in collaboration with:

Erasmus School of Health Policy & Management

Maastricht University

External Assessment Group Report

Upadacitinib for previously treated moderately to severely active Crohn's Disease [ID4027]

Produced by	Kleijnen Systematic Reviews (KSR) Ltd, in collaboration with Erasmus University Rotterdam (EUR) and Maastricht University
Authors	Mark Perry, Systematic Reviewer, KSR Ltd, United Kingdom (UK) Nigel Armstrong, Health Economist, KSR Ltd, UK Pim Wetzelaer, Health Economist, EUR, the Netherlands (NL) Eline Krijkamp, Health Economist, EUR, the NL Lisa Stirk, Information Specialist, KSR Ltd, UK Maiwenn Al, Health Economist, EUR, the NL Robert Wolff, Managing Director, KSR Ltd, UK
Correspondence to	Robert Wolff, Kleijnen Systematic Reviews Ltd Unit 6, Escrick Business Park Riccall Road, Escrick York, YO19 6FD United Kingdom
Date completed	08/12/2022

Source of funding: This report was commissioned by the NIHR Evidence Synthesis Programme as project number NIHR135745.

Declared competing interests of the authors None

Acknowledgements None

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:

Perry M, Armstrong N, Wetzelaer P, Krijkamp E, Stirk L, Al M, Wolff R. Upadacitinib for previously treated moderately to severely active Crohn's disease [ID4027]: a Single Technology Assessment. York: Kleijnen Systematic Reviews Ltd, 2022

Contributions of authors

Mark Perry acted as a systematic reviewer, critiqued the clinical effectiveness methods and evidence and contributed to the writing of the report. Nigel Armstrong acted as project lead and health economist/review manager on this assessment, critiqued the clinical effectiveness methods and evidence and contributing to the writing of the report. Maiwenn Al acted as health economic project lead, critiqued the company's economic evaluation and contributed to the writing of the report. Pim Wetzelaer and Eline Krijkamp acted as health economists on this assessment, critiqued the company's economic evaluation and contributed to the writing of the report. Lisa Stirk critiqued the search methods in the submission and contributed to the writing of the report. Robert Wolff critiqued the company's definition of the decision problem and their description of the underlying health problem and current service provision, contributed to the writing of the report and supervised the project.

1. Summary of the EAG's view of the company's cost-comparison case

The Evidence Assessment Group (EAG) does not think that the company has demonstrated that upadacitinib (UPA) is equivalent to other technologies in the treatment of Crohn's Disease (CD), and therefore a cost-comparison case is not appropriate.

Upadacitinib is targeted as a second-line treatment in the advanced therapy stage, and therefore needs to be compared to other treatments that are targeted as second-line advanced therapy (Figure 1 of the company response to the request for clarification[1] shows the proposed treatment pathway, and is helpful to understand the issues discussed below).

The main problem is that not all appropriate second-line comparators have been included in the network meta-analyses (NMA), and so it is unknown if UPA is equivalent to all relevant second-line advanced therapy comparators. The comparators listed in the National Institute for Health and Care Excellence (NICE) final scope are the tumour necrosis factor alpha (TNF-alpha) inhibitors infliximab (IFX) and adalimumab (ADA), the biologics vedolizumab (VDZ) and ustekinumab (UST) and best supportive care. However, in the decision problem the comparators are restricted to the biologics VDZ and UST. The company justification for the exclusion of best supportive care as a comparator is strong: even if all biologics fail, the least ineffective of these will be used in the next line, making best supportive care a highly unlikely subsequent approach. In contrast, the rationale for the removal of TNF-alpha inhibitors is less robust. The company justifies the removal of TNF-alpha inhibitors (IFX and ADA) as comparators on the basis that the target population would not use TNF-alpha inhibitors for second-line advanced therapy. The company explains that this is because TNF-alpha inhibitors would have already been used for first-line advanced therapy, and that if a TNF-alpha inhibitor has failed it is not used again for second-line advanced therapy. The EAG does not accept this argument, because it is not true to say that first-line advanced therapy drugs will always be a TNF-alpha inhibitor. As shown in the company's response to clarification (Figure 1, company response to clarification[1]) UST is not a TNF-alpha inhibitor and yet it is used as a NICE-recommended first-line advanced therapy (as well as second-line advanced therapy) drug. This means that some patients will not receive TNF-alpha inhibitors for first-line advanced therapy, and so TNF-alpha-inhibitors can be regarded as appropriate second-line advanced therapy comparators for these patients. Failure to include all appropriate comparators recommended in the NICE scope means that it is unknown if UPA was comparable to all the appropriate comparators.

It could be argued that if UPA is shown to be comparable to UST, which has been shown to be cost-effective in this population, then this confirms that UPA is also cost-effective, and there is no need to involve other comparators. However, this ignores the fact that the TNF-alpha inhibitors, which are a different class of drug, may be more cost-effective than UPA, and therefore more appropriate for use in

this population. Therefore, excluding the TNF-alpha inhibitors from the comparators means that there is a risk that the non-optimal technology could be recommended.

Over and above the fact that not all the appropriate comparators are included in the NMA, additional issues remain for the NMA analyses concerning UPA and the two included comparators. Two principal issues are described below, which call into question the company's conclusion that equivalence exists between UPA and those particular comparators:

- Firstly, there was some doubt that clinical harm was equivalent between UPA and the two comparators. Although the outcome of serious adverse events demonstrated comparability between UPA and the two comparators, the NMA for discontinuation due to adverse events yielded point estimates that favoured the comparators. The credible intervals straddled the null line but demonstrated greater probability of a population effect denoting benefit for the comparators, rather than UPA.

- Secondly, heath-related quality of life (HRQoL) was not included as an outcome in the NMA, despite this being a highly relevant clinical effectiveness outcome for patients. The company's argument that there is no prior precedent from previous Single Technology Assessments (STAs) for including HRQoL is not relevant, because previous STAs are not necessarily reference standards of good practice and might reflect some discussions relevant to the respective intervention of interest (which might not apply to this submission). There is a strong methodological rationale for utilising HRQoL, because it is the most patient-focussed effectiveness outcome. The company also argues that HRQoL data were sparse in the included trials. This may be true, but the company should have used all available data, in order to allow the committee to interpret it. In addition, the NICE scope outcomes of mucosal healing, surgery, and hospitalisation rates were not subjected to NMA analysis. The failure to evaluate all the NICE scope outcomes is a serious limitation because it means that comparability cannot be assured; true comparability between treatments can only be confirmed if all relevant health outcomes are considered, particularly those that are patient-related such as HRQoL, surgery or hospitalisation.

2. Critique of the decision problem in the company's submission

In terms of population, the decision problem focuses only on a stratum of those previously treated – those who have previously failed on biologics or for whom TNF-alpha inhibitors were deemed unsuitable - even though the NICE scope makes no distinction between previous failure on biologics (BF) or conventional care failure (CCF) in its definition of the population. This narrowing of the scope was planned pre-hoc, and so it cannot be regarded as a biased response to initial results on an unstratified population. Nevertheless, because of the very different efficacy in the two strata, with the NMAs demonstrating inferior efficacy for the CCF stratum, it is vital not to base recommendations for both strata on the data from the BF stratum.

Only participants achieving a clinical response in U-EXCEL and U-EXCEED were eligible for inclusion in U-ENDURE. This would be non-representative of the target population in this submission, who are not people who have previously responded to the study drug. The EAG understands that it might be considered unethical for patients who did not respond, to continue to be followed up on the arm to which they were originally randomised. It might also be of interest to understand whether there is benefit to maintenance treatment (as opposed to curtailment of treatment) on achieving induction. However, the fact remains that the population are not relevant to the decision problem. In addition, the populations in the various maintenance trial comparisons are intrinsically different in terms of the drug to which they have responded. This level of clinical heterogeneity across comparisons may make an NMA approach inappropriate, and therefore makes any results derived from an NMA potentially invalid. The EAG therefore thinks that maintenance data should not be considered in this submission.

As argued in the previous Section, the inappropriate exclusion of appropriate comparators in the decision problem means that the NMA results cannot demonstrate that UPA is equivalent to all relevant comparators.

In terms of outcomes, mucosal healing is not included as an outcome in the decision problem despite being in the NICE scope. Instead, the outcome 'endoscopic outcomes' is used, which is supposed to include multiple outcomes indicative of mucosal healing. The EAG does not agree that 'endoscopic outcomes' is a useful term to encompass the construct of 'mucosal healing', as it appears to be an overly non-specific term. Similarly, 'surgery' is not included as an outcome in the decision problem despite being in the NICE scope. No surgery data were available in the UPA trials. This is a limitation in the submission because the need for surgery is a highly relevant clinical outcome. Its omission means that a full evaluation of UPA and its comparators is not possible.

The NICE scope suggested that stratification for CD location should be carried out. However, the company did not include stratification for CD location in the decision problem. This was partially because the studies were not powered for such an analysis. However, other sub-grouping analyses were

carried out without the study being powered for them either and so underpowering appears to represent a weak rationale. The company also referenced expert clinical opinion deeming CD location not clinically relevant, but this is not the opinion of NICE who stipulated that CD location should be a sub-grouping criterion. The company provided a sub-group analysis for CD location in response to clarification questions. This analysis did not reflect expert opinion, showing that location of CD was a potential outcome modifier, with ileal CD responding less well to UPA (relative to placebo) than other locations. No NMA was carried out for this but given the available evidence suggesting no benefit over placebo, it appears that UPA is not effective in this region. It is important that this is considered when making recommendations.

3. Summary of the EAG's critique of clinical effectiveness evidence submitted

The company does not detect any evidence of risk of bias in the three UPA trials, nor the seven trials involving the two comparators. The EAG has looked at the clinical study report (CSR) for each of the three studies[2-4] and agrees that the risk of bias is likely to be low. However, there is a lack of clarity around allocation concealment, because it is not made clear that those recruiting participants were unaware of the randomisation sequence, even though this is implied by the randomisation schedule being generated by the statistics department at AbbVie.

The evidence synthesis conducted by the company was of a good standard. Identified studies were assessed by two blinded, independent researchers in parallel using the pre-defined inclusion/exclusion criteria. Any discrepancies were resolved by a third party. Data from included studies were extracted into a pre-defined Excel-based template by a single analyst and all results were checked for accuracy by a senior reviewer.

Network meta-analyses were only conducted for clinical remission, clinical response, serious adverse events and discontinuation due to adverse events. The NICE scope outcomes of mucosal healing, surgery, hospitalisation rates and HRQoL were not subjected to NMA analysis. The failure to evaluate all the NICE scope outcomes is a serious limitation because it means that comparability cannot be assured; true comparability can only be confirmed if all relevant health outcomes are considered, particularly those that are patient-related such as HRQoL, surgery or hospitalisation.

The induction NMAs conducted for clinical remission, clinical response, serious adverse events and discontinuation due to adverse events demonstrated varying results.

- For clinical remission, there was fairly clear evidence of superiority of UPA over the two included comparators, but this was only observed in the BF stratum. This was conducted with a fixed effect (FE) NMA analysis, which was appropriate given the similarity of Deviance Information Criteria (DIC) values in the FE and random effects (RE) models.

- For clinical response, an FE NMA also demonstrated evidence of an advantage to UPA versus the two included comparators, although again this was only seen in the BF stratum. However, an RE NMA approach may have been more appropriate for the outcome of clinical response because of clinical heterogeneity between comparisons, combined with a DIC value that was 2.91 lower for the RE model than the FE model. Spiegelhalter et al. 2002[5] state that lower DIC values are preferred and typically differences of at least 3 points are considered meaningful. As the DIC difference is very close to 3, and the difference in Dbar is also over 6 points, the EAG would question the decision to use an FE model for this outcome. Use of the RE approach no longer demonstrated a clear benefit of UPA over the comparators for clinical response, but did show evidence of comparability, with a

point estimate favouring UPA, and most of the credible interval lying in the zone in favour of UPA. Therefore, it could be argued that if a FE model is believed to be more appropriate for this outcome, then the company have been conservative for this outcome in the NMA in assuming equivalence. If the RE model is believed to be appropriate for this outcome, then the result would still be consistent with equivalence, although only for this particular outcome.

- Both the induction safety outcomes were appropriately analysed with an FE model. For the outcome of serious adverse events, comparability was evident.

- However, for discontinuation due to adverse events, the point estimates in both the RE and FE NMAs favoured the comparators, and the credible intervals were consistent with a higher probability that the true population effect would favour the comparators.

- Maintenance NMAs were similar, but because the population for these was outside the decision problem (as argued previously) the results from these are not regarded by the EAG as relevant.

About 20% of patients were excluded from the U-EXCEL and U-EXCEED trials in the NMA. This restriction was aimed at increasing coherence between comparisons in terms of Crohn's Disease Activity Index (CDAI) score. However, this methodology may also have had the potential to affect the external validity of the NMA results. For the restriction of participants to adversely affect external validity two conditions would need to be fulfilled:

- Firstly, the restricted cohort would need to be shown to be different to the United Kingdom (UK) target population. It is conceivable that the unrestricted cohort could be closer to the UK target population in terms of CDAI score than the restricted population, on the simple grounds that the UK target population are also unrestricted. However, no data are available on the CDAI scores of the UK target population, and so this assumption cannot be confirmed.

- Secondly, a clear difference in results between restricted and unrestricted analyses would be needed. This would demonstrate that if the UK target population were more akin to the unrestricted population, then results derived from a restricted population would be less applicable to them. There was a trend for the efficacy results to be more beneficial towards UPA in the restricted analysis than the unrestricted analysis, but this effect was not large and did not change interpretations: in both restricted and unrestricted efficacy analyses there was either clear evidence of superiority for UPA over comparators, or a demonstration of equivalence. Therefore, the EAG concludes that it is unlikely that the exclusion of participants will have affected external validity to any great extent, and so the benefits accrued from improved coherence between comparisons in the NMA are unlikely to be significantly affected.

4. Summary of the EAG's critique of cost evidence submitted

4.1 Decision problem for cost comparison

As outlined in Section 2, the current analysis only considers one of the two sub-populations that were defined in the NICE scope, and in that regard, the current cost comparison can be considered as incomplete.

The analysis compares UPA with UST and VDZ. As stated in Section 1 above, first-line biologic failure does not necessarily involve a TNF-alpha-inhibitors, as UST may be used as first-line biologic as well. Hence, the TNF-alpha inhibitors IFX and ADA can also be regarded as appropriate comparators second-line. This means that the current cost comparison is incomplete.

4.2 Cost comparison model

The Excel model that was developed for the cost comparison has a time horizon of 1 year, with the option to also include the costs in each year of treatment beyond year 1. It is important to note though, that no clinical effectiveness data are available to inform the relative effectiveness and safety in the second year of treatment.

The model calculates the induction and maintenance costs for patients receiving UPA, UST, or VDZ. In this calculation the patient is assumed to have responded to induction treatment and proceed to receive maintenance treatment. The base case includes induction and maintenance treatment, and reflects the cost of the patient's first year on treatment, while the Year 2+ scenario reflects the cost of additional years on maintenance treatment only (these maintenance costs are assumed to be the same in all years after year 1)).

Alternative pathways, such as patients not responding to induction treatment or patients discontinuing treatment due to adverse events, relapse, or death are not incorporated in the model. This is in contrast to some previous appraisals – TA521, TA596, TA723 and TA803 - where a cost comparison was considered.[7, 8, 9, 10]

Presumably this modelling choice is based on the assumption that UPA, UST and VDZ can be considered equivalent in terms of efficacy and safety. However, by only including the pathway of patients successfully treated over the time horizon, the differences in costs will appear larger than when also less successful pathways are included (assuming costs of a potential next treatment are the same for all three treatments being compared).[1] When interpreting the magnitude of the result of the cost comparison it is important to keep this in mind.

[1] For example, assume the costs of full treatment (induction and maintenance) are £500 for treatment X and £1,000 for treatment Y. If we assume 70% of patients follow this pathway, whereas 30% does not respond to induction (at costs of £300 and £500 for X and Y, respectively), then the total average costs for treating patients with X are £440 and with Y £850.

Not only does the focus on successfully treated patients lead to estimated savings that cannot be extrapolated to all patients starting treatment with UPA, UST and VDZ, it also disregards the issue of treatment sequencing and thus the downstream costs. Due to the various mechanisms of action of the three drugs considered here, it is unclear what treatment would be given as the third-line option and how this would impact the cost comparison.

4.3 Model parameters

The parameter values used in the company's cost-comparison analysis are presented in the CS,[6] Tables 62-65. A summary of the key model parameters is presented in Table 67 of the CS. The main model assumptions are summarised in Table 68 of the CS.

i. *Weight*

For patients receiving UST, the IV dosage depends on the weight of the patient. Thus, the company did a post-hoc analysis of BF patients in U-EXEL and U-EXCEED to find the distribution of patients in the ≤55 kg, >55 kg and ≤85 kg, >85 kg weight bands (see Table 62, CS[6]).

ii. *Distribution high and low dose maintenance*

For UPA, UST, and IV VDZ patients may receive a low or a high dose during the maintenance phase of the treatment, and the distribution varies by treatment. For the cost comparison, the company has sought expert opinion regarding the distribution of patients between low and high dose (see Table 63, CS[6]). According to the experts on the Health Technology Assessment (HTA) Advisory Board for Risankizumab (RZB),[11] UST is mostly given in a high dose, 92.5%, whereas for UPA a high dose is given to 30% of the patients, for VDZ intravenous (IV) a high dose to 22% of the patients. For VDZ subcutaneous (SC) this is 0%. This is in line with the company submission, where only a fixed dose for SC VDZ is applied.

It should be noted that in later expert interviews, it was suggested that for VDZ IV high dose maintenance would be given to 30% of the patients. The company has used the latter value for the base-case analysis but has provided a scenario analysis using 22% in their response to the clarification letter (Question B8).

iii. *Acquisition costs*

Upadacitinib is administered orally, during induction (12 weeks) at 45 mg per day and during maintenance at 15 mg or 30 mg per day.

Thus, when only looking at successful patients a savings of £500 would be anticipated, but based on the mixture of more and less successfully treated patients, a savings of £410 would be achieved.

For the price of UPA a simple Patient Access Scheme (PAS) was agreed with National Health Service (NHS) England leading to the following prices: 45 mg = [redacted]; 30 mg = [redacted]; 15 mg = [redacted].

Ustekinumab is administered by IV during induction, with a single dose of, on average, 3*130 mg. During maintenance (starting at week 8, patients receive 90 mg SC either once per 12 weeks or once per 8 weeks. The list prices are: 130 mg (IV) = £2,147; 90 mg (SC) = £2,147.

Vedolizumab is given by IV during induction, as a dosage of 300 mg in weeks 0, 2 and 6. During maintenance (starting in week 14), it may be given by IV at a dose of 300 mg either once per 8 weeks or once per 4 weeks, or it can be administered via SC injections at a dose of 108 mg once every 2 weeks. The list prices are: 300 mg = £2,050 and 108 mg = £512.50.

Note that the confidential prices for the comparators are presented in the confidential appendix.

In the dosing schedules presented above the standard induction period has been used. Depending on the level of response to the induction treatment, the induction period may be extended. However, the company expects this to concern a minority of patients based on the clinical response rate in the first 12-week induction period in the UPA trials.[2, 3] Furthermore, clinical experts indicated that patients with an inadequate response would be more likely to switch to a different advanced therapy/biologic than receive extended induction.[11] Thus, the company excluded the extended induction from the base-case and instead included it in a scenario analysis.

For UPA, extended induction is 30 mg administered once daily for an additional 12 weeks (i.e., to Week 24) following inadequate response to standard induction therapy. The VDZ extended induction includes an additional 300 mg IV dose at week 10. The extended induction dose of UST is 90 mg and is administered at week 8. Since the maintenance dose of UST of 90 mg is also administered at week 8, the company has assumed in the model that any patients requiring extended induction of UST effectively receive a double dose (twice 90 mg) at week 8.

It is not clear to the EAG that this approach to extending induction with UST by giving a double dose at week 8 is indeed used in clinical practice, as the CS did not provide any references nor did the EAG find any confirmation that this dosing schedule may be used to extend induction.

iv. Administration costs

It was assumed that oral therapy is not associated with any administration costs. For IV treatment, the company assumed that the HRG code FD02H Inflammatory Bowel Disease without Interventions, with CC Score 0 would apply, at £291 per administration.[12]

For SC costs it was assumed that costs would only be incurred at the first administration, since patients will self-administer the subsequent injections. These initial administration costs were estimated at £44.

The EAG concurs with the assumption that SC administration will only incur costs the first time. For the costs of IV administration, the EAG compared the current approach with that used in previous STAs. For example, in TA633 (UST for treating moderately to severely active ulcerative colitis)[13] the tariff for an outpatient visit was used, which amounted to £142.

Recently a paper was published looking into the costs of IV and SC administration of biologics.[14] In that paper it was pointed out how various studies use different tariffs for the IV administration of biologics, as no specific tariff code is available for this procedure. It was put forward that most often tariffs for IV chemotherapy administration are used, with tariffs ranging from £142 to £426, with the latter value for 'Deliver Complex Chemotherapy, including Prolonged Infusional Treatment, at First Attendance'.

Based on interviews with various stakeholders, a micro-costing approach was used to estimate the IV administration costs, which amounted to £414 if only in-tariff costs were included.

From the above, it is clear that the estimates for administration costs for IV biologics can differ between studies and no clear unique tariff is currently being used. If the cost estimate from the (expert opinion based) micro-costing study is used, the 1-year administration costs for UST and VDZ, will be higher than estimated in the current model. In an EAG scenario, we will explore how the costs change when the lower value of £142 is used.

4.4 EAG model check

The EAG conducted a range of checks on the company's cost-comparison model. This included a verification that the dosing scheme of the treatments in Excel matched the described scheme in the CS[6] and verification that the costs are in line with the costs described in the CS[6] (see CS, Table 64). We also performed an inspection of the formulae used in Excel.

Main observations:

- The model does not have any input parameters related to efficacy and safety that are informed by data from the trials.
- The calculated dosing scheme for UPA is in line with what is stated in the CS[6] (page 10, CS[6]).
- The dosing scheme for the comparators is described in less detail in the CS[6] (page 16, CS[6]). Although the CS[6] states that patients on a low dose can switch during treatment to a high dose this is not modelled. In the model it is assumed that those that end up with a high dose in the maintenance phase will do so since the start of the maintenance phase.

Other observation:

- In the CS[6] it is stated that for patients over 65 years UPA should only receive the 15 mg dose in the maintenance phase. As age is not part of the model this dose recommendation is not explicitly taken into account. The EAG acknowledges that it is possible that this dose limitation has implicitly been taken into account in the applied distribution between 15 mg (70%) and 30 mg (30%) during maintenance treatment.

Minor model errors, none of which affected the results:

- There is a hardcoding error in the calculation of the number of subsequent administrations for doses of VDZ SC. However, since only the first SC administration incurs costs, this error has no effect on the results.

- There is a reference error in the formula to estimate the number of dosages for UPA for standard and extended induction, in order to estimate the administration costs. However, the administration costs per dose are 0, because UPA is an oral drug. In addition, despite using the wrong cell reference, the value that is return is still correct, so the results are not affected.

4.5 Company's model results

The company base-case cost comparison results compare the 1-year results for UPA, UST, and VDZ both IV and SC. For UPA the PAS price was used whilst list prices were used for UST and VDZ (see CS,[6] Table 69). Results using discounted prices for UST and VDZ as well can be found in the confidential appendix.

Uncertainty over model assumptions was assessed with a range of scenario analyses (CS, Tables 70-76, response to clarification letter Tables 12 and 13). No subgroup analyses were performed.

The results of the company's base-case analysis indicated that UPA is a cost saving strategy compared to UST and both versions of VDZ (IV and SC) (see CS,[6] Table 69). The estimated base-case costs by the company are [redacted] for UPA, £19,336 for UST, £22,942 VDZ IV and £16,805 VDZ SC.

The conclusion that UPA is a cost saving strategy compared to UST and VDZ (IV and SC) applies also to all the sensitivity analyses performed by the company (see CS,[6] Tables 70–75). A complete overview of all results is presented in Table 1 below.

Table 1: Company base-case and scenario results

	Costs UPA (PAS price)	Costs UST (list price)	Costs VDZ IV (list price)	Costs VDZ SC (list price)
Company base-case	[redacted]	£19,336	£22,942	£16,805
Scenario results from CS				
Scenario 1: Year 2+ costs	[redacted]	£13,607	£19,781	£13,325
Scenario 2a: 100% on low dose maintenance of UPA 15 mg	[redacted]	£19,336	£22,942	£16,805
Scenario 2b: 0% on low dose maintenance of UPA 15 mg	[redacted]	£19,336	£22,942	£16,805
Scenario 3a: 0% on UST standard maintenance dose	[redacted]	£19,658	£22,942	£16,805
Scenario 3b: 20% on UST standard maintenance dose	[redacted]	£18,799	£22,942	£16,805
Scenario 3c: 30% on UST standard maintenance dose	[redacted]	£18,370	£22,942	£16,805
Scenario 4: Extended induction	[redacted]	£21,527	£24,581	£19,146
Additional scenario results from clarification response				
Scenario CR1: Extended induction with 100% on high maintenance dose	[redacted]	£21,849	£32,774	£19,146
Scenario CR2: 22% on VDZ IV high maintenance dose[a]	[redacted]	£19,336	£21,818	£16,805

	Costs UPA (PAS price)	Costs UST (list price)	Costs VDZ IV (list price)	Costs VDZ SC (list price)
EAG scenario results				
Scenario EAG1a: 100% on low dose maintenance of VDZ IV	[redacted]	£19,336	£18,728	£16,805
Scenario EAG1b: 0% on low dose maintenance of VDZ IV	[redacted]	£19,336	£32,774	£16,805
Scenario EAG2: Cost IV administration £142	[redacted]	£19,187	£21,482	£16,358

CR = clarification response; CS = company submission; EAG = External Assessment Group; IV = intravenous; mg = milligram; PAS = Patient Access Scheme; SC = subcutaneous; UPA = upadacitinib; UST = ustekinumab; VDZ = vedolizumab

[a]These results were corrected by the EAG, because the results as reported by the company in their clarification results were erroneously based on the "Extended induction" setting.

4.6 EAG exploratory analysis

The EAG undertook three additional exploratory analysis using the company's original submitted Excel model. The analysis presented in this Section reflects the PAS discount price for UPA whilst list prices were used for UST and VDZ. Results using discounted prices for UST and VDZ as well can be found in the separate confidential appendix.

Since the company only changed the percentage of patients receiving VDZ high dose maintenance to 22%, and not the more extreme limits of 100% and 0% as was done in Scenarios 2a and 2b for UPA, the EAG explored the impact of these more extreme values.

In addition, the model was amended to assess the impact of using a lower estimate of IV administration costs, £142, on the results.

For all these scenarios UPA remains cost-saving.

4.7 EAG conclusion

The EAG considers the current cost comparison incomplete as the the TNF-alpha inhibitors IFX and ADA can also be regarded as appropriate comparators in the second-line for the BF population.

In addition, compared to the NICE scope the cost comparison may be regarded as incomplete as only the BF population is regarded. However, the only input estimated from the trials is the weight

distribution of the patients, so the impact of limiting the population on the overall conclusions reading costs will be minimal.

In the current model, only the pathway of patients successfully treated over the time horizon is included, alternative pathways, such as patients not responding to induction treatment or patients discontinuing due to adverse events, relapse, or death are not incorporated in the model. Even if all treatments can be considered equivalent in terms of efficacy and safety, by only including the pathway of patients successfully treated over the time horizon, the differences in costs will appear larger than when also less successful pathways are included (assuming costs of a potential next treatment are the same for all three treatments being compared). When interpreting the magnitude of the cost difference resulting from the cost comparison it is important to keep this in mind.

With list prices for all treatments, UPA is estimated to be cost saving compared to the comparators UST and VDZ. This applies for the company's base-case analysis and for all company and EAG scenario analyses. Results with discounted prices for all treatments are shown in a confidential appendix to this report.

5. EAG commentary on the robustness of evidence submitted by the company

The company's evidence is not robust enough to confirm comparability of efficacy and safety between UPA and all appropriate comparators. To summarise points made previously:

- Not all the appropriate comparators have been included. The company's justification for not including TNF-alpha inhibitors as second-line comparators (because TNF-alpha inhibitor comparators would be used first-line, and so would not be able to be used second-line) was insufficient because it ignored the fact that TNF-alpha inhibitors are not the only biologics given first-line. Without all appropriate comparators included it is impossible to know if UPA is comparable to all such comparators.

- Network meta-analyses were not conducted for all the relevant outcomes. In particular HRQoL should have been included as it is the key clinical effectiveness outcome. Justification for the omission of relevant outcomes was weak. Without inclusion of all appropriate outcomes, it is impossible to ascertain true comparability between UPA and its comparators.

- The NMAs that were carried out were not all conducted optimally. The NMA for induction clinical response used an FE model when an RE model would have been more appropriate.

- The results from the NMA for discontinuation due to adverse events did not suggest comparability.

- Results for the maintenance data are not relevant to the decision problem population, as they comprised responder data only. Though such data were inevitable for ethical and pragmatic reasons, the use of responder data does mean that the data are not applicable to the decision problem in this submission.

6. References

[1] AbbVie. Upadacitinib for previously treated moderately to severely active Crohn's disease [ID4027]: Response to request for clarification from the ERG: AbbVie, 2022. 105p.

[2] AbbVie. A multicenter, randomized, double-blind, placebo controlled induction study of the efficacy and safety of upadacitinib (ABT-494) in subjects with moderately to severely active Crohn's disease who have inadequately responded to or are intolerant to conventional and/or biologic therapies (U-EXCEL [M14-433]). Clinical study report, 2022

[3] AbbVie. A multicenter, randomized, double-blind, placebo-controlled induction study of the efficacy and safety of upadactinib (ABT-494) in subjects with moderately to severely active Crohn's disease who have inadequately responded to or are intolerant to biologic therapy (U-EXCEED [M14-431]). Clinical study report, 2022

[4] AbbVie. A multicenter, randomized, double-blind, placebo-controlled maintenance and long-term extension study of the efficacy and safety of upadacitinib (ABT-494) in subjects with Crohn's disease who completed the studies M14-431 or M14-433 (U-ENDURE [M14-430]). Clinical study report, 2022

[5] Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. J R Stat Soc Series B Stat Methodol 2002;64:583-639.

[6] AbbVie. Upadacitinib for previously treated moderately to severely active Crohn's disease [ID4027]: Submission to National Institute of Health and Care Excellence. Single technology appraisal (STA): Document B - Company evidence submission: AbbVie, 2022. 125p.

[7] National Institute for Health and Care Excellence. Guselkumab for treating moderate to severe plaque psoriasis [TA521]. London: NICE, 2018 Available from: https://www.nice.org.uk/guidance/ta521

[8] National Institute for Health and Care Excellence. Risankizumab for treating moderate to severe plaque psoriasis [TA596]. London: NICE, 2019 Available from: https://www.nice.org.uk/guidance/ta596

[9] National Institute for Health and Care Excellence. Bimekizumab for treating moderate to severe plaque psoriasis [TA723]. London: NICE, 2021 Available from: https://www.nice.org.uk/guidance/ta723

[10] National Institute for Health and Care Excellence. Risankizumab for treating active psoriatic arthritis after inadequate response to DMARDs [TA803]. London: NICE, 2022 Available from: https://www.nice.org.uk/guidance/ta803

[11] AbbVie. Crohn's disease HTA advisory board report (MAP ID: 2021-GB-MNP-00091) [Data on file], 2021

[12] NHS England. National tariff payment system [Internet]. Available from: https://www.england.nhs.uk/pay-syst/national-tariff/national-tariff-payment-system/

[13] National Institute for Health and Care Excellence. Ustekinumab for treating moderately to severely active ulcerative colitis [TA633]. London: NICE, 2020 Available from: https://www.nice.org.uk/guidance/ta633

[14] Heald A, Bramham-Jones S, Davies M. Comparing cost of intravenous infusion and subcutaneous biologics in COVID-19 pandemic care pathways for rheumatoid arthritis and inflammatory bowel disease: a brief UK stakeholder survey. Int J Clin Pract 2021;75(9):e14341.