

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Final draft guidance

Seladelpar for previously treated primary biliary cholangitis

1 Recommendation

1.1 Seladelpar can be used, within its marketing authorisation, as an option to treat primary biliary cholangitis, including pruritus, in adults. It can be used:

- with ursodeoxycholic acid (UDCA), if UDCA alone has not worked well enough, or
- alone, if UDCA cannot be tolerated.

What this means in practice

Seladelpar must be funded in the NHS in England for the condition and population in the recommendations, if it is considered the most suitable treatment option. Seladelpar must be funded in England within 90 days of final publication of this guidance.

There is enough evidence to show that seladelpar provides benefits and value for money, so it can be used routinely across the NHS in this population.

Why the committee made this recommendation

Usual treatment for primary biliary cholangitis is UDCA. If this does not work well enough, or is not tolerated, add-on treatments include obeticholic acid or elafibranor.

Clinical-trial evidence shows that seladelpar reduces liver enzymes, which are raised in primary biliary cholangitis, more than placebo. This suggests that seladelpar could

delay the condition getting worse. The evidence also shows that seladelpar reduces pruritus (itch) compared with placebo.

Seladelpar has not been directly compared in a clinical trial with obeticholic acid or elafibranor. But the results of indirect comparisons suggest that seladelpar may reduce itch more than obeticholic acid. And it may reduce liver enzymes more than obeticholic acid or elafibranor.

The cost-effectiveness estimates are below what NICE considers an acceptable use of NHS resources. So, seladelpar can be used.

2 Information about seladelpar

Marketing authorisation indication

- 2.1 Seladelpar (Livdelzi, Gilead) is indicated for 'the treatment of primary biliary cholangitis (PBC), including pruritus, in adults in combination with ursodeoxycholic acid (UDCA) who have an inadequate response to UDCA alone, or as monotherapy in those unable to tolerate UDCA'.

Dosage in the marketing authorisation

- 2.2 The dosage schedule is available in the [summary of product characteristics for seladelpar](#).

Price

- 2.3 The list price of seladelpar is £3,155.00 per 30-tablet pack of 10-mg tablets (excluding VAT; company submission).
- 2.4 The company has a commercial arrangement (simple discount patient access scheme). This makes seladelpar available to the NHS with a discount. The size of the discount is commercial in confidence.

Sustainability

- 2.5 Information on the Carbon Reduction Plan for UK carbon emissions for Gilead will be included here when guidance is published.

3 Committee discussion

The [evaluation committee](#) considered evidence submitted by Gilead, a review of this submission by the external assessment group (EAG), and responses from stakeholders. See the [committee papers](#) for full details of the evidence.

The condition

Primary biliary cholangitis

- 3.1 Primary biliary cholangitis (PBC) is a chronic, progressive autoimmune condition that leads to a build-up of bile in the liver. This happens because the body's immune system destroys bile ducts in the liver, causing cholestasis. This means the flow of bile through the liver and biliary system is reduced. Over time, cholestasis leads to scarring of the liver (fibrosis and cirrhosis) and liver failure, and can ultimately lead to death. The cause of PBC is not known but is thought to be a mix of environmental and genetic factors. PBC is typically diagnosed by testing for biochemical indicators of liver function such as alkaline phosphatase (ALP). Many people do not have symptoms until they have significant liver damage. Common symptoms include itchy skin (pruritus) and fatigue. Around 20,000 people in the UK have PBC, with an annual incidence of 2 to 3 new cases per 100,000. PBC is more common in women (90%) and in people over 40 (75%). The patient group submissions described the challenges of living with PBC, such as severe fatigue and severe itching. The patient expert elaborated that there is both the physical fatigue felt in the body, affecting movement and the ability to do daily activities, and the mental and physical exhaustion that comes with itch. The patient expert and patient group submissions emphasised that the chronic symptoms greatly affect people with PBC and their families and carers. They explained that symptoms affect quality of life, sleep and the ability to work or manage daily activities. The patient experts reported that people with PBC often experience a delayed diagnosis, which can lead to feelings of isolation, confusion and frustration because of unexplained symptoms. Although having a diagnosis can bring relief, the rarity of the condition

means that many people have never heard of it, and this can reinforce feelings of isolation. PBC can progress unpredictably to cirrhosis or liver cancer, with some people eventually needing a liver transplant. This adds to feelings of anxiety and uncertainty. The committee concluded that PBC has a substantial effect on people's lives.

Clinical management

Treatment pathway and positioning of seladelpar

3.2 There are no NICE guidelines specifically for treating PBC. The most relevant guidelines are the [British Society of Gastroenterology/UK-PBC primary biliary cholangitis treatment and management guidelines](#). These were developed before the evaluation for [NICE's technology appraisal guidance on elafibranor for previously treated primary biliary cholangitis](#) (from now, TA1016). The clinical expert explained that first-line treatment for PBC is UDCA. People whose PBC has an inadequate response to UDCA (defined as ALP above 1.67 times the upper level of normal [ULN]) can have second-line obeticholic acid (OCA), elafibranor, or fibrates (off-label treatment), with or without UDCA. People who cannot tolerate UDCA have OCA, elafibranor, or fibrates (off-label). OCA is recommended in [NICE's technology appraisal guidance on OCA for treating primary biliary cholangitis](#) (from now, TA443) and elafibranor is recommended in TA1016. The care pathway is structured in a way that to start licensed second-line treatments, people must first be seen by a multidisciplinary team. The patient organisations explained there are frustrations with variation in care and difficulties accessing specialist teams, especially for about 40% of people who need second-line treatment after UDCA. Access to these treatments can vary and some, like OCA, have more side effects. Treatments for symptoms, such as colestyramine for itching, can also be difficult to take. The company proposed that seladelpar be positioned as a second-line treatment for people who have an inadequate response to, or cannot tolerate, UDCA. It said that it may also be considered as a third-line option for people who cannot tolerate, or whose condition has not responded, to OCA. The clinical expert noted that seladelpar and

elafibranor have similar mechanisms of action. The reported benefit of seladelpar in normalising ALP levels means the clinical expert would be keen to use it as early in the treatment pathway as possible. The clinical expert noted OCA and seladelpar may benefit people differently: people with high transaminase levels may benefit from having OCA, and people with pruritus may benefit from having seladelpar. This is because OCA has an increased risk of worsening pruritus, and seladelpar may reduce pruritus. The committee noted that the clinical effectiveness of seladelpar in later lines of treatment is uncertain because data on third-line use is limited. The committee concluded that seladelpar would be positioned as a second-line treatment and that it may also be used at third line.

Excluding fibrates as comparators

3.3 Fibrates are anti-pruritic agents, but they are not licensed for treating PBC. They were not included as comparators in NICE's final scope for this evaluation. In TA443 and TA1016, fibrates were not considered to be second-line treatments. Instead, they were viewed as add-on treatments for pruritus. In TA1016, bezafibrate was included in the economic model only for its role in treating pruritus. It was not considered to be a standalone second-line treatment in that context. The EAG advised that fibrates were a relevant comparator. It explained that 1 of its 3 clinical experts had stated that fibrates were their preferred second-line treatment in clinical practice. It also noted the [results of a UK-wide, population-based evaluation of care delivery \(2024\)](#), which found that 50% of people whose PBC had responded inadequately to UDCA had fibrates. The clinical expert at the first committee meeting agreed that fibrates were used as a second-line treatment. They noted that 20% to 25% of people use fibrates as a second-line treatment, including a small proportion who use them as monotherapy at this position in the treatment pathway. The company disagreed that fibrates were a comparator, noting the inconsistency with previous appraisals and that fibrates are not licensed for treating PBC. The committee acknowledged that:

- a particular benefit of seladelpar is the reduction in pruritus. The marketing authorisation for seladelpar states that it is for treating PBC including pruritus, but the marketing authorisations for elafibranor and OCA state that these treatments are for treating PBC only
- people are having fibrates at second line in the NHS, as an alternative to OCA
- NICE's methods allow consideration of off-label treatments as comparators if there is evidence of their use in clinical practice.

At the first meeting, the committee decided it would be informative to see a comparison of benefits and costs of seladelpar compared with fibrates. The scope was re-issued, listing fibrates as a potential comparator to allow exploration of this.

After consultation, the company maintained its arguments around the inconsistency with previous appraisals, lack of licensing of fibrates for the treatment of PBC and lack of clear evidence suggesting fibrates are part of established clinical practice. It noted that it would not be feasible to compare the clinical benefits of seladelpar and fibrates using a network meta-analysis (NMA), because of heterogeneity between the trials identified in its systematic review. Differences include the inclusion criteria, baseline characteristics and outcome definitions. A committee member noted that an indirect treatment comparison (ITC) including fibrates could be informative if adverse-event data was available and suitable for comparison within an NMA. But they supported the company's position that an NMA comparing fibrates with seladelpar may not be appropriate because of limitations in the available data. The EAG maintained that fibrates are a potentially appropriate comparator because its clinical experts had stated that fibrates are used in some centres in the NHS. It advised that fibrates have a similar mechanism of action to seladelpar because they are also peroxisome proliferator-activated receptor agonists. So, fibrates could be used to treat PBC and not just as an add-on treatment for pruritus.

In response to consultation, both the patient and professional groups stated concerns about the inclusion of fibrates as a comparator. These included a lack of comparative evidence with current PBC treatments, safety concerns, their unlicensed status and varied access. They noted that including fibrates could risk reducing access to licensed treatments such as seladelpar, for which more robust evidence is available. In response to consultation, NHS England stated that fibrate use is variable across the UK and is not part of standard clinical practice. It also advised that including fibrates as a comparator could limit access to licensed treatments. [NICE's manual on health technology appraisal and highly specialised technologies guidance](#) states that unlicensed treatments may be considered as comparators if they are part of established NHS clinical practice. The committee concluded that fibrates are not an appropriate comparator, because stakeholders advised that fibrates are not routinely used in the NHS in England.

Clinical effectiveness

The RESPONSE trial

3.4 The main source of clinical-effectiveness evidence for seladelpar was from RESPONSE. This was a phase 3, randomised, double-blind, placebo-controlled trial that lasted for 12 months. It evaluated seladelpar in people with PBC who had an incomplete response to, or could not tolerate, UDCA. A total of 193 people were enrolled in the trial. Of these, 128 people had seladelpar 10 mg, with a reduced 5 mg dose used in cases of intolerance. The other 65 people had placebo. Both the seladelpar and the placebo arms could include UDCA use, and about 94% of people in each arm had UDCA. There were 17% of people who had previously had OCA or fibrates. The primary outcome was the proportion of people having a composite biochemical response at month 12. This was defined as:

- ALP less than 1.67 times the ULN
- a reduction in ALP of at least 15%, and

- total bilirubin at or below 1.0 times the ULN.

Key secondary outcomes included the proportion of people with ALP normalisation (1.0 times at or below the ULN) at 12 months. The clinical and patient experts emphasised that an ALP level above normal, even if it was less than 1.67 times the ULN, was still associated with disease progression. So, they advised that ALP normalisation is an important outcome even if it is considered 'mild'. Another key secondary outcome was change from baseline in the weekly averaged pruritus numerical rating scale (NRS) score. The committee noted that although ALP levels were an important outcome in terms of disease progression, they may not always reflect the symptoms people with PBC experience. The committee concluded that the trial outcomes were informative for decision making.

RESPONSE results

3.5 In RESPONSE, 61.7% of people in the seladelpar arm met the primary outcome of composite biochemical response at 12 months compared with 20% in the placebo arm. This composite outcome was mostly driven by improvements in ALP-related measures. The EAG noted that baseline bilirubin levels were already low in both the treatment and placebo arms. This suggests that people in RESPONSE were probably at an earlier stage of disease, when changes in bilirubin would be less pronounced. At 12 months, 25% of people in the seladelpar arm and no people in the placebo arm had normalised ALP levels. Among people with moderate to severe pruritus at baseline, having seladelpar statistically significantly reduced the pruritus NRS score compared with placebo. There was a least-squares mean change from baseline to -3.2 with seladelpar, compared with -1.7 with placebo. The committee concluded that seladelpar is clinically effective at improving ALP levels and reducing pruritus compared with placebo.

Positive treatment response with placebo

3.6 In RESPONSE, 20% of people in the placebo arm had a meaningful clinical improvement in ALP response, despite having no active treatment other than background treatments. The committee noted that this placebo response was high. And it was not explained by changes in background treatment or UDCA dosing, which remained consistent with peoples' previous use. The EAG's clinical experts had noted that such placebo responses are common in PBC trials but they vary in size. A clinical expert suggested that improved adherence to UDCA during clinical trials may contribute to these effects. Better adherence could lead to better outcomes, even in the absence of new treatments. The committee suggested it was possible that the positive treatment response was because of regression to the mean. The EAG noted that this placebo response could introduce uncertainty when comparing data from RESPONSE with data from trials for other treatments, especially if placebo effects differ across trials. The EAG advised that there may also be uncertainty about real-world adherence over time and how well the trial results reflect NHS clinical practice. The committee agreed it was plausible that the placebo effect was caused by increased adherence to UDCA. But it noted that similar adherence to UDCA would also be expected in the seladelpar arm. The committee decided that the relative treatment-effect estimates were not likely to be biased by adherence to UDCA, because the same effect would be seen in both trial arms. It concluded that the reason for the observed placebo response was unknown and that this likely contributes uncertainty when comparing outcomes across the trials.

Indirect comparison approach

Bayesian NMA and MAIC

3.7 There are no head-to-head trials directly comparing seladelpar with the comparators included by the company in the final scope (OCA and elafibranor). Instead, the company did ITCs using data from the following trials:

- RESPONSE for comparing seladelpar plus UDCA with UDCA plus placebo
- ELATIVE for comparing elafibranor plus UDCA with UDCA plus placebo
- POISE, COBALT and NCT03633227 for comparing OCA plus UDCA with UDCA plus placebo.

The company used different methodological approaches for comparing seladelpar with each comparator. To compare seladelpar with OCA, the company used a Bayesian NMA. For the comparison with elafibranor, it used an anchored matching-adjusted indirect comparison (MAIC). The company decided that a Bayesian NMA was unsuitable for comparing seladelpar with elafibranor. This was because of differences in baseline bilirubin and cirrhosis rates between the RESPONSE and ELATIVE trials, which the company suggested violated the transitivity assumption. That is, the trials differed in ways beyond the treatments being compared, limiting the validity of indirect comparisons. The company noted there were differences in the definitions of ULN for the ALP measures and the specific ULN cut-offs by sex. The company recalculated the outcomes for seladelpar to adjust for differences in the sex-specific cut-offs between RESPONSE, POISE and ELATIVE. The MAIC method adjusted individual-patient data from RESPONSE to match the baseline characteristics of the ELATIVE population. Four treatment-effect modifiers were used for matching: age, baseline ALP, bilirubin and cirrhosis. The company explained at clarification that these modifiers were consistent with those used in TA1016 and were supported by literature and company expert opinion. The EAG did not agree it was justifiable to use a MAIC for comparing seladelpar and elafibranor. It explained that the forest plot supplied for ALP normalisation indicated that the most significant effect modifiers were cirrhosis and total bilirubin. But neither of these outcomes were statistically significant, because the confidence intervals overlapped.

Also, the forest plot did not provide clear evidence of effect modification

for either age or baseline ALP. The EAG also stated that it was important to consider if differences in the effect modifiers between the trial populations were clinically meaningful. The EAG's clinical experts had advised that the differences in baseline ALP between the 2 trials were small and unlikely to favour any treatment. The EAG also noted that the MAIC resulted in a small effective sample size (36% of the original sample), suggesting that baseline differences between the populations in RESPONSE and ELATIVE were difficult to reconcile with matching. This may introduce an additional source of bias and uncertainty into the treatment-effect estimates. So, the EAG preferred to use Bayesian NMAs for the indirect comparisons ([see section 3.8](#)).

NMA uncertainty

3.8 The company considered the treatment-effect estimates produced by the NMA to be confidential, so they cannot be reported here. But, during the first committee meeting the committee noted the extremely large credible intervals around the relative treatment-effect estimates for seladelpar compared with elafibranor and with OCA in the ALP outcomes. The intervals were particularly wide for the comparison with OCA. This indicated there was high uncertainty around the results and raised concerns about the validity of the model. In response to consultation the company updated its ITC report, which showed no patients had ALP normalisation in the placebo arms of ELATIVE, RESPONSE or POISE. The company explained it had included a continuity correction to account for zero values. The committee acknowledged that comparisons using a connection to placebo would be very uncertain, but this did explain the wide credible intervals for ALP normalisation.

At the first meeting, the committee noted that in the results reported in the individual trials (the naive results), seladelpar had a smaller estimated treatment effect than those for OCA or elafibranor. But in the Bayesian NMA, seladelpar had a larger treatment effect. The committee decided this lacked face validity and it had concerns about the analyses. In

response to these concerns and the committee's request to use Bayesian NMAs for all indirect comparisons, the company submitted an updated NMA exploring the comparative efficacy of seladelpar, elafibranor and OCA. The company did not incorporate the findings from this updated NMA into its base-case analysis. The company stated that it was not appropriate to update the base case because of differences in the definitions of ULN for ALP across the 3 key trials (RESPONSE, ELATIVE and POISE). This prevented the use of a single consistent network. The company reiterated its approach to address these differences by recalculating outcomes in RESPONSE (see [section 3.7](#)). The company explained that this led to the NMA results differing from the naive trial data and to a reversal in the direction of both absolute and relative treatment effects. The committee acknowledged that the definition of ALP response outcomes differed across the trials. It agreed with both the company and the EAG that this had an impact on the relative treatment-effect estimates. The EAG and the company agreed it was not appropriate to create a single network of all 3 comparators because of the differences in definitions. The committee also acknowledged the reason for the differences in the naive results and the NMA. The company maintained its preference for an ITC using a MAIC to adjust for baseline differences between RESPONSE and ELATIVE. The EAG preferred to do 2 separate NMAs; one comparing seladelpar with elafibranor, and another comparing seladelpar with OCA. The committee concluded that it preferred the EAG's approach of using NMAs. It noted that this better reflected the underlying heterogeneity in clinical trial populations and outcome definitions.

Outcome recalculation

- 3.9 During the second committee meeting, the committee asked for clarification on how the outcome definitions differ between the trials (see section 3.8). The clinical expert explained that comparing ALP values across clinical trials is unreliable without individual-patient data analysed on a single platform. This is because of differences in laboratory assays

and thresholds. They said that this introduced uncertainty and limited the validity of the indirect comparisons. They explained that in their experience it is difficult to compare ALP biochemical response outcomes between trials and there is no method to resolve this scientifically. The committee noted that recalculating outcomes to align ULN threshold definitions across trials introduced further complexity. The committee queried whether it was more appropriate to do the treatment comparisons without outcome recalculation. The EAG noted that analyses with and without recalculation of ULN thresholds across trials showed limited impact on the relative treatment effects within trials. The committee noted that doing comparisons without outcome recalculation did not meaningfully impact the cost-effectiveness results. The committee concluded that, in the absence of individual-patient data and a robust method for resolving the differences in how the ULN threshold was defined across trials, it preferred that treatments should be compared without outcome recalculation.

Adverse events and patient-reported outcomes

- 3.10 For adverse events and patient-reported outcome measures (PROMs) the company used the Bayesian NMA results. Although the EAG noted there were potential issues with the NMA methodology and its reporting, the credible intervals around the adverse events and PROM results were narrower than around the ALP outcomes. The NMA showed that seladelpar was associated with less pruritus at 12 months compared with both the 5-mg and 10-mg doses of OCA and with placebo. Seladelpar also had lower odds of upper respiratory tract infections compared with placebo and elafibranor. For the patient-reported outcomes, seladelpar showed a numerical reduction in pruritus at 12 months compared with placebo and OCA. But this was not statistically significant. The EAG noted that no minimum clinically important difference had been established for either the 5-D itch or PBC-40 scales. So, it was unable to determine if differences between treatments for each outcome were clinically meaningful. In response to consultation, a stakeholder stated that the

PBC-40 has been validated in a large PBC population and clinically meaning improvements are possible to observe. Although the company had presented results for elafibranor PROMs, these came from a subset of people in ELATIVE with moderate or severe pruritus at baseline for which it was unable to identify baseline characteristics. So, the EAG advised it was not possible to compare population characteristics across the trials to determine transitivity. This introduced uncertainty into the comparability of the overall study populations. The committee concluded that the results from the NMA support clinical opinion that seladelpar improves pruritus compared with OCA, but it could not be certain of an improvement compared with elafibranor.

Economic model

Company's modelling approach

- 3.11 The company used a cohort-level Markov state-transition model to evaluate the cost effectiveness of seladelpar, with or without UDCA, compared with OCA or elafibranor with or without UDCA. The model had 2 components, with health states defined by ALP levels and liver-disease progression. Transitions between health states were driven by ALP levels, which were used as a proxy for increased or decreased risk of disease progression (see [section 3.4](#)). Disease progression was defined as the transition to the compensated cirrhosis or elevated bilirubin health state. After this point, the condition can no longer improve and will continue to worsen over time. The model incorporated both costs and the impact on health-related quality of life (utilities) for people in ALP and liver-disease states. It also accounted for the burden of pruritus at varying levels of severity within these health states. The model was run over a lifetime horizon of up to 50 years to capture long-term outcomes and costs. The committee noted that the model had a similar structure to the models used in the appraisals of OCA and elafibranor. But it included an additional health state for ALP normalisation, separate from the mild-normalisation health state (see section 3.4). The committee agreed that this was appropriate and the model structure was appropriate for decision making.

Treatment discontinuation at 0 to 12 months

3.12 For the first 12 months, the company's base case modelled treatment discontinuation directly from the clinical trials for seladelpar and each comparator. These were RESPONSE for seladelpar, ELATIVE for elafibranor and POISE for OCA. But the EAG preferred using ITC-derived rates, anchored to the RESPONSE data. These ITC-derived rates were notably higher for comparators, especially for OCA, which had a cumulative discontinuation of 26.95% at 12 months compared with 9.59% in the company's base case. The cumulative rate of discontinuation for seladelpar was 6.73% across both approaches. The EAG advised that anchoring discontinuation to RESPONSE was appropriate because RESPONSE informed the transition probabilities for the seladelpar arm in the model. So, the anchoring maintained a link between discontinuation and clinical effectiveness. The clinical expert advised that the EAG's rates seemed plausible. At the first meeting, the committee concluded that it preferred the EAG's approach, with the caveat that the 12-month rates were derived from the ITC, which introduced uncertainty around these estimates ([see section 3.7](#)).

In response to consultation, the company maintained its base case for the period of 0 to 12 months but proposed an alternative scenario deriving rates from the ITC anchored to real-world data for OCA ([Abbas et al. 2023](#)). It stated that there was no clinical justification to assume differences in response rates between seladelpar and elafibranor and so applied an equal response rate of 8.4% at 12 months for both treatments. The EAG maintained that discontinuation should remain anchored to the RESPONSE data rather than real-world data. This is to maintain internal consistency with how much of the medicine was taken and treatment effectiveness in the modelling. The EAG also considered the assumption of identical discontinuation rates for seladelpar and elafibranor. It noted that differences in pruritus outcomes suggest that people having seladelpar may be more likely to continue the treatment. The committee agreed that discontinuation should remain anchored to the RESPONSE

data to maintain internal consistency and did not consider there was justification to assume an equal response rate between seladelpar and elafibranor. It noted that it is typical to ensure consistency between the evidence sources of treatment discontinuation and clinical effectiveness because these are interdependent. At the second meeting, the committee concluded that it still preferred the EAG's approach to treatment discontinuation from 0 to 12 months.

At the third meeting, the company presented data on discontinuation from the ITC and descriptive statistics from the clinical trials for seladelpar, elafibranor and OCA. It noted the trial data showed that OCA had the highest relative discontinuation rate compared with placebo at 12 months. The company explained that this was driven by much lower discontinuation in the UDCA plus placebo arm of POISE relative to the other trials in the ITC. So, using the EAG's preferred approach would mean that the discontinuation rate applied for OCA in the economic model is much higher (26.1%) than the rates applied for both seladelpar (7.8%) and elafibranor (12.7%). The company explained that the low UDCA plus placebo discontinuation in POISE may be because of the exclusion of patients with severe pruritus from the trial, which artificially inflated the relative discontinuation rate of OCA in the ITC. It said that this also produced counter-intuitive cost-effectiveness results. The company explained that when seladelpar's discontinuation rate from the ITC doubles, costs fall disproportionately more than QALYs, which lowers the incremental cost-effectiveness ratio (ICER). The company's opinion was that this contradicts clinical evidence showing seladelpar is clinically superior to OCA, meaning that people remaining on an effective treatment for longer makes it less cost-effective. The EAG commented that it could find no clear reason why UCDA plus placebo discontinuation in POISE would be lower than in RESPONSE or ELATIVE, and that it would expect equal discontinuation across all 3 trials. But it emphasised high uncertainty because of low event counts in these trials. On the company's claim that using results from the ITC causes counter-intuitive cost-

effectiveness results, the EAG explained that this can occur for high-cost medicines when incremental costs accrue faster than incremental benefits.

The company presented sources of real-world data for discontinuation rates for seladelpar, elafibranor and OCA. These included IQVIA Longitudinal Access and Adjudication Data. They showed that estimates differ widely between sources and from the clinical studies. The company emphasised that IQVIA showed a discontinuation rate of 34% for elafibranor and 46% for seladelpar at 6-month follow up, which is much higher than the rates predicted when using the ITC results applied to the observed rate in RESPONSE. The company explained that because of high uncertainty in both the ITC results and in the real-world data, together with the sensitivity of the model to discontinuation, all 3 comparators should have the same discontinuation rate. The company stated its preference for 15.6%, which is the average of the rates from the studies included in the ITC. The EAG agreed with the company that equal discontinuation rates may be reasonable, given that there were no differences in baseline characteristics of the control arms for the pivotal trials. But it recommended exploring higher rates (20%, 30%, 40%) that would align more closely with the higher discontinuation rates seen in the real-world data. The EAG explained that it had chosen 40% for its base case, but that this was not explicitly preferred because of the high degree of uncertainty. The committee agreed with the company and the EAG that the results of the ITC comparing discontinuation across studies were inherently uncertain and that the confidence intervals were wide. It concluded that it would be reasonable to assume that discontinuation rates could be equal, given the quality of the available evidence. It agreed that it would consider a range of discontinuation rates in the cost-effectiveness analysis.

Treatment discontinuation after 12 months

3.13 After 12 months, the company initially applied a discontinuation rate ratio of 0.28 based on ELATIVE and its open-label extension study. This was used for seladelpar and OCA. The EAG preferred a lower ratio of 0.12, derived from RESPONSE and ASSURE (a long-term open-label trial of seladelpar). At the first committee meeting the committee preferred the EAG's approach, noting that the 12-month rates were based on the ITC and subject to uncertainty. At the second committee meeting, the clinical expert verbally referenced a follow-up paper from [Abbas et al. \(2025\)](#). But they did not provide a reference to the publication in response to consultation. In response to consultation, the company accepted the committee's preferred approach to apply a ratio of 0.12.

Utility values

Source of utility values

3.14 To derive utility values, the company used disease-specific PBC-40 data collected in RESPONSE. It mapped this to EQ-5D-5L using real-world data from 90 people included in the ITCH-E study. It then also mapped from EQ-5D-5L to EQ-5D-3L using the Hernández-Alava algorithm. The company used a mixed model for repeated measures (MMRM) based on RESPONSE data to apply disutilities for ALP health states and pruritus. But in its model the company instead used disutility values associated with pruritus from [Smith et al. \(2022\)](#). These values were based on EQ-5D-5L data from the 16-week GLIMMER study of linerixibat in people with moderate to severe pruritus. Disutilities from Smith et al. were larger than those from RESPONSE. The EAG favoured the MMRM-derived disutilities, advising that they were more appropriate because of their internal consistency, alignment with UK cohort data, and direct use of RESPONSE data. The clinical expert at the first meeting advised they would expect a utility loss from pruritus. The committee noted that the EAG's approach might underestimate the burden of pruritus because the disutility values were very small. The committee considered the EAG's comment that the model included costs for pruritus treatments, such as fibrates, but did not account for treatment benefits. It noted that utility

values decreased across ALP health states, although the reason for this was unclear. It decided it may be possible that the decrease in utility seen across health states could be partly because of the impact of pruritus. This made it difficult to determine whether the addition of another disutility from pruritus would lead to double counting. The committee acknowledged that pruritus is associated with elevated bilirubin, which was only measured in the highest ALP health state. It noted that other symptoms of PBC, such as fatigue, were not explicitly modelled but may also affect utility values. At the first meeting, the committee concluded that it preferred the EAG's approach to utility values because it was based on trial data from people like those who would have treatment in clinical practice. Also, it used data from a consistent source for all health states. But there remained uncertainty about whether this approach underestimated the disutility associated with pruritus. The committee decided that additional evidence from the literature may be of value, and asked for more information on how well the model reflects quality of life with PBC.

Alternative disutility values

- 3.15 In response to consultation, the company explored 2 alternative scenarios for modelling pruritus disutility but maintained its base case from the first committee meeting. The EAG had concerns about the face validity of the re-anchored disutility for severe pruritus (-0.32) from one of the company's alternative scenarios, which re-anchored values from Smith et al. The EAG noted this was comparable to disutilities used in previous NICE appraisals for hospitalisation because of heart failure (see [NICE's technology appraisal guidance on dapagliflozin for treating chronic heart failure with reduced ejection fraction](#)). But the EAG's clinical expert had advised that the estimate was implausible, because people with severe pruritus may still be able to do many daily activities including working. So, the company's disutilities suggest that the impact on quality of life is more severe than that associated with heart failure or advanced Parkinson's. The patient expert emphasised the severity of pruritus from PBC. They

described the impact pruritus has on the family and relatives of people with PBC. In the second scenario the company referenced [Hussain et al. \(2023\)](#), which reported EQ-5D utility values for people with primary sclerosing cholangitis (PSC). Because both PSC and PBC lead to cholestatic pruritus, it argued that there is no clinical rationale to assume differences in the experience or impact of pruritus between the 2 conditions. The EAG found these alternative disutility values to be more clinically plausible than those reported by Smith et al. But the data was only available in an abstract, so this limited the ability to fully assess its methodology and robustness. The EAG advised that using higher disutility values from less credible sources could overestimate the impact of pruritus. The EAG still preferred utility values to be derived from the MMRM2 mapping model based on EQ-5D-3L data from RESPONSE. It noted that these estimates aligned with findings from [Rice et al. \(2021\)](#), a UK study of 2,240 people with PBC. Rice et al. reported only a small disutility for itch (-0.018), which was not statistically associated with impaired quality of life. The committee recalled its concerns about how quality of life had been reflected in the model (see [section 3.14](#)). The company did not provide any new evidence to explain how the model incorporated quality of life considerations across all ALP health states. So, it was difficult to determine what was driving the differences in utility values between ALP states and uncertainty about potential double counting of pruritus disutility remained. The committee considered the scenarios submitted by the company, but concluded that the EAG's choice of disutility values associated with pruritus remained the most appropriate, recalling the data was based on trial data from people likely to have the medicine in clinical practice and considering the risk of double counting the disutility. The committee concluded it was satisfied with the quality of life across ALP health states for decision making, especially as it had not been presented with any new information in response to consultation.

At the third meeting, the company presented new data describing the

extent of disutility associated with pruritus. It explained that it had commissioned a survey done by the UK PBC foundation (n=152), which showed the considerable impact of pruritus on daily life. The survey showed that in the preceding 4 weeks, 47% of respondents reported itch-disturbed sleep and a third had scratched their skin raw. The company restated its view that the scale of the impact experienced by patients is not reflected in the values preferred by the EAG and committee at the second meeting. It explained that there remained fundamental methodological limitations in the MMRM2 mapping model, which mean that the mapped disutilities for pruritus from ITCH-E were invalid. This was because the economic model captured severity of pruritus on the numerical rating scale (NRS), but there was little overlap between NRS 'severe' scores and the PBC-40 'clinically severe' scores. The company noted that this is evidenced by analysis of the data from RESPONSE, which measured pruritus on both scales. The company explained that the disutilities for severe pruritus from ITCH-E would likely under-represent the magnitude of the disutility of severe pruritus on the NRS. This is because the utility values would have been sampled from patients with mild and moderate pruritus on the NRS. The EAG commented that the UK PBC foundation survey demonstrated high itch impact and prevalence, but that it does not quantify disutility and so does not justify rejection of ITCH-E data.

The company proposed an alternative source of disutility values for pruritus. It provided analysis of an Adelphi Disease Specific Programmes (DSP) study in PBC to capture severity on the NRS, to align with the measure used in the economic model without the need for mapping. DSP is a large, multinational observational study of clinical practice for PBC. It collected various patient-reported outcomes including pruritus on the NRS, the PBC-40 and EQ-5D. It also collected data from medical records. The company estimated adjusted differences in EQ-5D utility using a multivariable linear regression model, with EQ-5D utility as the dependent variable and pruritus severity category (using the NRS) as the primary

explanatory variable. The resulting values are considered confidential by the company and cannot be reported here. The model adjusted for available demographic and clinical covariates in the Adelphi DSP dataset (including age, gender, ethnicity, insurance status, BMI, fatigue, use of anti-pruritus medications and ALP levels). The company explained that it had adjusted for all known covariates reported to have an impact on the quality of life of people with pruritus, if possible. But the EAG had concerns about the validity of the company's Adelphi DSP analysis. In reviewing the coefficients and standard errors of all explanatory variables, the EAG's analyses suggested that ethnicity is a much bigger determinant of health-state utility than having severe pruritus, which the EAG and committee agreed was implausible. So, the EAG advised that the Adelphi DSP analysis lacked face validity, which suggests that the model may be mis-specified. The EAG noted that ITCH-E was designed to answer the relevant question, and so should not be dismissed. But it also acknowledged that the disutility for severe pruritus may be underestimated in ITCH-E (-0.0345), so it explored a range of scenarios (-0.05, -0.1 and -0.15). The committee considered the robustness of the company's Adelphi DSP analysis, and particularly the high impact of white ethnicity on health utility in the model. It agreed with the EAG that it lacked face validity. The committee acknowledged that the disutilities from ITCH-E likely underestimate the true disutility associated with pruritus. But it concluded that it had seen no robust disutility data that was methodologically preferable to the use of disutilities mapped from the ITCH-E data.

Cost-effectiveness estimates

Acceptable ICER

- 3.16 [NICE's manual on technology appraisal and highly specialised technologies guidance](#) notes that, above a most plausible ICER of £25,000 per quality-adjusted life year (QALY) gained, judgements about the acceptability of a technology as an effective use of NHS resources will take into account the degree of certainty around the ICER. The committee

will be more cautious about recommending a technology if it is less certain about the ICERs presented. But it will also take into account other aspects including uncaptured health benefits. The committee noted the high level of uncertainty, specifically about the:

- results of the ITC, including the wide credible intervals for comparisons about ALP outcomes, and the impact of outcome definition (see [section 3.8](#))
- positive treatment response for people having placebo in the seladelpar trials (see [section 3.6](#))
- differences in utility values across ALP health states (see [section 3.14](#)).

The committee acknowledged that the company had provided more information to try and overcome the uncertainties in its response to consultation, including a clearer ITC report and scenario analyses for disutilities associated with pruritus. It also acknowledged the additional information provided for the third committee meeting. But the committee concluded that considerable uncertainty remained, largely because of unresolvable limitations with the data for seladelpar and the comparators. So, it concluded that an acceptable ICER would be at the lower end of the threshold.

Company and EAG cost-effectiveness estimates

3.17 The ICERs cannot be reported here because the comparators have confidential patient access schemes. In the company's base case, the ICERs for seladelpar compared with elafibranor and OCA are below the threshold that NICE considers a cost-effective use of resources. In the committee's preferred base case, and in all plausible scenarios for equal discontinuation rates up to 12 months, the ICERs were also below what NICE considers a cost-effective use of NHS resources. The committee's preferred assumptions are:

- the baseline distribution of people across the modelled health states should reflect people who would have seladelpar in clinical practice,

that is, people who have ALP elevation more than 1.67 times the ULN (see [section 3.4](#))

- Bayesian NMAs should be done and validated for the ITC for seladelpar and all its comparators, without outcome recalculation (see [section 3.8](#) and [section 3.9](#))
- treatment discontinuation rates derived from the ITC should not be applied (see [section 3.12](#))
- the same data source (RESPONSE) should be used for utility values for ALP health states and pruritus (see [section 3.14](#))
- disutilities for pruritus mapped from the ITCH-E data (see [section 3.15](#)).

The committee noted that both the company and the EAG presented fully incremental cost-effectiveness analyses for the comparators. It decided these analyses are appropriate when considering a recommendation for the whole population. This approach is consistent with section 4.10.8 of the NICE manual, which states that decisions should be based on results from fully incremental analyses.

Other factors

Equality

3.18 The company noted that people with PBC may have long waiting times for care, often between 3 and 4 months. They also have higher mortality rates while on the liver transplant waiting lists compared with people with other liver diseases. Stakeholders for this appraisal noted that the [UK evaluation of care delivery \(2024\)](#) showed geographical disparities in access to specialist teams and second-line treatments for PBC, driven by differences in local resource availability. The committee acknowledged its duties under the Equality Act 2010. It acknowledged that previous technology appraisals (for example, TA1016) identified other factors that should also be considered, such as:

- there is a particularly high prevalence of this condition in women, with around 90% of cases occurring in women

- men are more likely to present with advanced disease that responds poorly to treatment, and have poorer outcomes
- age influences outcomes with people diagnosed under 50 experiencing more severe and progressive disease, poorer treatment response and potentially poorer outcomes
- younger women may have concerns about fertility.

The committee decided that, although reducing differences in access and liver transplant waiting times were outside of its remit, it would consider current clinical practice and its impact on patients' experiences. For example, it acknowledged that including fibrates as a comparator may have negatively impacted access to licensed treatments for patients, which contributed to the committee accepting excluding fibrates as a comparator. But the committee concluded that because its recommendation does not restrict access to treatment for some people over others, these were not potential equalities issues that it could address.

Conclusion

Recommendation

- 3.19 The committee considered the clinical-trial evidence for seladelpar in treating PBC, including its potential to improve liver biochemistry and pruritus compared with OCA and elafibranor, with or without UDCA. The cost-effectiveness estimates were below the range NICE considers an acceptable use of NHS resources. So, seladelpar can be used as an option for treating PBC.

4 Implementation

- 4.1 Section 7 of the [National Institute for Health and Care Excellence \(Constitution and Functions\) and the Health and Social Care Information Centre \(Functions\) Regulations 2013](#) requires integrated care boards, NHS England and, with respect to their public health functions, local

authorities to comply with the recommendations in this evaluation within 90 days of its date of publication.

- 4.2 The Welsh ministers have issued directions to the NHS in Wales on implementing NICE technology appraisal guidance. When a NICE technology appraisal guidance recommends the use of a drug or treatment, or other technology, the NHS in Wales must usually provide funding and resources for it within 60 days of the first publication of the final draft guidance.
- 4.3 When NICE recommends a treatment ‘as an option’, the NHS must make sure it is available within the period set out in the paragraphs above. This means that, if a patient has primary biliary cholangitis and the healthcare professional responsible for their care thinks that seladelpar is the right treatment, it should be available for use, in line with NICE’s recommendations.

5 Evaluation committee members and NICE project team

Evaluation committee members

The 4 technology appraisal committees are standing advisory committees of NICE. This topic was considered by [committee D](#).

Committee members are asked to declare any interests in the technology being evaluated. If it is considered there is a conflict of interest, the member is excluded from participating further in that evaluation.

The [minutes of each evaluation committee meeting](#), which include the names of the members who attended and their declarations of interests, are posted on the NICE website.

Chair

Dr Megan John

Chair, technology appraisal committee D

Final draft guidance – Seladelpar for previously treated primary biliary cholangitis [ID6429] Page 26 of 27

Issue date: June 2026

© NICE 2026. All rights reserved. Subject to [Notice of rights](#).

NICE project team

Each evaluation is assigned to a team consisting of 1 or more health technology analysts (who act as technical leads for the evaluation), a technical adviser, a project manager and an associate director.

Madiha Adam and Luke Cowie

Technical leads

Mary Hughes and Emily Leckenby

Technical advisers

Kate Moore

Project manager

Elizabeth Bell

Principal technical adviser

ISBN: [to be added at publication]