Elafibranor for treating primary biliary cholangitis [ID6331]

Evidence Assessment Group Report

Produced by Newcastle University

Authors

Eugenie Evelynne Johnson, Research Assistant, Population Health Sciences Institute, Newcastle University, Newcastle Upon Tyne, UK

Giovany Orozco Leal, Research Associate, Population Health Sciences Institute, Newcastle University, Newcastle Upon Tyne, UK

Aalya Al-Assaf, Research Associate, Population Health Sciences Institute, Newcastle University, Newcastle Upon Tyne, UK

Hangjian Wu, Research Associate, Population Health Sciences Institute, Newcastle University, Newcastle Upon Tyne, UK

Opeyemi Agbeleye, Research Assistant, Population Health Sciences Institute, Newcastle University, Newcastle Upon Tyne, UK

Sedighe Hosseini Jeblei, Research Associate, Population Health Sciences Institute, Newcastle University, Newcastle Upon Tyne, UK

Tumi Sotire, Research Assistant, Population Health Sciences Institute, Newcastle University, Newcastle Upon Tyne, UK

Emma Dobson, Research Associate, Population Health Sciences Institute, Newcastle University, Newcastle Upon Tyne, UK

Sheila Wallace, Research Fellow, Population Health Sciences Institute, Newcastle University, Newcastle Upon Tyne, UK

Fiona Pearson, Principal Research Associate, Population Health Sciences Institute, Newcastle University, Newcastle Upon Tyne, UK

Stephen Rice, Senior Lecturer, Population Health Sciences Institute, Newcastle University, Newcastle Upon Tyne, UK

Correspondence to Stephen Rice
Baddiley Clark Building,
Newcastle University,
Newcastle upon Tyne
NE2 4BN, UK

Date completed 26 July 2024

Source of funding: This report was commissioned by the National Institute for Health and Care Research (NIHR) Evidence Synthesis Programme as project number NIHR166388.

Declared competing interests of the authors: None

Contributions of authors: Stephen Rice acted as project lead. Eugenie Evelynne Johnson and Fiona Pearson acted as lead effectiveness reviewers. Giovany Orozco Leal acted as lead health economist. Sheila Wallace acted as lead reviewer of the literature search methods. Aalya Al-Assaf and Opeyemi Agbeleye acted as assistant effectiveness reviewers. Hangjian Wu, Sedighe Hosseinijebeli and Tumi Sotire acted as assistant health economics reviewers. Emma Dobson assisted in reviewing the literature search methods.

Acknowledgements: We thank Dr Ryan Kenny and Dr Nick Meader for their assistance in the Evidence Assessment Group's network meta-analyses. We thank Dr Aaron Wetten, a Fellow at Newcastle University, for providing clinical expert advice to the Newcastle TAR Group.

Rider on responsibility for the 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: Johnson EE, Orozco Leal G, Al-Assaf A, Wu H, Agbeleye O, Hosseini Jeblei S, Sotire T, Dobson E, Wallace S, Pearson F, Rice S. Elafibranor for treating primary biliary cholangitis [ID6331]: A Single Technology Appraisal. Newcastle upon Tyne: Population Health Sciences Institute, Faculty of Medical Sciences, Newcastle University 2024.

Report key: Commercial in confidence (CiC) data are highlighted in blue throughout the report.

Any de-personalised data are highlighted in pink throughout the report.

Abbreviations

AE	Adverse events
AIC	Akaike's Information Criterion
ALP	Alkaline phosphate
ANA	Antinuclear antibodies
BIC	Bayesian information criterion
BSC	Best supportive care
CE	Cost-effectiveness
CEM	Cost-effectiveness model
CI	Confidence interval
CRD	Centre for Reviews and Dissemination
CrI	Credible interval
CS	Company's submission
DCC	Decompensated cirrhosis
DSU	Decision Support Unit
EAG	Evidence Assessment Group
EQ-5D	European Quality of Life-5 Dimensions
HCC	Hepatocellular carcinoma
HRQoL	Health-related quality of life
HR	Hazard ratio
HST	Highly Specialised Technology
HTA	Health technology assessment
HUI	Health utility index
ICEP	Incremental cost-effectiveness plane
ICER	Incremental cost-effectiveness ratio
INAHTA	International Network of Agencies for Health Technology Assessment
ITT	Intention-to-treat
LT	Liver transplant
LYG	Life years gained
MeSH	Medical subject headings
N/A	Not applicable
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health Research
NMA	Network meta-analysis
NMB	Net monetary benefit
OCA	Obeticholic acid
OR	Odds ratio
OWSA	One-way sensitivity analysis
PAS	Patient access scheme
PBC	Primary biliary cholangitis
PBC-40	Primary biliary cholangitis-40 questionnaire
PSA	Probabilistic sensitivity analysis
PSS	Personal Social Services
QALY	Quality adjusted life year
QoL	Quality of life
RCT	Randomised controlled trial
RR	Risk ratio
SD	Standard deviation
SE	Standard error
SIGN	Scottish Intercollegiate Guidelines Network
SLR	Systematic literature review
SUCRA	Surface Under the Cumulative RAnking curve

TA	Technology Assessment
TB	Total bilirubin
TEAE	Treatment-emergent adverse event
U/L	Units per litre
UDCA	Ursodeoxycholic acid
UK	United Kingdom
UK-PBC	United Kingdom Primary Biliary Cholangitis
ULN	Upper limit of normal
VAS	Visual analogue scale

Table of Contents

Table of Tables

Table of Figures

1 EXECUTIVE SUMMARY

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

Section 1.1 provides an overview of the key issues identified by the EAG. Section 1.2 presents the model outcomes. Section 1.3 summarises all key issues identified by the EAG relating to clinical effectiveness and cost-effectiveness. Section 1.4 summarises the EAG's preferred assumptions and ICERs.

Further detail regarding key and non-key issues are described in the main EAG Report (Sections 2 to 6).

All issues identified represent the EAG's view, not the opinion of the National Institute for Health and Care Excellence (NICE).

1.1 Overview of the EAG's key issues

Table 1.1: Summary of EAG's key issues

Issue number	Brief summary of issue	Report section(s)
1	Uncertainty in the results of the network meta-analysis (NMA)	Section 3.3.4, 3.5
2	Uncertainty and lack of validation in the economic model's survival predictions	Section 4.3.3
3	All-cause discontinuation predictions for OCA determining cost-effectiveness in the economic model.	Section 4.3.4.2
4	Appropriate utility value for the high-risk biomarker health state in the economic model.	Section 4.3.5.1
Abbreviations: EAG = Evidence Assessment Group; NMA = network meta-analysis; OCA = obeticholic acid		

1.2 Overview of key model outcomes

NICE technology appraisals compare how much a new technology improves length (overall survival) and quality of life in a quality-adjusted life year (QALY). An ICER is the ratio of the extra cost for every QALY gained. A technology is considered absolutely dominant when it improves quality of life (measured in QALYs gained) and reduces costs (measured in £GBP) relative to its best alternative treatment.

Overall, the technology is modelled to affect QALYs by:

- Improving the primary biliary cholangitis (PBC) biomarker risk category (i.e. reducing the risk of liver disease): There were no treatment-specific differences in quality of life. Instead, reducing the PBC biomarker risk of liver disease improved quality of life, then treatment response differences led to differences in quality of life across treatment arms.

- Treatment discontinuation leading to a deterioration in the PBC biomarker risk categories (i.e. increased risk of liver disease): After one year on treatment elafibranor and OCA patients are assumed to stay on their risk category unless they discontinue and move to UDCA and best supportive care. Patients receiving UDCA are assumed to not have PBC biomarker risk improvements after the first year of treatment. Therefore, differences in treatment discontinuation are an important driver of quality of life differences.

Overall, the technology is modelled to affect costs by:

- [redacted]Improving the PBC biomarker risk category: Patients categorised at high-risk of liver disease are assumed to receive more intensive care than patients at mild or moderate risk. Patients as high risk of disease can transition to more severe disease stages such as liver failure leading to transplant (LT), decompensated cirrhosis (DCC), and hepatocellular cancer (HCC) at a higher rate than moderate-risk of disease patients.
- Treatment maintenance: Discontinuation of elafibranor or OCA leads to an increase in the risk of liver disease, and it only continues to increase under UDCA with best available care. Due to the lifelong duration of treatment, assumptions around long-term maintenance differences also affect differences in total treatment costs.
- Compliance differences: There are small differences in treatment compliance between OCA and elafibranor. This has an impact on treatment costs but there is a lack of evidence on the impact of compliance in effectiveness, so this only affects cost differences.

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

- Duration of the treatment differences in discontinuation: Data is only available comparing all-cause discontinuation between elafibranor and OCA over the first year. The economic model assumes this difference is sustained over the lifelong duration of treatment.
- Relative effectiveness parameters to model treatment with OCA in the economic model: To derive cholestasis response, occurrence of pruritus as an adverse event, and all-cause discontinuation for OCA, the model uses 12-month odds ratios from the network meta-analysis anchored to 12-month elafibranor baseline risks to derive 12-month risk ratios. A constant risk ratio is assumed and applied to 3-month elafibranor probabilities to generate 3-month probabilities for OCA. An alternative approach suggested by the EAG is to assume a constant hazard ratio calculated from the 12-months odds ratio.
- Treatment effectiveness definition: The company's base-case analysis follows the cholestasis response definition from the POISE and ELATIVE trials, and the PBC biomarker risk definitions from NICE TA443. More strict treatment response definitions such as alkaline phosphate (ALP) normalisation, the Barcelona criteria, or the PARIS II criteria require low risk-of-progression patients to achieve lower ALP thresholds. These thresholds are more difficult to attain and lead to reduced treatment effectiveness estimates.
- Utility values in the high-risk PBC biomarker health-state: The economic model used utility values from the published literature to calculate quality of life. Utility values collected from patients in the ELATIVE trial covered all the PBC biomarker risk

categories; however, only the mild and moderate-risk utility values from the trial were explored in a scenario analysis. There was a noticeable discrepancy between the PBC high risk value elicited from the ELATIVE trial ([redacted]) and the published utility value used in the economic model (0.55) from NICE TA330. [1,2] The company considers that selection bias and a small sample size make the trial utilities less reliable; however, the omission of this scenario many not accurately represent the parameter uncertainty in the economic model.

- UDCA after the first year: Patients discontinuing second-line treatment are assumed to move to UDCA and best supportive care. At this stage their risk level is assumed to continue deteriorating and is assumed not to improve (unlike elafibranor and OCA patients who stay in their risk category after the first year) less severe assumptions of risk progression at third-line UDCA treatment reduce the effectiveness estimates.
- Pruritus: Pruritus differences play a minor role in the cost-effectiveness results from the economic model compared to all-cause discontinuation, cholestasis response, and treatment cost differences. The difference in pruritus between elafibranor and OCA is assumed to remain constant over the lifelong duration of treatment, impacting both quality of life and total costs.

1.3 Description of the EAG's key clinical and economic issues

Table 1.2: Key issue 1: Uncertainty in the results of the network meta-analysis (NMA)

Report section	Sections 3.3.4, 3.5
Description of issue and why the EAG has identified it as important	The company state in their submission that the NMA results show elafibranor 80 mg to be superior to the comparators from the POISE trial, including OCA 5-10 mg. However, the EAG note that the NMA, used as the company's base case for dichotomous outcomes, is subject to methodological limitations and that the 95% CrIs are substantially wide. As a result, the EAG considers the outcome estimates obtained from NMAs to be highly uncertain.
What alternative approach has the EAG suggested?	The EAG undertook sensitivity analyses exploring alternative methodological approaches and assumptions for both dichotomous and continuous NMA outcomes included within the company's economic model (see Section 3.5). These sensitivity analyses included conducting both random-effects and fixed-effect frequentist analyses instead of using a Bayesian approach, as well as conducting a fixed-effect Bayesian approach with RR as the summary statistic instead of an OR for dichotomous outcomes. The EAG were unable to satisfactorily run a random-effects Bayesian model with RR as the summary statistic. The EAG's sensitivity analyses did not change the NMAs findings; rather they highlighted the 95% CIs were still substantially wide and that the estimates are highly uncertain.
What is the expected effect on the cost effectiveness estimates?	Due to the uncertainties and lack of additional available data, the EAG are unable to comment on whether more evidence would either increase or reduce the ICERs.

Report section	Sections 3.3.4, 3.5
What additional evidence or analyses might help to resolve this key issue?	Further comparable evidence between elafibranor and OCA 5-10 mg or UDCA monotherapy, either within direct head-to-head comparisons or to add to indirect treatment comparisons, may potentially increase the certainty of the clinical effectiveness of elafibranor.
Abbreviations: CI = confidence interval; CrI = credible interval; EAG = Evidence Assessment Group; ICER = incremental cost-effectiveness ratio; NMA = network meta-analysis; OCA = obeticholic acid; UDCA = ursodeoxycholic acid	

Table 1.3: Key issue 2: Uncertainty and lack of validation in the economic model's survival predictions

Report section	Section 4.3.3
Description of issue and why the EAG has identified it as important	The EAG is concerned that survival predictions from the economic model, whether it is liver-disease free, transplant free, or overall survival were not validated with clinical experts during the company submission. Moreover, there was little validation of survival with published evidence. The EAG is concerned the model is under-predicting the proportion of patients who are free of liver disease.
What alternative approach has the EAG suggested?	The EAG has suggested changing the approach to including the excess mortality risk for high-risk patients and testing scenario analyses removing transitions from moderate risk to liver disease or reducing the treatment effect on discontinuation (which is a primary factor in the transitions towards high-risk of liver disease).
What is the expected effect on the cost effectiveness estimates?	If the proportion of patients who develop liver disease is lower than the predictions from the model, the potential QALY gain and costs saved associated with the treatment may be lower than those currently predicted.
What additional evidence or analyses might help to resolve this key issue?	The model predictions could be validated with clinical experts and published literature.
Abbreviations: EAG = Evidence Assessment Group	

Table 1.4: Key issue 3: All-cause discontinuation predictions for OCA determining cost-effectiveness in the economic model

Report section	Section 4.3.4.2
Description of issue and why the EAG has identified it as important	Treatment discontinuation is the primary driver of cost and QALY outcomes in the economic model. The proportion of patients stopping OCA treatment predicted by the model were considered high when compared to clinical expert opinion and external data. The EAG believes the cause of this could partly be down to the assumption that the difference in discontinuation rates between elafibranor and OCA continues indefinitely.

Report section	Section 4.3.4.2
What alternative approach has the EAG suggested?	Opting for a one-year duration in the difference in discontinuation rates between OCA and elafibranor rather than a lifetime duration led to better discontinuation predictions for OCA. The EAG also evaluated a scenario with different outcomes for third-line UDCA after discontinuation of second-line treatment.
What is the expected effect on the cost effectiveness estimates?	Assuming a 1-year duration of a difference in discontinuation rates would increase the cost of OCA treatment, decrease the cost of liver disease in the OCA arm, and increase the discounted QALYs of OCA because patients would remain on OCA for longer.
What additional evidence or analyses might help to resolve this key issue?	Observational evidence on treatment discontinuation for both elafibranor and OCA.
Abbreviations: EAG = Evidence Assessment Group; OCA = obeticholic acid; UDCA = ursodeoxycholic acid	

Table 1.5: Key issue 4: Appropriate utility value for the high-risk biomarker health state in the economic model

Report section	Section 4.3.5.1
Description of issue and why the EAG has identified it as important	A utility value was elicited using the EQ-5D questionnaire in the ELATIVE trial for the high risk biomarker state, while the company base-case analysis used a value from the published literature. There was a noticeable difference between the values obtained in the trial compared to the literature for the high-risk of liver disease state in the model. There was also considerable variation in utility estimates for compensated cirrhosis in the literature. The EAG is concerned that trial utility values were only explored for the mild-risk and moderate-risk patients as a scenario analysis, particularly when utility values at the high-risk had an impact on overall results while being highly uncertain. Therefore, exploring the full parametric uncertainty for the high-risk utility value may be informative for decision-making.
What alternative approach has the EAG suggested?	The EAG has presented a scenario analysis using trial values across all the PBC biomarker risk of liver disease states. Moreover, the EAG adopted a high-risk utility value from a more recent published source, a systematic review with a meta-analysis that includes the study referenced for the estimate used in the CS.
What is the expected effect on the cost effectiveness estimates?	A higher utility for the high risk biomarker state would increase discounted QALYs more in the OCA arm than in the elafibranor arm, thus decreasing the cost-effectiveness of elafibranor.

Report section	Section 4.3.5.1
What additional evidence or analyses might help to resolve this key issue?	Observational evidence for the utility of patients in the high risk biomarker state.
Abbreviations: EAG = Evidence Assessment Group; PBC = primary biliary cholangitis; QALY = quality-adjusted life year	

1.4 Summary of the EAG's preferred assumptions and ICER

Table 1.6: Summary of EAG's preferred assumptions and ICER

Technologies	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER (£/QALY)
CS base-case – Probabilistic					
Elafibranor	[redacted]	[redacted]	[redacted]	[redacted]	Elafibranor dominating
OCA	[redacted]	[redacted]	[redacted]	[redacted]	
Fixing errors (1-8) – Probabilistic					
Elafibranor	[redacted]	[redacted]	[redacted]	[redacted]	Elafibranor dominating
OCA	[redacted]	[redacted]	[redacted]	[redacted]	
EAG base-case – Probabilistic					
Elafibranor	[redacted]	[redacted]	[redacted]	[redacted]	Elafibranor dominating
OCA	[redacted]	[redacted]	[redacted]	[redacted]	
Abbreviations: CS = company submission; EAG = Evidence Assessment Group; ICER = incremental cost-effectiveness ratio; QALY = quality-adjusted life year; OCA = obeticholic acid					

Table 1.7: Summary of key EAG scenario analysis results – deterministic analysis: elafibranor versus OCA

Scenario #	EAG base-case input	Alternative input	Incremental costs (£)	Incremental QALYs	ICER (£/QALY)
	EAG base-case	N/A	[redacted]	[redacted]	Elafibranor Dominating
2	Treatment difference on discontinuation for 1 year	No treatment difference on discontinuation	[redacted]	[redacted]	Elafibranor Dominating
4	Literature values for PBC biomarker state utilities	Trial values for PBC biomarker state utilities	[redacted]	[redacted]	Elafibranor Dominating
7	All-cause discontinuation risk function: lognormal	All-cause discontinuation risk function: Gompertz	[redacted]	[redacted]	Elafibranor Dominating
9		All-cause discontinuation risk function: Exponential	[redacted]	[redacted]	Elafibranor Dominating
12	UDCA probabilities after one year	UDCA probabilities after one year	[redacted]	[redacted]	Elafibranor Dominating

Scenario #	EAG base-case input	Alternative input	Incremental costs (£)	Incremental QALYs	ICER (£/QALY)
	follow the probabilities seen in months 9-12	follow the average probabilities of the first 12 months including probabilities to improve PBC risk			
13	Treatment effectiveness definition: Cholestasis response	Treatment effectiveness definition: ALP normalisation	[redacted]	[redacted]	Elafibranor Dominating
14		Treatment effectiveness definition: Barcelona criteria	[redacted]	[redacted]	Elafibranor Dominating
15		Treatment effectiveness definition: Paris II	[redacted]	[redacted]	Elafibranor Dominating

Abbreviations: EAG = Evidence Assessment Group; ICER = incremental cost-effectiveness ratio; QALY = quality-adjusted life year; OCA = obeticholic acid

2 CRITIQUE OF COMPANY'S DEFINITION OF DECISION PROBLEM

Table 2.1: Statement of the decision problem (as presented by the company)

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	EAG comment
Population	Adults with primary biliary cholangitis (PBC) whose disease has an inadequate response to, or who are unable to tolerate, ursodeoxycholic acid (UDCA).	As per the final scope	N/A	**Some concerns** The inclusion criteria of the ELATIVE trial does not prohibit participants who have received prior treatments other than UDCA, although prior OCA treatment is unlikely. This means some participants may have been receiving elafibranor as a third-line treatment in the trial; e.g. if they have potentially previously used OCA. Furthermore, clinical advice to the EAG noted that it may not be appropriate to combine those who are intolerant to UDCA and those who do not respond to UDCA in a single analysis, as these are two clinically heterogenous populations, though almost all data are for the non-response population. See Section 2.1 for further details.
Intervention	Elafibranor alone or in combination with UDCA.	As per the final scope	Elafibranor treatment with and without UDCA (determined according to tolerability to UDCA) are not considered separately in the company submission as the ELATIVE trial population is representative of the distribution of patients treated with and without UDCA in clinical practice.	**Appropriate** The EAG's concerns surrounding the stratification of those who do not respond to UDCA and those intolerant to UDCA are described in Section 2.1. The EAG find the intervention in the ELATIVE trial to be in line with the NICE scope.

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	EAG comment
Comparator(s)	For people whose disease has an inadequate response to UDCA: Obeticholic acid (OCA) in combination with UDCA UDCA monotherapy For people who are unable to tolerate UDCA: OCA monotherapy Best supportive care	As per the final scope	As stated above, subgroups according to patient response to UDCA and/or tolerability to UDCA are not considered separately in the company submission as the ELATIVE trial population is representative of the distribution of patients treated with and without UDCA in clinical practice. Thus, the comparators presented are UDCA and OCA 5-10mg dose with UDCA (where a proportion of both arms do not receive UDCA, which represents the cohorts receiving OCA only and no treatment). To note, only approximately 5% of patients are unable to tolerate UDCA, as reflected in the proportions of patients in the elafibranor and OCA trials. [3-6] Any best supportive care treatment other than OCA 5-10 mg has not been recommended by NICE and therefore will not be considered in the submission.	**Some concerns** The EAG asked the company to clarify the meaning behind "*best supportive care treatment other than OCA 5-10 mg*" and has concerns that relevant comparators used within clinical practice may have been missed from the submission as a result. See Section 2.3 for further details.
Outcomes	The outcome measures to be considered include: mortality liver function based on markers of liver biochemistry symptoms including pruritus, fatigue, and abdominal pain time to liver transplantation	As per the final scope	All outcomes have been addressed throughout the company submission, as follows: As outcomes of the ELATIVE trial, including outcomes based on liver function biomarkers, occurrence of pruritus symptoms and adverse events, and health-related quality-of-life (Section B.2.3 and B.2.6).	**Appropriate** The company further clarified the reasoning for a composite of surrogate outcomes to measure the primary outcome in the ELATIVE trial in their response to the points for clarification (PfCs); the EAG were satisfied with their response. Furthermore, although mortality was not measured as an outcome measure in the ELATIVE trial,

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	EAG comment
	PBC-related events, including ascites, varices, encephalopathy, and hepatic cell carcinoma adverse effects of treatment health-related quality-of-life		As outcomes of the cost-effectiveness model, which captures patient mortality, outcomes according to liver function biomarkers, pruritus, adverse events, liver transplantation, health-related quality-of-life, and PBC disease-specific health states, including hepatocellular carcinoma and decompensated cirrhosis [including PBC-related events such as ascites, varices, encephalopathy] (Section B.3.3).	deaths were reported as an adverse event and mortality was considered in the economic model by using life years and QALYs.
Economic analysis	The reference case stipulates that the cost-effectiveness of treatments should be expressed in terms of incremental cost per quality-adjusted life year. If the technology is likely to provide similar or greater health benefits at similar or lower cost than technologies recommended in published NICE technology appraisal guidance for the same indication, a cost comparison may be carried out. The reference case stipulates that the time horizon for estimating clinical and cost-effectiveness should be sufficiently long to reflect any differences in costs or outcomes	As per the final scope	N/A	**Some concerns** The company presented an incremental cost-utility analysis using QALYs in accordance with the reference case and the final scope. The population in the scope is limited to patients who have not responded to UDCA or are intolerant to UDCA. The company's economic analysis was consistent with this population. The EAG assumes that cost-effectiveness of elafibranor was not evaluated at third-line treatment as it was outside the scope and there was no effectiveness evidence at third-line. Elafibranor and OCA could have been included as third-line treatments in the evaluation of cost-effectiveness of elafibranor at second-line, but the same evidence issues apply.

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	EAG comment
	between the technologies being compared. Costs will be considered from an NHS and Personal Social Services perspective.			
Subgroups to be considered	None	None	As stated above, subgroups according to patient response to UDCA and/or tolerability to UDCA are not considered separately in the company submission as the ELATIVE trial population is representative of the distribution of patients treated with and without UDCA in clinical practice.	None
Other considerations	Guidance will only be issued in accordance with the marketing authorisation. Where the wording of the therapeutic indication does not include specific treatment combinations, guidance will be issued only in the context of the evidence that has underpinned the marketing authorisation granted by the regulator.	As per the final scope	N/A	**Appropriate** As per the NICE scope.

Source: CS Section B.1.1, Table 1, p.12-3[1]; PfC response[7]

Abbreviations: CS = company submission; EAG = Evidence Assessment Group; N/A = not applicable; NHS = National Health Service; NICE = National Institute for Health and Care Excellence; OCA = obeticholic acid; PBC = primary biliary cholangitis; PfC = points for clarification; QALY = quality-adjusted life year; UDCA = ursodeoxycholic acid

2.1 Population

2.1.1 Lines of therapy

The clinical advisor to the EAG indicated that elafibranor could plausibly be positioned as a third-line treatment. The company are positioning elafibranor as a second-line treatment for PBC in people who do not respond to or are intolerant to UDCA see company submission (CS) Section B.1.3.5, Figure 14, p.41.[1] However, the inclusion criteria of the ELATIVE trial does not prohibit participants from receipt of prior second-line therapy, such as OCA (CS Section B.2.3.1, Table 6, p.48-9). As such, the EAG cannot be certain that the participants in the ELATIVE trial are all receiving either elafibranor or placebo with or without UDCA as a second-line treatment, as opposed to third-line treatment, although prior OCA treatment seems unlikely.

2.1.2 Handling of participants who do not respond to, or are intolerant to, UDCA

The EAG asked the company to provide additional information on the distribution of participants who had an inadequate response to UDCA or who were intolerant to UDCA (PfC A8). The company responded that the population enrolled in ELATIVE trial was representative of a typical population of patients with PBC who have inadequate response and/or intolerance to UDCA and that, at baseline, 95% of participants in the trial were on concurrent UDCA and the remaining 5% were intolerant to UDCA.[7] As clinical advice to the EAG suggested that around 3-5% of patients with PBC are intolerant to UDCA, the EAG are satisfied that the proportion of participants intolerant to UDCA in ELATIVE is representative of clinical practice.

The clinical advisor to the EAG agreed that the trial population seemed reflective of the population seen within UK clinical practice. However, the participants unresponsive to UDCA and participants intolerant to UDCA can be considered two clinically different populations. The company did not stratify analyses of the ELATIVE trial by whether participants respond to treatment with UDCA or whether they could not tolerate UDCA. The majority of participants in the ELATIVE trial were taking UDCA at baseline (95.0%; CS Section B.2.3.2, p.50).[1] Although both the EAG and the clinical advisor appreciate that the overall sample size in ELATIVE may have prohibited stratification, pooling both populations means the effect of elafibranor on those intolerant to UDCA compared with those who do not respond to UDCA is uncertain.

2.2 Comparators

In the CS, the company stated: "*Any best supportive care treatment other than OCA 5-10 mg has not been recommended by NICE and therefore will not be considered in the submission*" (CS Section B.1.1, Table 1, p.12).[1] Clinical advice to the EAG noted that the term "best supportive care" (BSC) was not usually used in reference to PBC patients and, if used, that it is likely done in the context of end of life care (for example, when treating people with decompensated cirrhosis). The EAG asked the company to clarify their statement from the decision problem (PfC C2). The company responded: *"The current wording of this statement is incorrect as it does imply that OCA 5-10mg is considered a supportive treatment which the company does not agree with. The statement should be amended to: "OCA 5-10 mg as a second-line treatment is the standard of care for patients with PBC. Any treatment used in best supportive care has not been recommended by NICE nor does it provide the standard of care; therefore, any best supportive care will not be considered in the submission"."*[7] The EAG are satisfied that this clarification confirms that OCA was not considered BSC in the submission.

Following the NICE scope, the company did not include alternate fibrates within their submission (CS Section B.1.1, Table 1, p.13).[1] However, in terms of use of fibrates, a recent UK-wide audit suggested that, of the 1074 participants with PBC who received second-line treatment, 571 received either bezafibrate or fenofibrate.[8] Clinical advice to the EAG suggested that fibrates can be used to treat people with PBC who also experience itch, meaning a small number of people may take a combination of UDCA, OCA and bezafibrate within specialist centres.

3 CLINICAL EFFECTIVENESS

3.1 Critique of the methods of review(s)

The CS describes a systematic literature review (SLR) conducted to identify evidence on the effectiveness and safety of elafibranor and relevant comparators for treating PBC. A summary of the EAG's critique is presented in Table 3.1 below. The EAG's assessments (detailed in bold) are on a three-point Likert scale (key issue, some concerns or appropriate).

Table 3.1: Summary of the EAG's critique of the clinical effectiveness systematic literature review

Systematic review stage	Section in CS where methods are reported	EAG's assessment of the robustness of methods
Data sources	Appendix D 1.1, p. 1-10	**Some concerns** The range of sources searched by the company was appropriate but the reason given for restricting the years for which conference proceedings were searched was unconvincing. The ClinicalTrials.gov results were restricted to those with results and conference proceedings were excluded from Embase searches which could have missed relevant studies. See Section 3.1.1.1 for further details.
Search strategies	Appendix D 1.1, p. 1-10	**Some concerns** The search strategies were appropriate but focusing thesaurus headings increases specificity to the detriment of sensitivity. See Section 3.1.1.2 for further details.
Search filters	Appendix D 1.1, p. 1-10	**Appropriate** Search filters adequately captured the decision problem. (In response to the clarification letter the company stated that the search filters used were those designed by the Scottish Intercollegiate Guidelines Network (SIGN)).
Eligibility criteria	Appendix D.1, p.1; D.1.1, Table 9 (p. 10-2)	**Some concerns** The EAG has some concerns about the review question, the eligible study designs, and other eligibility criteria listed in Appendix D, Table 9. The EAG also note that the protocol for the SLR was not provided within the company submission and asked the company to clarify this. See Section 3.1.2 below for further details.
Screening	Appendix D1.2, p.12; D1.3, p.14	**Some concerns** The EAG have some concerns regarding the company's screening process, particularly surrounding the handling of studies that lacked available information and RCTs with mixed lines of treatment. Please see Section 3.1.3 for further details.
Data extraction	Appendix D1.2, p.12-3	**Some concerns** It is unclear from the CS whether the data extraction form was piloted. Furthermore, the EAG have some concerns about the data extraction process and the

Systematic review stage	Section in CS where methods are reported	EAG's assessment of the robustness of methods
		company did not provide a copy of the data extraction form. See Section 3.1.4 for further details.
Quality appraisal	Appendix D1.2, p.13	**Some concerns** The EAG have concerns about the appropriateness of the chosen quality appraisal tool in the review for all types of study designs. Moreover, the EAG have some concerns surrounding how the quality appraisal was performed. See Section 3.1.5 for further details.
Abbreviations: CS = company submission; EAG = Evidence Assessment Group; RCT = randomised controlled trial; SIGN = Scottish Intercollegiate Guidelines Network; SLR = systematic literature review		

3.1.1 Search methods for the clinical effectiveness SLR

The company conducted separate searches for clinical effectiveness studies (presented in the CS Appendix D),[9] and cost effectiveness, HRQoL and cost and resource use studies (CS Appendix G).[10] The EAG used the PRESS checklist to appraise the search strategies.[11] In this section, we present the critique of the search methods for clinical effectiveness studies. The critique of searches for cost effectiveness studies, HRQoL studies and cost and resource use is presented in Section 4.1. As some of the issues were the same for all the searches there are cross references to the relevant section to avoid repetition. The searches were based on terms related to the condition with the application of a study design (or research type) search filter(s) in some of the electronic bibliographic databases. The searches were first run in November 2022 and were updated in December 2023, so are considered up to date.

3.1.1.1 Data Sources

The company excluded conference proceedings in Embase, which could have led to the exclusion of relevant records. However, the EAG's clinical advisor confirmed that the two main conferences in the area had been covered by the company's hand searches. The company stated that: "*the exclusion of abstracts from conferences prior to 2021 was justified under the assumption that high-quality research would since have been published in a peer-reviewed journal*" (Appendix D.1.1, p.10).[9] This may not have been the case, as there may be other reasons for non- or slow-publication (such as results not being perceived as 'positive', direction of effect of result(s), lack of statistical significance of results and non-English language) as well as the potential effect of the COVID-19 pandemic on speed of publication of non-COVID-19 related results. However, conference searching could have been more expansive in this respect.[12,13] The ClinicalTrials.gov search was limited to those records with study results posted, which may have resulted in the exclusion of some relevant studies.

3.1.1.2 Search strategies

Bibliographic searches typically focussed on five terms related to the condition: biliary liver cirrhosis; primary biliary cholangitis; primary biliary cholestasis; primary biliary cirrhosis; and PBC. The EAG have some concerns regarding this, as search terms were not as broad as they could have been, resulting in a search strategy that was more specific than sensitive. Search terms need to be as comprehensive as possible to avoid missing potentially relevant

studies. Truncation could have been used when searching to capture plurals. The company did not always include MeSH terms as free-text terms (e.g. the MeSH heading 'Liver Cirrhosis, Biliary' was not translated into free-text terms). Focused MeSH (MEDLINE) and Emtree (Embase) terms were used in the search string. For example, in MEDLINE there was a focus on the heading 'Liver Cirrhosis, biliary.' Focusing this heading could lead to the exclusion of any studies that discuss liver cirrhosis but where this heading was not identified as the focus of the paper. The five terms used for bibliographic database searching were not used consistently when searching conference abstracts and health technology assessment (HTA) websites. The term 'primary biliary cholestasis' was not included in conference and HTA search strings, which could have resulted in these searches missing key reports.

3.1.2 Eligibility criteria

3.1.2.1 SLR protocol

The EAG have concerns about whether a pre-defined eligibility criteria within the SLR were adhered to. Firstly, the company state: "*The SLR was performed in accordance with a pre-specified protocol*" (CS Appendix D.1, p.1).[9] However, the company did not provide a copy of the protocol and it is not stated within the CS whether the protocol was published or registered on a database (e.g. PROSPERO). Having sight of an a-priori published review protocol is usually the only possible way to assess whether pre-defined eligibility criteria have been adhered to.[14] As such, it is difficult for the EAG to assess whether predefined eligibility criteria were adhered to during the review process and therefore, there is possibility of selection bias in the SLR.[15]

3.1.2.2 Included study designs

In Appendix D (Section D.1, p.1), the review question is stated as: "What randomised control trials (RCTs) have been conducted that evaluate the efficacy and/or safety of elafibranor and other comparators of interest in patients with PBC?"[9] However, within the inclusion criteria for the SLR, non-randomised interventional studies and observational studies are listed as included (Appendix D, Table 9, p.11).[9] As a well-formulated review question guides all aspects of the SLR, including setting the eligibility criteria,[16] the EAG asked the company to clarify this point (PfC A4). The company responded: "*Ultimately, only RCT study designs were included for data extraction; […] The only exception to this was data for studies of elafibranor itself, wherein all study designs containing summary clinical data were eligible for inclusion*." [7] The EAG have concerns regarding this, as the review questions and eligibility criteria should be clearly defined before starting the review and adhered to throughout the review process unless there is a justifiable reason to deviate from these criteria, which should be transparently stated.

3.1.2.3 Interventions and comparators

The list of comparators in Appendix D (Table 9) describes "*any other comparators (or none)*" as eligible.[9] As it was unclear to the EAG what these other comparators were, the EAG asked the company to clarify this point (PfC A5). The company responded: "*It was also anticipated that some interventional studies may look to compare different dosing regimens of the same investigational drug, hence the breadth. Under this definition, both UDCA and OCA would be covered under "any other comparators" (either could also be argued under "standard of care").*"[7] As these treatments are the current standard of care for PBC, [6,17] the EAG are satisfied with this response. However, it is unclear to the EAG what other interventions were

eligible for the SLR and whether the listed interventions were eligible to be included in the broader SLR only, or only in the SLR submitted for NICE.

3.1.2.4 Other limitations in the eligibility criteria

Eligible studies in the company's SLR were those published in English (Appendix D, Table 9, p.12).[9] As it has been suggested that studies conducted in non-English speaking countries are more likely to be published in English journals if they have statistically significant results than studies with insignificant results,[18] it is possible that potentially eligible studies may have been excluded from the SLR.

Furthermore, the company limited the included RCTs identified on ClinicalTrials.gov to trials with results only (Appendix D, Tables 7 and 8, p.10).[9] The Cochrane Handbook states: "*Searches for studies should be as extensive as possible in order to reduce the risk of publication bias and to identify as much relevant evidence as possible*." [19] Therefore, there is a chance of publication bias in the SLR if relevant RCTs were excluded.

3.1.3 Screening

In Appendix D (Section D.1.2, p.12), it is stated that: "*In cases where the article did not give enough information to be sure it met the inclusion criteria; it was excluded to ensure that only relevant articles were ultimately included in the SLR*".[9] It was unclear to the EAG whether the company attempted to contact authors of studies that lacked enough information; therefore, the EAG asked the company to clarify this point (PfC A6). The company responded: "*This was not conducted, though no instances where this may have been the only option to obtain missing data were noted with the data that were ultimately extracted in this review."*[7] However, this is not consistent with what is reported in Appendix D (Section D.1.2, p.12). Not including some studies that might be relevant due to limited information in the publication may lead to reporting bias in an SLR. Contacting authors of the primary studies is, therefore, important to enhance the precision and completeness of the review and decrease the chance of missing information and the consequential impact of reporting bias.[20]

In Appendix D (Section D.1.3, p.14), it is stated that: "*181 records reporting on observational studies in a first-line or mixed treatment line setting being deprioritised*."[9] Furthermore, in Appendix D (Table 9, p.12) it is stated that: "*Any studies of elafibranor, and RCTs in the second-line or later treatment setting were then prioritised for extraction".*[9] It is unclear to the EAG whether the company included RCTs with mixed lines of treatments where results of eligible treatment lines were reported separately. If the results of any such studies were not considered in the SLR, this may have led to potentially eligible studies being excluded from the SLR and consequently, this could possibly have impacted the NMA and subsequently the economic model.

3.1.4 Data extraction

It is unclear if the data extraction form was piloted; moreover, the company did not provide a copy of the data extraction form (Appendix D, section D.1.2, p.12-13).[9] It is mentioned in Appendix D (Section D.1.2, p.12) that a single individual extracted the data and a second individual verified the extracted data independently and checked that no relevant information was missing. Although this is considered an acceptable minimum, this approach could lead to significantly higher chance of error than two researchers extracting the data independently.[18]

3.1.5 Quality appraisal

Eligible studies in the company's SLR included both observational studies and RCTs (Appendix D, Table 9, p.11). However, the company's quality appraisal of included studies focused on the Centre for Reviews and Dissemination's (CRD's) quality assessment tool, which is mainly used for interventional studies. It is therefore unclear how the quality of observational studies would have been assessed. Furthermore, it is unclear whether the quality assessment tool was piloted and it is reported that the quality assessment was performed by one individual and verified by another reviewer independently.[9] As critical appraisal can be open to subjectivity, the CRD's guidance recommends piloting the use of the quality appraisal tool and by having two researchers perform the process independently.[18] This helps minimise the error in quality assessment and the influence of individual preconceptions.[21] As such, the EAG believes the quality appraisal process may have been open to greater subjectivity or error and, consequently, the judgements regarding the included studies and the interpretations of findings might be inaccurate or inappropriate.[22]

3.2 Critique of trials of the technology of interest, their analysis and interpretation (and any standard meta-analyses of these)

A summary of the EAG's critique of the design, conduct and analysis of the ELATIVE trial is presented in Table 3.2.

Table 3.2: Summary of EAG's critique on the design, conduct and analysis of the ELATIVE trial

Trial design or conduct concept	Section in CS where methods are reported	EAG's assessment
Intervention	B.2.3.1, Table 6, p.49	**Appropriate** The intervention in the ELATIVE trial was elafibranor 80 mg with or without concomitant UDCA therapy. The EAG agrees this is in line with the NICE decision problem.
Comparator	B.2.2.1, Table 6, p.48; B.1.1, Table 1, p.8; PfC A8	**Appropriate** According to the CS, the comparator in the trial was a placebo, with or without UDCA; as 95% of participants were taking UDCA, the EAG is satisfied that the ELATIVE trial adequately matches the NICE decision problem.
Randomisation	B.2.3.1, Table 6, p. 48	**Appropriate** The CS reported that the ELATIVE trial was randomised but did not report on the method of randomisation. However, a journal article associated with the ELATIVE trial describes the randomisation method.[4] As such, the EAG is satisfied that randomisation in the ELATIVE trial was appropriate.
Allocation concealment	B.2.3.1, Table 6, p.177	**Some concerns** There was limited information on the method or process of allocation concealment in the trial; allocation method but not concealment method was reported in

Trial design or conduct concept	Section in CS where methods are reported	EAG's assessment
		the protocol within an associated article.[4] The EAG have concerns that inadequate allocation concealment can introduce selection bias and a possible overestimation of effects. See section 3.2.1 for further details
Eligibility criteria	B.2.3.1, Table 6, p.48-9	**Appropriate** Eligible participants were adults aged 18 to 75 years with PBC who had an inadequate response to, or were unable to tolerate, UDCA. The demographic variables are similar to the UK population and key prognostic factors were captured. The EAG believes the eligibility criteria for the trial was in line with the NICE decision problem.
Blinding	B.2.3.1, Table 6, p.48-9	**Appropriate** The CS notes that the trial was double blinded but there were no details of the approach used. However, an article associated with the ELATIVE trial reported that the investigator, participants and study personnel were blinded to treatment.[4] Accordingly, the EAG is satisfied that blinding in the ELATIVE trial was adequate.
Baseline characteristics	B.2.3.2, Table 7, p.50-1	**Appropriate** The company stated that: "*Treatment arms were well balanced for each key demographic and baseline variable*." The clinical advisor agreed that the population characteristics were well balanced across the treatment groups. The EAG therefore considers that this is appropriate.
Dropout rate	B.2.3.3.1, Figure 17, p.52	**Appropriate** Across both arms of the ELATIVE trial, the discontinuation rate was under 20%. As such, the EAG has no concerns about the dropout rate.
Statistical analyses	B.2.4.2, p.53-6	**Appropriate** The statistical analyses were appropriate to detect effects.
Outcome measures	B.2.3.1, Table 6, p. 48-9; PfC A9	**Appropriate** The company used surrogate composite endpoints (ALP ≤ 1.67 x ULN, TB ≤ ULN, and ALP decrease ≥15%) indicative of cholestatic response as the primary outcome. The EAG asked the company to justify this choice of primary outcome and comment on whether alternative measures could have been used. The company clarified the reasoning for a composite of surrogate outcomes to measure the primary outcome in the ELATIVE trial in the PfCs; the EAG were satisfied with their response. Furthermore, although mortality was not measured as an outcome measure in the ELATIVE trial, deaths were reported as an adverse event and mortality was considered in the economic model by using life years and QALYs.

Trial design or conduct concept	Section in CS where methods are reported	EAG's assessment
Results: Efficacy outcomes	B.2.6, p.56-61	**Appropriate** The EAG has no concerns about the reporting of efficacy outcomes in the ELATIVE trial.
Results: Adverse events	B.2.10, Tables 25-26, p.90-92	**Appropriate** The EAG has no concerns about the reporting of adverse events in the ELATIVE trial.
Results: Subgroup analyses	B.2.7, p.66-9. PfC A8	**Some concerns** The subgroups 'inadequate response' and 'unable to tolerate UDCA' were not considered in the ELATIVE trial subgroup analysis. The EAG asked the company to provide additional information on the distribution of participants who had an inadequate response to UDCA or were unable to tolerate UDCA in the trial. See section 3.2.5 for further details.

Abbreviations: ALP = alkaline phosphatase; CS = company submission; EAG = Evidence Assessment Group; NICE = National Institute for Health and Care Excellence; OCA = obeticholic acid; PBC = primary biliary cholangitis; PfC = points for clarification; TB = total bilirubin; UDCA = ursodeoxycholic acid; ULN = upper limit of normal

3.2.1 Allocation concealment

The process and methods for concealing allocation was not reported in the CS, though the protocol provided as supplementary material to an associated article describes how allocation took place (protocol section 7.4).[44] However, the method of concealing allocation is unclear.[4] The EAG is concerned whether treatment allocation was adequately concealed, as this is important in preventing any potential bias in the reporting of subjective outcomes, such as for the PBC-40 questionnaire.[23]

3.2.2 Results: Subgroup analyses

Subgroups stratifying the population between those who had an inadequate response to UDCA and those who were unable to tolerate UDCA were not considered in the ELATIVE trial as the company stated that trial population was representative of the distribution of patients treated with and without UDCA in clinical practice (CS Section B.1.1, Table 1, p.12) The EAG asked the company to provide additional information on the distribution of participants who had an inadequate response to UDCA or were unable to tolerate UDCA in the trial (PfC A8). The company responded that the ELATIVE trial enrolled a population representative of a typical population of patients with PBC in need of second-line therapy, reporting that 95% of participants in the trial (153/161) were using UDCA concurrently, while the 5% (8/161) of participants who were intolerant to UDCA received elafibranor monotherapy or placebo.[7] Clinical advice to the EAG noted those who are intolerant to UDCA and those who do not respond to UDCA are two clinically different populations, but appreciated that numbers in these subgroups may have been too small to facilitate stratified analysis. The EAG agrees that, although such subgroup analyses may have facilitated understanding of the effectiveness of elafibranor for these two population groups, the overall numbers of participants who were unable to tolerate UDCA was too low to be able to facilitate this stratification.

3.3 Critique of trials identified and included in the indirect comparison and/or multiple treatment comparison

The company conducted an network meta-analysis (NMA) between the ELATIVE and POISE trials to indirectly compare the effectiveness of elafibranor against obeticholic acid (OCA). A summary of the EAG's critique of the NMA is provided in Table 3.3.

Table 3.3: Summary of the EAG's critique of the company's indirect comparisons

Aspect of NMA design or conduct	Section in CS where methods are reported	EAG's assessment
Statistical methods	B.2.9.1, p.70-1; Appendix D.1.5, p.87-101; PfCs A10, A16, A17, A18	**Some concerns** The company conducted an NMA to assess the effectiveness of elafibranor and OCA. The EAG have concerns with regards the appropriateness of conducting an NMA when only two studies were included. Furthermore, the EAG asked the company to clarify multiple points surrounding the statistical methods, including the choice of priors, the presentation of the DICs and SUCRAs, and the methods used to assess heterogeneity. See Section 3.3.1 for further details.
Included and excluded studies	Appendix D.1.4.2 and Appendix D.1.4.2.1, p.63-82; PfC A13 and A14	**Some concerns** The EAG questioned why the COBALT trial had been excluded from the NMA and whether the company had contacted the study authors to obtain information to ascertain eligibility. Furthermore, the EAG also questioned whether data from the phase II trial of elafibranor and of OCA could have been used to inform analyses where the time-point used was the earliest measured within the trials.[24,25] See Section 3.3.2 for further details.
Included study characteristics and demographics	Appendix D.1.4.2.2, Table 27, p.86	**Appropriate** Clinical advice to the EAG considered that the baseline demographics of both trials were balanced enough to feasibly permit pooling.
Transitivity assumption	PfC A15	**Some concerns** The EAG asked the company to comment on the transitivity assumption within the NMA in the PfCs. The company stated that the distribution of treatment effect modifiers within ELATIVE and POISE were similar, meaning the transitivity assumption was not violated. However, the EAG note that data were not available for a key effect modifier ANA positive status. See Section 3.3.3 for further details.
Results	B.2.9.1.1 to B.2.9.1.11, p.72-86	**Key issue 1** Results of the NMAs feeding into the company's economic model were highly uncertain due to wide CrIs and the EAG have some concerns that the use of ORs to assess dichotomous outcomes instead of RRs may have overestimated effectiveness.

Aspect of NMA design or conduct	Section in CS where methods are reported	EAG's assessment
		See Section 3.3.4 for further details. **Some concerns** Although included in the decision problem, fatigue was not included as an outcome within the company's NMAs. See Section 3.3.4 for further details.
Subgroup analyses	PfC A21	**Appropriate** The EAG asked the company to clarify whether subgroup analyses were planned or performed, the company confirmed that no subgroup analyses were performed for the NMAs. Given that the decision problem did not specify any subgroups, the EAG considers this appropriate.
Sensitivity analyses	Appendix D, PfC A22	**Some concerns** Only a sensitivity analysis changing the NMA structure from random effects to fixed effects was undertaken in the CS. The EAG asked the company to clarify whether any other sensitivity analyses were planned or performed. The company noted that no further sensitivity analyses were performed, therefore the EAG conducted their own sensitivity analyses by changing dichotomous outcomes from ORs to RRs. See Sections 3.3.4.3 and 3.4 for further details.
Abbreviations: ANA = antinuclear antibodies; CrI = credible interval; CS = company submission; DIC = Deviance Information Criterion; EAG = Evidence Assessment Group; NMA = network meta-analysis; OCA = obeticholic acid; OR = odds ratio; PfC = points for clarification; RR = risk ratio; SUCRA = Surface Under the Cumulative RAnking curve		

3.3.1 Statistical methods

3.3.1.1 Rationale for conducting an NMA

In the absence of head-to-head evidence between elafibranor and OCA, the company conducted an NMA between the ELATIVE and POISE trials. The EAG asked the company to clarify the rationale behind conducting an NMA when only two studies were considered relevant to the decision problem (PfCs, A10). The company responded that the rationale for choosing an NMA was provided in Appendix D (Sections D.1.4.1 to D.1.4.3) and briefly summarised in the CS (Section B.2.9).[1,7,9] The company noted that the NICE Decision Support Unit (DSU) recommends a Bayesian approach to NMA and, because the network connected elafibranor to OCA 5-10 mg and ELATIVE and POISE were considered sufficiently homogenous by the company's clinical experts, an NMA was permissible.[26]

The EAG acknowledge that a Bayesian NMA is an appropriate methodology recommended by the NICE DSU.[26] Additionally, the EAG's clinical advisor confirmed that the POISE and ELATIVE trials seemed sufficiently homogenous to pool. However, the company noted that there were issues with convergence in the random-effects NMA (Appendix D, Section D.1.5,

p.89).[9] This is exemplified by the substantial number of burn-ins and iterations reported for each random-effects Bayesian NMA conducted. For example, the NMA for the primary outcome of 'Odds of achieving cholestasis response at 12 months' had a burn-in of [redacted], followed by [redacted] iterations with a thinning interval of [redacted] (CS Section B.2.9.1.1, Figure 30, p.72).[1] Furthermore, there is also a large amount of uncertainty in the results of the NMA (see Section 3.3.3). Given this, the EAG believe that the company could have explored other methodologies to compare the clinical effectiveness of elafibranor versus OCA 5-10 mg (such as an anchored matching-adjusted indirect comparison).

3.3.1.2 Choice of informative priors for pruritis outcomes

The company derived their choice of informed priors for the NMA from Turner et al (2015).[27] For the outcomes surrounding pruritis, the company chose to use the "Subjective outcomes" informative prior for the following outcomes (Appendix D.1.5, p.89).[9]

- Mean change from baseline in pruritis (5-D Itch) at 12 months
- Mean change from baseline in pruritis (5-D Itch) at 2-4 weeks
- Mean change from baseline in pruritis (PBC-40 Itch domain) at 12 months
- Mean change from baseline in pruritis (PBC-40 Itch domain) at 2-4 weeks

The EAG asked the company to explain the rationale behind applying this informative prior as opposed to the "Signs/symptoms reflecting continuation/end of condition" informative prior, given that pruritis can be seen as a symptom of PBC continuation (PfC A19).[28] The company responded that the measurement methods for assessing pruritis are not objective and that there are no biomarkers associated with PBC that directly correlate to the presence or severity or pruritis and, as such, subjective measurements (the PBC-40 Itch and 5-D Itch questionnaires) were used in both ELATIVE in POISE.[7] However, the EAG note that the paper by Turner et al (2015) lists "Signs/symptoms reflecting continuation/end of condition" as a subjective outcome (Table 3).[27] As such, the EAG believe that conducting a sensitivity analysis on these four outcomes using the alternative informative prior may have been justifiable.

As such, the EAG requested that the company re-run the NMAs for the pruritis outcomes listed above using the "Signs/symptoms reflecting continuation/end of condition" informative prior from Turner et al (2015; PfC A20).[27] The company responded by providing new analyses for these outcome measures using the "Signs/symptoms reflecting continuation/end of condition" informative prior from Turner et al (2015).[7,27] The EAG present a comparison of the results for elafibranor versus OCA 5-10 mg between the different priors in Table 3.4 below. The EAG are satisfied that there is little to no difference between results dependent on the choice of priors for the pruritis outcomes.

Table 3.4: Comparison of results for pruritis outcomes using company and alternative choice of priors

Outcome measure	Company prior: Subjective outcomes	Alternative prior: Signs/symptoms reflecting continuation/end of condition
Mean change from baseline in pruritis (5-D Itch) at 12 months	[redacted]	[redacted]

Mean change from baseline in pruritis (5-D Itch) at 2-4 weeks	[redacted]	[redacted]
Mean change from baseline in pruritis (PBC-40 Itch domain) at 12 months	[redacted]	[redacted]
Mean change from baseline in pruritis (PBC-40 Itch domain) at 2-4 weeks	[redacted]	[redacted]

Source: created by the EAG using data from CS Sections B.2.9.1.4, B.2.9.1.5, B.2.9.1.6 and B.2.9.1.7; and PfC A20.[1,7]
Abbreviations: CS = company submission; EAG = Evidence Assessment Group; PBC-40 = primary biliary cholangitis-40; PfC = points for clarification

3.3.1.3 Reporting of Deviance Information Criterion and Surface Under the Cumulative Ranking curve

The company did not report the Deviance Information Criterion (DIC) or Surface Under the Cumulative Ranking (SUCRA) curves for the NMAs in the original CS. The EAG asked the company to clarify why the DIC or SURCRA had not been presented (PfCs A16 and A17). For the query regarding the DICs (PfC A16), the company responded by providing a table outlining the DIC values for all NMAs, comparing the fit of the random effects model with the fixed effects model.[7] As also noted by the company in their response, a difference of less than three points suggests there is little difference between the models.[29] The EAG are satisfied that the DICs presented by the company suggest there is little difference in fit between the random and fixed effect models.

Regarding the SUCRAs (PfC A17), the company stated that they did not report these within the CS as it is not a requirement of the NICE DSU and they would not provide any additional information to differentiate beyond the summary statistics already presented.[7] The EAG acknowledge that SUCRAs are not mentioned in NICE DSU Technical Support Document 2 and appreciate that, in light of the small amount of treatments being compared within the NMAs, presenting SUCRAs may not have had any further benefit to presenting the summary statistics and posterior probabilities.[26]

3.3.1.4 Methods used to assess heterogeneity

In Appendix D (Section D.1.5, p.89), the company state: "*In order to truncate the priors on the continuous outcomes, different informative priors were identified to enable assessment of the between-study standard deviation on the standardised mean difference scale*." Given that this is not a standard method for assessing heterogeneity within NMAs, the EAG asked the company to provide a reference and explain the rationale behind using this method (PfC A18). The company responded by stating that the priors were truncated according to methodology recommended by Ren et al (2018) to prevent simulation of excessively large between-study variance and aid convergence in the NMAs.[7,30] The EAG believe that the use of the Ren et al (2018) methodology was appropriate.[30]

3.3.2 Included and excluded studies

3.3.2.1 Exclusion of the COBALT trial

Although most of the reasons for studies' exclusion from the NMAs given within Appendix D were deemed appropriate (Section D.1.4.2, Table 22), the company noted that the COBALT trial was excluded and stated: "*Study not published in full to facilitate balanced evaluation in the feasibility assessment*" (p.64).[9] As such, the EAG asked the company to clarify whether they had attempted to contact the authors of the COBALT trial to obtain the information required to ascertain its suitability for the NMA (PfC A13). The company responded that the COBALT trial was terminated early due to feasibility challenges where the data monitoring committee noted that the objectives of the trial were not feasible; the trial did not demonstrate a statistically significant difference in clinical endpoints between OCA and placebo and results were only reported as an abstract.[7,31]

The EAG acknowledge that the limited detail in the abstract would have made assessing the similarity of COBALT with ELATIVE and POISE difficult. However, the EAG asked the company to clarify whether they had asked the authors of the COBALT study for further information to be able to assess the suitability of the trial for NMA. The company did not provide this information for this question and, in a previous PfC response (A6), noted that contacting authors for information was not conducted.[7] As such, the EAG believe that the company could have contacted the authors of the COBALT study to ask about key effect modifiers, which may have allowed assessment of suitability for the NMA. In not doing so, it is possible that the company excluded a potentially eligible study from the NMA.

3.3.2.2 Exclusion of Schattenberg et al (2021) and Hirschfield et al (2015)

A phase II trial of elafibranor (Schattenberg et al 2021) and a trial of OCA (Hirschfield et al 2015) were also excluded from the company's NMAs.[24,25] The company excluded both studies because the ELATIVE trial was "*designed to evaluate efficacy after 12 months of treatment, studies which provide only 12 weeks of treatment would not be comparable in their outcomes*" (Section D.1.4.2.1, p.77).[9] However, the EAG asked the company to provide further rationale (PfC A14), given that some outcomes used within the NMAs were measured at earlier time-points than 12 months (e.g. 'Change from baseline in pruritus according to the 5-D Itch score questionnaire using the earliest reported data after commencement of treatment').

For Hirschfield et al 2015, the company responded that neither PBC-40 Itch nor 5-D Itch were reported and did not assess OCA at its licensed dose of 5-10 mg; therefore, the study could not have been included in the relevant analyses.[7] The EAG are satisfied with this rationale. For Schattenberg et al (2021), the company noted that 5-D Itch was not reported and PBC-40 Itch data were only reported as a median percentage change from baseline without population size, SDs or standard errors (SEs) provided. The EAG acknowledge that the data regarding pruritis measured using PBC-40 was limited within Schattenberg et al (2021), with data not shown within the paper.[24] However, as with the COBALT study (see section 3.3.2.1), it may have been possible for the company to contact the authors of the Schattenberg et al (2021) trial in order to obtain further information needed to fully assess its suitability for NMA.

3.3.3 Transitivity assumption

Following details of the feasibility assessment for the NMA presented in Appendix D, the EAG asked the company to provide further comment on the transitivity assumption (PfC A15). The company responded that five key effect modifiers were considered as part of their feasibility assessment: age at diagnosis, ALP levels, TB level, cirrhosis and antinuclear antibody (ANA) positive status. ALP levels and TB levels were deemed sufficiently similar in ELATIVE and POISE. Age at diagnosis was not directly reported by ELATIVE but was calculated using participants' age and time since diagnosis to compare with POISE, which was also found to be sufficiently homogenous. Cirrhosis was not directly reported in either ELATIVE or POISE, though the company noted that their clinical experts stated that liver stiffness of 17 kPa or more could be used as a proxy; the difference between the two trials in terms of liver stiffness was deemed to be within a reasonable margin of error.[7]

However, ANA positive status was not reported in either ELATIVE nor POISE, and so an assessment of homogeneity could not be conducted.[7] Given that the company have noted that ANA positive status was a key effect modifier, it is not possible to know whether the participants in POISE and ELATIVE were sufficiently similar for this variable. As such, it is not known whether the transitivity assumption has been violated in terms of ANA positive status, which may threaten the validity of the indirect estimates.[32]

3.3.4 Results

3.3.4.1 Exclusion of fatigue as an outcome measure in the NMA

The final NICE scope listed fatigue as an outcome of interest for the submission.[33] Furthermore, information provided by patient organisations to NICE noted that fatigue was a key symptom of PBC that impacts on quality of life,[34,35] while the British Society of Gastroenterology/UK-PBC guidelines state: "*Fatigue is a significant problem in up to half of patients and is complex in nature. Social isolation is an important factor in poor QoL in fatigued patients with PBC*."[17] As such, the EAG asked the company to justify why fatigue had not been considered for an NMA, given that data from the ELATIVE trial on the PROMIS Fatigue Short Form 7a Score was available (CS Section B.2.6.1.3.7, p. 66; PfC A12).[1]

The company responded that fatigue was measured by both ELATIVE and POISE using the PBC-40 fatigue domain but that there was no evidence from either trial that there was a significant impact of treatment on the symptoms compared with placebo. Therefore, the company concluded that including fatigue in the economic model would not significantly influence the results and, as such, did not pursue an NMA for this outcome.[7] However, further post-hoc ELATIVE trial results are being planned to understand the impact of elafibranor on fatigue. The EAG believe that a lack of evidence of effect in individual trials does not preclude pooling within an NMA. In not pooling the data on fatigue from ELATIVE and POISE, it is not possible to determine the indirect treatment effect of elafibranor compared with OCA 5-10 mg for this outcome.

3.3.4.2 Uncertainty in NMA results used within the economic model

Results for 'Odds of achieving cholestasis response at 12 months' (CS Section B.2.9.1.1, p.72-3), 'Mean change in pruritis (PBC-40 Itch) from baseline at 12 months' (CS Section B.2.9.1.6, p.78-9), 'Mean change in PBC-40 Itch from baseline using the earliest reported data after commencement of treatment' (CS Section B.2.9.1.7, p.80-1), 'Odds of occurrence of pruritis

TEAE of any severity within 12 months' (CS Section B.2.9.1.8, p.81-2) and 'Odds of discontinuation (all-case) within 12 months' (CS Section B.2.9.1.10, p.84-5) against OCA 5-10 mg were all used within the economic model. The results were as follows.

[redacted]

[redacted]

[redacted]The company noted that there were issues with convergence in the random-effects NMA (Appendix D, Section D.1.5, p.89), and with less than five studies used within the NMA, a fixed-effects model may have been preferable. However, the lack of any significant change in the results for the fixed-effects models compared with the random-effects models suggests this choice would not change the uncertainty in the NMA results for outcomes used to inform the cost-effectiveness.

3.3.4.3 Use of odds ratios as opposed to risk ratios for dichotomous outcome measures

The EAG note that the company used ORs instead of RRs to assess the effectiveness of elafibranor within the NMA. When there is an association between the exposure and outcome, ORs tend to overestimate the effects of interventions if misinterpreted as RRs, which could present an issue; however, the qualitative direction of effect will not be changed.[23,36] Furthermore, NICE DSU Technical Document 1 states: "*A clear discussion of the underlying statistical and clinical assumptions implied by the model, and their impact on the final decision should also be provided. In particular, reasons for choosing to model the outcomes on a particular scale (e.g. odds ratio, hazard ratio, risk difference etc) and the assumptions implied in any transformation from the relative to the absolute effects should be clearly presented*" (p.17-8). This rationale was not provided in the CS. As such, the EAG conducted their own analyses to estimate the relative effect of elafibranor versus OCA 5-10 mg for the dichotomous outcomes 'Odds of achieving cholestasis response at 12 months', 'Odds of occurrence of pruritis TEAE of any severity within 12 months' and 'Odds of discontinuation (all-case) within 12 months' using RRs (see Section 3.4 below).

3.4 Additional work on clinical effectiveness undertaken by the EAG

The EAG asked the company to provide the datasets used to perform the NMAs within the CS (PfC A11), which the company provided.[7] The EAG used data already provided in the CS (Document B) and data from the POISE trial, obtained via TA443,[1,6] as well as the data provided by the company in response to the PfCs, to perform additional analyses.

Given the potential issues of using ORs, as noted in Section 3.3.3 above, the EAG aimed to conduct sensitivity analyses for the outcomes 'Odds of achieving cholestasis response at 12 months,' 'Odds of occurrence of pruritis TEAE of any severity within 12 months' and 'Odds of discontinuation (all-case) within 12 months' by conducting NMAs where the effects were presented as RRs with associated 95% CrIs. However, the EAG were unable to satisfactorily run these analyses using OpenBUGS or the gemtc R package.

Frequentist NMAs using both random and fixed effects models using both RRs and ORs were run by the EAG for all binary outcome measures used within the economic model, listed above. This was done to check that the results were plausible given the difficulty in achieving convergence using the Bayesian approach. To run the frequentist NMAs, the EAG used the online application MetaInsight, which uses code from the netmeta R package to generate results.[37,38] To facilitate running the NMAs for continuous outcomes, the EAG transformed the standard errors (SEs) provided by the company in their datasets into standard deviations (SDs) using the following formula in Microsoft Excel: SD=SE*sqrt(n).

3.4.1 Dichotomous outcomes

3.4.1.1 Random-effects analyses

Results comparing the company's random-effects NMA results with the EAG's alternative results for dichotomous outcomes used in the economic model are presented in Table 3.5. As noted in Section 3.4 above, the EAG were not able to satisfactorily run a random-effects Bayesian NMA using RR for dichotomous outcomes. When using a fixed-effects frequentist model with OR, the result for the odds of pruritis TEAE of any severity at 12 months became statistically significant. However, the remaining confidence intervals derived from the frequentist analyses were wide and still not statistically significant, suggesting uncertainty in the underlying data and the overall effectiveness of elafibranor.

Table 3.5: NMA results comparing elafibranor 80 mg versus OCA 5-10 mg using different approaches (dichotomous outcomes, random-effects models)

Outcome measure	Company base-case (Bayesian OR, random-effects)	Frequentist OR, random-effects	Frequentist RR, random-effects
Cholestasis response at 12 months			
Odds of pruritis TEAE of any severity at 12 months			
All-cause discontinuatio			

Outcome measure	Company base-case (Bayesian OR, random-effects)	Frequentist OR, random-effects	Frequentist RR, random-effects
n at 12 months			

Source: created by the EAG and using data from CS Sections B.2.9.1.1 (p.72-3), B.2.9.1.6 (p.78-9), B.2.9.1.7 (p.80-1), and B.2.9.1.10 (p.84-5)[1]
Abbreviations: CI = confidence interval; CrI = credible interval; CS = company submission; EAG = Evidence Assessment Group; NMA = network meta-analysis; OCA = obeticholic acid; OR = odds ratio; RR = risk ratio; TEAE = treatment-emergent adverse event

3.4.1.2 Fixed-effects analyses

Results comparing the company's fixed-effects NMA results with the EAG's alternative results for dichotomous outcomes used in the economic model are presented in Table 3.6. When using a fixed-effects frequentist model with RR, the result for the odds of pruritis TEAE of any severity at 12 months was no longer statistically significant compared with the company's approach. However, the remaining confidence intervals derived from the frequentist analyses were wide and still not statistically significant, suggesting uncertainty in the overall effectiveness of elafibranor.

Table 3.6: NMA results comparing elafibranor 80 mg versus OCA 5-10 mg using different methodologies (dichotomous outcomes, fixed-effects models)

Outcome measure	Company base-case (Bayesian OR, fixed-effects)	Bayesian RR, fixed-effects	Frequentist OR, fixed-effects	Frequentist RR, fixed-effects
Cholestasis response at 12 months	[redacted]	[redacted]	[redacted]	[redacted]
Odds of pruritis TEAE of any severity at 12 months	[redacted]	[redacted]	[redacted]	[redacted]
All-cause discontinuation at 12 months	[redacted]	[redacted]	[redacted]	[redacted]

Source: created by the EAG and data from CS Appendix D (Sections D.1.6.1, D.1.6.6, D.1.6.8 and D.1.6.11)[9]
Abbreviations: CI = confidence interval; CrI = credible interval; CS = company submission; EAG = Evidence Assessment Group; NMA = network meta-analysis; OCA = obeticholic acid; OR = odds ratio; RR = risk ratio; TEAE = treatment-emergent adverse event

3.4.2 Continuous outcomes

Results comparing the company's random and fixed-effects frequentist NMA results with the EAG's alternative results for continuous outcomes used in the economic model are presented in Table 3.7. When the EAG used a frequentist, random-effects approach, the NMA result for the change in the PBC-40 Itch domain at 12 months became statistically significant in favour of elafibranor. However, there was little difference between the company's and EAG's analyses for the remaining analyses surrounding change in PBC-40 Itch domain score at 12 months and at 2-4 weeks.

Table 3.7: NMA results comparing elafibranor 80 mg versus OCA 5-10 mg using different approaches (continuous outcomes)

Outcome measure	Company NMAs		EAG analyses	
	Bayesian, median difference in mean change, random-effects	Bayesian, median difference in mean change, fixed-effects	Frequentist MD, random-effects	Frequentist MD, fixed-effects
Change in PBC-40 Itch domain at 12 months	[redacted]	[redacted]	[redacted]	[redacted]
Change in PBC-40 Itch domain at 2-4 weeks	[redacted]	[redacted]	[redacted]	[redacted]

Source: created by the EAG and data from: CS Sections B.2.9.1.6 and B.2.9.1.7 (p.78-81); CS Appendix D, Sections D.1.6.6 and D.1.6.7; and PfC A11.[1,9]
Abbreviations: CI = confidence interval; CrI = credible interval; CS = company submission; EAG = Evidence Assessment Group; MD = mean difference; NMA = network meta-analysis; OCA = obeticholic acid;

3.5 Conclusions of the clinical effectiveness section

An SLR was conducted to identify evidence on the effectiveness and safety of elafibranor and relevant comparators for treating PBC. The EAG have some concerns surrounding multiple aspects of the SLR process, such as the literature search, eligibility criteria, screening, data extraction, and quality appraisal. The review question suggests that the aim of the review is to identify RCTs that evaluate the efficacy and/or safety of elafibranor and other comparators of interest whereas the inclusion criteria list observational studies among the eligible study designs to be included in the review. No protocol nor a reference to a published or registered protocol was provided. Therefore, it was not possible to judge if prespecified eligibility criteria were adhered to throughout the process. The company did not attempt to contact study authors regarding missing or unclear information, which could have led to excluding eligible studies from the SLR. Moreover, RCTs of mixed-line treatments were excluded, which could have resulted in excluding RCTs where results of eligible treatment lines were reported separately. If the results of any such studies were not considered in the SLR, this may have led to potentially eligible studies being excluded from the SLR and, consequently, this could possibly have impacted the NMA and subsequently the economic model.

The main clinical evidence was based on the ELATIVE trial, a multinational RCT investigating the efficacy and safety of elafibranor 80 mg with or without UDCA versus placebo with or without UDCA in 161 adult patients with PBC who have had an inadequate response to or

were unable to tolerate UDCA, followed up over a 52-week period. The EAG believe the ELATIVE trial was mainly conducted appropriately, though the process for allocation concealment was unclear. Moreover, the subgroup analyses performed did not include the subgroups “inadequate response” and “unable to tolerate UDCA.” Though clinical advice to the EAG noted that these are two clinically different populations, the EAG appreciates that the overall numbers of participants who did not tolerate UDCA was too low to be able to facilitate this stratification. No data were reported from the trial on long-term outcomes listed in the scope, such as mortality and liver transplantation, but the EAG appreciate that these are considered in the economic analyses.

To compare the relative efficacy of elafibranor with OCA, the company performed a series of indirect treatment comparisons in the form of NMAs. The EAG had several concerns surrounding the choice of methodology and presentation of the methods, though the company provided additional information and rationale for many of these queries during the points for clarification process. However, the EAG noted that the width of the 95% CrIs, including when elafibranor is compared against OCA 5-10 mg, were substantially wide and the company noted that there was difficulty in achieving convergence within the model. As such, the EAG performed multiple additional NMA analyses to explore the effect of changing the model on the results for outcomes used within the economic model. In general, the results of these analyses would not change the overall conclusions of the NMAs. Although the results of the EAG analyses mostly aligned with those of the company’s base case and fixed-effect sensitivity analyses, it should be noted that the results were still open to substantial uncertainty and it is therefore difficult to draw any conclusions regarding the clinical effectiveness of elafibranor versus OCA 5-10 mg.

4 COST EFFECTIVENESS

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

This section pertains mainly to the review of cost-effectiveness analysis studies. However, the search section also contains summaries and critiques of other searches related to cost-effectiveness presented in the company submission. Therefore, the following section includes searches for the cost-effectiveness analysis review, measurement and evaluation of health effects as well as for cost and healthcare resource identification, measurement and valuation.

Table 4.1 presents an overview of the EAG's critique of the methods used to identify studies for the review of cost-effectiveness.

Table 4.1: Summary of the EAG's critique of the methods for the review of cost-effectiveness

Aspect of cost-effectiveness SLR	Section in CS where methods are reported	EAG's assessment
Data sources for cost-effectiveness analysis review	Appendix G.1, p. 1	**Some concerns** A systematic review was carried out by the company in November 2022 with a December 2023 update. Searches were conducted simultaneously for cost-effectiveness studies, HRQoL, and cost and resource use. An appropriate range of electronic bibliographic databases, HTA websites and conference proceedings were searched alongside hand-searching bibliographies to identify additional relevant studies, but the reason given for restricting the years for which conference proceedings were searched was unconvincing. Conference proceedings were excluded from Embase searches, which could have missed relevant studies. See Section 4.1.1 for further details.
Search strategies	Appendix G.1.1, p. 2; Appendix H.1.1, p. 1; Appendix I.1.1, p. 1	**Some concerns** The search strategy used to find cost-effectiveness studies is generally fit for purpose; however, the use of focussed MeSH headings may have increased specificity to the detriment of sensitivity. See Section 4.1.2 for further details.
Search filters	Appendix G.1.1, p. 2	**Appropriate** The search filters adequately captured the decision problem.
Data sources for model input	Appendix G.1.3, p. 21; Appendix I.1.3, p. 4; Appendix H.1.3, p. 3	**Appropriate** Eight cost-effectiveness studies were identified, of which four were health technology appraisal submissions for OCA using the same model structure.
Eligibility criteria for inclusion of	Appendix G.1.2, p. 19	**Appropriate** The eligibility criteria were appropriate to capture cost-effectiveness studies in this area.

Aspect of cost-effectiveness SLR	Section in CS where methods are reported	EAG's assessment
economic evaluations		
Eligibility criteria for inclusion of health state utility value studies	Appendix H.1.2, p. 2	**Appropriate** The eligibility criteria were appropriate to capture quality of life data in this area.
Eligibility criteria for inclusion of resource use and cost studies	Appendix I.1.2, p. 2	**Appropriate** The eligibility criteria were appropriate to capture resource use and costs in this area.
Abbreviations: CS = company submission; EAG = Evidence Assessment Group; HRQoL = health-related quality of life; HTA = health technology assessment; OCA = obeticholic acid		

4.1.1 Data sources for cost-effectiveness analysis review

The range of sources searched by the company for the cost-effectiveness, health-related quality of life (HRQoL), and cost and resource SLR was appropriate: electronic bibliographic databases (MEDLINE, Embase, INAHTA); conference proceedings and hand searching of reference lists. The company excluded conference proceedings in Embase, which could have led to the exclusion of relevant records. However, clinical advice to the EAG confirmed that the two main conferences in the area had been covered by the company's hand searches.

4.1.2 Search strategies

The company's search included filters to identify cost-effectiveness studies, cost and resource use (direct and indirect), HRQoL and utilities developed by SIGN. This was amended to include additional search terms which may have increased the scope of the search. Whilst these adaptations may have increased the sensitivity of the search, the use of focussed MeSH and Emtree headings within the search strategy have the opposite effect and may have limited the sensitivity of the search. Please see section 3.1.1.2 where this issue is discussed further along with the restriction of conference abstracts searched to 2021 onwards.

4.2 Conclusions of the cost effectiveness review

The SLR was unable to find previously published economic evaluations assessing the cost-effectiveness of elafibranor as a treatment for PBC. A total of eight economic evaluations were identified by the company, four of which were HTAs for OCA using the same model structure, one published micro-simulation for OCA, two publications focused on UDCA, and one publication on liver transplantation (CS Section B.3.1, Table 29).[1] The structure used to evaluate OCA across the four health technology appraisals [6,39-41] informed the structure in the clinical pathway for the current submission.

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

4.3.1 NICE reference case checklist

Table 4.2 summarises the NICE reference case checklist and the EAG's assessment on the company's submission in relation to their base-case analysis.

Table 4.2: NICE reference case checklist

Element of health technology assessment	Reference case	EAG comment on company's submission
Defining the decision problem	From the scope: Adults with PBC whose disease has an inadequate response to, or who are unable to tolerate UDCA.	**Appropriate** The population included patients who are unable to tolerate UDCA and patients with an inadequate response to UDCA. There were a small number of patients unable to tolerate UDCA. The EAG considered this approach appropriate.
Comparators	For people whose disease has an inadequate response to UDCA: • OCA in combination with UDCA • UDCA monotherapy For people who are unable to tolerate UDCA: • OCA monotherapy • Best supportive care	**Some concerns** OCA and UDCA are presented as the second-line treatment alternatives to elafibranor, in line with the scope developed together with NICE. As patients who discontinue move to UDCA as third-line treatment, the EAG is concerned that elafibranor could be used with OCA in sequence as an alternative treatment strategy to the elafibranor to UDCA sequence. Consultations with a clinical expert indicated this is a possibility given the different mechanisms of action between elafibranor and OCA but clinical effectiveness data may be scarce. The scope also omits the use of fibrates in second-line treatment. These are typically used off-label and could make up a sizeable share of second-line treatment, even if they do not have regulatory approval. See sections 2.1.1 and 4.3.2.1 for further details.
Perspective on outcomes	Outcome measures from the final scope considered to be included: • Mortality	**Some concerns** Outcomes included in the cost-effectiveness model were: • Mortality (life years gained)

Element of health technology assessment	Reference case	EAG comment on company's submission
	• Liver function based on markers of liver biochemistry • PBC symptoms including pruritus, fatigue, and abdominal pain • Time to liver transplantation • PBC related events including ascites, varices, encephalopathy, and hepatic cell carcinoma (HCC) • Adverse effects of treatment • Health-related quality of life (HRQoL).	• Liver function biomarkers • Pruritus • Time to liver transplantation • PBC-related events included HCC and decompensated cirrhosis (DCC) • Adverse events: pruritus, fatigue, and urinary tract infections • HRQoL measured in QALYs The economic model does not explicitly parametrise the impact of ascites, varices, encephalopathy or abdominal pain from PBC onto each health state. See section 4.3.5.2 for further details.
Perspective on costs	NHS and personal social services (PSS)	**Appropriate** The EAG considers the perspective on costs was adequately captured.
Type of economic evaluation	Cost-utility analysis with a fully incremental analysis	**Appropriate** The company presented a full cost-utility analysis using QALYs over an ICER for OCA.
Time horizon	Long enough to reflect all important differences in costs and outcomes between the technologies being compared	**Appropriate** A lifetime horizon was used for the cost-effectiveness analysis.
Synthesis of evidence on health effects	Based on a systematic review	**Appropriate** An NMA was performed including the ELATIVE and POISE trials. A systematic review was used to search for quality-of-life data.
Measuring and valuing health effects	Quality of life to be presented in QALYs. The EQ-5D is the preferred measure of health-related quality of life in adults.	**Appropriate** As per the NICE reference case.
Source of data for	Reported directly by the patients or carers or both.	**Key issue 4**

Element of health technology assessment	Reference case	EAG comment on company's submission
measurement of health-related quality of life		EQ-5D-5L patient data were collected from the pivotal ELATIVE trial, but not used in the model. Utility values from NICE TA443 for the target population were used in the economic model instead. See section 4.3.5.1 for further details.
Source of preference data for valuation of changes in health-related quality of life	Representative sample of the UK population.	**Appropriate** EQ-5D values were scored in accordance with NICE guidelines.
Equity considerations	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit.	**Appropriate** No decision modifiers were applied on the results.
Evidence on resource use and costs	Costs should relate to NHS and PSS resources and should be valued using the prices relevant to the NHS and PSS.	**Some concerns** Costs and resource use mostly sourced from NHS reference costs, PSS and the established trials, which largely consistent with NICE perspective, although a few costs were extracted from the literature based on their systematic review. Evidence from the NHS cost tariffs were not clearly referenced, posing a transparency concern to the EAG. See section 4.3.6 for further details.
Discounting	The same annual rate for both costs and health effects (3.5%)	**Appropriate** Discounting of costs and outcomes was in line with NICE guidelines

Source: Company submission document B, Table 1[1]
Abbreviation: EAG = Evidence Assessment Group; DCC = decompensated cirrhosis; HCC = hepatocellular carcinoma; HUI = health utility index; NHS = National Health Service; NICE = National Institute for Health and Care Excellence; NMA = network meta-analysis; OCA= obeticholic acid; PBC = primary biliary cholangitis; PSS = Personal Social Services; QALY = quality adjusted life-year; UDCA = ursodeoxycholic acid; VAS = visual analogue scale

4.3.2 Decision problem

Table 4.3: Summary of EAG's critique on the design of the decision problem

Analysis feature	Section in CS where methods are reported	EAG's assessment
Population	Document B.3.2.1, p. 110	**Appropriate** Patient baseline characteristics were based on the ELATIVE trial intention-to-treat (ITT) population.
Interventions	Document B.3.2.3, p. 116	**Appropriate** The intervention was elafibranor 80 mg, which is in line with the NICE decision problem.
Comparators	Document B.3.2.3, p. 116	**Some concerns** Elafibranor is compared to OCA and UDCA alone as alternatives for second-line treatment, while third-line treatment after both elafibranor and OCA consists of UDCA and best supportive care. This is consistent with the initial scope developed together with NICE. Beyond this, the EAG considers it may be possible to use elafibranor and OCA together in sequence as an alternative treatment strategy considering the different mechanisms of action from each. This was further confirmed after consultation with a clinical expert in the field, even though there is an evidence gap in the clinical effectiveness of any sequence strategy combining elafibranor and OCA. The use of fibrates was also not considered part of the initial scope since they are typically used off-label. However, fibrates may make an important share of the second-line treatment in this patient population. See sections 2.1.1 and 4.3.2.1 for further details
Perspective	Document B.3.2.2, p. 115	**Appropriate** The company used NHS and PSS perspective in costs and all direct health effects for patients, which is appropriate for the submission.
Time horizon	Document B.3.2.2, p. 116	**Appropriate** The company used a lifetime horizon, which the EAG finds appropriate.
Discounting	Document B.3.2.2, p. 116	**Appropriate** The company used a 3.5% annual discount for cost and health outcomes, which the EAG finds appropriate.
Severity modifier	Document B.3.6, p. 144	**Appropriate** The company concluded that the population did not meet the severity modifier criteria.

Source: EAG outputs
Abbreviations: CS = company submission; EAG = Evidence Assessment Group; ITT = intention-to-treat; OCA= obeticholic acid; NICE = National Institute for Health and Care Excellence; NHS = National Health Service; OCA= obeticholic acid; PSS = personal social services

4.3.2.1 Comparators

- **Potential use of OCA in sequence from elafibranor**

The model assumes a clinical pathway where both second-line elafibranor and OCA are followed by UDCA and best available care after second-line discontinuation. After consultation with an expert clinician, the EAG would like to highlight the fact that, due to the different mechanisms of actions of elafibranor or OCA, sequential treatment strategies can be proposed where OCA is offered to patients discontinuing elafibranor, and vice versa.

The EAG further considers that the positioning of OCA or elafibranor as third-line treatments could affect their cost-effectiveness but acknowledges that more evidence on the effectiveness of OCA or elafibranor as sequential treatments would be required to obtain more concrete results.

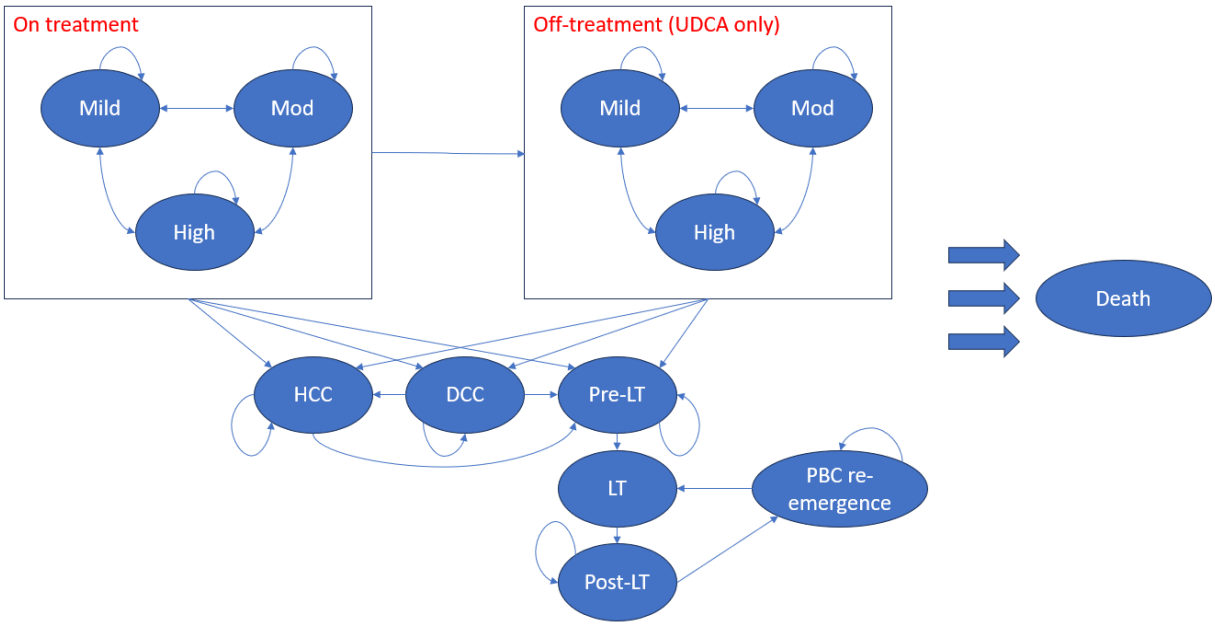
- **Use of fibrates as third-line treatment**

The company was asked about the use of fibrates as a potential treatment within elafibranor's anticipated positioning.[7] In their response, the company highlighted that: fibrates were not included in the NICE scope, since their use is off-label; they have not been studied to regulatory standards for PBC patients; and there are concerns of tolerability issues for patients with cardiovascular disease. [redacted]

4.3.3 Model structure and assumptions

The company model diagram is reproduced in Figure 4.1. The company used a cohort Markov model, with the model structure based on the model developed in NICE TA443.[6] In this model, patients transition between mild, moderate and high risk biomarker states. Responders to treatment, while they are on treatment, have a higher probability of being in a lower risk biomarker state. When a patient discontinues treatment, they effectively return to their initial state pre-second-line treatment, implemented in a manner appropriate to a cohort analysis. There is a probability of transitioning from moderate and high biomarker states to liver disease states. Once a patient has transitioned to a liver disease state, they may progress to other liver disease states, pre/undergoing/post liver transplant (pre-LT/LT/post-LT), liver disease states, PBC re-emergence, and death.[1]

Figure 4.1: Model structure



Source: CS Document B, Section 3.2.2, Figure 41[1]
Abbreviations: CS = company submission; DCC = decompensated cirrhosis; HCC = hepatocellular carcinoma; LT = liver transplant, PBC = primary biliary cholangitis; UDCA = ursodeoxycholic acid

Table 4.4: Summary of EAG's critique on the design of the economic modelTable 4.4 summarises the EAG's critique on the model structure adopted by the company.

Table 4.4: Summary of EAG's critique on the design of the economic model

Analysis feature	Section in CS where methods are reported	EAG's assessment
Type of model	Document B.3.2.2, p.110	**Appropriate** A Markov state-transition model was used. The structure aligned with NICE TA443 assessing OCA treatment for PBC.
Health states/events and transitions	Document B.3.2.2, p.110	**Some concerns** Moderate-risk patients are assumed to progress directly to the liver disease health state without moving through the high-risk. The EAG expects this is likely to be a very small risk, however there was a lack of clarity in the methods used to calculate this parameter. Although this assumption was validated with clinical experts, the EAG is concerned the model predictions were not and are likely to present a scenario where fewer than expected patients remain free from liver disease in the long-term. It was not clear to the EAG how the excess mortality risk at high-risk parameter was obtained, and although this assumption was agreed with clinicians, the EAG

Analysis feature	Section in CS where methods are reported	EAG's assessment
		saw no evidence that the survival predictions were validated by clinicians. The model structure includes a pre-LT state capturing patients with moderate risk, high risk, DCC, and HCC and allows them to stay in this state over their lifetime. The EAG believes there is structural uncertainty around whether transitions direct to LT or through pre-LT should be modelled from each disease state. The EAG is also concerned about the parameters and approach to calculate excess mortality. Although different approaches led to similar survival predictions, comparisons with the literature suggest that model predictions of survival for HCC and DCC may be lower than the survival expected in clinical practice. See section 4.3.3.1 for further details.
UDCA transitions		**Some concerns** The EAG considers it is still a matter of uncertainty what happens to the biomarker risk distribution after patients discontinue second-line treatment. The company's base-case approach was considered appropriate, but the EAG has explored an alternative scenario. The model also makes a strong assumption that after the first year, there are no transitions from moderate risk to mild risk for patients receiving UDCA. [redacted] The model only used the 9-12 month transitions from the placebo arm of ELATIVE to predict the long-term transitions for third-line treatment with UDCA. The EAG questions that the full 12 month transitions during the trial duration were only used as a scenario analysis. See section 4.3.3.2 for further details.
Model predictions		**Key issue 2** The EAG is concerned that the survival predictions made in the model (from liver disease-free, liver transplant-free, and overall survival), were not validated with expert clinicians or the published literature. A point of concern is the potential

Analysis feature	Section in CS where methods are reported	EAG's assessment
		overprediction of patients moving from the PBC biomarker risk states to liver disease in the model. See section 4.3.3.3 for further details.

Source: EAG output
Abbreviations: CS = company submission; DCC = decompensated cirrhosis; EAG = Evidence Assessment Group; HCC = hepatocellular carcinoma; LT = liver transplant; OCA = obeticholic acid; NICE = National Institute for Health and Care Excellence; PBC = primary biliary cholangitis; TA = technology appraisal; UDCA = ursodeoxycholic acid

4.3.3.1 Health states/events and transitions

- **Transitions from the moderate risk category to liver disease**

A deviation from the model structure in NICE TA443 was the possibility of moderate risk patients to transition directly to the liver disease health-states (CS Section B.3.3.2.3, Table 35).[1,6] The company in the current submission argued that the assumption that moderate-risk patients would remain in the moderate-risk health state for the rest of their life in NICE TA443 was criticised by the EAG that reviewed TA443 (CS Section B.3.3.2.3).[1] Therefore, the company assumes that moderate-risk patients can transition directly to liver disease without moving through the high-risk of progression stage.

The EAG Report from NICE TA443,[6] made a critique about the PBC biomarker risk stage in the model, where after 12 months OCA patients are assumed to stay in the moderate or high-risk stage and not move to other risk stages. The current model makes this assumption for both elafibranor and OCA. However, we found no mention of the transitions from moderate-risk to the liver-disease stage of the model in the EAG Report for TA443.[6,43]

The base-case model submitted assumes that moderate-risk patients develop decompensated cirrhosis (DCC) without developing compensated cirrhosis (CC) which is part of the definition of the high-risk level. [redacted]. The EAG acknowledges there can be a minority of patients with a missed CC diagnosis developing DCC. However, this risk is expected to be small compared the risk of CC for moderate risk category patients in the model.

From the information provided in Document B from the company submission[1] it was unclear to the EAG how the cycle probabilities from moderate risk to liver disease were derived. At the factual accuracy check, the company clarified that during the clinical validation of inputs meeting one of the clinicians noted that 6% of the moderate to high-risk health state patients are rapid progressors.[44] This estimate is used to derive the transition probability of being a rapid progressor using a method presented in the FAC the EAG has not yet critiqued. [redacted][44][redacted] Therefore, while it is plausible that a small number of patients in a moderate risk group develop liver

disease either directly or after progressing to a high risk biomarker state first, yet the risk of this transition is still uncertain and may not have been fully explored, hence the EAG has run a scenario excluding this assumption, see scenario 10, section 6.1.2.

- **Excess mortality for high-risk patients**

A difference in the model structure in this submission compared to NICE TA443 was the application of disease-related excess mortality on the cohort of patients categorised as at high-risk of PBC progression (CS Section B.3.2.2, Table 32; Section B.3.3.5, Table 40).[1,6] The company stated that an assumption of a 1.2% excess from the general population mortality was applied upon the advice of clinical experts. During the factual accuracy check, the company indicated that a 5% annual excess mortality risk was provided by the clinicians in the clinical validation of inputs meeting, slide 23 (FIECON 2024, page 16[44]). The EAG was not able to find the quoted estimate but found that one clinician estimated annual excess mortality to be 10% to 15% for high-risk patients, and 2% to 3% for moderate-risk (FIECON 2024, page 16[44]). These inputs were not discussed or assessed on the base-case analysis presented by the company, but more importantly, the survival predictions resulting from these inputs were not validated by neither published evidence nor expert opinion.

The EAG has confirmed with a clinical expert that the primary reason of excess mortality is progression to liver disease. Other than disease progression, fatigue and cardiovascular conditions may lead to a higher risk of mortality relative to the general population. From the perspective of the company, no other clinical explanation for excess mortality was offered other than liver disease progression. The EAG investigates the impact of assuming excess mortality on the cost-effectiveness estimates in a scenario analysis with no excess mortality, see scenario 1, section 6.1.2.

The model prediction for the proportion of patients free from liver disease over time may be too low when compared to GLOBE and UK-PBC predictions. Excess mortality for patients in the high-risk PBC biomarker health state is a key parameter determining the survival predictions in the model. To illustrate, Table 4.5 presents survival at different time points with and without excess mortality in the high-risk PBC biomarker health state. Removing excess mortality assumption in this health state increases the median overall survival of OCA by two years as shown in Table 4.5. No excess mortality risks were applied in low and moderate-risk health states.

Table 4.5: Overall survival for OCA in the company base-case model after correcting for errors 1-12: different approaches to excess mortality at high-risk PBC

	Mortality excess in high-risk (absolute approach)	Mortality excess in high-risk (proportional approach)	No mortality excess in high-risk
Timepoints	**OCA**	**OCA**	**OCA**
1 year	[redacted]	[redacted]	[redacted]
5 years	[redacted]	[redacted]	[redacted]
10 years	[redacted]	[redacted]	[redacted]
20 years	[redacted]	[redacted]	[redacted]
40 years	[redacted]	[redacted]	[redacted]
Median (years)	[redacted]	[redacted]	[redacted]
Source: CS economic model, EAG output			

Abbreviations: CS = company submission; EAG = Evidence Assessment Group; OCA = obeticholic acid

- **Calculations used for excess mortality in high risk state**

The company implemented excess mortality in the economic model by adding the values reported in Table 4.6 to the cycle-adjusted mortality risk for the general population. It is unclear to the EAG whether this is an appropriate interpretation of the methodology used in NICE TA443 (which is where excess mortality parameters are sourced from), as excess mortality is additive to (independent of) the general population mortality risk.[6]

Table 4.6: Calculations used for excess mortality by the company

Health state	Excess mortality	Source
High-risk PBC	1.20%	Expert opinion
DCC	4.20%	TA443, 2017[6]
HCC	10.20%	TA443, 2017[6]
Pre-LT	2.20%	TA443, 2017[6]
LT	18.90%	TA443, 2017[6]
Post-LT	1.50%	TA443, 2017[6]
Re-emergence of PBC	2.20%	TA443, 2017[6]

Source: CS Document B, Table 40, p. 124[1]
Abbreviations: CS = company submission; DCC = decompensated cirrhosis; HCC = hepatocellular carcinoma; LT = liver transplant; PBC = primary biliary cholangitis; TA = technology appraisal

However, further comparisons with the published literature suggest that liver disease survival, as predicted using the model parameters, is potentially lower than what would be expected in clinical practice both for HCC[45,46] and DCC.[47]

One of the reasons that excess mortality for high-risk of liver disease has a large impact on the results is because the company opted for an additive approach increasing the per-cycle mortality risk by 1.2% (e.g. 2% in general population + 1.2% excess). The EAG opted for an approach that reinterprets the 1.2% excess mortality as a percentage of the age-specific general population mortality probability (0.02*1.012 rather than 0.02+0.012) because of uncertainty in excess mortality in the high-risk group, and to increase the survival predictions of the model. The survival predictions using this approach are reported in Table 4.5.

- **The presence of the pre-liver transplant as an absorbing state**

The EAG acknowledges that the model structure was based on a previous submission accepted by NICE following the clinical pathway of PBC patients.[6] In this regard, one of the criticisms raised by the EAG assessing TA443 was the use of a pre-liver transplantation health state that a patient must enter before subsequently making the transition to the liver transplant state or dead state.[43]

As the current submission followed a similar model structure. The EAG thinks there is structural uncertainty associated with the inclusion of this pre-LT state. The cost of the pre-LT state (£5297) is significantly higher than the high-risk biomarker state (£2081), DCC state (£4161), and the HCC state (£3053), while the utility is equal to the DCC state (0.38) and lower than the HCC state (0.45) and the high-risk biomarker state (0.55). Patients may transition to the pre-liver disease state from each of these states, and the three-month probability of a liver

transplant from the pre-LT state was 0.1, meaning that patients can remain in the state for a long time. It is unclear whether the higher average cost and lower average utility is representative for a long period of time.

The EAG reiterates the concern from TA443 that patients in different biomarker risk and liver disease stages of liver disease are captured in the same health state and can stay there for the duration of their lifetime, sharing the same HRQoL, costs, and time to liver transplant probability.[43] The company did not discuss these implications in the company submission.

The EAG base-case excluded the pre-LT state from the model structure, allowing patients to transition directly to liver-transplantation from their liver disease state following the EAG critique for TA443 (see Section 6.1.1).[43]

4.3.3.2 UDCA transitions

- **Uncertainty around biomarker risk categories after treatment discontinuation**

The company state that patients who discontinue elafibranor were assumed to return to their biomarker risk level state at baseline (CS Section B.3.3.2.1, p. 166).[1] Within the model file, patients in the PBC biomarker stage who discontinue second-line treatment (elafibranor or OCA) are distributed across the biomarker risk categories based on the biomarker risk distribution from ELATIVE at baseline.[1] The EAG is concerned that this approach does not account for changes in the risk distribution within the cohort over time, and whether the biomarker risk distribution from the ELATIVE baseline population was representative of the risk distribution after discontinuation of second-line treatment. The EAG has therefore tested a scenario (provided as an option in the model submitted by the company), where patients remain in their risk stage after discontinuation, which is a stronger and less likely assumption where moving to third-line treatment does not immediately make the cohort's risk distribution more severe (see scenario 6, Section 6.1.2). This alternative implicitly assumes that the initial benefit if treatment is maintained.

- **Uncertainty around moderate to mild risk transitions after 12 months**

The change in PBC risk status for patients receiving third-line UDCA after 12 months is assumed to only deteriorate (progress to higher risk) without the possibility of improving, which implies changes in PBC biomarkers are permanent and only deteriorate after this stage. The probability of risk progression is taken from the placebo arm of the ELATIVE trial,[4] if the probabilities from months 9 to 12 in the trial stay the same over the patient lifetime.

[redacted]

The EAG considers this is a strong assumption based on highly uncertain evidence, especially since allowing for risk improvements still means an overall risk progression from mild to

moderate, only at a lower rate. Therefore, the EAG explored a scenario relaxing this assumption by using a scenario set up in the CS, allowing for temporary improvements or progression across risk of progression stages after 12 months (see Section 6.1.2).

- **Uncertainty around the use of 9 to 12 month probabilities from the placebo arm to predict long-term transitions for UDCA**

For patients treated with UDCA only, the transition probabilities from mild risk to medium risk, and from medium risk to high risk, are assumed to be the same over the patient lifetime from month 12 (with no chance of risk improvement, see above). These probabilities were assumed to be the same as the probabilities of the placebo arm of ELATIVE between months 9 and 12.

The company did not provide any clinical justification for this assumption, neither from published evidence nor expert opinion. Therefore, the EAG is concerned that the model might be making inefficient use of the dataset by ignoring the first nine months of data. Patients in the placebo arm had an inadequate response or no response to UDCA at entry to the trial.

The EAG has tested different scenario analyses with different approaches to obtain long-term data using the 12-month data (see scenario 12 in Section 6.1.2).

4.3.3.3 Model predictions

- **Uncertainty around the lack of validation in the model survival predictions**

[redacted]

Overall, the EAG is concerned the model is potentially under-predicting liver disease-free survival. There are many assumptions that the company makes that could potentially contribute to this, including: the transitions from moderate risk to liver disease; the increase in mortality for high-risk patients; the immediate deterioration of biomarker risk stage after discontinuation; the assumption that UDCA patients cannot transition from moderate to low risk; and the uncertainty around long-term transitions. The latter three assumptions also increase the weight of treatment discontinuation assumptions relative to cholestasis response in being the determinant factor of outcomes in the economic model.

Moreover, the EAG was concerned that the overall survival predictions were not further validated by clinical experts or the use of published literature. Communications between the company and clinical experts requested by the EAG provided insights into how some of the parameters of survival for the HCC and DCC states might not be fully reflective of advances in clinical care. This was partially corroborated by the EAG when comparing both survival predictions (median HCC survival of 1.5 years, and median DCC survival of four years) with the literature (e.g. HCC overall survival estimates after five years varied between 43% and 69%).[45-47] The EAG conducted scenario analyses adapting some of the scenarios proposed in the company submission to assess the structural uncertainty from assumption around treatment discontinuation.

4.3.4 Treatment effectiveness, adverse effects and outcome probabilities

Table 4.4: Summary of EAG's critique on the design of the economic modelTable 4.7 summarises the EAG's critique on the treatment effectiveness, adverse effects and outcome probabilities within the economic model.

Table 4.7: Summary of EAG's critique on the design of the economic model

Analysis feature	Section in CS where methods are reported	EAG's assessment
Treatment effectiveness and extrapolation	Document B.3.3.2, p. 117	**Some concerns** The EAG is concerned that the ALP thresholds in the definition of cholestasis response used for the NMA does not match the ALP thresholds used for the effectiveness in the economic model, especially since the NMA has a stricter definition it is unclear whether ALP threshold differences are clinically meaningful. The implementation of NMA data for cholestasis response, pruritus occurrence as a TEAE, and all-cause discontinuation relied on deriving a 12-month RR and assuming a constant RR across different time periods was considered suboptimal. The EAG preferred to assume a constant hazard ratio (HR) across time periods. There is a potential for using elafibranor and OCA in sequence, as suggested in Sections 2.1.1 and 4.3.2.1. However, although the clinical expert consulted by the EAG suggested this is feasible, there is a lack of effectiveness data at third line. Suboptimal modelling of treatment effectiveness in the economic model. Inappropriate parametric distribution for the OR. The assumption of constant RRs across widely varying baseline risks and different time periods. See Section 4.3.4.1 for further details.
Time-to-event analysis and extrapolation methods	Document B.3.3.4, p.177	**Key issue 3** The EAG has concerns that the approach used by the company to model treatment discontinuation overpredicts the proportion of OCA patients stopping treatment. The EAG prefers the conservative assumption that the difference in treatment discontinuation rates has a 1-year duration. This is of particular concern as differences in discontinuation are the primary factor driving the incremental QALY gain of elafibranor in the economic model. See Section 4.3.4.2 for further details
Conceptualisation of pruritus in the economic model	B.3.3.6, p. 124	**Some concerns** The EAG is concerned about the conceptualisation of pruritus outcomes in the economic model. The

Analysis feature	Section in CS where methods are reported	EAG's assessment
		company used outcomes from the PBC-40 to calculate clinically significant pruritus, which is a questionnaire that is not typically used in clinical practice to assess pruritus. Moreover, the economic model included pruritus identified as a Grade ≥ 2 AE, and pruritus identified from the PBC-40 questionnaire. It is not clear to the EAG whether the definition of both pruritus outcomes was mutually exclusive or how the company accounted for potentially double-counting the impact of pruritus. See Section 4.3.4.3 for further details.
Source: CS Document B[1] Abbreviations: CS = company submission; EAG = Evidence Assessment Group; NMA = network meta-analysis; OCA= obeticholic acid; OR = odds ratio; PBC = primary biliary cholangitis; TEAE = treatment-emergent adverse events		

4.3.4.1 Treatment effectiveness and extrapolations

- **Definitions of effectiveness between the NMA and the economic model**

The primary outcome in the ELATIVE and POISE trials used for the NMA was cholestasis response, defined as ALP ≤ 1.67x ULN, with a reduction of ≥ 15% from baseline, and total bilirubin ≤ ULN.[1,4,5] In the economic model, relative treatment effectiveness was presented as an increase in the transition probabilities from moderate risk or high risk to mild risk biomarker states, where the risk categories were stratified as follows.

- Mild risk: ALP ≤ 200 u/L and TB ≤ 20 µmol/L
- Moderate risk: ALP > 200 u/L and TB ≤ 20 µmol/L
- High risk: TB > 20 µmol/L or compensated cirrhosis (defined as kPa > 15)

It is unclear to the EAG what the impact on cost-effectiveness of using different ALP thresholds to define mild and moderate risk in the economic model compared to the cholestasis response definition used in ELATIVE and POISE is.[1,4,5]

In the points for clarification letter, the EAG asked the company to provide the cholestasis treatment response thresholds used.[7] In the response, the company specified that the 1.6x ULN ALP threshold for women was 174 U/L for and for men was 215 U/L, while the ULN for TB was defined as 20.5 µmol/l for all the population.[7]

The mild to moderate risk threshold in the economic model (defined by ALP levels), diverges from the more conservative cholestasis definition used in the ELATIVE trial. Although the 200 U/L in the economic model may align more with clinical practice in the UK,[7] the EAG considers that applying a less conservative threshold on ELATIVE data could translate into a larger proportion of moderate-risk patients returning to mild-risk. However, it is not clear to the EAG whether a narrower ALP threshold (e.g. ALP > 174 U/L) would make a significant difference on patients transitioning between moderate and mild risk states.

The EAG acknowledges that using the cholestasis response definition from ELATIVE is appropriate to have consistent definitions of response in estimating the odds ratio of response for OCA versus elafibranor. However, there is an inconsistency in the response definitions used to estimate the OR and to estimate the baseline risks.

After consultation with an expert clinician, the EAG is concerned that the risk categories presented for the economic model are not typically used in clinical practice. Although ALP and TB are strongly related to progression to liver disease, response is usually seen as a dichotomous variable while the risk is usually assessed using the UK-PBC risk score or the GLOBE scoring system, which includes more variables related to progression.

The EAG asked the company to assess structural uncertainty through assessing alternative definitions for the mild, moderate and high-risk health states.[7] The company provided scenario analyses using different definitions of treatment response showing that stricter definitions of uncertainty lead to reduced incremental QALY estimates, these analyses were replicated for the EAG base-case in Section 6.1.2.

The EAG was not able to produce a scenario analysis changing the ALP thresholds between the moderate and mild risk health states in the economic model. Therefore, the EAG considers the company base-case to be a potentially favourable scenario towards the intervention, as stricter treatment response criteria tended to reduce the incremental benefit of elafibranor.[7] However, the EAG considers it unlikely that narrower thresholds will change the results.

- **Sampling distributions of OR for the probabilistic analysis**

The EAG requested the company to review and correct the formula for the lognormal distribution sampling the OR from the NMA for: cholestasis response; all-cause discontinuation; and occurrence of pruritus as a TEAE. The lognormal distribution was incorrectly specified in the CS Excel model for the OR of response and the OR of likelihood of pruritus as a TEAE, and no parametric distribution was specified for the OR of treatment discontinuation. In the response to the EAG's request, the company updated the economic model by changing the sampling distribution of the OR parameters mentioned from the lognormal to the gamma distribution.[7]

For the OR of cholestasis response, the mean, median, lower limit and upper limit of the 95% CrI for the gamma distribution (α = [redacted], β = [redacted]) used in the company submission and the lognormal distribution (μ=[redacted][ln(median)], SD = [redacted]) are presented in Table 4.8 along with the NMA results presented in the CS.[1] The mean OR for cholestasis response was not reported in the CS and the EAG could not replicate the Bayesian random-effects NMA results due to difficulties in convergence. As such, the comparison of the mean estimate from a lognormal distribution to the mean estimate from the NMA for the fixed effect analysis is also presented in Table 4.8. After the factual accuracy check, the company provided the mean NMA OR for OCA cholestasis response, Table 4.8 has been updated to reflect this.

It appears that the gamma distribution was specified so that the gamma mean value was the NMA median value. Given the discrepancy between median values from the NMA and the median values from the gamma distribution, the EAG considers it an error to opt for the gamma distribution over the lognormal distribution, which provides a much better fit to the NMA results, even though the mean is slightly closer to 1 than the likely NMA mean estimate. This was

addressed by sampling OR parameters using the lognormal distribution in the EAG's base-case analysis cholestasis response, all-cause discontinuation, and occurrence of pruritus as a TEAE (see Section 6.1.1).

Moreover, the median OR NMA values (e.g. [redacted] for response) were used in the deterministic analysis in the CS. The EAG considers mean values should be used where possible in deterministic analysis. As the mean values from the NMA for the OR of cholestasis response, all-cause discontinuation, and pruritus occurrence as a TEAE were not reported in the CS (and the EAG could not replicate the Bayesian random-effects NMA), the EAG initially used the mean values associated with the lognormal distribution (e.g. [redacted] for response). After the factual accuracy check, the company provided the mean estimates from the NMA for: OCA odds of cholestasis response, OCA odds of pruritus occurrence as a TEAE, and OCA odds of all-cause discontinuation, therefore, the inputs of the EAG's base-case deterministic analysis have been updated to include the mean estimates from the company's NMA. The mean and CI results from the NMA and the mean and CI results used in the EAG analyses for each outcome are presented in Table 4.9.

Table 4.8: The fit of gamma and lognormal distributions for OCA OR cholestasis response

	Random effects			Fixed effect	
Statistic	**NMA results**	**Gamma distribution**	**Lognormal distribution**	**NMA results**	**Lognormal distribution**
Mean	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Median	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
CI lower limit	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
CI upper limit	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Source: CS economic model, EAG output
Abbreviations: CI = confidence interval; CS = company submission; EAG = Evidence Assessment Group; NMA = network meta-analysis; OCA = obeticholic acid
*From the EAG running the fixed effect NMA using the company code

Table 4.9: Median and 95% CI for the OR parameters from the NMA (random effects) versus lognormal parameters used by the EAG in the economic model

	NMA - random effects results			Lognormal distribution values for the EAG analyses		
OCA OR	**Mean**	**LL (95% CI)**	**UL (95% CI)**	**Expected value**	**LL (95% CI)**	**UL (95% CI)**
Cholestasis response	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Pruritus occurrence as an AE	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
All-cause discontinuation	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Source: CS economic model, EAG output
Abbreviations: AE= adverse event; CI = confidence interval; CS = company submission; EAG = Evidence Assessment Group; LL = lower limit; NMA = network meta-analysis; OCA =obeticholic acid; OR = odds ratio; UL = upper limit

- **How OR was used to determine OCA probabilities of response, pruritus and discontinuation**

The OR of response and discontinuation for OCA versus elafibranor were estimated using evidence after 12 months' follow-up. The model cycle length was three months. Transition probabilities to the low-risk biomarker state in the first four cycles were based on trial data where non-zero probabilities ranged from [redacted]. The elafibranor probability of response at 12 months was [redacted]. There was considerable uncertainty in the OR estimates.

Either a constant RR, OR, or hazard ratio (HR) can be assumed across time periods and baseline risks. While each of these may vary empirically with baseline risk due to various factors, the OR and HR are mathematically independent of the baseline risk. The mathematical dependence of RR on baseline risk makes certain RR estimates less plausible than associated HRs and ORs at different baseline risks.

The company chose to assume a constant RR. There was no apparent method in the model to ensure that the OCA probability of transitioning to mild risk would always be ≤ 1 when sampling from a lognormal or Gamma distribution for the OR parameter; and values > 1 would have occurred for moderate to mild risk in the first cycle when running the PSA.

There is no risk of deriving impossible probabilities when assuming a constant OR, or HR. For OR, this requires first deriving a RR from the OR using the transition probability in the model (see equation below[49]) before multiplying the same transition probability by that RR. For HR, this requires calculating the HR at 12 months and multiplying this with the elafibranor rate at three months derived from transition probability at three months.

$$RR_3 = \left(\frac{OR_{12}}{1 - E\_R_3 + E\_R_3 \times OR_{12}}\right)$$

Considering the different time periods (3 and 12 months) and repeated time periods, the EAG chose to assume a constant HR.

- **Effectiveness of elfibranor used in sequence, either as second-line followed by OCA or as third-line after OCA**

Elafibranor and OCA could have been included as third-line treatments in the evaluation of cost-effectiveness of elafibranor at second-line, but there was no effectiveness evidence at third-line. Treatment following discontinuation of second-line treatment is potentially not representative of what may occur in clinical practice given the use of these drugs as second-line.

4.3.4.2 Time-to-event analysis and extrapolation

- **All-cause discontinuation predictions for OCA**

The company's model took a different approach to including all-cause discontinuation compared with TA443.[6] As patients undergo treatment for the duration of their life, treatment discontinuation was also included over the long term. The company used time-to-event data for all-cause discontinuation of treatment from the elafibranor arm of ELATIVE,[4] applying a

12-month RR (derived from the OR estimate in the NMA) across each cycle to predict all-cause discontinuation for OCA.

The model selection to predict long-term discontinuation for elafibranor was chosen on the grounds of statistical fit to trial data. The company reported that clinical opinion recommended "*the flattest curve compared to other distributions*" (CS Section B.3.3.4, p. 122).[1] However, the Gompertz curve (having a flat tail) was deemed as having an unrealistically high retention rate over the long term and was therefore excluded from the base-case analysis.[1] An exponential distribution was selected in the CS as it was a good fit to the data. The exponential distribution assumes that the discontinuation rate is constant over time.

[redacted]. Furthermore, the EAG's clinical expert criticised the assumption that the risk of discontinuation is constant over time, as many patients who need to discontinue OCA due to pruritus or drug-induced liver damage will do so early in their treatment. It was considered that this is also likely to be the case with elafibranor. As such, this means that difference in discontinuation rates between elafibranor and OCA may change over time.

After the points for clarification, the company updated their base-case scenario to use the lognormal distribution function to model the long-term risk of all-cause discontinuation.[7] Previous communications between the company and clinical experts provided in the response to the EAG's points for clarification letter suggest that all-cause discontinuation occurs primarily early-on during treatment and, in the case of OCA, pruritus is a major cause for discontinuation (although other fibrates may present renal toxicity issues leading to discontinuation).[7] After the first couple of years, discontinuation occurs due to disease progression or lack of efficacy. For the long-term predictions, one of the clinicians suggested somewhere between the lognormal and the Gompertz models would be appropriate.

[redacted] This indicates to the EAG that the company base-case assumptions in the economic model are potentially overpredicting the proportion of patients who discontinue OCA. Figure 4.2 compares the discontinuation predictions for OCA from the company's base-case, the company's base-case after corrections, and the EAG's proposed analysis.

Figure 4.2: OCA all-cause treatment discontinuation predictions CS and EAG base-case

*CS base-case using the lognormal mean OR for discontinuation

Abbreviations: CS = company submission; EAG = evidence assessment group; OR = odds ratio

The EAG explored the impact on cost-effectiveness of alternative assumptions of all-cause discontinuation, particularly assessing the Gompertz function where treatment retention predictions align more closely to expert opinion, and the assumption that the treatment effect on discontinuation only lasts for the first year to represent patients who discontinue early on.

A summary of the EAG's view on the company's choice of each parametric survival model is summarised in Table 4.10 below.

Table 4.10: Comparison of the company and EAG's preferred choices of extrapolations

Survival measure	CS Section	Company choice of extrapolation	EAG's preferred choice of extrapolation
All-cause discontinuation: OCA versus elafibranor	B.3.3.4, p.121	Initial submission: Exponential with lifelong treatment effect After PfC response: Lognormal with lifelong treatment effect	Lognormal with 1-year treatment effect (Gompertz with 1-year treatment difference in the scenario analysis)

Source: EAG output

Abbreviations: CS = company submission; EAG = Evidence Assessment Group; OCA = obeticholic acid; PfC = points for clarification

4.3.4.3 Conceptualisation of pruritus in the economic model

- **Inclusion of pruritus as a TEAE and pruritus measured by the PBC-40**

The economic model includes ELATIVE data on Grade ≥ 2 TEAE occurring in ≥ 5% of participants during the trial, which includes pruritus for both the elafibranor and placebo arms.[4] The company used the OR from the NMA on pruritus recurrence as a TEAE to calculate the proportion of pruritus as a TEAE for OCA.

Independently from the analysis above, the company includes the impact of pruritus on costs and quality of life using the PBC-40, in particular for patients with PBC-40 scoring ≥ 7, classified as clinically significant pruritus.[50] The company uses NMA results for median PBC-40 score differences from baseline to 12 months to generate the proportion of OCA patients expected to have a clinically significant pruritus (PBC-40 score ≥ 7).

It is not clear to the EAG whether the definitions of pruritus as a Grade ≥ 2 TEAE and pruritus captured by the PBC-40 are mutually exclusive and, if not, what measures the company took to avoid double-counting the impact of pruritus. The EAG has applied a conservative assumption to the EAG base-case analysis where all pruritus AE differences between treatments are being captured by the PBC-40 scores, see section 6.1.1.

- **Use of the PBC-40 to generate proportions of patients with clinically significant pruritus**

After consultation with a clinical expert, the EAG is concerned that the use of the PBC-40 questionnaire to calculate the proportion of patients with clinically significant pruritus may not be an accurate approach, as clinicians in the NHS use different methods to assess whether a patient requires treatment for pruritus, including rating pruritus on 1-10 scales or using the 5D-Itch questionnaire. Uncertainty surrounding clinically significant pruritus calculated was compounded by clinical advice suggesting the proportion of OCA-treated patients presenting mild pruritus was too high. Therefore, the EAG ran a sensitivity analysis around the threshold value of clinically significant pruritus from the PBC-40 and its impact on cost-effectiveness, see scenario 3 in section 6.1.2.

4.3.5 Health-related quality of life

Table 4.11 summarises the EAG's critique on HRQoL within the economic model.

Table 4.11: Summary of EAG's critique on HRQoL

Analysis feature	Section in CS where methods are reported	EAG's assessment
HRQoL evidence used for Markov states	B.3.4.1; B.3.4.2, p.125-129. B.3.4.6.1, p.133-134;	**Some concerns** The EAG have concerns around the applicability of quality of life values for the pre-liver transplant state (capturing patients with multiple liver disease stages), across the sources gathered by the company, particularly with respect to Rice et al, 2021.[51] **Key issue 4** Utility values were elicited using the EQ-5D questionnaire in the ELATIVE trial, while the company

Analysis feature	Section in CS where methods are reported	EAG's assessment
		base-case analysis uses values from the published literature. The EAG is concerned that trial utility values were only included in scenario analysis for the mild-risk and moderate-risk patients, especially when there seems to be a large discrepancy between trial values and the parameters used for utility at the high-risk state. See section 4.3.5.1 for further comment.
Disutility for adverse effects	B.3.4.4, p.133.	**Some concerns** The EAG is uncertain whether the pruritus utility decrements adequately represent the difference in pruritus included in the model as a TEAE and as an adverse event. See section 4.3.4.3 for further comment.
Abbreviations: AE = adverse events; CEM = cost-effectiveness model; CS = company submission; EAG = Evidence Assessment Group; EQ-5D = European Quality of Life-5 Dimensions; HRQoL = health-related quality of life; LT = liver transplant; PBC = primary biliary cholangitis; SLR = systematic literature review; TEAE = treatment-emergent adverse event		

4.3.5.1 HRQoL evidence used in the cost-effectiveness model:

The utility values for each PBC biomarker risk state calculated using trial data are presented in Table 4.12. The utility values included in the CS economic model are reported in Table 4.13.

Table 4.12: Utility values for the biomarker states elicited from the ELATIVE trial

Health states	Utility values
Low-risk state	[redacted]
Moderate-risk state	[redacted]
High-risk state	[redacted]
Source: CS Document B, Table 45, p.128[1] Abbreviations: CS = company submission	

Table 4.13: Health state utility values in the CS economic model

	Utility value	Reference	Justification
Mild	0.84 (0.17)	Table 61, p. 157	Cholestatic disease utility reported in Younossi et al (2000)[52]
Moderate	0.84 (0.17)	Table 61, p. 157	Cholestatic disease utility reported in Younossi et al (2000)[52]
High	0.55 (0.11)	Table 61, p. 157	Previously reported value for compensated cirrhosis[2]
DCC	0.38 (0.08)	Table 83, p. 194	Previously reported value for DCC;[2] redacted utility decrement not applied
HCC	0.45 (0.09)	Table 61, p. 157	Previously reported value for HCC[2]
Pre-LT	0.38 (0.08)	Table 83, p. 194	Previously reported value for

	Utility value	Reference	Justification
			pre-LT;[2] redacted utility decrement not applied
LT	0.57 (0.11)	Table 83, p. 194	Previously reported value for LT;[2] redacted utility decrement not applied
Post-LT	0.67 (0.13)	Table 83, p. 194	Previously reported value for post-LT[2]
Re-emergence of PBC	0.67 (0.13)	Table 61, p. 157	Assumed equivalent to post-LT, without utility decrement provided according to KOL feedback[2]

Source: CS Section B.3.4.6.1, Table 49, p.134[1]
Abbreviations: AR = adverse reaction; CS = company submission; DCC = decompensated cirrhosis; HCC = hepatocellular carcinoma; HS = health state; KOL = key opinion leader; LT = liver transplant; PBC = primary biliary cholangitis; SE = standard error; TA = technology appraisal

- **Applicability of utility values**

There is a question around the applicability of the utility values in Table 4.13. There is significant variation in utility values in the literature referenced by the company. This CS draws upon utility estimates for similar states in hepatitis C patients. In a systematic review (6 studies, N=162) of chronic hepatitis C studies of utility values by Saeed et al (2020),[53] which was not reported in the CS, statistical heterogeneity was high (I^2 = 88%). The mean estimate 0.595 (SE = 0.062,) was higher than that reported in the CS (0.38).

- **Utility values for the high-risk health state in the trial versus the economic model**

HRQoL evidence was collected through the EQ-5D-5L in the ELATIVE trial, but utility values obtained from the literature (consistent with NICE TA443 for OCA) were used in the economic model due to concerns of small sample sizes for patients in the high-risk (biomarker) health state.[1,6] EQ-5D-5L scores from the trial were mapped to the EQ-5D-3L version using the mapping algorithm from Hernandez-Alava et al (2020).[54] A linear mixed effects model was then used to calculate the utility values of each PBC biomarker risk health state from the economic model.[1] The utility values for each PBC biomarker risk state calculated using trial data are presented in Table 4.13.

The company claimed that: "*the incremental difference in utility between the moderate and high risk health states is lower than expected from the regression analysis,*" which might be "*driven by the low sample size in the high risk state*" (CS Section B.3.4.2, p.128).[1] The company decided to use utility values obtained from the literature based on the above judgement, and conducted a scenario analysis using the utility values for the mild and moderate biomarker risk state using utility values from ELATIVE.[4] The EAG considers that the company could have explored a scenario using all the patient-reported data collected from the trial for all the health states covered in the model, including the high-risk of liver disease state.

The EAG acknowledges that the size of the high-risk sample was small but considers the data collected to still be informative for the analysis. The analysis relied on a sample of 78 observations at high-risk, which was only 10% of the overall sample of observations and was therefore considered unreliable.[7] Adding to this uncertainty, the company mentioned the

possibility that trial recruitment tends to favour representation from patients with better health than what would be expected for a particular health condition (PfC B16a).[7]

The utility values included in the CS economic model are reported in Table 4.13. The EAG notes that the use of utility values derived from Younossi et al (2000) for the low and moderate risk states, based on a hepatitis C population, was criticised by the EAG appraising NICE TA443.[6,52] The utility value of 0.55 for the high risk state (compensated cirrhosis) was for a chronic hepatitis C population.[55]

The EAG notes that the utility for the high-risk state used in the economic model (0.55) [redacted]. The EAG requested the company to comment on this difference in the points for clarification letter, to which the company reiterated that the decrement in utility scores observed between the moderate and the high-risk health states was lower than they expected and, therefore, was considered unreliable (PfC B16a).[7]

The EAG agrees with the company that disease symptoms are likely to be the drivers of quality of life in PBC and, particularly at the early stages, PBC displays relatively stable symptoms. The EQ-5D score analysis presented by the company in their response to the EAG's points for clarification reflects this through the small, non-statistically significant differences across biomarker health states (PfC B15b).[7]

Because the EAG considers patient-reported utility data to have value,[56] the EAG ran a scenario analysis using utilities elicited from the ELATIVE trial for all the biomarker health states.[4] Moreover, the EAG base-case adopted an alternative published EQ-5D-3L utility value (mean = 0.717, SE = 0.021, I^2 = 62%, 8 studies, N = 414) for the high risk state from a more recent source,[53] based on compensated cirrhosis for a chronic hepatitis C population (see Section 6.1.1). The value used for the high-risk state in the CS comes from Wright et al,[55] which was included in the meta-analysis in Saeed et al,[53] and which was the lowest value in the meta-analysis.

- **Utility value used for the pre-liver transplant health state**

Patients can move to the pre-LT state from the high-risk biomarker state, the DCC state and the HCC state with similar transition probabilities (1.02%, 1.53%, 1.02%) and there are significantly more patients in the high-risk state over time than in the DCC and HCC states. Utility values in these states were 0.55, 0.38, 0.45, respectively. As explained in Section 4.3.3.1, patients may stay in the pre-LT for a long time. It is not clear to the EAG that a utility of 0.38 is representative of the utility in this state. There is also the aforementioned issue of uncertainty in the applicability of several of the utility values to the PBC population. The EAG has removed the pre-LT state in its base-case analysis (see Sections 4.3.3.1 and 6.1.1).

4.3.5.2 Disutility values for adverse events

- **Disutility values for pruritus**

Pruritus was treated as a TEAE and a symptom of interest in the company's submission. The associated disutility values are presented in Table 4.14. It was not clear to the EAG how the company separated pruritus caused by TEAEs (and thus the disutility from pruritus) from pruritis as a symptom of PBC, so it is unclear to the EAG whether the differences in the pruritis

disutility values adequately represent the differences in pruritus as a TEAE and as an adverse event.

Table 4.14 Disutility values for pruritus

	Utility value	Justification
As an adverse event	0.11	Clinical expert opinion[42]
As a long-term disutility applied in the model	[redacted]	ELATIVE trial[4]
Source: CS Section B.3.4.6.2, Table 50, p.135[1] Abbreviations: CS = company submission		

4.3.6 Resources and costs

Table 4.15 summarises the EAG's critique on resources and costs within the economic model.

Table 4.15 Summary of EAG's critique on resources and costs

Analysis feature	Section in CS where methods are reported	EAG's assessment
Adverse event costs	B.3.5.3, p.142	**Appropriate** The EAG noticed there was divergence from NICE TA443[6] regarding the resources used to treat pruritus. The EAG checked with the company and an external clinical expert to ensure this reflected changes in current practice from the time of the previous submission. See Section 4.3.6.1 for further comment.
Treatment acquisition costs	B.3.5.1, p.136	**Some concerns** The company did not sufficiently justify the compliance rate used in treatment acquisition calculations for elafibranor and OCA. In response to the points of clarification, the company provided multiple estimates based on different calculation methods. See Sections 4.3.6.2 and 4.3.6.3 for further comment.
Health states costs	B.3.5.2, Table 56, P.138-41	**Some concerns** The company did not clearly reference the evidence regarding the NHS costs. See Section 4.3.6.3 for further comment.
End-of-life costs (terminal care costs)	B.3.5.5, p.143-44	**Some concerns** The company included end of life costs in the economic model for DCC and HCC, which the EAG believes may be a potential issue of double counting. See Section 4.3.6.4 for further comment.
Abbreviations: CS = company submission; EAG = Evidence Assessment Group; DCC = decompensated cirrhosis; HCC = hepatocellular carcinoma; NHS = National Health Service; OCA = obeticholic acid; PBC = primary biliary cholangitis; TA = technology appraisal; UDCA = ursodeoxycholic acid		

4.3.6.1 Adverse event costs

The EAG noted a difference in the proportions of patients having OCA and receiving medicines for pruritus in the company's current submission compared with NICE TA443;[6] these proportions are presented in Table 4.16. For example, 30% of patients treated with OCA or UDCA received colestyramine for pruritus in the current submission, yet this figure was 85% in the previous submission.[6] Colestyramine is the drug of choice for treating cholestatic pruritus recommended by NICE (see the BNF recommendation for colestyramine),[57] yet its prevalence and alternative options are not clearly stated. Clinical advice to the EAG suggested that the proportions presented for OCA/UDCA are appropriate. Moreover, after a request for comment in the EAG's points for clarification, the company confirmed that clinicians validated the resource use of managing pruritus and confirmed it reflects current practice.

Table 4.16: Percentage of patients who receive medicines for pruritus in the current and previous company submission

Drug	Percentage of patients cost applies to for patients treated with OCA or UDCA (current submission) [a]	Percentage of patients cost applies to for patients treated with OCA or UDCA (previous submission) [b]
Colestyramine	30%	85%
Rifampicin	30%	15%
Bezafibrate	20%	N/A
Gabapentin	15%	N/A
Naltrexone	5%	5%

Source: (a) CS Section B.3.5.4, Table 58, p.142;[1] (b) TA443, Section 5.5.4[6]
Abbreviations: CS = company submission; OCA = obeticholic acid; UDCA = ursodeoxycholic acid

4.3.6.2 Compliance rates

The compliance rate used in treatment acquisition calculations for elafibranor and OCA is [redacted], which was sourced from the ELATIVE study. The EAG believes that trials tend to overestimate compliance rates among patients and therefore asked the company to provide further clarification regarding the rationale for using this compliance rate.[7] In response, the company provided further estimates of treatment compliance for elafibranor ([redacted] depending on the method used to estimate compliance) and for OCA (93.55%). The company updated its base-case to have a [redacted] compliance rate for elafibranor and a 93.55% rate for OCA. The EAG base-case adopted a conservative scenario where compliance rates are 93.55% for both treatments, see section 6.1.1.

4.3.6.3 Health state costs

The costs for the health states are presented in Table 4.17. The EAG note that the costs for the HCC state are lower than the one for the DCC in the economic model. This may be because not everyone with HCC has DCC, and so some people with HCC do not incur the cost of treating complications associated with DCC. The company later corroborated this using expert views and published studies that this cost difference may occur due to less symptom severity for HCC patients relative to DCC patients.[44,51,55]

The EAG noticed that the NHS reference costs used in the economic model were not properly referenced (i.e. the service code and name of clinical procedure were not provided). As such, it is not possible for the EAG to check the consistency and appropriateness of the type of NHS costs used in the model. Hence, there is a lack of transparency in the evidence presented by the company submission.

Table 4.17: List of health states and associated costs in the economic model

Health state	Cost per cycle (GBP)	Source
Mild	106.67	National tariffs NHS England 2021/22; NICE TA443[6,58]
Moderate	154.72	National tariffs NHS England 2021/22; NICE TA443[6,58]
High	2080.52	National tariffs NHS England 2021/22; NICE TA443[6,58]
DCC	4161.05	National tariffs NHS England 2021/22; Wright et al (2006)[55,58]
HCC	3053.32	National tariffs NHS England 2021/22; Wright et al (2006)[55,58]
Pre-LT	5296.66	HST17[59]
LT	163,638.57	HST17; BNF records on azathioprine, tacrolimus and prednisolone[59-62]
Post-LT	919.57	BNF records on azathioprine, tacrolimus and prednisolone; Rice et al (2020)[51,60-62]
Re-emergence of PBC	2080.52	Assumption

Source: CS Section B.3.5.2, Table 56, p.138-41[1]
Abbreviations: BNF = British National Formulary; CS = company submission; DCC = decompensated cirrhosis; GBP = pounds sterling; HCC = hepatocellular carcinoma; HST = highly specialised technology; LT = liver transplant; NHS = National Health Service; PBC = primary biliary cholangitis; TA= technology appraisal

4.3.6.4 End of life costs

End of life costs were included in the economic model for DCC and HCC in the current submission as presented in Table 4.18. The EAG is concerned that there is potentially a double counting issue if no distinctive estimation was made between these end-of-life costs and the aforementioned health state costs for the DCC and HCC states especially since the cost values used cover a 12-month period, which is longer than the 3-month model cycle.[7] The EAG explored an alternative scenario testing the impact of removing end-of-life costs from the analysis, see scenario 5, section 6.1.2.

Table 4.18: End of life costs considered in the model

	End of life cost	Source
DCC	£10,902	Gola et al (2015)[63]
HCC	£8805	NICE TA666[64]

Source: CS Section B.3.5.5, Table 63, p.144 [1]
Abbreviations: CS = company submission; DCC = decompensated cirrhosis; HCC = hepatocellular carcinoma; TA = technology appraisal

5 COST EFFECTIVENESS RESULTS

5.1 Company's cost effectiveness results

The company's base-case deterministic cost-effectiveness results using the PAS discount for elafibranor are shown in Table 5.1 and Table 5.2. Table 5.1 shows the deterministic analysis for the combined PBC population of second-line patients that have inadequate response or that cannot tolerate UDCA. The analysis compares elafibranor, OCA and UDCA alone for this population and shows elafibranor dominating OCA by increasing health outcomes by [redacted] QALYs and decreasing costs by [redacted] per patient; and being more costly and more effective than UDCA alone (ICER = [redacted]). The incremental net monetary benefit of elafibranor versus OCA for a willingness to pay threshold of £30,000 was [redacted] (see Table 5.2).

Table 5.1: Company base-case deterministic results for elafibranor versus OCA and elafibranor versus UDCA, using the PAS price of elafibranor

Technology	Total costs (£)	Total LYs	Total QALYs	Incremental costs (£)	Incremental LYs	Incremental QALY	ICER (£)
Elafibranor	[redacted]	[redacted]	[redacted]	-	-	-	-
OCA	£242,656	12.67	8.27	[redacted]	[redacted]	[redacted]	Dominating
UDCA	£104,283	10.81	6.38	[redacted]	[redacted]	[redacted]	[redacted]

Source: CS Document B, Section 3.9.1

Abbreviations: CS = company submission; ICER = incremental cost-effectiveness ratio; LYs = life years gained; OCA = obeticholic acid; PAS = Patient Access Scheme; QALY = quality-adjusted life year; UDCA = ursodeoxycholic acid

Table 5.2: Company base-case deterministic results for net monetary benefit

Technology	Incremental costs (£)	Incremental QALY	ICER (£)	NMB at £20,000	NMB at £30,000
OCA	[redacted]	[redacted]	Dominating	[redacted]	[redacted]

Source: EAG outputs

Abbreviations: EAG = Evidence Assessment Group; ICER = incremental cost-effectiveness ratio; NMB = net monetary benefit; OCA = obeticholic acid

5.2 Company's sensitivity analyses

To explore uncertainty within their cost-effectiveness analysis, the company conducted a probabilistic sensitivity analysis over 10,000 iterations using the PAS price for elafibranor. After updating the model following the points for clarification process, the company reported the following probabilistic sensitivity analysis (PSA) results showing elafibranor as the dominant intervention over OCA, increasing QALYs by [redacted] and decreasing costs by [redacted]. Table 5.3 and Figure 5.1: ICEP for elafibranor versus OCA and elafibranor versus UDCA (10,000 iterations), using the PAS price of elafibranor show the probabilistic results reported by the company after the response to the EAG's points for clarification.

The EAG considers that the parametric distributions used to model uncertainty in the mean estimate were inappropriate (see section 4.3.4.1). These were corrected as errors in Section 6. The EAG also considers that arbitrary uncertainty has been introduced in the model by specifying a gamma distribution for the cost of OCA using variance based on an arbitrary 20%

of the cost of OCA. The EAG has not corrected for this, but is should not introduce much uncertainty relative to the uncertainty elsewhere in the model.

Table 5.3: PSA results for elafibranor versus OCA and elafibranor versus UDCA, using the PAS price of elafibranor

Technology	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALY	ICER (£)
Elafibranor	[redacted]	[redacted]	-	-	-
OCA	£243,132	7.997	[redacted]	[redacted]	Dominating
UDCA	£102,898	6.499	[redacted]	[redacted]	[redacted]

Source: CS Document B, Section B.3.10.1[1]
Abbreviations: CS = company submission; ICER = incremental cost-effectiveness ratio; OCA = obeticholic acid; PAS = Patient Access Scheme; PSA = probabilistic sensitivity analysis; QALY = quality-adjusted life year; UDCA = ursodeoxycholic acid

Figure 5.1: ICEP for elafibranor versus OCA and elafibranor versus UDCA (10,000 iterations), using the PAS price of elafibranor

Source: PfCs[7]
Abbreviations: GBP = pounds sterling; OCA = obeticholic acid; PfC = points for clarification; PSA = probabilistic sensitivity analysis; QALY = quality-adjusted life-year; UDCA = ursodeoxycholic acid

The EAG re-ran the PSA analysis in the same model file and obtained markedly different results, with marked differences in the CrIs as well. In the EAG's run, elafibranor was dominant over OCA, presenting an increment in QALYs of [redacted] and a change in costs of [redacted] Results obtained by the EAG are reported in Table 5.4 and Figure 5.2.

Table 5.4 PSA results for elafibranor versus OCA and elafibranor versus UDCA, using the PAS price of elafibranor

Technology	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALY	ICER (£)
Elafibranor	[redacted]	[redacted]	-	-	-
OCA	£286,862	8.80	[redacted]	[redacted]	Dominating
UDCA	£103,017	6.50	[redacted]	[redacted]	[redacted]

Source: CS Document B, Section B.3.10.1[7]
Abbreviations: CS = company submission; ICER = incremental cost-effectiveness ratio; OCA = obeticholic acid; PAS = Patient Access Scheme; PSA = probabilistic sensitivity analysis; QALY = quality-adjusted life year; UDCA = ursodeoxycholic acid

Figure 5.2: EAG re-run of ICEP for elafibranor versus OCA (10,000 iterations), using the PAS price of elafibranor

Source: CS model, EAG analysis
Abbreviations: EAG = Evidence Assessment Group; GBP = pounds sterling; ICEP = incremental cost-effectiveness plane; OCA = obeticholic acid; PAS = Patient Access Scheme; PSA = probabilistic sensitivity analysis; QALY =quality-adjusted life year

The base-case one-way sensitivity analysis (OWSA) presented by the company excluded the deterministic analysis of the OR parameters for cholestasis response, occurrence of pruritus, and all-cause discontinuation, as well as treatment compliance for elafibranor and OCA. The EAG considered these parameters to be an informative part of the analysis. Therefore, these were included and reported subsequently in Table 5.5 and Figure 5.4.

Table 5.5: OWSA results for elafibranor versus OCA (top 10 most sensitive parameters only)

Parameter name	Lower bound NMB (£)	Upper bound NMB (£)	Difference (£)
OCA odds ratio of all-cause discontinuation	[redacted]	[redacted]	[redacted]
OCA cost per cycle (10 mg cycle 3+) (GBP)	[redacted]	[redacted]	[redacted]
OCA compliance	[redacted]	[redacted]	[redacted]
Elafibranor compliance	[redacted]	[redacted]	[redacted]
OCA cost per cycle (5 mg up to cycle 2) (GBP)	[redacted]	[redacted]	[redacted]
Health state cost – High	[redacted]	[redacted]	[redacted]
OCA odds of cholestasis response	[redacted]	[redacted]	[redacted]
Health state cost – LT	[redacted]	[redacted]	[redacted]
Elafibranor clinically significant itch at month 12+	[redacted]	[redacted]	[redacted]
Mean difference in PBC-40 Itch relative to elafibranor (versus OCA 5-10 mg) at month 12	[redacted]	[redacted]	[redacted]

Source: CS Document B, Section B.10.2[1]
Abbreviations: CS = company submission; GBP = pounds sterling; LT = liver transplant; NMB = net monetary benefit; OCA = obeticholic acid; OWSA = one-way sensitivity analysis

Figure 5.3: OWSA results for elafibranor versus OCA in net monetary benefit (top 10 most sensitive parameters only)

Source: CS economic model
Abbreviations: CS = company submission; GBP = pounds sterling; LT = liver transplant; NMB = net monetary benefit; OCA = obeticholic acid; OWSA = one-way sensitivity analysis; PBC = primary biliary cholangitis

The OWSA suggests the OR parameter of discontinuation is the largest determinant of cost-effectiveness, as well as an important source of uncertainty. Other parameters included the cost of OCA, the impact of compliance differences on drug costs, the health state cost for high-risk of liver disease patients, and the OR of cholestasis response.

From the scenario analyses conducted by the company, changing the price of OCA had the potential to make it a less costly and less effective alternative to elafibranor. Scenario results of particular relevance to the EAG are reported in Table 5.6, full results are reported in the company response to the EAG's points for clarification (Appendix C).[7] Changing the assumption around discontinuation had the largest impact on relative efficacy, although elafibranor remained dominant over OCA. Excluding AEs had little impact on cost-effectiveness, while the use of more strict treatment effectiveness definition decreased the incremental effectiveness of elafibranor but this remained positive.

Table 5.6: Deterministic scenario analysis results for the company base-case submitted after the points for clarification (selection of scenarios considered relevant to the EAG analysis)

#	Model aspect	Base-case	Scenario analysis	Incremental costs OCA (£)	Incremental QALYs OCA	ICER versus OCA (£)
	Company base-case	**N/A**	**N/A**	[redacted]	[redacted]	**Dominating**
1	Time horizon	Lifetime	20 years	[redacted]	[redacted]	Dominating
4	OCA price per pack discount	0%	10%	[redacted]	[redacted]	Dominating
5			20%	[redacted]	[redacted]	Dominating
6			30%	[redacted]	[redacted]	[redacted]
7			40%	[redacted]	[redacted]	[redacted]
8			50%	[redacted]	[redacted]	[redacted]
9			[redacted]	[redacted]	[redacted]	[redacted]
11	AEs	Include	Exclude	[redacted]	[redacted]	Dominating
13	Definition of treatment response	Cholestasis response	ALP normalisation	[redacted]	[redacted]	Dominating
14			Reduction in ALP of ≥ 40%	[redacted]	[redacted]	Dominating
15			PARIS-II	[redacted]	[redacted]	Dominating
16	UDCA extrapolations	Improvements not possible	Improvements possible	[redacted]	[redacted]	Dominating
17	UDCA transition matrix extrapolation	Last observation carried forwards	Average of all transition matrices	[redacted]	[redacted]	Dominating
18	Moderate risk to liver disease transitions	Include	Exclude	[redacted]	[redacted]	Dominating
19	Duration of treatment effect of elafibranor relative to OCA on discontinuation	Lifetime	1 year	[redacted]	[redacted]	Dominating
20		Lognormal	Exponential	[redacted]	[redacted]	Dominating
21			Weibull	[redacted]	[redacted]	Dominating

#	Model aspect	Base-case	Scenario analysis	Incremental costs OCA (£)	Incremental QALYs OCA	ICER versus OCA (£)
22	Treatment discontinuation distribution		Log-logistic	[redacted]	[redacted]	Dominating
23			Gompertz	[redacted]	[redacted]	Dominating
25	Mild and moderate risk biomarker health states utilities	Younossi et al 2000 [52]	ELATIVE	[redacted]	[redacted]	Dominating
27	Compliance	Drug exposure ([redacted] versus 93.55%)	Mean cumulative ([redacted] versus 93.55%)	[redacted]	[redacted]	Dominating
28	Discontinuation	Return to baseline	Stay in state	[redacted]	[redacted]	Dominating

Source: response to the EAG's points for clarification, appendix c[7]
Abbreviations: AE = adverse event; ALP = alkaline phosphatase; ICER; = incremental cost-effectiveness ratio; N/A = not applicable; OCA = obeticholic acid; QALY = quality adjusted life years; UDCA= ursodeoxycholic acid

5.3 Model validation and face validity check

5.3.1 Face validity assessment and technical verification

The EAG has found multiple errors in the excel file calculating the model, these are listed in Section 6.1.1.

5.3.2 Comparison with external data

No expert opinion was elicited to validate overall survival predictions from the model (see Section 4.3.3.3). UK-PBC data and expert opinion were used to validate OCA discontinuation data.

6 EVIDENCE ASSESSMENT GROUP'S ADDITIONAL ANALYSES

6.1 Exploratory and sensitivity analyses undertaken by the EAG

Based on the considerations in the preceding sections of this EAG report, the EAG defined an EAG base-case. This EAG base-case included several adjustments to the company base-case presented in Section 5. These adjustments have been subdivided into three categories (derived from Kaltenthaler 2016).[65]

- Fixing errors (correcting the model where the company's submitted model was unequivocally wrong)
- Fixing violations (correcting the model where the EAG considered that the NICE reference case, scope or best practice had not been adhered to)
- Matters of judgement (amending the model where the EAG considers that reasonable alternative assumptions are preferred)

6.1.1 EAG base-case

Adjustments made by the EAG to derive the EAG base-case (using the CS base-case as starting point) are listed below.

Fixing errors

1. High-risk to DCC risk parameter, Excel file: the company stated that the source of the annual transition probability between high risk and DCC of 1% was NICE TA443.[6] The value in NICE TA443 was actually 10%. "Data Store!" Sheet, Cell E127, changed value 1% for 10%.

2. Distribution sampling the OR parameters (see section 4.3.4.1), Excel file: The EAG changed the distribution sampling the OR parameters for cholestasis response, all-cause discontinuation, and pruritus recurrences as a TEAE for the probabilistic analysis to the lognormal distribution. "Model Parameters!" Sheet, Cell F33 update formula to "=(LN(M33)-LN(L33))/3.92", Cell J33 update formula to "=IFERROR(EXP(NORM.INV(RAND(),LN('Clinical Inputs'!D20),F33)),E33)". Cell F34 update formula to "=(LN(M34)-LN(L34))/3.92", Cell J34 update formula to "=IFERROR(EXP(NORM.INV(RAND(),LN('Clinical Inputs'!D127),F34)),E34)". Cell F79 update formula to "=(LN(M79)-LN(L79))/3.92", Cell J79 update formula to "=IFERROR(EXP(NORM.INV(RAND(),LN('Clinical Inputs'!D143),F79)),E79)"

3. Mean and median OR parameters (see section 4.3.4.1), Excel file: the company's base-case analysis uses the median parameters from the NMA for the OR of OCA on cholestasis response, all-cause discontinuation, and pruritus recurrence as TEAE, as the mean value in the deterministic analysis. Initially the EAG used the mean values from the distributions sampling each OR parameter in the model, after the factual accuracy check, the company provided the mean estimates from the NMA results. "Model Parameters!" Sheet, Cells E33, E34, E79 values were replaced by the mean values in Table 4.9, respectively.

4. The discounting is implemented from cycle 0 rather than 1: The EAG reconstructed the Markov trace to make it easier for the EAG to remove the pre-LT state in a scenario analysis described below, and initial cycle for discounting became apparent. See "EAG elafibranor engine!" Sheet, and "EAG OCA engine!" Sheet.

5. Excel file: The upper and lower values for the OR of all-cause discontinuation are not consistent with the credible intervals reported in the submission:[1] Model parameters! Sheet, Cells L79 and M79, changed values to [redacted]

6. Excel file: The OR parameter for ALP normalisation is 0.07 in the economic model rather than [redacted] stated in the company's response to the EAG's points for clarification, question B2.c, page 31[7]: "Data Store!" sheet, cell P65, changed value 0.07 to [redacted].

7. Excel file and CS Document B (Section B.3.10.2):[1] The submitted model did not include the NMA parameters for OCA odds of cholestasis response, odds of pruritus occurrence, and odds of discontinuation in the OWSA: "Model Parameters!" sheet, column K, cells K33, K34, and K79, replaced value with "0".

8. Excel file: The model does not include OCA and elafibranor compliance rates in the OWSA: "Model Parameters!" sheet, column K, K86 and K89 replaced value with "0".

9. Excel file: Compliance parameters sampled from a normal distribution can go above 100% in the sensitivity analyses. The sampling distribution was changed to the Beta distribution: "Model Parameters!" sheet; changed formula in G86 to "=IFERROR(E86*((E86*(1-E86)/F86^2)-1),"")"; changed formula in H86 to "=IFERROR((1-E86)*((E86*(1-E86)/F86^2)-1),"")"; changed formula in J86 to "=IFERROR(BETA.INV(RAND(),G86,H86),E86)"; updated lower and upper distribution values. In row 89 changed formula in G89 to "=IFERROR(E89*((E89*(1-E89)/F89^2)-1),"")"; changed formula in H89 to "=IFERROR((1-E89)*((E89*(1-E89)/F89^2)-1),"")"; changed formula in J89 to "=IFERROR(BETA.INV(RAND(),G89,H89),E89)".

10. The model does not include UDCA cost parameters in the OWSA: "Model Parameters!" sheet, cell K90, replaced value with "0".

11. Although the analysis in the company submission tests multiple risk distributions for the baseline risk of elafibranor, the model does not include the uncertainty around the parameter inputs for the probabilistic distribution of the baseline risk of all-cause discontinuation of elafibranor. The EAG used the Gompertz, log-logistic, and exponential distributions in the scenario analysis to account for this.

12. The model cycle length was three months. The cycle treatment discontinuation probabilities were calculated from 84-day time periods rather than 91.25 day time periods from the parametric time to discontinuation curves. The EAG changed the cells in: Data Store! D624:D924 to '=C625*cycle_length_days' from '=C625*(12*7)'.

13. Upper interval level of the Odds of cholestasis response parameter: after the factual accuracy check the company corrected that the upper interval levels of the OR parameter for OCA cholestasis response had been erroneously reported in the company submission document B and the response to the points for clarification. The value in sheet "Model Parameters!" cell M33 was updated to [redacted].

Fixing violations

The EAG did not identify a clear violation of the NICE guidelines, or the scope agreed between NICE and the company. There are questions surrounding the potential omission of fibrates

from the scope, as well as the potential inclusion of elafibranor and OCA in sequence as an alternative strategy (see section 4.3.2.1).

Matters of judgement

1. *Assumption of constant 12-month hazard ratios*

The OR of response and discontinuation for OCA versus elafibranor were estimated using evidence after 12 months follow-up (see Section 4.3.4.1). The model cycle length was three months. Transition probabilities to the low-risk biomarker state in the first four cycles were based on trial data where non-zero probabilities ranged from [redacted]. The elafibranor probability of response at 12 months was [redacted]. There was considerable uncertainty in the OR estimates.

As explained in Section 4.3.4.1, the EAG preferred to assume a constant HR over different time periods and baseline risks rather than a constant RR that the company assumed.

2. The pre-liver transplant health state in the model

The economic model structure proposed by the company included a pre-LT state through which a patient must pass before transitioning to the LT state. The three-month probability of a LT from the pre-LT state was 0.1, and the patient can die before a transplant. Patients can move to the pre-LT state from: the PBC biomarker states of moderate risk and high risk; the DCC state; or the HCC state (see Section 4.3.3.1). There is structural uncertainty associated with the inclusion of this state.

The approach taken by the EAG to address this issue is borrowed from the appraisal of NICE TA443, where the then EAG decided towards eliminating the pre-liver transplant state to allow the cohort to transition from health states at risk of liver failure directly to the liver transplant state.[6,43] Moreover, although changing this assumption is expected to reduce the survival predictions of PBC patients in the short-term, the change in transplant-free survival was very small, Table 6.1 compares the transplant-free survival predictions for the company base-case model (after fixing for errors 1 to 12) between including the pre-LT state versus allowing direct transitions to LT.

Table 6.1 Transplant-free survival predictions for OCA in the company base-case model after errors 1-12

	Including pre-LT	Direct transitions to LT
Timepoints	**OCA**	**OCA**
1 year	[redacted]	[redacted]
5 years	[redacted]	[redacted]
10 years	[redacted]	[redacted]
20 years	[redacted]	[redacted]
40 years	[redacted]	[redacted]
Median (years)	[redacted]	[redacted]

Source: CS economic model, EAG output
Abbreviations: CS = company submission; EAG = Evidence Assessment Group; LT = liver transplant; OCA = obeticholic acid

3. Discontinuation assumptions

After considering the opinion of clinicians consulted by the EAG and by the company, the EAG considers that the primary driver of outcomes in the model is the difference in all-cause discontinuation between elafibranor and OCA. The results of the NMA suggest elafibranor offers an improvement in all-cause discontinuation over OCA and the model assumes this difference is maintained during the complete lifetime duration of treatment (see Section 4.3.4.2).

Pruritus is a primary factor driving differences in discontinuation rates between OCA and elafibranor; it is expected that patients are more likely to discontinue treatment early on if pruritus is the primary cause, based on the clinical opinion received by the company.[42] Therefore, the EAG base-case model assumes that the difference between elafibranor and OCA in discontinuation rates is only maintained over the first year. This is likely to be a conservative assumption, based on the 12-month data informing the NMA.

The EAG has explored the use of a Gompertz function to model long-term treatment discontinuation in the scenario analysis (see Sections 4.3.4.2 and 6.1.2).

4. High risk state excess mortality as a percentage of general population mortality

One of the reasons that excess mortality for high-risk of liver disease has a large impact on the results is because the company opted for an additive approach increasing the per-cycle mortality risk by 1.2% (e.g. 2% in general population + 1.2% excess). The company reported that a clinical expert stated that the excess mortality for the high-risk state could be between 0% and 4%. Considering the uncertainty around an excess mortality risk in the high-risk biomarker state and the lower than expected survival estimates in the model, the EAG opted for an approach that reinterprets the 1.2% excess mortality as a percentage of the age-specific general population mortality probability (0.02*1.012 rather than 0.02+0.012; see Section 4.3.3.1).

5. Alternative utility values for the high-risk state

The scenarios presented by the company only include ELATIVE trial data for the mild and moderate PBC biomarker states, as pointed out in Section 4.3.5.1. The EAG has included a scenario analysis using ELATIVE data for all biomarkers (see Section 6.1.2). Moreover, the EAG has adopted utility values for the high-risk state from a more up-to-date study in its base-case equivalent to 0.72,[53] which is in-between the trial data estimate and the utility value proposed by the company based on published evidence.

6. Pruritus as a symptom or as a TEAE

It remained unclear to the EAG how patients with pruritus as a Grade ≥ 2 AE were differentiated from pruritus patients identified using the PBC-40 scale. Therefore, the EAG base-case makes the conservative assumption that pruritus differences (both as a symptom and treatment expected AE - TEAE) are being captured by PBC-40 score differences. Hence, the frequency of pruritus as a TEAE was assumed to be equal between elafibranor and OCA. This was done to avoid double-counting in the calculation of the impact of pruritus on quality of life, as highlighted in Sections 4.3.4.3 and 4.3.5.2.

7. Compliance rates

Differences in compliance rates between elafibranor and OCA can vary according to the method used to calculate them (see the company's response to the points for clarification and Section 4.3.6).[7] As described in Section 4.3.6.2, since the administration methods of elafibranor and OCA are similar the EAG base-case makes the conservative assumption that both treatments follow the OCA compliance rate provided by the company of 93.6%.

6.1.2 EAG exploratory scenario analyses

The EAG performed the following scenario analyses to explore the impact of alternative assumptions conditional on the EAG base-case.

EAG scenarios

1. Reduce the excess mortality risk in the high-risk biomarker state by changing the excess mortality at high-risk from 1.2% to 0% (see Section 4.3.3.1).
2. Extrapolation of discontinuation: Assume no difference in discontinuation (set all-cause discontinuation OR for OCA versus elafibranor to 1; see Section 4.3.4.1).
3. Adverse events: Change the PBC-40 threshold for clinically significant pruritus from scores ≥ 7 to scores ≥ 8 (see Sections 4.3.4.3 and 4.3.5.2).
4. HRQoL: Use ELATIVE trial health-utility values for all biomarker states (see Section 4.3.5.1 and Table 4.12).
5. Resources and costs: Remove the palliative care costs from the HCC and DCC states (see Section 4.3.6.4).

Scenarios from the CS

6. Assume patients do not change biomarker risk after moving to third-line (see Section 4.3.3.2).
7. Change the discontinuation distribution to the Gompertz function from the lognormal function (see Section 4.3.4.1).
8. Change the discontinuation distribution to the log-logistic function from the lognormal function (see Section 4.3.4.1).
9. Change the discontinuation distribution to the exponential function from the lognormal function (see Section 4.3.4.1).
10. Remove the moderate risk transitions to liver disease (see Section 4.3.3.1).
11. Remove the restriction that UDCA patients cannot improve (see Section 4.3.3.2).
12. Use average biomarker risk transition probabilities for UDCA after 12 months rather than only 9-12 month probabilities and remove the restriction that UDCA patients cannot improve (see Section 4.3.3.2).
13. Other definitions of treatment response: ALP normalisation (see Section 4.3.4.1).
14. Other definitions of treatment response: 40% reduction in ALP (see Section 4.3.4.1).
15. Other definitions of treatment response: PARIS-II (see Section 4.3.4.1).

16. OCA unit price reduced by 20%

17. OCA unit price reduced by 50%

6.1.3 EAG subgroup analyses

No additional subgroup analyses were conducted by the EAG.

6.2 Impact on the ICER of additional clinical and economic analyses undertaken by the EAG

6.2.1 The EAG base-case, scenario and sensitivity analyses

The EAG base-case was described in Section 6.1.1. Table 6.2 reports the cost-effectiveness results of updating the company base-case model correcting for errors found by the EAG, and the individual impact of the matters of judgement by the EAG to generate the EAG base-case results.

Treatment discontinuation has the biggest impact on the cost-effectiveness results. Increasing treatment discontinuation rates decreases treatment cost, increases liver disease cost, and decreases total QALYs. Treatment costs dwarf liver disease costs. When fixing errors in the company model, the mean OR for discontinuation was used instead of the median value. That increased the discontinuation rate for OCA, significantly reducing the total cost of the OCA arm. OCA is not cost-effective compared to UDCA at a threshold of £30,000 with an ICER of £67,707/QALY. This indicates that the reduction in OCA cost is more significant than the reduction in OCA QALYs. The deterministic ICER for elafibranor increased from elafibranor dominating to an ICER of £1,528 after fixing errors.

The cost-effectiveness of elafibranor + UDCA versus OCA + UDCA with UDCA as third-line treatment is a combination of the cost-effectiveness of OCA + UDCA and the cost-effectiveness of UDCA. The ICER for elafibranor versus OCA is £1,528, while the ICER for elafibranor versus UDCA was £25,643. Hence, the greater percentage of patients in the OCA arm receiving UDCA only, the less cost-effective elafibranor is. In a full incremental analysis, OCA is dominated by extension, and would be eliminated from the analysis. However, given that OCA is recommended for use in the NHS, the pairwise results for elafibranor versus OCA are presented in this section.

Likewise, when the difference in discontinuation rates between OCA and elafibranor was assumed to only last for 1 year, more patients continued receiving OCA, and this increased the ICER for elafibranor from £1,528 to elafibranor dominating after making that assumption. OCA cost increases significantly.

Making the assumption of a constant hazard ratio has the next biggest impact, followed by a higher utility value for the high risk biomarker; while the other preferred assumptions in the EAG base case have little impact on cost-effectiveness outcomes.

Table 6.2: Deterministic and probabilistic EAG base-case

Technologies	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER (£/QALY)
CS base-case – deterministic					
Elafibranor	[redacted]	[redacted]	[redacted]	[redacted]	
OCA	[redacted]	[redacted]	[redacted]	[redacted]	Elafibranor Dominating
CS base-case – probabilistic					
Elafibranor	[redacted]	[redacted]	[redacted]	[redacted]	
OCA	[redacted]	[redacted]	[redacted]	[redacted]	Elafibranor Dominating
Fixing errors (1-12) – deterministic					

Technologies	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER (£/QALY)
Elafibranor	[redacted]	[redacted]	[redacted]	[redacted]	
OCA	[redacted]	[redacted]	[redacted]	[redacted]	£1,528
Fixing errors (1-12) – probabilistic					
Elafibranor	[redacted]	[redacted]	[redacted]	[redacted]	
OCA	[redacted]	[redacted]	[redacted]	[redacted]	Elafibranor Dominating
Constant hazard ratio for response and discontinuation					
Elafibranor	[redacted]	[redacted]	[redacted]	[redacted]	
OCA	[redacted]	[redacted]	[redacted]	[redacted]	£6391
Removing the pre-liver transplant state					
Elafibranor	[redacted]	[redacted]	[redacted]	[redacted]	
OCA	[redacted]	[redacted]	[redacted]	[redacted]	£1295
Reducing the difference in discontinuation rates to 1 year					
Elafibranor	[redacted]	[redacted]	[redacted]	[redacted]	
OCA	[redacted]	[redacted]	[redacted]	[redacted]	Elafibranor Dominating
Changing the formula of excess mortality at high-risk					
Elafibranor	[redacted]	[redacted]	[redacted]	[redacted]	
OCA	[redacted]	[redacted]	[redacted]	[redacted]	£1524
Alternative utility at high-risk					
Elafibranor	[redacted]	[redacted]	[redacted]	[redacted]	
OCA	[redacted]	[redacted]	[redacted]	[redacted]	£1657
Pruritus differences using PBC-40 scores only					
Elafibranor	[redacted]	[redacted]	[redacted]	[redacted]	
OCA	[redacted]	[redacted]	[redacted]	[redacted]	£1553
Equivalent compliance rates					
Elafibranor	[redacted]	[redacted]	[redacted]	[redacted]	
OCA	[redacted]	[redacted]	[redacted]	[redacted]	£748
EAG base-case (errors 1-12, matters of judgment 1-7) – deterministic*					
Elafibranor	[redacted]	[redacted]	[redacted]	[redacted]	
OCA	[redacted]	[redacted]	[redacted]	[redacted]	Elafibranor Dominating
EAG base-case (errors 1-12, matters of judgment 1-7) – probabilistic*					
Elafibranor	[redacted]	[redacted]	[redacted]	[redacted]	
OCA	[redacted]	[redacted]	[redacted]	[redacted]	Elafibranor Dominating

*EAG results updated after the FAC, see fixing errors 3 and 13
Abbreviations: CS = company submission; EAG = Evidence Assessment Group; ICER = incremental cost-effectiveness ratio; QALY = quality-adjusted life year

6.3 Overall conclusions of the EAG's cost-effectiveness analysis

The estimated probabilistic results from the EAG base-case suggest that elafibranor dominates OCA using the PAS price for elafibranor. Incremental QALYs for elafibranor versus OCA were [redacted] and incremental costs were [redacted]. The probabilistic EAG base-case analyses indicated cost effectiveness probabilities of [redacted]% and [redacted]% at willingness to pay thresholds of £20,000 and £30,000 per QALY gained, respectively.

The incremental cost-effectiveness plane showing the incremental costs and QALYs for elafibranor compared to OCA is presented in Figure 6.1. The cost-effectiveness acceptability curves for elafibranor and OCA is presented in Figure 6.2.

Figure 6.1 Incremental cost-effectiveness plane elafibranor versus OCA (EAG base-case)

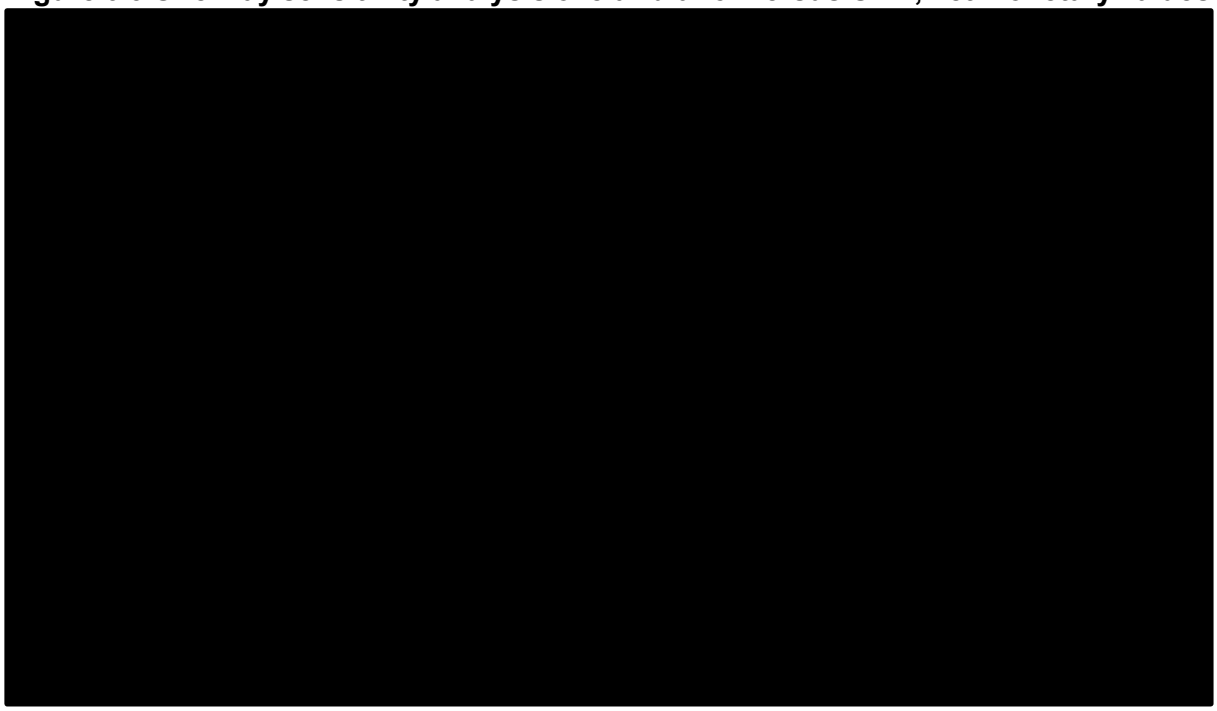
Source: CS model, EAG's base-case
Abbreviations: CS = company submission; EAG = Evidence Assessment Group; GBP = pounds sterling; OCA = obeticholic acid; PSA = probabilistic sensitivity analysis; QALY = quality-adjusted life year

Figure 6.2 Cost-effectiveness acceptability curve (CEACs) elafibranor versus OCA (EAG base-case)

Source: CS model, EAG's base-case
Abbreviations: CEAC = cost-effectiveness acceptability curve; CS = company submission; EAG = Evidence Assessment Group; GBP = pounds sterling; OCA = obeticholic acid

The most influential parameters in the deterministic OWSA were: the unit cost of OCA; the odds ratio of all-cause discontinuation; differences in compliance rates; and differences in clinically significant pruritus. Treatment response had a minor impact relative to these parameters in the cost-effectiveness results. Results using net monetary values are illustrated in Figure 6.3 and reported in Table 6.3.

Figure 6.3 One-way sensitivity analysis of elafibranor versus OCA, net monetary values



Source: CS model, EAG's base-case
Abbreviations: CS = company submission; EAG = Evidence Assessment Group; GBP = pounds sterling; LT = liver transplant; NMB = net monetary benefit; OCA = obeticholic acid

Table 6.3 One-way sensitivity analysis of elafibranor versus OCA, net monetary values

Parameter name	Lower bound NMB (£)	Upper bound NMB (£)	Difference (£)
OCA cost per cycle (10 mg cycle 3+) (GBP)	[redacted]	[redacted]	[redacted]
OCA odds ratio of all-cause discontinuation	[redacted]	[redacted]	[redacted]
OCA compliance	[redacted]	[redacted]	[redacted]
Elafibranor compliance	[redacted]	[redacted]	[redacted]
OCA cost per cycle (5 mg up to cycle 2) (GBP)	[redacted]	[redacted]	[redacted]
Elafibranor clinically significant itch at Month 12+	[redacted]	[redacted]	[redacted]
Mean difference in PBC-40 Itch relative to elafibranor (vs OCA 5-10 mg) at Month 12	[redacted]	[redacted]	[redacted]
Health state cost – LT	[redacted]	[redacted]	[redacted]

Parameter name	Lower bound NMB (£)	Upper bound NMB (£)	Difference (£)
OCA odds of cholestasis response	[redacted]	[redacted]	[redacted]
Health state cost – High	[redacted]	[redacted]	[redacted]

Source: CS Document B, Section B.10.2[1]
Abbreviations: CS = company submission; GBP = pounds sterling; LT = liver transplant; NMB = net monetary benefit; OCA = obeticholic acid

The results of the EAG scenario analyses are reported in Table 6.4. The scenarios with the largest impact on the cost-effectiveness of elafibranor versus OCA assessed by the EAG were: assuming no treatment difference in discontinuation (more cost-savings but less incremental QALYs), changing the risk function of treatment discontinuation (higher risks led to lower cost-savings and lower incremental QALYs), changing the assumptions around third-line treatment with UDCA (less severe disease progression meant less cost-savings and fewer incremental QALYs), and using more strict definitions of treatment effectiveness (leading to less cost-savings and fewer incremental QALYs).

Table 6.4: EAG scenario analysis

Scenario #	EAG base-case input	Alternative input	Incremental costs (£)	Incremental QALYs	ICER (£/QALY)
	EAG base-case	N/A	[redacted]	[redacted]	Elafibranor Dominating
1	1.2% excess mortality at high-risk	No excess mortality at high-risk	[redacted]	[redacted]	Elafibranor Dominating
2	Treatment difference on discontinuation for 1 year	No treatment difference on discontinuation	[redacted]	[redacted]	Elafibranor Dominating
3	Clinically significant itch if PBC-40 ≥ 7	Clinically significant itch if PBC-40 ≥ 8	[redacted]	[redacted]	Elafibranor Dominating
4	Literature values for PBC biomarker state utilities	Trial values for PBC biomarker state utilities	[redacted]	[redacted]	Elafibranor Dominating
5	Palliate care costs for HCC and DCC	Removing palliative care costs for HCC and DCC	[redacted]	[redacted]	Elafibranor Dominating
6	Risk distribution after discontinuation based on ELATIVE baseline	Risk distribution after discontinuation does not change	[redacted]	[redacted]	Elafibranor Dominating
7	All-cause discontinuation	All-cause discontinuation	[redacted]	[redacted]	Elafibranor Dominating

Scenario #	EAG base-case input	Alternative input	Incremental costs (£)	Incremental QALYs	ICER (£/QALY)
	risk function: lognormal	risk function: Gompertz			
8		All-cause discontinuation risk function: Log-logistic	[redacted]	[redacted]	Elafibranor Dominating
9		All-cause discontinuation risk function: Exponential	[redacted]	[redacted]	Elafibranor Dominating
10	Risk of progression from moderate risk to liver disease	No risk of progression from moderate risk to liver disease	[redacted]	[redacted]	Elafibranor Dominating
11	UDCA treated patients cannot improve their risk category after year 1	UDCA treated patients can improve their risk category	[redacted]	[redacted]	Elafibranor Dominating
12	UDCA probabilities after one year follow the probabilities seen in months 9-12	UDCA probabilities after one year follow the average probabilities of the first 12 months including probabilities to improve PBC risk	[redacted]	[redacted]	Elafibranor Dominating
13	Treatment effectiveness definition: Cholestasis response	Treatment effectiveness definition: ALP normalisation	[redacted]	[redacted]	Elafibranor Dominating
14		Treatment effectiveness definition: Barcelona criteria	[redacted]	[redacted]	Elafibranor Dominating
15		Treatment effectiveness definition: PARIS-II	[redacted]	[redacted]	Elafibranor Dominating
16	List price for OCA 5-10 mg	20% price reduction for OCA 5-10 mg	[redacted]	[redacted]	Elafibranor Dominating

Scenario #	EAG base-case input	Alternative input	Incremental costs (£)	Incremental QALYs	ICER (£/QALY)
17		50% price reduction for OCA 5-10 mg	[redacted]	[redacted]	[redacted]

Source: EAG outputs
Abbreviations: ALP = alkaline phosphate; DCC = decompensated cirrhosis; EAG = Evidence Assessment Group; HCC = hepatocellular carcinoma; ICER = incremental cost-effectiveness ratio; N/A = not applicable; OCA = obeticholic acid; PBC = primary biliary cholangitis; QALY = quality-adjusted life year; UDCA = ursodeoxycholic acid

6.4 Overall conclusions of the EAG's critique

The EAG consider that the CS met the NICE scope to an appropriate degree. The EAG had comments regarding the positioning of elafibranor and OCA as the third-line treatment after discontinuation of elafibranor or OCA as second-line treatment. It is plausible for OCA to follow discontinuation of elafibranor in sequence and vice versa, due to the different mechanisms of action in each. The assumption that UDCA is the only possible third-line treatment may not accurately reflect the clinical pathway in either the OCA or elafibranor arms of the decision model. However, treatment effectiveness at third-line is uncertain due to a lack of evidence.

The company conducted an SLR to identify evidence surrounding the effectiveness and safety of elafibranor and relevant comparators for treating PBC. The EAG have some concerns surrounding multiple aspects of the SLR methodology, including: the literature search; eligibility criteria; screening; data extraction; and quality appraisal. The main clinical evidence was based on the ELATIVE trial. In general, the EAG believes that the ELATIVE trial was well-conducted and relevant to the decision problem but the method of allocation concealment was not reported and there was a lack of subgrouping by participants intolerant to and those non-responsive to UDCA. However, the EAG appreciate that the numbers of participants who were intolerant to UDCA in the ELATIVE trial was low and is reflective of clinical practice.

To compare the relative efficacy of elafibranor with OCA, the company performed a series of NMAs. Although the company provided additional information and a rationale for many of the EAG's queries regarding the NMA methodology during the points for clarification process, the EAG still have concerns about the substantial width of the 95% CrIs, including when compared against OCA 5-10 mg. Additionally, it was noted by the company that there was difficulty in achieving convergence within the model. The EAG performed multiple additional NMA analyses for outcomes used within the economic model; the results of these analyses did not change the overall conclusions. The results of the EAG analyses were still open to substantial uncertainty and it is therefore difficult to draw any conclusions regarding the clinical effectiveness of elafibranor versus OCA 5-10 mg.

The company conducted SLRs with searches aimed at identifying cost-effectiveness studies, HRQoL and cost and resource use data to inform the economic model. The search strategy used was considered fit for purpose but the use of focused MESH headings may have increased the specificity of the search to the detriment of specificity. Moreover, conference proceedings were excluded from Embase searches, which may have missed relevant studies.

Regarding the economic model, the posterior distributions of the ORs estimated in the NMA were skewed with considerable variance and the company inadequately specified the parametric distributions for the ORs in their base-case. Median values were used in the CS, which the EAG replaced these with mean values. The company assumed a constant RR, while the EAG preferred to assume a constant HR.

The lack of external validation, whether from clinical experts or from the published literature, of the survival predictions in the model for OCA or elafibranor (liver-disease free, LT-free, OS, etc.) was noted as a key issue by the EAG due to concerns that the model was underpredicting liver disease-free survival for elafibranor compared to the predictions from UK-PBC scores and GLOBE scores from ELATIVE.[48] The EAG thinks this may partly be a consequence of strong assumptions in the model structure including: the risks of progression from moderate risk to liver disease; the excess mortality at high risk parameter; the assumption that biomarker

risk categories continue to deteriorate in third-line after elafibranor or OCA; and the assumption that biomarker risk cannot improve in third-line, accelerating its deterioration. The use of elafibranor trial data as the baseline in the economic model, with current practice (OCA) response and discontinuation derived by multiplying baseline risks with the effectiveness statistics, makes the development of a model with plausible predictions harder. The EAG was also concerned about whether the mortality parameters for liver disease were reflective of advances in clinical practice.

Another key issue the EAG raised regarding the economic analysis is the uncertainty around treatment discontinuation, particularly since the difference in treatment discontinuation rates between elafibranor and OCA is the primary driver of cost-effectiveness estimates. Consultation with a clinical expert and additional data from UK-PBC provided by the company suggested that treatment discontinuation predictions for OCA in the model may be too high.[66] The economic model assumes that the difference in treatment discontinuation rates between OCA and elafibranor are maintained over a lifelong treatment duration. However, the patterns of discontinuation can shift after the first year or two with OCA, as patients appear to discontinue at a higher rate early on (in part due to the effect of OCA on pruritus). Furthermore, uncertainty surrounding the risk of treatment discontinuation over the long term for OCA and elafibranor is a key cause of uncertainty in the cost-effectiveness results. The EAG has suggested limiting the difference in treatment discontinuation rates between OCA and elafibranor to one year, which leads to better predictions for OCA. Nevertheless, uncertainty in treatment discontinuation rates continued to have a significant impact on outcomes.

The next key issue highlighted by the EAG was the use of utility values from the published literature for the PBC biomarker risk states in the economic model, rather than using the patient-reported values elicited from the ELATIVE trial.[1] The most impactful quality of life parameter was utility at the PBC high-risk of liver disease biomarker state, where the utility values selected for the base-case were noticeably lower than the moderate-risk health state, and lower than the value elicited for this population from the ELATIVE trial data.[1] The EAG explored an alternative utility value for the high-risk state from the published literature in between the trial value and the company's base-case value informed by NICE TA443.[6,53]

On the subject of how the economic model calculated quality of life, the EAG was concerned about the applicability of utilities from NICE TA443, since they include a confidential decrement based on expert opinion; the implications of this assumption were not discussed in the CS.[1,6] Furthermore, the model included different disutility values from different sources for pruritus: as a TEAE; and as a symptom of PBC. It was not clear how each definition of pruritus was mutually exclusive, or how any potential overlap was accounted for.

The company considered that this condition did not meet the severity modifier criteria.

The approach taken to calculate costs and resource use was considered fit for purpose. The EAG only raised concerns on two issues. Firstly, the EAG were concerned with transparency in the use of NHS tariffs from a previous NICE submission,[6] as the current submission lacked clarity around the specific cost codes being used. The second issue surrounded the differences in treatment compliance rates between elafibranor and OCA, which is an area of uncertainty feeding directly into the total cost differences, as different approaches to calculating compliance rates led to different estimates.[7]

The EAG base-case assumed a constant HR for cholestasis response and discontinuation, removed the pre-LT state, reduced the duration of the difference in discontinuation risk rates between OCA and elafibranor to one year, updated the high-risk utility value, changed the approach to high-risk mortality, and assumed PBC-40 differences in pruritus also capture treatment-emergent exacerbations.

After updating for errors found by the EAG, the company base-case suggested that, after applying the PAS discount to the unit cost of elafibranor, elafibranor was the dominant strategy over OCA by increasing QALYs by [redacted] and decreasing costs by [redacted] with credible intervals showing substantial uncertainty around the cost-effectiveness estimates. After applying the PAS discount to the unit cost of elafibranor, the EAG base-case also suggested that elafibranor was the dominant intervention over OCA by increasing QALYs by [redacted] and decreasing costs by [redacted] with a [redacted]% probability of being cost-effective at a £20,000 willingness to pay threshold.

The cost of OCA, the difference in treatment discontinuation rates, the assumption of a constant HR, treatment compliance differences, and differences in pruritus were found by the EAG to be the parameters with the largest impact on the cost-effectiveness results. Further structural assumptions were tested using scenario analyses proposed by the EAG and recreating scenarios from the CS. Assuming no difference in treatment discontinuation rates, changing the parametric time-to-discontinuation model, and changing the treatment effect definitions had the largest impact on cost-effectiveness estimates. Nonetheless, elafibranor remained dominant over OCA across most of the scenarios after the PSA discount for elafibranor was applied. Although the dominance of elafibranor over OCA remained robust after the analyses proposed by the EAG, large uncertainties from the NMA results were translated into large uncertainties in the incremental costs and benefits of elafibranor. Moreover, the model structure strongly emphasises the impact of differences on treatment discontinuation over differences in treatment effectiveness[redacted]The EAG would be interested to see how alternative treatment strategies, such as the use of elafibranor and OCA in sequence, could affect the treatment landscape for this cohort of PBC patients and how this could be further explored.

7 REFERENCES

1 Ipsen. *Elafibranor for treating primary biliary cholangitis [ID6331]. Single technology appraisal. Company evidence submission: Document B*. Paris, France: Ipsen; 2024.

2 NICE. *Sofosbuvir for treating chronic hepatitis C [TA330]. Technology appraisal guidance.* London: National Institute for Health and Care Excellence (NICE); 2015. Available from: https://www.nice.org.uk/guidance/ta330.

3 Invernizzi P, Floreani A, Carbone M, Marzioni M, Craxi A, Muratori L, et al. Primary Biliary Cholangitis: advances in management and treatment of the disease. *Digestive and Liver Disease*. 2017;49(8):841-6. Available from: https://doi.org/https://doi.org/10.1016/j.dld.2017.05.001.

4 Kowdley K, V., Bowlus C, L., Levy C, Akarca U, S., Alvares-da-Silva M, Reis., Andreone P, et al. Efficacy and Safety of Elafibranor in Primary Biliary Cholangitis. *New England Journal of Medicine*. 2024;390(9):795-805. Available from: https://doi.org/10.1056/NEJMoa2306185.

5 Nevens F, Andreone P, Mazzella G, Strasser SI, Bowlus C, Invernizzi P, et al. A placebo-controlled trial of obeticholic acid in primary biliary cholangitis. *N Engl J Med*. 2016;375(7):631-43. Available from: https://doi.org/10.1056/NEJMoa1509840.

6 NICE. *Obeticholic acid for treating primary biliary cholangitis [TA443]. Technology appraisal guidance*. London: National Institute for Health and Care Excellence (NICE); 2017. Available from: https://www.nice.org.uk/guidance/ta443 [Date accessed: 18/04/2024].

7 Ipsen. *Elafibranor for treating primary biliary cholangitis [ID6331]. Single technology appraisal. Company evidence submission: Points for Clarification Response*. Paris, France: Ipsen; 2024.

8 Abbas N, Smith R, Flack S, Bains V, Aspinall RJ, Jones RL, et al. Critical shortfalls in the management of PBC: Results of a UK-wide, population-based evaluation of care delivery. *JHEP Reports*. 2024;6(1):100931. Available from: https://doi.org/https://doi.org/10.1016/j.jhepr.2023.100931.

9 Ipsen. *Elafibranor for treating primary biliary cholangitis [ID6331]. Single technology appraisal. Company evidence submission: Appendix D*. Paris, France: Ipsen; 2024.

10 Ipsen. *Elafibranor for treating primary biliary cholangitis [ID6331]. Single technology appraisal. Company evidence submission: Appendix G*. Paris, France: Ipsen; 2024.

11 McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS peer review of electronic search strategies: 2015 guideline statement. *Journal of Clinical Epidemiology*. 2016;75:40-6. Available from: https://doi.org/10.1016/j.jclinepi.2016.01.021.

12 Scherer RW, Meerpohl JJ, Pfeifer N, Schmucker C, Schwarzer G, von Elm E. Full publication of results initially presented in abstracts. *Cochrane Database of Systematic Reviews*. 2018;(11). Available from: https://doi.org/10.1002/14651858.MR000005.pub4.

13 Hopewell S, Loudon K, Clarke MJ, Oxman AD, Dickersin K. Publication bias in clinical trials due to statistical significance or direction of trial results. *Cochrane Database of Systematic Reviews*. 2009;(1). Available from: https://doi.org/10.1002/14651858.MR000006.pub3.

14 Whiting P, Savović J, Higgins JP, Caldwell DM, Reeves BC, Shea B, et al. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. *J Clin Epidemiol*. 2016;69:225-34. Available from: https://doi.org/10.1016/j.jclinepi.2015.06.005.

15 Drucker AM, Fleming P, Chan A-W. Research Techniques Made Simple: Assessing Risk of Bias in Systematic Reviews. *Journal of Investigative Dermatology*.

2016;136(11):e109-e14. Available from: https://doi.org/https://doi.org/10.1016/j.jid.2016.08.021.

16 Thomas J, Kneale D, McKenzie JE, Brennan SE, Bhaumik S. Chapter 2: Determining the scope of the review and the questions it will address. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, eds. *Cochrane Handbook for Systematic Reviews of Interventions version 6.4 (updated August 2023). Cochrane*. 2023.

17 Hirschfield GM, Dyson JK, Alexander GJM, Chapman MH, Collier J, Hubscher S, et al. The British Society of Gastroenterology/UK-PBC primary biliary cholangitis treatment and management guidelines. *Gut*. 2018;67(9):1568-94. Available from: https://doi.org/10.1136/gutjnl-2017-315259.

18 CRD. *Systematic Reviews: CRD's guidance for undertaking reviews in health care*. York: Centre for Reviews and Dissemination (CRD), University of York; 2009. Available from: https://www.york.ac.uk/media/crd/Systematic_Reviews.

19 Lefebvre C, Glanville J, Briscoe S, Featherstone R, Littlewood A, Metzendorf M-I, et al. Chapter 4: Searching for and selecting studies. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, eds. *Cochrane Handbook for Systematic Reviews of Interventions version 6.4 (updated October 2023) Cochrane*. 2023.

20 Li T, Higgins JPT, Deeks JJ, eds. Chapter 5: Collecting data. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, eds. *Cochrane Handbook for Systematic Reviews of Interventions version 6.4 (updated August 2023). Cochrane*. 2023.

21 Boutron I, Page MJ, Higgins JPT, Altman DG, Lundh A, Hróbjartsson A. Considering bias and conflicts of interest among the included studies. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, eds. *Cochrane Handbook for Systematic Reviews of Interventions, Version 6.4 (updated August 2023)*. The Cochrane Collaboration; 2023.

22 Tod D, Booth A, Smith B. Critical appraisal. *International Review of Sport and Exercise Psychology*. 2022;15(1):52-72. Available from: https://doi.org/10.1080/1750984X.2021.1952471.

23 Higgins JPT, Savović J, Page MJ, Elbers RG, Sterne JAC. Assessing risk of bias in a randomized trial. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, eds. *Cochrane Handbook for Systematic Reviews of Interventions, Version 6.4*. The Cochrane Collaboration; 2023.

24 Schattenberg JM, Pares A, Kowdley KV, Heneghan MA, Caldwell S, Pratt D, et al. A randomized placebo-controlled trial of elafibranor in patients with primary biliary cholangitis and incomplete response to UDCA. *J Hepatol*. 2021;74(6):1344-54. Available from: https://doi.org/10.1016/j.jhep.2021.01.013.

25 Hirschfield GM, Mason A, Luketic V, Lindor K, Gordon SC, Mayo M, et al. Efficacy of obeticholic acid in patients with primary biliary cirrhosis and inadequate response to ursodeoxycholic acid. *Gastroenterology*. 2015;148(4):751-61.e8. Available from: https://doi.org/10.1053/j.gastro.2014.12.005.

26 Dias S, Welton NJ, Sutton AJ, Ades AE. *NICE DSU Technical Support Document 2: A generalised linear modelling framework for pairwise and network meta-analysis of randomised controlled trials; last updated September 2016*. Sheffield: NICE Decision Support Unit (DSU); 2011. Available from: https://www.sheffield.ac.uk/nice-dsu/tsds/full-list [accessed 26 April 2024].

27 Turner RM, Jackson D, Wei Y, Thompson SG, Higgins JPT. Predictive distributions for between-study heterogeneity and simple methods for their application in Bayesian meta-analysis. *Statistics in Medicine*. 2015;34(6):984-98. Available from: https://doi.org/https://doi.org/10.1002/sim.6381.

28 NHS. *Primary biliary cholangitis (primary biliary cirrhosis): Symptoms (Page last reviewed: 02 February 2024;*

Next review due: 02 February 2027). London: National Health Service (NHS). Available from: https://www.nhs.uk/conditions/primary-biliary-cholangitis-pbc/symptoms/ [accessed 15 May 2024].

29 Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2002;64(4):583-639. Available from: https://doi.org/https://doi.org/10.1111/1467-9868.00353.

30 Ren S, Oakley JE, Stevens JW. Incorporating Genuine Prior Information about Between-Study Heterogeneity in Random Effects Pairwise and Network Meta-analyses. *Medical Decision Making*. 2018;38(4):531-42. Available from: https://doi.org/10.1177/0272989X18759488.

31 Kowdley KV, Brookhart MA, Hirschfield GM. Efficacy of Obeticholic Acid (OCA) vs Placebo and External Control (EC) on Clinical Outcomes in Primary Biliary Cholangitis (PBC). *Hepatology*. 2022;77(Abstract 5032. AASLD Meeting):E116. Available from: https://doi.org/10.1097/HEP.0000000000000347.

32 Chaimani A, Caldwell DM, Li T, Higgins JPT, Salanti G. Chapter 11: Undertaking network meta-analyses. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, eds. *Cochrane Handbook for Systematic Reviews of Interventions version 6.4 (updated August 2023). Cochrane, 2023. Available from www.training.cochrane.org/handbook*. 2023.

33 NICE. *Elafibranor for treating primary biliary cholangitis [ID6331]. Final scope. Health technology evaluation*. London: National Institute for Health and Care Excellence (NICE); 2024. Available from: https://www.nice.org.uk/guidance/gid-ta11378/documents/final-scope [Date accessed: 8 April 2024].

34 Ipsen. *Elafibranor for treating primary biliary cholangitis [ID6331]. Patient organisation submission - British Liver Trust.* Single Technology Appraisal (STA). Ipsen; 2024.

35 Ipsen. *Elafibranor for treating primary biliary cholangitis [ID6331]. Patient organisation submission - Liver4Life*. Single Technology Appraisal (STA). Ipsen; 2024.

36 Balasubramanian H, Ananthan A, Rao S, Patole S. Odds ratio vs risk ratio in randomized controlled trials. *Postgrad Med*. 2015;127(4):359-67. Available from: https://doi.org/10.1080/00325481.2015.1022494.

37 Owen RK, Bradbury N, Xin Y, Cooper N, Sutton A. MetaInsight: An interactive web-based tool for analyzing, interrogating, and visualizing network meta-analyses using R-shiny and netmeta. *Research Synthesis Methods*. 2019;10(4):569-81. Available from: https://doi.org/https://doi.org/10.1002/jrsm.1373.

38 Balduzzi S, Rücker G, Nikolakopoulou A, Papakonstantinou T, Salanti G, Efthimiou O, Schwarzer G. netmeta: An R Package for Network Meta-Analysis Using Frequentist Methods. *Journal of Statistical Software*. 2023;106(2):1 - 40. Available from: https://doi.org/10.18637/jss.v106.i02.

39 CADTH. *Pharmacoeconomic Review Report: Obeticholic Acid (Ocaliva): Intercept Pharma Canada, Inc.: Indication: For the treatment of primary biliary cholangitis.* Ottawa: Canadian Agency for Drugs and Technologies in Health (CADTH); 2017. Available from: https://pubmed.ncbi.nlm.nih.gov/30525351/.

40 NCPE. *Cost-effectiveness of obeticholic acid (Ocaliva®) for the treatment of primary biliary cholangitis in combination with ursodeoxycholic acid (UDCA) in adults with an inadequate response to UDCA or as monotherapy in adults unable to tolerate UDCA*. Dublin: National Centre for Pharmacoeconomics, Ireland (NCPE); 2017. Available from: https://www.ncpe.ie/drugs/obeticholic-acid-ocaliva/.

41 SMC. *Obeticholic acid (Ocaliva®). SMC No: 1232/17*. Glasgow: Scottish Medicines Consortium (SMC); 2017. Available from: https://www.scottishmedicines.org.uk/medicines-advice/obeticholic-acid-ocaliva-fullsubmission-123217/.

42 Ipsen. *Elafibranor for treating primary biliary cholangitis [ID6331]. Clinical expert opinion. Single Technology Appraisal (STA)*. 2024.

43 Fayter D, Ramaekers BLT, Grimm S, Pouwels X, Armstrong N, Ryder S, et al. *Obeticholic acid for treating primary biliary cirrhosis: a Single Technology Assessment [TA443]. Evidence Assessment Group Report*. York: Kleijnen Systematic Reviews Ltd; 2016.

44 Ipsen, FIECON. *Clinical consultation meeting report: economic model validation (Clinician validation findings report merged_v2.0_FINAL_10May2024)*. France, Paris: Ipsen; 2024.

45 Drefs M, Schoenberg MB, Börner N, Koliogiannis D, Koch DT, Schirren MJ, et al. Changes of long-term survival of resection and liver transplantation in hepatocellular carcinoma throughout the years: A meta-analysis. *European Journal of Surgical Oncology*. 2024;50(3). Available from: https://doi.org/10.1016/j.ejso.2024.107952.

46 Sun Q, Zhang X, Gong X, Hu Z, Zhang Q, He W, et al. Survival analysis between laparoscopic and open hepatectomy for hepatocellular carcinoma: a meta-analysis based on reconstructed time-to-event data. *Hepatology International*. 2021;Oct(5):1215-35. Available from: https://doi.org/10.1007/s12072-021-10219-1.

47 Tantai X, Liu Y, Yeo YH, Praktiknjo M, Mauro E, Hamaguchi Y, et al. Effect of sarcopenia on survival in patients with cirrhosis: A meta-analysis. *Journal of Hepatology*. 2022;76(3):588-99. Available from: https://doi.org/https://doi.org/10.1016/j.jhep.2021.11.006.

48 Ipsen. *Clinical Study Report: A Double-Blind, Randomized, Placebo-Controlled Study and Open-Label Long Term Extension to Evaluate the Efficacy and Safety of Elafibranor 80 Mg in Patients with Primary Biliary Cholangitis with Inadequate Response or Intolerance to Ursodeoxycholic Acid*. Paris, France: Ipsen Pharma; 2023.

49 Holger J, Schünemann GEV, Higgins JPT, Santesso N, Deeks JJ, Glasziou P, et al. Interpreting results and drawing conclusions In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, eds. *Cochrane Handbook for Systematic Reviews of Interventions, Version 6.4 (updated August 2023)*. The Cochrane Collaboration; 2023.

50 Mayo MJ, Carey E, Smith HT, Mospan AR, McLaughlin M, Thompson A, et al. Impact of Pruritus on Quality of Life and Current Treatment Patterns in Patients with Primary Biliary Cholangitis. *Digestive Diseases and Sciences*. 2023;68(10):995-1005. Available from: https://doi.org/10.1007/s10620-022-07581-x.

51 Rice S, Albani V, Minos D, Fattakhova G, Mells GF, Carbone M, et al. Effects of primary biliary cholangitis on quality of life and health care costs in the United Kingdom. *Clin Gastroenterol Hepatol*. 2021;19(4):768-76.e10. Available from: https://doi.org/10.1016/j.cgh.2020.06.025.

52 Younossi ZM, Kiwi ML, Boparai N, Price LL, Guyatt G. Cholestatic Liver Diseases and Health-Related Quality of Life. *Official journal of the American College of Gastroenterology | ACG*. 2000;95(2). https://journals.lww.com/ajg/fulltext/2000/02000/cholestatic_liver_diseases_and_health_related.28.aspx.

53 Saeed YA, Phoon A, Bielecki JM, Mitsakakis N, Bremner KE, Abrahamyan L, et al. A Systematic Review and Meta-Analysis of Health Utilities in Patients With Chronic Hepatitis C. *Value in Health*. 2020;23(1):127-37. Available from: https://doi.org/10.1016/j.jval.2019.07.005.

54 Hernandez Alava M, Pudney S, Wailoo A. The EQ-5D-5L Value Set for England: Findings of a Quality Assurance Program. *Value in Health*. 2020;23(5):642-8. Available from: https://doi.org/https://doi.org/10.1016/j.jval.2019.10.017.

55 Wright M, Grieve R, Roberts J, Main J, Thomas HC, U.K. Mild Hepatitis C Trial Investigators. Health benefits of antiviral therapy for mild chronic hepatitis C: randomised controlled trial and economic evaluation. *Health technology assessment (Winchester, England)*. 2006;10(21):1-113, iii. Available from: https://doi.org/10.3310/hta10210.

56 NICE. *NICE health technology evaluations: the manual. Process and methods [PMG36]. Published: 31 January 2022. Last Update Date: 31 October 2023.* London: National Institute for Health and Care Excellence (NICE); 2022. Available from: https://www.nice.org.uk/process/pmg36/chapter/introduction-to-health-technology-evaluation [Date accessed: 08/04/2024].

57 BNF. *Treatment summaries: Cholestasis*. National Institute for Health and Care Excellence; 2023. Available from: https://bnf.nice.org.uk/treatment-summaries/cholestasis/.

58 NHS England. *2021/22 National Cost Collection Data Publication*. 2023. Available from: https://www.england.nhs.uk/publication/2021-22-national-cost-collection-data-publication/.

59 NICE. *Odevixibat for treating progressive familian intrahepatic cholestasis [ID1570, HST17]: Committee papers.*; 2021. Available from: https://www.nice.org.uk/guidance/hst17/documents/committee-papers.

60 BNF. *Medicinal forms : Azathioprine*. British National Formulary (BNF). Available from: https://bnf.nice.org.uk/drugs/azathioprine/medicinal-forms/.

61 BNF. *Medicinal forms: Tacrolimus* British National Formulary (BNF). Report No.: 23/05/2024. Available from: https://bnf.nice.org.uk/drugs/tacrolimus/medicinal-forms/.

62 BNF. *Medicinal forms: Prednisolone*. British National Formulary (BNF). Available from: https://bnf.nice.org.uk/drugs/prednisolone/medicinal-forms/.

63 Gola A, Davis S, Greenslade L, Hopkins K, Low J, Marshall A, et al. Economic analysis of costs for patients with end stage liver disease over the last year of life. *BMJ Supportive & Palliative Care*. 2015;5(1):110. Available from: https://doi.org/10.1136/bmjspcare-2014-000838.23.

64 NICE. *Atezolizumab with bevacizumab for treating advanced or unresectable hepatocellular carcinoma [TA666]. Technology appraisal guidance*. London: National Institute for Health and Care Excellence (NICE); 2020. Available from: https://www.nice.org.uk/guidance/ta666?fromsource=mas.

65 Kaltenthaler E, Carroll C, Hill-Mcmanus D, Scope A, Holmes M, Rice S, et al. The use of exploratory analyses within the National Institute for Health and Care Excellence single technology appraisal process: an evaluation and qualitative analysis. *Health Technology Assessment*. 2016;20(26):1-48. Available from: https://doi.org/10.3310/hta20260.

66 Ipsen. Ipsen UK-PBC Data Requests. Paris, France: Ipsen Pharma; 2024.