA828]*. https://www.nice.org.uk/guidance/ta828.
7. National Institute for Health and Care Excellence (NICE). *Filgotinib for treating moderately to severely active ulcerative colitis [TA792]*. https://www.nice.org.uk/guidance/ta792.
8. . *National Institute for Health and Care Excellence (NICE). Upadacitinib for treating moderately to severely active ulcerative colitis [GID-TA10866]. Available at: https://www.nice.org.uk/guidance/indevelopment/gid-ta10866. [Accessed 05 October 2022].*
9. National Institute for Health and Care Excellence. NICE Fast track appraisals - guiding notes for ERGs: NICE, 2017.
10. National Institute for Health and Care Excellence. Single Technology Appraisal. Ustekinumab for treating moderately to severely active ulcerative colitis [ID1511]. Committee Papers. 2020.
11. Feagan BG, Rutgeerts P, Sands BE, et al. Vedolizumab as Induction and Maintenance Therapy for Ulcerative Colitis. *New England Journal of Medicine* 2013;369(8):699-710.
12. Sands BE, Sandborn WJ, Panaccione R, et al. Ustekinumab as Induction and Maintenance Therapy for Ulcerative Colitis. *New England Journal of Medicine* 2019;381(13):1201-14.
13. Eli Lilly (Data on file): LUCENT-1 Clinical Study Report.
14. Eli Lilly (Data on file). LUCENT-2 Clinical Study Report.
15. Naegeli AN, Hunter T, Dong Y, et al. Full, Partial, and Modified Permutations of the Mayo Score: Characterizing Clinical and Patient-Reported Outcomes in Ulcerative Colitis Patients. *Crohn's & Colitis 360* 2021;3(1):otab007.
16. Center for Drug Evaluation and Research. Ulcerative Colitis: Clinical Trial Endpoints Guidance for Industry. Available at: https://www.fda.gov/files/drugs/published/Ulcerative-Colitis--Clinical-Trial-Endpoints-Guidance-for-Industry.pdf. [Accessed 15 November 2022].
17. Eli Lilly (Data on file). NMA report appendices.
18. Kalita N, Lord J, Pickett K, et al. Ustekinumab for treating moderately to severely active ulcerative colitis: A Single Technology Appraisal. Southampton: Southampton Health Technology Assessments Centre (SHTAC), 2019.
19. Centre for Reviews and Dissemination. Systematic reviews: CRD's guidance for undertaking reviews in health care. York Publishing Services Ltd.: CRD, 2009.

20. Lu X, Jarrett J, Sadler S, et al. Comparative efficacy of advanced treatments in biologic-naïve or biologic-experienced patients with ulcerative colitis: a systematic review and network meta-analysis. *Int J Clin Pharm* 2022;Online ahead of print.
21. Stan Development Team. *Stan Modeling Language Users Guide and Reference Manual*. https://mc-stan.org.
22. Dias S, Welton NJ, Sutton AJ, et al. NICE DSU technical support document 2: a generalised linear modelling framework for pairwise and network meta-analysis of randomised controlled trials. Sheffield: ScHARR, University of Sheffield, 2016.
23. Dias S, Sutton AJ, Welton NJ, et al. Evidence Synthesis for Decision Making 3: Heterogeneity—Subgroups, Meta-Regression, Bias, and Bias-Adjustment. *Medical Decision Making* 2013;33(5):618-40.
24. Dias S, Welton NJ, Sutton AJ, et al. NICE DSU technical support document 4: inconsistency in networks of evidence based on randomised controlled trials. Sheffield: Decision Support Unit, ScHARR, University of Sheffield, 2014.
25. Essat M, Tappenden P, Ren S, et al. Vedolizumab for the treatment of adults with moderately to severely active ulcerative colitis: A Single Technology Appraisal: School of Health and Related Research (ScHARR), 2014.
26. National Institute for Health and Care Excellence. Single Technology Appraisal. Tofacitinib for previously treated active ulcerative colitis [ID1218]. Committee Papers., 2018.
27. *British National Formulary (BNF). Available at: https://bnf.nice.org.uk/. [Accessed 22 September 2022].*

10 Appendices

Appendix 1 Comparator mechanisms of action and modes of administration

Table 10 Mechanisms of action and modes of administration for the comparators listed in the NICE scope for this PATT

NICE TA	Biologic therapy	Mechanism of action	Mode of administration [a]
TA329	Infliximab [b]	Monoclonal antibodies that inhibit the activity of TNF-α which is a key component in the inflammation process.[3]	Either by intravenous infusion, or initially by intravenous infusion followed by subcutaneous injection.
	adalimumab		Subcutaneous injection
	golimumab		Subcutaneous injection
TA342	vedolizumab	A humanised monoclonal antibody that binds to the α4β7 integrin expressed on certain gut homing T helper lymphocytes. When bound to α4β7 integrin vedolizumab inhibits adhesion of these cells to mucosal addressing cell adhesion molecule-1 (MAdCAM-1). Vedolizumab therefore selectively targets the gut and reduces gut inflammation by preventing the selective migration of pathogenic gut-homing lymphocytes.[4]	Induction by intravenous infusion followed by subcutaneous maintenance doses
TA547	tofacitinib	Janus kinase (JAK) inhibitor (similar in structure to adenosine triphosphate (ATP) and competes with ATP at target sites).[26]	Oral
TA633	ustekinumab	Fully human IgG1κ monoclonal antibody that binds to the p40 subunit of IL-12 and IL-23 cytokines thereby dampening	Induction infusion followed by

		the inflammatory cascade underlying UC.[10]	subcutaneous maintenance doses.
TA792	filgotinib	JAK1 inhibitor[27]	Oral
TA828	ozanimod	A sphingosine-1-phosphate receptor modulator thought to inhibit inflammation by preventing lymphotcyte movement to sites including the intestine.[6,27]	Oral
TA856	upadacitinib	JAK1 inhibitor[27]	Oral

[a] Information on mode of administration has been taken from the BNF[27] for each drug in the relevant indication; [b] and biosimilars

Appendix 2 EAG's risk of bias assessments of the LUCENT-1 and LUCENT-2 trials

The EAG's risk of bias assessment of the pivotal mirikizumab LUCENT-1 and LUCENT 2 trials, in comparison to the company's assessment, is shown in Table 11.

Table 11 EAG and company's risk of bias assessments of the LUCENT-1 and LUCENT-2 trials

Study question (Yes/No/Unclear)	LUCENT-1 COMPANY ASSESSMENT	LUCENT-1 EAG ASSESSMENT	LUCENT-2 COMPANY ASSESSMENT	LUCENT-2 EAG ASSESSMENT
Was randomisation carried out appropriately?	Yes	Yes Assignment to treatment groups determined by a computer-generated random sequence using an interactive web-response system (LUCENT-1 Trial Protocol section 7.2)	Yes	Yes Assignment to treatment groups for clinical responders determined by a computer-generated random sequence using an interactive web-response system (LUCENT-2 Trial Protocol section 7.2)
Was the concealment of treatment allocation adequate?	Yes	Yes Interactive web-response system used	Yes	Yes Interactive web-response system used
Were the groups similar at the outset of the study in terms of prognostic factors?	Yes	Yes Disease location, severity (total Mayo score), endoscopic Mayo subscore of severe disease,faecal calprotectin, and prior biologic or tofacitinib failure were similar between arms (CS Table 9)	Yes	Yes Disease location, severity (total Mayo score), endoscopic Mayo subscore of severe disease,faecal calprotectin, and prior biologic or tofacitinib failure were similar between the randomised arms (CS Table 11)
Were the care providers, participants and outcome assessors blind to treatment allocation?	Yes	Yes Double-blind study. (investigator, site personnel performing assessments and patients were blinded) Blinded study personnel prepared investigational product.	Yes	Yes Double-blind study. (investigator, site personnel performing assessments and patients were blinded) Blinded study personnel prepared investigational product.

		Mirikizumab visually indistinguishable from placebo. (LUCENT-1 Trial Protocol 7.1.1 and 7.3)		Mirikizumab visually indistinguishable from placebo. (LUCENT-2 Trial Protocol 7.1.1 and 7.3)
Were there any unexpected imbalances in drop-outs between groups?	Unclear	Unclear There were imbalances, but not necessarily unexpected, with a greater proportion discontinuing due to adverse events (most common event was ulcerative colitis), withdrawal by subject and lack of efficacy in the placebo arm compared to the mirikizumab arm. (LUCENT-1 CSR Table 8.1. CS Appendix F.4)	Unclear	Unclear There were imbalances, but not necessarily unexpected, with a greater proportion discontinuing due to adverse events (most common event was ulcerative colitis), withdrawal by subject and lack of efficacy in the placebo arm compared to the mirikizumab arm (LUCENT-2 CSR Table 8.1, CS Table 36)
Is there any evidence to suggest that the authors measured more outcomes than they reported?	No	No Objectives and endpoints in protocol match those reported in the CSR	No	No Objectives and endpoints in protocol match those reported in the CSR
Did the analysis include an intention-to-treat analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	Yes	Yes Analysis was based on modified intention to treat This was due to baseline errors in electronic data collection devices. This approach was agreed with FDA. Appropriate methods used to impute missing data for primary outcome (LUCENT-1 SAP 5.3.1 and 5.4)	Yes	Yes Analysis was based on modified intention to treat This was due to baseline errors in electronic data collection devices. This approach was agreed with FDA. Appropriate methods used to impute missing data for primary outcome (LUCENT-2 SAP 5.3.1 and 5.4)

Source: The company risk of bias assessments were extracted from CS Appendix D.3 Table 46.

Appendix 3 EAG update to Company's one-way sensitivity results using list prices

Figure 4 Tornado plot with results from the one-way sensitivity analysis – mirikizumab (list price) versus vedolizumab IV in the biologic-naïve population

Figure 5 Tornado plot with results from the one-way sensitivity analysis – mirikizumab (list price) versus vedolizumab IV in the biologic-failed population

Figure 6 Tornado plot with results from the one-way sensitivity analysis – mirikizumab (list price) versus vedolizumab SC/IV in the biologic-naïve population

Figure 7 Tornado plot with results from the one-way sensitivity analysis – mirikizumab (list price) versus vedolizumab SC/IV in the biologic-failed population

Figure 8 Tornado plot with results from the one-way sensitivity analysis – mirikizumab (list price) versus ustekinumab in the biologic-naïve population

Figure 9 Tornado plot with results from the one-way sensitivity analysis – mirikizumab (list price) versus ustekinumab in the biologic-failed population