Etranacogene dezaparvovec for treating moderately severe or severe haemophilia B [ID3812]:

A Single Technology Appraisal

Produced by

Peninsula Technology Assessment Group (PenTAG)
University of Exeter Medical School
South Cloisters
St Luke's Campus
Heavitree Road
Exeter
EX1 2LU

Authors

Caroline Farmer[1]
Elham Nikram[1]
Laura A Trigg[1]
Sophie Robinson[1]
Jason Coppell[2]
Madhusubramanian Muthukumar[1]
G.J. Melendez-Torres[1]
Edward CF Wilson[1]

[1] Peninsula Technology Assessment Group (PenTAG), University of Exeter Medical School, Exeter

[2] Royal Devon and Exeter NHS Foundation Trust

Correspondence to

Caroline Farmer

3.09 South Cloisters, St Luke's Campus, Heavitree Road, Exeter, EX1 2LU; c.farmer@exeter.ac.uk

Date completed	04/04/2023
Source of funding	This report was commissioned by the NIHR Evidence Synthesis Programme as project number NIHR135784.
Declared competing interests of the authors	None
Acknowledgments	The authors acknowledge the administrative support provided by Mrs Sue Whiffin and Ms Jenny Lowe, and additional input in to the EAG's analyses by Brian O'Toole (all PenTAG).
Rider on responsibility for document	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	Farmer C, Nikram E, Trigg L, Robinson S, Coppell J, Muthukumar M, Melendez-Torres, G.J., Wilson ECF. Etranacogene dezaparvovec for treating moderately severe or severe haemophilia B [ID3812]: A Single Technology Appraisal. Peninsula Technology Assessment Group (PenTAG), 2023.

Author Contributions:

Caroline Farmer	Project lead, critical appraisal of the company submission, writing and editorial input
Elham Nikram	Critical appraisal of the economic evidence and analysis submitted by the company, conducted additional economic analyses and drafted sections of the report
Laura A Trigg	Critical appraisal of the economic evidence and analysis submitted by the company and drafted sections of the report
Sophie Robinson	Critical appraisal of the company's literature search strategies
Jason Coppell	Expert clinical advice about haemophilia B and its treatment
Madhusubramanian Muthukumar	Review and QC of model and EAG analyses
G.J. Melendez-Torres	Critique of the company's indirect treatment comparison

Edward CF Wilson	Health economic lead, critical appraisal of the company submission and analysis, conducted additional economic analyses and drafted sections of the report. Report Guarantor.

Table of Contents

List of key issues

List of tables

List of Figures

1. EXECUTIVE SUMMARY

This summary provides an overview of the key issues identified by the external 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. Section 1.2 provides an overview of key model outcomes and the modelling assumptions that have the greatest effect on the ICER. Sections 1.3 to 1.6 explain the key issues in more detail. Background information on the condition, technology and evidence and information on non-key issues are in the main EAG report.

All issues identified represent the EAG's view, not the opinion of NICE.

1.1. Overview of the EAG's key issues

A brief overview of the key issues identified by the EAG in their appraisal of the company submission (CS) is provided in Table 1. Further detail of the issues is provided in Sections 1.3, to 1.6.

Broadly speaking the key clinical issues stemmed from the lack of a randomised trial for ED (specifically a lack of a reliable comparison between ED and current treatment options), and gaps in the evidence base submitted by the company in their submission. Notably, the EAG considered there to be a risk that the effect of ED in the key study, HOPE-B, may be overstated. In terms of cost effectiveness issues, the EAG noted that the definition of ED treatment failure was set at a very low FIX activity level (<2%), the durability extrapolation excluded non-responders and was associated with a great deal of uncertainty due to small numbers and limited follow-up, and that the treatment-related utility of ED vs. IV FIX may be overestimated. Most significantly, the likelihood that ED was cost effective was highly impacted by assumptions surrounding the durability of ED treatment response in the model.

Table 1: Summary of key issues

ID	Summary of issues	Report sections
Key Issue 1	The company did not report evidence for the true change in FIX levels following treatment with ED in the HOPE-B	3.2.2.5

ID	Summary of issues	Report sections
Key Issue 2	Clinical outcomes in the HOPE-B study may overstate the potential benefits of ED	3.2.2.3; 3.2.2.4; 3.2.2.6 and 6.2.2; 6.2.3; 6.2.10.1
Key Issue 3	Comparative efficacy estimates of ED and prophylactic FIX treatments were unreliable	3.3; 3.4
Key Issue 4	Definition of treatment failure was at a low FIX activity level	6.2.2
Key Issue 5	The durability extrapolation model was based on limited data and excluded non-responders	4.2.6.1; 5.2.3.16.2.3; 6.2.10.1; 6.3.1
Key Issue 6	Health state utilities were associated with treatment rather than health states, and the difference may be overestimated.	4.2.7.1; 6.2.5

The key differences between the company's preferred assumptions and the EAG's preferred assumptions are outlined in Table 2.

Table 2: Key differences between the company's preferred assumptions and EAG's preferred assumptions

	Company's preferred assumption	EAG preferred assumption	Report Sections
IV FIX taken alongside ED and post ED failure	Various for pairwise comparisons, Refixia for fully incremental analysis	Only fully incremental analyses conducted, assuming Refixia alongside ED in all cases	4.2.4
FIX activity threshold at which prophylactic IV FIX is resumed ("treatment failure")	2%	5%	4.2.6.2; 6.2.3
Time to steady state	3 weeks	6 months	4.2.6.3; 6.2.4
Disutility of IV FIX treatment compared with ED	[redacted]	0.042	4.2.7.1; 6.2.5
Duration of adverse event costs and consequences from ED.	1-year post-ED administration	Whilst durability of ED continues	6.1; 6.2.9

Abbreviations: ED, etranacogene dezaparvovec; FIX, Factor IX; IV, intravenous

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 extra QALY gained.

Overall, the technology is modelled to affect QALYs by:

- Increased health state utility associated with receiving a once-only injection of ED compared with (once or twice weekly) IV injections of FIX.
- Reduced risk of bleeds with ED compared with IV FIX.

Overall, the technology is modelled to affect costs by:

- Lower lifetime acquisition cost of ED versus other FIX products.

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

- Durability of ED (i.e. time before prophylactic IV FIX is resumed)
- 'Treatment associated' health utility bonus from a single injection of ED versus repeated IV FIX injections.

1.3. The decision problem: summary of the EAG's key issues

The EAG did not identify any key issues related to the company's definition of the decision problem.

1.4. The clinical effectiveness evidence: summary of the EAG's key issues

Key Issue 1: The company did not report evidence for the true change in FIX levels following treatment with ED in the HOPE-B study

Report sections	3.2.2.5
Description of issue and why the EAG has identified it as important	The HOPE-B study was a single-arm study that compared outcomes following treatment with ED with participants' outcomes during a baseline lead-in phase of 6-months. As there was no control arm, outcomes assessed during the lead-in phase were the only data to represent participant outcomes while receiving a comparator treatment (prophylactic FIX replacement). One of the key study outcomes, levels of circulating FIX following treatment, was an important outcome for determining the effect of the treatment, and how the condition affects people's lives, including the need for

	additional FIX replacement therapies. However, the company did not report FIX levels during the lead-in phase, and therefore it was not possible for the EAG to determine to what extent FIX levels changed following treatment with ED. The company calculated change in FIX levels from baseline, however the baseline data used for these calculations were not based on data from the lead-in phase, but were rather an estimate of what participants' FIX levels would be if they were receiving no treatment at all (i.e. they used the severity of their condition to impute a FIX level). The EAG considered this approach to be inconsistent with the decision problem for this appraisal, and that the presentation of these findings could potentially be misleading.
What alternative approach has the EAG suggested?	The company did not provide a rationale for why true FIX levels from the lead-in phase of the HOPE-B study were not used to calculate the change in FIX following treatment with ED. The EAG was aware that FIX levels following prophylactic FIX replacement fluctuate with high levels following treatment administration and low levels in the 'trough' before another dose is administered. For this reason, it may be that the company were uncertain how to select a representative FIX level for the lead-in phase from which to calculate the change outcome. However, the EAG considered that the company could have provided descriptive data for the lead-in phase and provided an analysis of change in FIX levels as compared to mean, highest and lowest FIX levels during the lead-in phase. This would have given an indication of the extent to which ED affected circulating FIX, and would be useful given limitations in bleeding and FIX replacement outcomes described in Key Issue 2.
What is the expected effect on the cost-effectiveness estimates?	FIX levels were not directly included as part of the decision modelling, although failure of ED was defined as a circulating FIX activity level of <2% in the company's durability extrapolation. Understanding of the difference in FIX levels between the lead-in phase and following treatment with ED may reduce uncertainty in the reliability of bleeding outcomes that were used in the model (Key Issue 2).
What additional evidence or analyses might help to resolve this key issue?	The company should provide true baseline FIX levels for the lead-in phase of HOPE-B. Useful analyses would be the difference in FIX levels following treatment with ED as compared to mean, minimum and maximum FIX levels during the lead-in phase.

Abbreviations: EAG, External Assessment Group; ED, etranacogene dezaparvovec; FIX, factor IX

Key Issue 2: Clinical outcomes in the HOPE-B study may overstate the potential benefits of ED

Report sections	3.2.2.3; 3.2.2.4; 3.2.2.6**;** 6.2.2; 6.2.3; 6.2.10.1
Description of issue and why the EAG has identified it as important	HOPE-B was a single-arm study that compared outcomes following treatment with ED with participants' outcomes during a baseline lead-in phase of 6-months. As there was no control

	arm, the comparability of outcomes measured in the lead-in phase and following treatment were crucial for determining the clinical effectiveness of ED. The EAG had two main concerns about the comparability of outcomes may affect the reliability of the study evidence: 1. The COVID-19 pandemic began after study participants had received ED and resulted in major disruption to the daily activities of people in the UK. Those with HIV, hepatitis or those receiving immunosuppression were included in advice to shield, whereas others may have experienced significant reductions in activities outside of their homes, including sports and travel. The EAG expected that these changes may have reduced the level of circulating FIX people with haemophilia B needed to do their daily activities, which may have therefore reduced the need for study participants to receive additional FIX replacement during the study. They may also have had a lower risk of bleeding during this time, due to their reduced activity. 2. The study procedures prohibited participants from receiving routine FIX replacement when they had circulating FIX levels of ≥5%. In these circumstances, investigating clinicians were permitted to administer ad hoc FIX replacement at their discretion, though the EAG considered that clinicians may be less likely to do this within the clinical study than they may do in practice, so as to adhere as closely as possible to the preferred study procedures. This requirement was not in place during the lead-in phase, and the EAG considered it plausible that rates of prophylactic FIX replacement would be higher in clinical practice than in the HOPE-B study.
What alternative approach has the EAG suggested?	The EAG explored the impact of changes to the clinical efficacy of ED and of increasing prophylactic FIX replacement in the EAG model through a number of scenarios (sections 6.2.2, 6.2.3 and 6.2.10.1)
What is the expected effect on the cost-effectiveness estimates?	Overstating the effectiveness of ED will both overestimate QALYs gained and underestimate cost through underestimation of IV FIX ultimately consumed by patients in the ED arm of the model.
What additional evidence or analyses might help to resolve this key issue?	As lives return towards normal in the years following the COVID-19 pandemic, study participants' daily activities may become more comparable with those during the lead-in phase. This may mean that subsequent data cuts of the HOPE-B study may provide a more representative view of the potential benefit of ED. To inform if and to what extent the use of FIX replacement therapy would be higher in clinical practice than in the HOPE-B study, the EAG would be interested to see the proportion of participants in HOPE-B with circulating FIX levels at alternative thresholds. The EAG would then seek clinical opinion on how many people at each threshold may choose to receive

	additional FIX therapy according to safety and/or personal preference.

Abbreviations: EAG, External Assessment Group; ED, etranacogene dezaparvovec; FIX, factor IX

Key Issue 3: Comparative efficacy estimates of ED and prophylactic FIX treatments were unreliable

Report sections	3.3; 3.4
Description of issue and why the EAG has identified it as important	The company identified four studies that reported outcomes for the main comparators to ED and used these along with outcomes from the HOPE-B study to indirectly compare treatment outcomes. There were no head-to-head comparisons of different FIX therapies, and most comparative studies compare prophylactic vs. on-demand treatment. Moreover, differences between the methods used in the studies seriously undermined the comparability of the outcomes. The company used matching of population characteristics to improve the quality of their ITC, but this process was itself highly limited due to the information available to them in the comparator studies. Overall, while the EAG considered that the company's methods for the ITC were the best available to them, the results were nevertheless unreliable and it therefore had little confidence in the results. The findings were most unreliable for BeneFIX, which [redacted].
What alternative approach has the EAG suggested?	The main difficulty with the company's ITC was the poor quality of evidence for prophylactic FIX and the differences in methods between the HOPE-B and comparator studies, including the definition and measurement of bleeding outcomes. This could not be resolved by the EAG. On the combined evidence of the HOPE-B study and the company's ITC, the EAG considered it plausible that treatment with ED would result in lower bleeding rates than FIX replacement. However, the EAG considered that the magnitude of that reduction was uncertain.
What is the expected effect on the cost-effectiveness estimates?	Overstating the effectiveness of ED will both overestimate QALYs gained and underestimate cost through underestimation of IV FIX ultimately consumed by patients in the ED arm of the model.
What additional evidence or analyses might help to resolve this key issue?	The company's methods were the best available to them with the current evidence. New, high-quality, comparative evidence to compare outcomes following treatment with ED vs. prophylactic FIX therapy was needed to resolve this issue.

Abbreviations: EAG, External Assessment Group; ED, etranacogene dezaparvovec; FIX, factor IX; ITC, indirect treatment comparison

1.5. The cost effectiveness evidence: summary of the EAG's key issues

Key Issue 4: Definition of treatment failure was at a low FIX activity level

Report sections	6.2.2
Description of issue and why the EAG has identified it as important	Treatment failure in the company model was effectively defined as resumption of prophylactic IV FIX. The company's base case durability extrapolation model was based on a resumption of IV FIX at <2% FIX activity level, however clinical advice to the EAG was that IV FIX was more likely to be reintroduced once FIX activity dropped below 5% rather than 2%. Durability of treatment effect (i.e. time to resumption of IV FIX) was fundamental to estimation of incremental costs and QALYs gained from ED.
What alternative approach has the EAG suggested?	The EAG base case utilises 5% as the threshold for reintroducing IV prophylactic FIX.
What is the expected effect on the cost-effectiveness estimates?	The use of a 2% threshold was considered to underestimate the ICER and thus overstate the cost-effectiveness of ED compared with IV FIX.
What additional evidence or analyses might help to resolve this key issue?	Wider consultation with clinical experts as to FIX activity levels at which they would reinstate prophylactic IV FIX would be informative.

Abbreviations: EAG, External Assessment Group; ED, etranacogene dezaparvovec; QALY, Quality Adjusted Life Year

Key Issue 5: The durability extrapolation model was based on limited data and excluded non-responders

Report sections	4.2.6.1; 5.2.3.16.2.3; 6.2.10.1; 6.3.1
Description of issue and why the EAG has identified it as important	Durability of the ED treatment effect was fundamental to the cost-effectiveness of ED. The extrapolation model used was based on small sample sizes and a very short follow-up relative to the extrapolation period.
What alternative approach has the EAG suggested?	The EAG conducted a threshold analysis to determine the minimum durability of ED required to yield an ICER below £20,000 and below £30,000 per QALY gained.
What is the expected effect on the cost-effectiveness estimates?	Overstating durability will overestimate incremental QALYs gained and underestimate incremental cost.
What additional evidence or analyses might help to resolve this key issue?	Due to the rarity of the disease, sample size limitations were unsurmountable. However, longer follow up of existing cohorts was considered essential to reducing uncertainty in durability.

Abbreviations: EAG, External Assessment Group; ED, etranacogene dezaparvovec; QALY, Quality Adjusted Life Year

Key Issue 6: Health state utilities were associated with treatment rather than health states, and the difference may be overestimated.

Report sections	4.2.7.1; 6.2.5
Description of issue and why the EAG has identified it as important	As a general principle, the EAG preferred health state utilities attached to states of health rather than treatment received because allowing treatment-driven utilities as well as differences in transition probabilities in a model risks double counting the impact of a treatment and thus overstating cost-effectiveness. However, the EAG agreed with the company that there may be a difference in utility by treatment over and above that associated with bleed rates and which was not otherwise captured in the decision model, namely a psychological benefit from receiving a once-in-a-lifetime treatment compared with frequent, repeat IV treatments. Nevertheless, the EAG considered the value applied to be overly optimistic.
What alternative approach has the EAG suggested?	The EAG considered that a lower treatment-related utility difference was more appropriate
What is the expected effect on the cost-effectiveness estimates?	Overestimating the utility difference would overestimate incremental QALYs and therefore underestimate the ICER
What additional evidence or analyses might help to resolve this key issue?	More evidence was needed to support the use and magnitude of a treatment-specific utility. Health state utilities based on EQ-5D collected alongside a randomised comparison of ED versus IV FIX would be the most appropriate evidence, though given the lack of an existing randomised study of ED the EAG considered that this was unlikely within the timeline of the appraisal. The EAG was unaware of an indirect population that would be suitable.

Abbreviations: EAG, External Assessment Group; ED, etranacogene dezaparvovec; QALY, quality-adjusted life-year

1.6. Other key issues: summary of the EAG's views

The EAG did not identify any other key issues.

1.7. Summary of EAG's preferred assumptions and resulting ICER

The EAG submitted a revised model correcting a number of errors. This also included a number of undocumented changes to the company base case made by the company at clarification. The results of the corrected company base case and the EAG preferred assumptions incorporating a patient access scheme (PAS) discount for ED of [redacted] are shown in Table 3.

Modelling errors identified and corrected by the EAG are described in section 6.1. For further details of the exploratory and sensitivity analyses done by the EAG, see section 6.2

Table 3: Summary of EAG's preferred assumptions and ICER (probabilistic results)

Preferred assumption	Section in EAG report	Comparators	Costs	QALYs	ICERs	NMB @ £20k	NMB @ £30k
EAG corrected company base case (excl. ED+mkt share)	6.1	ED+Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Benefix	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Alprolix	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Idelvion	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
EAG preferred base case assumptions							
5% FIX activity definition of failure	6.2.2	ED+Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Benefix	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Alprolix	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Idelvion	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
6 month time to steady state	6.2.4	ED+Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Benefix	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Alprolix	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Idelvion	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Disutility of IV FIX treatment of 0.042	6.2.5	ED+Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Benefix	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Alprolix	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Idelvion	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Adding AE cost and disutility to ED after first year	6.2.9	ED+Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Benefix	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Alprolix	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Idelvion	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Cumulative		ED+Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Benefix	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Alprolix	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Idelvion	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Abbreviations: ED, etranacogene dezaparvovec; FIX, factor IX; IV, intravenous; mkt, market

2. INTRODUCTION AND BACKGROUND

2.1. Introduction

In this report, the External Assessment Group (EAG) provides a review of the evidence submitted by CSL Behring '(the company') for etranacogene dezaparvovec (ED) for the treatment of severe and moderately severe haemophilia B. This report is accompanied by an appendix that contains the company and EAG analyses using confidential prices for comparators to ED. As these prices are not included in the analyses within this report, the findings are indicative only and do not represent current NHS funding for comparators to ED.

2.2. Critique of the company's description of the underlying health problem

The company's description of the condition highlighted key areas for understanding the humanistic burden of severe and moderately severe haemophilia B. For the most part, the EAG considered the company's description to be appropriate, though noted the following additional points:

- The EAG noted a minor typo on p.25 of the CS – "in rare cases, women can have [*severe and moderately severe*] haemophilia B". The EAG agreed with the company's description about the role of gender in the condition and received feedback from its clinical expert that the few females who experience severe and moderately severe haemophilia B would be affected similarly as males.

- The company described the incidence and impact of joint bleeding and arthropathy, which significantly impacts the lives of people with haemophilia B and is associated with delayed or insufficient treatment to maintain sufficient FIX levels and prevent bleeds. As an addition to the company description, the EAG noted that the younger cohort of people with haemophilia B in England will have a much lower risk of joint bleeds and arthropathy in their lifetimes due to earlier access to prophylactic FIX replacement. Clinical advice to the EAG was that the older cohort who did not have access to prophylactic treatment typically have received at least one joint replacement and experience significant disability, whereas the majority of those in the younger cohort are much less likely to have severe joint problems and require replacements. This is likely to lead to higher lifetime health-related quality of life (HRQoL) in the younger cohort.

- Further to the information provided by the company about the risk of mortality in those with haemophilia B, the EAG noted that mortality risk for the condition will have changed over the past several decades. This will be due in part to increased access to prophylactic FIX replacement, but also because those treated in the 1970-80s may have been exposed to contaminated blood products during FIX replacement and were at a higher risk of mortality due to infections such as HIV and hepatitis. The EAG considered it likely that mortality rates in the younger cohort of people in England with haemophilia B are likely to be much lower. Clinical advice to the EAG was that life expectancy in England may now be similar to the general population.

- The EAG considered that the company's description of the way in which the condition affects people's HRQoL lacked evidence for how HRQoL varies across the population. The EAG was aware that HRQoL is poorest for people with haemophilia B who develop inhibitors to FIX, meaning that they cannot receive FIX replacement therapy and their health outcomes and the impact of the condition on their life will be much greater. Overall, evidence suggests that HRQoL is worse for those with higher disease severity, though people of all disease severities can report high levels of HRQoL[2]. This may be because of differences in the impact of the condition on people's preferred lifestyles, and/or because people adapt to their condition and its management. Experiencing joint pain is also associated with poorer HRQoL[2]. The EAG understood that the condition does not cause people to feel unwell on a daily basis, and that therefore deficits in HRQoL are primarily driven by the impact of the condition on their joints (e.g. chronic pain), the psychological impact associated with the risk of bleeds and the lifestyle modifications required to manage the disease safely, and by the burden of treatment.

- The company stated that carers may experience both humanistic and economic burden (CS section B.1.3.3). Four studies[3-6] cited by the company reported that carers experience financial expenses due to their loved one's condition, though three studies were based in the US and not directly applicable, while the other reported overall indirect costs associated with people with either haemophilia A or B and did not separate out costs incurred by carers as compared to other indirect costs (e.g. loss of earnings). The EAG further noted that the two studies[7 8] cited by the company to support the humanistic burden of haemophilia B for carers were restricted to considering the burden amongst the carers of children with haemophilia, who are outside the target indication for this appraisal. The EAG therefore concluded that the company had not provided evidence to support its assertions concerning

the HRQoL impacts for the carers of those with haemophilia B. The EAG considered it plausible that carers would experience a detrimental impact from their loved one's condition, though expected that the detriment would be much greater for the carers of children and young adults who may be required to facilitate access to healthcare appointments and have greater responsibility in ensuring that their loved one is safe. For adults with haemophilia, the EAG considered it plausible that limitations on the lifestyles of those with the condition would be expected to have some impact on the carers of adults with haemophilia, and anxiety related to the risk and impact of bleeds is also likely to be felt by a person's carers.

- The company estimated that there were 242 people with severe disease registered in the UK and 271 with moderate disease, of whom a sub-population will have moderately severe disease (defined within the HOPE-B study as ≤2% FIX levels; p. 24 & p.33) The company did not report a breakdown of the number of people considered to have moderately severe disease, though the figures provided by the company suggested that this would be a small population. The EAG was unable to identify other figures for the incidence of severe vs. moderately severe disease to validate the company figures, though clinical advice to the EAG agreed that this would be a small population. The EAG noted that 'moderately severe' was not an established threshold in NHS practice, though it has been used in studies of IV FIX replacement therapies.
- The company estimated there to be [redacted] people in England who would be eligible to receive ED. This calculation was based on UKHCDO data for the number of registered people in England with severe and moderately severe disease minus those who would not have been eligible for treatment with ED in the HOPE-B study. The EAG noted that the numbers reported in the CS (p.34) did not tally with the final numbers, however the difference was minimal and was assumed to be due to a typo.

2.3. Critique of the company's overview of current service provision

Overall, the EAG considered the company's description of the current treatment pathway for the target population to be accurate.

2.4. Critique of company's definition of decision problem

The EAG considered that the CS was consistent with the NICE scope and decision problem for this appraisal. The approved product licence for ED aligned with the population in the key study

for ED (HOPE-B) and was therefore considered to be representative of the target population. The EAG also clarified that ED was intended to be delivered alongside standard care, which would include routine prophylactic FIX replacement if/when the treatment effect of ED wanes, and on-demand FIX replacement as required. The EAG also considered it plausible that some people may receive additional prophylactic FIX replacement, depending on their response to ED. The EAG appraisal of the company's definition of the decision problem is provided in Table 4.

Table 4: Summary of decision problem

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	EAG comment
Population	People with moderately severe* or severe haemophilia B	As per final scope	Not applicable	The principal clinical evidence for ED was the HOPE-B study, which included evidence from people with both moderately severe and severe haemophilia B, as consistent with the expected licence for ED. However, only 18.5% of people in the study had moderately severe disease. The company presented some data that suggested that this was reflective of the true population, though the data was not provided in full for validation. On the basis of the evidence presented, the EAG concluded that the study was likely representative, though noted that the generalisability of the evidence would be in question if this was not the case. This was because clinical advice to the EAG was that the relative treatment effect of ED would likely be smaller in those with moderately severe disease compared to severe disease.
Intervention	Etranacogene dezaparvovec (ED)	As per final scope	Not applicable	The company's evidence was consistent with the NICE scope and decision problem for this appraisal, though the EAG noted that ED would be administered in conjunction with standard care, including Factor IX (FIX) replacement therapy (the comparator). The evidence presented by the company suggested that FIX replacement would be administered at a lower rate than in the comparator arm, though the EAG noted some uncertainty about the magnitude of this difference

				(Key Issue 2). The way that FIX replacement therapy would be expected to be delivered alongside ED is discussed in Section 3.2.2.3
Comparator(s)	Established clinical management (including prophylaxis and on-demand treatment)	As per final scope, comparator was IV prophylaxis with on-demand option used in some patients	FIX prophylaxis was the most relevant comparator used in clinical practice. A very small cohort of patients using on-demand FIX treatment may be eligible for ED, i.e. those who are eligible for prophylaxis but continue to treat on-demand due to patient choice or clinical challenges	The EAG agreed that prophylactic FIX replacement was the most appropriate comparator, as this was considered to be the best available treatment for the target population. As stated by the company, a small number of people in practice choose to use on-demand treatment due to personal preference or clinical issues with administering prophylactic treatment, however this was not permitted by the clinical study inclusion criteria and was not considered within the company's model. Nearly half of participants in HOPE-B were receiving standard-life FIX replacement for prophylaxis at baseline. A market share report provided by the company suggested [redacted] clinical advice to the EAG was that more people may begin to use extended life products in future. Short-life products would be associated with more instability in FIX levels and higher resource use, though extended life products have a much greater cost. This issue is discussed in Section 3.2.2.4.
Outcomes	The outcome measures to be considered include: • change in FIX levels • need for further treatment with FIX injections • annualised bleeding rate	As per final scope	Not applicable	The EAG considered that the company had presented evidence for all of the scoped outcomes. Data tables for the CSR of the HOPE-B study were not provided with the CS, and therefore full data for all outcomes were not available. Notably, baseline data from HOPE-B for change in FIX levels (which represented the

	• durability of response to treatment • complications of the disease (e.g., joint problems and joint surgeries) • adverse effects of treatment • health-related quality of life.			comparator to ED) were not reported in the CS.
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. 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 between the technologies being compared. Costs will be considered from an NHS and Personal Social Services perspective. The availability of any commercial arrangements for the intervention, comparator and subsequent treatment technologies will be taken into account. The use of ED is conditional on the presence of a specific biomarker (currently considered confidential by the company). The economic	As per final scope, noting that the use of ED is conditional on the test result for a biomarker.	The clarification included in the previous column intends to flag that patients will require to undertake a specific biomarker test for neutralising antibodies before receiving ED. Clinicians will consider the use of ED based on the test result (no cut-off values defined). The company will provide the test free of charge, which is not routinely performed in the NHS, and therefore its costs are not included in the cost-effectiveness model. The company assumes that indirect costs associated with testing patients (e.g., staff time) will not be substantial, as testing will take place as part of routine clinic follow-up.	The economic analysis broadly followed the NICE reference case. The company presented a series of pair-wise comparisons against each comparator in its deterministic base case, with fully incremental analyses presented in the PSA. The EAG considered pairwise comparisons inappropriate for decision making and therefore presented fully incremental analyses as a modification of the company base case. However, this was complicated by the choice of FIX once ED fails (i.e. people return to routine prophylaxis). This is discussed further in Section 6.

	modelling should include the costs associated with diagnostic testing for biomarkers in people with haemophilia B who would not otherwise have been tested. A sensitivity analysis should be provided without the cost of the diagnostic test.			
Subgroups	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 final scope	Not applicable	ED had received CMA by the EMA and, more recently, from the MHRA for use in the UK. The company presented evidence from a number of subgroup analyses, and a further subgroup analysis (relating to the use of corticosteroids during the study) was identified from the EMA report. As the sample size from ED studies was low, there was limited power to explore potential variation in treatment effect across participants.
Special considerations including issues related to equity or equality	None in the final scope.	None in the final scope	Not applicable	The clinical studies for ED included only male participants. The EAG understood that females with haemophilia B typically have mild disease, and very few females would meet the criteria for moderately severe or severe disease. Clinical advice to the EAG was that disease characteristics were similar between males and females, and that the study evidence was generalisable to females who met the eligibility criteria for ED.

Abbreviations: CMA, conditional marketing authorisation; CSR, clinical study report; EAG, External Assessment Group; ED, etranacogene dezaparvovec; FIX, Factor IX replacement therapy; NICE, National Institute for Health and Care Excellence; PSA, probabilistic sensitivity analysis

3. CLINICAL EFFECTIVENESS

3.1. Critique of the methods of review(s)

The company undertook a systematic literature review (SLR) to identify evidence for the clinical effectiveness and safety of treatments used for haemophilia B. A literature search strategy was used to capture evidence published between March 2013 and October 2022, and two published SLRs[9 10] were used to capture evidence published before these dates.

Table 5: Summary of EAG's critique of the methods implemented by the company to identify evidence relevant to the decision problem

Systematic review step	Section of CS in which methods are reported	EAG assessment of robustness of methods
Searches	Appendix D	Searches were well conducted with a variety of keywords and subject headings used in a range of databases. A variety of grey literature sources were also searched. There was some discrepancy in the sources searched in the original searches of August 2021 and then in the update searches of October 2022; most notably the Cochrane Database of Systematic Reviews (CDSR) was searched in the update searches but not in the original searches. Adverse reactions were searched for at the same time as clinical effectiveness evidence (Appendix F).
Inclusion criteria	Appendix D1.1	The inclusion criteria were clear and appropriate to the review aims
Screening	Appendix D1.1	The methods used were consistent with best practice.
Data extraction	Appendix D1.1	The methods used were consistent with best practice.
Tool for quality assessment of included study or studies	Appendix D1.3	Quality assessment of RCTs was conducted using the modified CRD checklist recommended by NICE[11] and quality assessment of uncontrolled studies was conducted using the Downs and Black checklist[12] both of which were appropriate. However, the checklist was not completed to a high standard and the EAG did not consider it to be useful for determining the presence of bias in the studies.
Evidence synthesis	CS Doc B, Section B.2.9 and Appendix D1.2	The company did not pool data from the HOPE-B study with the other phase IIb and I/II studies as it considered the latter studies not to be relevant for decision-making. As the company nevertheless reported data from these studies and referred to the data to support some of its assertions about ED, the EAG considered that the company should have provided a qualitative comparison between the study outcomes. However, as AMT-060 was with a different formulation of ED, and the other study included only 3 participants, the EAG did not consider that any quantitative pooling of data would have been meaningful. The company conducted ITCs to compare ED with routine FIX replacement strategies where evidence for these was

		identified by its SLR. There was a high degree of heterogeneity between the methods used by the ED and FIX replacement studies, including variation in the definition and measurement of outcomes. The company used the best possible approach to account for the available data, though the EAG considered that the underlying evidence resulted in unreliable results (see Key Issue 3).

Abbreviations: CRD, centre for reviews and dissemination; CS, Company submission; EAG, External Assessment Group; ED, etranacogene dezaparvovec; FIX, factor IX; ITC, indirect treatment comparisons; RCT, randomised controlled trial; SLR, systematic literature review

3.2. Critique of trials of the technology of interest, the company's analysis and interpretation (and any standard meta-analyses of these)

3.2.1. Studies included in the clinical effectiveness review

The CS described three studies (Table 6). These comprised a phase I/II single arm study of AMT-060 (N=10), an early form of ED using the same vector and cassette design, but with a wild-type Factor IX transgene instead of the hFIXco-Padua gene variant later incorporated into ED. Two clinical studies of ED were reported, including one Phase IIb single arm study (N=3) and Hope-B, which was a Phase III single arm study (N=67).

Table 6: Clinical evidence for ED included in the CS

Study name and acronym	Study design	Population	Intervention	Comparator	Follow-up	Study type
HOPE-B **NCT03569891**[13]	Phase III, open label, single arm, multicentre	Adult patients with moderately severe or severe haemophilia B with Factor IX level ≤2% (N=67)	ED (single dose, 2×10^{13} GC)	Lead in study phase while participants received prophylactic Factor IX treatment (≥26 weeks)	2 years*	Clinical efficacy, safety, utility
CT-AMT-061-01 **NCT03489291**[14,15]	Phase IIb, open label, single arm	N=3	ED (single dose, 2×10^{13} GC)	NA	3 years*	Dose-comparison. Clinical efficacy, safety
CT-AMT-060-01 **NCT02396342**[16]	Phase I/II, open-label trial with ongoing extension	N=10	AMT-060 • Cohort 1: 5×10^{12} GC • Cohort 2: 2×10^{13} GC	NA	5 years*	Clinical efficacy, safety

Abbreviations: ED, etranacogene dezaparvovec; GC, genome copy; RCT, NA, not applicable; randomised controlled trial

*Latest follow-up available for this submission. Up to five years' data collection is planned. The next data cut will be 36-months' follow-up, which will become available Q3-Q4 2023.

The company did not submit a complete Clinical study report (CSR) for HOPE-B; while the main body of an updated CSR (to 24-month follow-up) was supplied[13], the data tables accompanying this document were not provided. The EAG assumed that the main body of the CSR was complete, though as it was labelled as an amendment to the original CSR, the EAG considered it possible that some information was retained only in the original. Moreover, while the company supplied CSRs for CT-AMT-060 and CT-AMT-061, these were also supplied without full data tables. As a consequence, the EAG did not have access to the full clinical effectiveness outcome data from the company studies. The EAG was uncertain if data identified as missing from the CS, such as FIX levels during the lead-in phase (Key Issue 1) and adverse events occurring following ≥1 year follow-up (Section 3.2.3.1), were reported in those data tables.

3.2.2. Description and critique of the design of the studies

3.2.2.1. Design of the studies

At the time of the appraisal there was a limited evidence base available for ED. Given the difference in composition between AMT-060 and ED, and the small sample sizes of the Phase I/II and Phase IIb studies, the company and the EAG each considered the principle evidence base for this appraisal to be the HOPE-B study. However, the EAG noted that data from the Phase IIb study (N=3) was used alongside that from HOPE-B in the Shah et al[17] 2022 analysis to predict the durability of the ED effect.

HOPE-B study was a small, single arm prospective cohort study with limited follow-up currently available. While up to five years of follow-up was planned, data in the submission was based on 24-months' of follow-up only (follow-up was a minimum of 24-months in all participants, though the company did not report the average follow-up across participants). The next planned data cut, which would provide 36-months follow-up, was not expected by the company to be available until Q3-Q4 2023. To support assumptions regarding the long-term durability of ED, the company referenced data from the Phase I/II of AMT-060, which measured outcomes at up to five years following treatment (CS p.48). The EAG had concerns about the reliability of these data as used for this purpose. This issue is discussed further in Section 4.2.6.1 and in Key Issue 5.

Given the rare nature of moderately severe and severe haemophilia B, the EAG acknowledged that the small sample size of HOPE-B was to be expected. Nevertheless, it presented challenges for interpreting the clinical efficacy of treatment, particularly given that treatment response appeared to vary across the population (see Section 3.2.3.1). The study was conducted internationally at 33 sites, including 3 sites in the UK, 17 sites in the US and 13 sites in the EU. Clinical advice to the EAG was that international variation in health outcomes for people with haemophilia B was largely due to the poor availability of FIX replacement therapies in low- and middle-income countries.[18] Given the procedures within the study, the EAG did not identify any reason why outcomes could not be generalised to the UK population.

As HOPE-B was a single-arm study, the company compared outcomes following treatment with ED with outcomes assessed during a lead-in phase of 26-weeks. This approach was generally preferred by the EAG as compared to no comparison or a naïve comparison with the findings in other samples; however, variations in care between the lead-in phase and following treatment with ED meant this comparison had limitations (see Section 3.2.2.4). In brief, protocols for the

use of standard care varied between the two time periods, and the EAG also considered that the onset of the COVID-19 pandemic following the lead-in phase would have likely impacted the lifestyles and treatments received by participants. These issues are discussed in Section 3.2.2.3 and Key Issue 2.

Following administration of ED, the company appeared to differentiate between the subsequent 6-months' of follow-up, during which time they stated that FIX levels needed to stabilise following treatment. Throughout the CS, the company generally reported clinical outcomes limited to data collected outside of this period, such as during months 7-18 and 7-24 after treatment only. Where data for the 0-6 month period following treatment were reported, this was not merged with the later follow-up timepoints (e.g. 0-24 months' following treatment). The EAG considered this to be an unusual approach, since outcomes immediately following treatment are just as meaningful to people receiving treatment and to understanding the treatment effect of ED. Excluding this time period also reduced the comparability of ED with comparator treatments in the company's indirect treatment comparison (ITC; Sections 3.3 – 3.4). Where feasible, the EAG sought to identify clinical data for the 0-6 month period following treatment from the study CSR, though this was rarely possible. On the basis of the evidence available, the EAG considered it plausible that bleed rates would be higher during months 0-6 after ED as compared to subsequent time periods.

3.2.2.2. Population

Study eligibility criteria

Adult males with severe or moderately severe haemophilia B (as indicated by ≤2% of normal circulating FIX) who were receiving continuous routine prophylactic FIX therapy and without a history of FIX inhibitors were eligible for inclusion. Inclusion criteria also specified that participants who showed high compliance with outcome measures during the lead-in phase were included. A number of exclusion criteria that may be relevant to evaluating the efficacy and safety of ED were also specified, and are summarised in Table 7. The EAG made the following observations:

- The study targeted people with severe and moderately severe haemophilia B, which corresponded with the anticipated licence for ED in England. Clinical advice to the EAG was that this was the most appropriate population since the expected benefit of

treatment may not be meaningful for those with mild or moderate disease who have a lower frequency of spontaneous bleeds.

- Clinical advice to the EAG was that the restriction of the clinical studies to male participants was acceptable. As haemophilia B is a recessive disorder linked to the X chromosome, females almost always have one healthy copy of the gene. Where women have haemophilia, they typically have mild or moderate disease. The EAG was advised that the study evidence could be generalised to any females who met the eligibility criteria.
- Clinical advice to the EAG was that the exclusion of people with a history of inhibitors to FIX was appropriate and unlikely to exclude many people, since the presence of inhibitors in haemophilia B was rare (estimated to be 1% - 3%). The EAG was advised that if a person were to develop inhibitors, this would be identified during childhood, before ED would be considered.
- The exclusion of people with active HIV, hepatitis B or C may affect the population of people with haemophilia B affected by contaminated blood products during the 1970s-1980s, some of whom were children at the time. While many of those affected have now died and the EAG assumed most others have received treatments to manage the conditions, this may still affect an unknown minority of people in the UK with haemophilia B. Clinical advice to the EAG was that this was likely to be a relatively small number of patients in 2023.
- The conditional marketing authorisation (CMA) for ED awarded by the European Medicines Agency (EMA) did not preclude the use of ED in those with elevated liver transaminase, however the SmPC for ED notes the potential risks of treatment with ED for liver function. That EAG therefore noted that outcomes related to liver function reported in the study were based on a sample without any pre-existing impairment in liver function, and that these outcomes may not be generalisable to those who would not have been eligible for inclusion in the study. Clinical advice to the EAG stated that liver complications were not more prevalent amongst patients with haemophilia B *per se*, but that around 90% of older patients exposed to contaminated blood products in the 1970s and 80s developed hepatitis C.
- The exclusion of participants who were anticipated to require chronic treatment with steroids was relevant as a minority of participants in the study required corticosteroids to treat injection site reactions and elevated transaminase levels. As treatment with corticosteroids following ED may affect treatment response (see Section 3.2.3.1), the

study findings may not be generalisable to those participants eligible to receive ED in practice but who would have been excluded from the study due to corticosteroid use. The EAG also considered that unexpected use of corticosteroids during the lifetime of the person treated may affect treatment response(Section 3.2.3.2), although clinical advice to the EAG was that in general, corticosteroid use amongst patients with haemophilia B would not be any different from that in the general population.

- In addition, participants were not permitted to have received a previous gene therapy treatment. The EAG understood that receipt of a gene therapy may not automatically prevent people from receiving a subsequent gene therapy, however it expected that this may be a requirement of gene therapy clinical study eligibility or future medical licences. Moreover, people may develop resistance to the vector used to deliver the gene therapy (as was the case following receipt of ED, CS p.128), which may mean that they are unable to receive gene therapies using the same vector or may experience reduced benefit.
- The study eligibility criteria did not exclude people on the basis of pre-existing neutralising antibodies to AAV, which may be present in 30-50% of the general population[19] and may interfere with vector administration. The company assessed levels of AAV antibodies at baseline and considered the impact of this on treatment outcomes in subgroup analyses.

Table 7: Selected participant exclusion criteria from HOPE-B

- ALT >2 times upper normal limit (i.e., upper limit of normal [ULN])
- AST >2 times ULN
- Total bilirubin >2 times ULN (except if caused by Gilbert disease)
- ALP >2 times ULN
- Creatinine >2 times ULN
- Hepatitis B or C infection requiring treatment
- Uncontrolled HIV infection
- Another known coagulation disorder
- Thrombocytopenia
- Known history of allergy to corticosteroids
- Known medical condition that would require chronic administration of steroids
- Known medical condition that may impact the intended transduction of the vector and/or expression of the protein

- Known severe infection or medical disorder that may interfere with tolerance or adherence to the study procedures

Abbreviations: ALT, alanine transaminase; AST, aspartate aminotransferase; HIV, human immunodeficiency virus

A breakdown of why eight people screened for HOPE-B were deemed ineligible for participation was not provided in the CS or other submitted documents. Likewise, a breakdown of the reasons for why a further 13 participants discontinued following screening was not provided: a list of reasons was provided in the CS, though the number of participants discontinuing for each reason was not given (the reasons included ineligible liver test as assessed by FibroScan, concomitant medications and comorbidities). In the CSR[20], the text noted that [redacted] participants discontinued following a positive FibroScan result for liver fibrosis. Overall, given the uncertainty in the reasons for participants not entering the study, the EAG concluded that up to 28% of people (21/75) who were interested in participating in HOPE-B were ineligible due to the study eligibility criteria. At the time of writing, any forthcoming licence and conditions for the use of ED in England was unknown, though if the licence were to match the study eligibility criteria, then a significant minority of people with severe and moderately severe haemophilia B would not be eligible. If more relaxed eligibility criteria were used, this minority population would not have been included in the HOPE-B study and therefore their outcomes may vary. The conditional licence awarded by the EMA specified that treatment would be contraindicated in those with active infections (acute or chronic) and people with known advanced liver fibrosis or cirrhosis (SmPC, p. 3-4), and would not be recommended for us in those with other significant hepatic disorders (p.3).

Comparable population inclusion criteria were used in both of the previous Phase IIb and Phase I/II studies.

Baseline characteristics

Select baseline characteristics for HOPE-B were reported in Table 8 of the CS (p.62-63). The EAG made the following key observations:

- The sample included a broad age range, from 19 – 75 years (mean 41.5 years). The EAG was aware that health outcomes would typically vary between younger and older participants due to younger participants having had earlier access to routine FIX replacement, including long-acting prophylactic therapies, that reduce the risk of joint

damage. As joint damage is irreversible, treatment with ED may be expected to have a lower impact on the broader HRQoL of older people.

- The majority of participants had severe haemophilia (44/54; 81.5%), with only ten participants (18.5%) classified as having moderately severe disease. The EAG struggled to identify recent epidemiological data for the incidence of severe vs. moderately severe disease in the UK, though clinical advice was that this was likely to be a small population and therefore the clinical study sample may be representative. Clinical advice to the EAG was that the anticipated benefits of treatment may be less for those with mild or moderate severe, and so any discrepancy between the study make-up and the target population may be affect the generalisability of study effect estimates.

- The EAG noted that nearly half of all participants were receiving standard half-life prophylactic FIX replacement therapy prior to screening. Clinical advice to the EAG was that at the time of the HOPE-B study, standard half-life products were the most commonly used treatment for prophylaxis though this may not be representative of current practice as more people in the NHS are now receiving the longer acting therapies. Within their practice, the last person switched to a longer acting treatment within the past 12-months. In practice, there may be geographical variation in the availability of treatments, but preferences and lifestyle choices of participants also affects choice of half-life product.

- A sizeable minority (38.9%) of participants exhibited neutralising antibodies to AAV5 at baseline, which were found in the company's subgroup analyses to affect treatment response (Section 3.2.3.1).

Further baseline characteristics were provided in additional documentation,[13] which revealed that:

- Only a small minority of participants were receiving on-demand FIX replacement therapy at baseline compared to prophylactic (CS, p.82), which the EAG understood to be representative of current practice in the NHS.

- [redacted] of participants ([redacted]%) were experiencing 0 – 3 bleeding episodes per year prior to screening, with remaining participants distributed at rates between [redacted] to [redacted] (p.82).

- Of bleeding episodes experienced in the year prior to screening, 55.6% were joint bleeds, 59.3% were spontaneous bleeds, and 37.0% were traumatic bleeds (p.82).

Clinical advice to the EAG was that the annual incidence of bleeding episodes was consistent with that observed in their clinical practice, albeit with the caveat that small patient numbers lead to a lot of uncertainty in percentage estimates.

3.2.2.3. Intervention

ED was intended to be administered as a single intravenous (IV) infusion of 2×10^{13} GC/kg of body weight, which corresponds to 2mL/kg of body weight (CS, p.18). Administration occurred as planned for the vast majority of participants in HOPE-B, though 3/54 (5.6%) of participants required a dose interruption due to an infusion reaction, of whom 1/54 (1.9%) were unable to receive the full dose. The same dose and administration were used in study AMT-061-01. In study AMT-060-01, the treatment administered was AMT-060, an earlier formulation of ED that according to the company used the same protein capsid and cassette deign, but a different amino acid to the Padua FIX variant. Two doses of AMT-060 were evaluated, including the same dose as used in the other studies (2×10^{13} GC/kg) and a higher dose of 5×10^{12} GC.

In the HOPE-B study, participants could continue with their routine FIX replacement on the day of treatment with ED and in the following weeks to ensure that FIX levels were sufficiently high. This allowed time for FIX levels to increase and stabilise following treatment with ED. During follow-up visits, FIX levels were assessed, and the use of continuous routine FIX replacement was withdrawn if participants' FIX levels were >5% of normal activity. Clinical advice to the EAG was that FIX levels of 3-5% of normal activity would be the lowest acceptable level while receiving treatment, and that clinicians would likely treat before FIX levels reached 5% where possible. Moreover, whilst a FIX level of 5% would be adequate for most 'normal' activities, it would be insufficient for higher-risk sport and physical activity, and therefore some people with haemophilia B would seek FIX replacement to maintain a higher FIX level, (or alternatively schedule their dosing to coincide with the high-risk activities to ensure they are at a peak level). The CS stated that while routine continuous FIX replacement was discouraged in those with FIX levels >5% of normal, "further management was based on the Investigator's clinical judgement and subject preference" (CS p.54). The EAG therefore considered it plausible that on-demand FIX replacement therapy was administered to those with FIX levels >5% of normal, though data to confirm that were not provided in the CS, and it was therefore unclear whether any on-demand treatment was comparable with the FIX replacement that would have been

administered in practice. Overall, the EAG considered it plausible that clinicians would administer more FIX replacement in clinical practice than they did in the clinical study, either because of patient preference and/or because they would typically seek to attain a higher FIX threshold than was permitted during the study.

Moreover, the EAG noted that the COVID-19 pandemic began during study follow-up, which drastically limited the daily activities of people in the UK. People with haemophilia B who were receiving immunosuppression or had hepatitis or HIV infections were also advised to shield, whereas others will have experienced various levels of lockdown. The company did not present any data that the EAG could use to determine whether study outcomes may have been affected by the onset of the pandemic, though a market share analysis commissioned by the company [redacted]. In addition, clinicians may have been reluctant to administer FIX replacement unless absolutely necessary due to concerns about the risk of thrombogenic events linked to COVID-19. Finally, due to reduced activities, people may also have been at a lower risk of bleeds. Overall, the EAG considered it plausible that the onset of the pandemic may have resulted in a reduced risk of bleeding and fewer doses of prophylactic FIX being administered to participants than would be normal in clinical practice.

The EAG considered whether participants who responded to ED would nevertheless continue to receive FIX replacement to further ‘top-up’ their FIX levels. While the need to attend once or twice weekly appointments to receive IV FIX was associated with a notable treatment burden for people with haemophilia B, the EAG considered it plausible that some people may choose to receive further IV FIX on a regular or semi-regular basis. The choice to do this may allow people to engage in activities not typically recommended for those with haemophilia B, such as more active sports. As part of Key Issue 2, the EAG has identified the potential benefit of further evidence for the number of people in HOPE-B with circulating FIX levels at different thresholds of normal. With this information, clinicians may be able to advise whether they think a proportion of people in clinical practice would request further FIX replacement therapy in addition to ED.

FIX infusions were also not recommended by the investigators if FIX levels were ≥40% of normal activity (i.e. the threshold for non-haemophilic levels of FIX). Clinical advice to the EAG was that a threshold of 40% was sufficient to protect against bleeding events for most everyday

activities and concurred with the company that higher rates of FIX were associated with thrombosis. The EAG therefore concluded that this requirement was not a concern for the generalisability of the study.

3.2.2.4. Comparator

The comparator to ED was participants' outcomes during the ≥6-month lead-in phase prior to dosing with ED. Prior to the start of the lead-in phase, participants underwent a washout period for their usual FIX replacement therapy: this was 3 days for normal half-life products and 10 days for extended half-life products. The EAG did not consider it to be clear in the CS whether treatment during the lead-in phase was aligned with the care participants were receiving prior to study participation, or whether the lead-in phase was also subject to the same controls over use of FIX replacement as applied following treatment with ED. However, on the basis of a statement in the study CSR[20] ("with standard of care continuous routine FIX prophylaxis", CSR p. 44), the EAG concluded that it was the former: i.e. participants received regular prophylactic treatment as per usual care. As noted in Section 3.2.2.2, nearly half of participants were receiving standard half-life therapy at baseline. The EAG assumed that participants continued to receive their usual FIX replacement treatment during the lead-in phase (as was stated to be the case for following treatment with ED). Clinical advice to the EAG was that whilst probably representative of clinical practice at the time of the HOPE-B study, the use of standard half-life products in the NHS was declining, as the extended half-life products can reduce treatment burden and prolong the treatment effect, thus providing people with improved coverage (though these are significantly more expensive). The company provided a report[1] of the market share of FIX replacement products in the UK from 2020 that suggested that a standard half-life product accounted [redacted]. The EAG therefore considered it possible that treatment outcomes during the lead-in phase could be conservative.

3.2.2.5. Outcomes

A broad number of clinical outcomes were evaluated in the HOPE-B study, including a variety of bleeding outcomes that accounted for different bleed types, outcomes specific to joint health, use of FIX replacement therapies, circulating FIX activity, safety and pharmacokinetic outcomes, health-related quality of life (HRQoL) and various patient-reported outcomes (PROs). The EAG considered that the outcomes included the principle metrics for determining the

efficacy of ED. However, the EAG identified two major concerns with the way outcomes were defined and measured:

- Clinical outcomes (i.e. not safety) typically excluded data measured during the 6-month time period following treatment with ED, which the company described as the period during which time stable FIX levels were established following treatment. The EAG strongly disagreed with this approach. The EAG attempted to identify clinical outcomes that included the initial 6-month time period after treatment, though this was not feasible for some outcomes. It was not clear to what extent this would affect the interpretation of the clinical results, however while ED begins to exert its mechanism of action, the EAG considered it plausible that circulating FIX levels would be lower, people may be at an increased risk of bleeding, and people would be more likely to receive FIX replacement therapy.
- The company reported change from baseline in circulating FIX levels, however did not use data from the lead-in phase in these outcomes. Instead, the company calculated an estimate of FIX activity to represent a comparison as if participants were not receiving any treatment for their condition. To do this, for each participant they imputed a baseline FIX level based on their condition severity (i.e. <1% of normal activity for those with diagnosed severe disease). The EAG considered these analyses to be inconsistent with the decision problem for this appraisal, and that the presentation of these data was potentially misleading. This issue is addressed in Key Issue 1.

In addition, the EAG noted the following minor issue:

- The company reported the proportion of participants with FIX levels <12%. This threshold was specified *a priori* in the study protocol, though no rationale was given for the choice. The Clinical advice to the EAG was that mild haemophilia B was defined as FIX levels between 5% and 40%. The EAG considered that the proportion of participants at different thresholds of disease would be a useful outcome; i.e. the number of people with severe haemophilia who became moderately severe, moderate and mild etc. Clinical advice was that approximate thresholds can be used to guide the minimum FIX levels for safe engagement in certain activities (e.g. certain sports), on the basis that thresholds are understood to represent varying risk of bleeding and the likely impact of the condition on people's lives. This data was identified as potentially useful to reduce uncertainty in Key Issue 2.

3.2.2.6. Critical appraisal of the design of the studies

Full critical appraisal for HOPE-B was reported in the CS Appendix D. The company used an appropriate checklist for considering the potential for bias in the study, though item responses lacked detail and consideration. Standardised critical appraisal checklists are intended to capture the most common types of bias present in the relevant study designs, though they are not intended to be comprehensive, and researchers are expected to consider potential risks of bias that may exist beyond those covered by the tool, or explicitly prompted in signalling questions. This was clearly not done in this appraisal.

Key points noted by the company appraisal included a lack of information about participants who dropped out from the study and a lack of information about the population from which the participants were recruited. It was also reported that it was not possible to assess whether the statistical tests used to assess the main outcomes were appropriate or if the study had sufficient power. The company appraisal did not provide insight into whether the company considered the population to be representative of the target population and usual treatment in the NHS. The appraisal focused on procedures following treatment with ED, without consideration of potential bias during the lead-in phase and in the comparability of the two periods. Moreover, there was no evidence that the company considered the potential for bias to vary across outcomes and (where relevant) subgroups.

The HOPE-B study was a single-arm, open-label study where change in outcomes was based on a historical comparison, and as such this was low-quality evidence[21]. Historical comparisons are always challenging because of the potential for change in factors other than the administration of treatment to influence participant outcomes. In this case, and as addressed in Key Issue 2, study procedures varied between the lead-in phase and following treatment, and the onset of the COVID-19 pandemic may have influenced outcomes. Without a concurrent comparison arm, the true effect of treatment was therefore uncertain.

As an open-label study, outcomes were also subject to performance bias, meaning that the care participants received may have been different because of knowledge that they had received ED. In addition, the assessment of outcomes can be affected by study participation and knowledge of the intervention being received (or in the case of the lead-in phase, not being received). This type of bias particularly affects subjective outcomes, such as diary entries of bleeding events completed by participants and the assessment of adverse events.

Overall, the EAG considered that the treatment effects reported were of a significant magnitude to suggest that they represented a true benefit of ED as compared to standard care with prophylactic FIX replacement. However, the study design and the unexpected start of the COVID-19 pandemic was considered by the EAG to introduce a risk of bias in favour of ED. The true treatment effect of ED may therefore be smaller than shown, and so study findings should be interpreted with caution.

3.2.3. Description and critique of the results of the studies

3.2.3.1. Clinical effectiveness results

Annualised bleeding rate (ABR)

The company reported a series of outcomes to assess bleeding rates in HOPE-B, which are reported in Table 11 of the CS (p.70), and included various types of bleeding (all, joint, spontaneous, traumatic, FIX-treated, new and true) and two types of analysis (unadjusted and adjusted). With the exception of bleeding episodes in people who tested positive for anti-AAV5 NAb (discussed later in this section under Subgroup Analyses, p44), bleeding rates were lower following treatment with ED than they were during the lead-in phase. Rate ratios across bleeding outcomes (i.e. adjusted ABR / lead-in ABR) for the ITT population ranged between 0.13 – 0.36, and 95% Cis around these were generally all within the range that the EAG considered a meaningful average reduction. The number of people experiencing bleeds (any bleed) also reduced following treatment with ED, from 74.1% during the lead in phase to 37% during months 7-18 and 50% for months 7-24.

Table 8: Annualised bleeding rates in HOPE-B

	All bleeds	**Joint bleeds**	**Spontaneous**	**FIX-treated**
Lead-in phase ABR	4.19 (3.22, 5.45)	2.35 (1.74, 3.16)*	1.52 (1.01, 2.30)	3.65 (2.82, 4.74)*
Total bleeds	136	77	50	118
People who had bleeds	40 (3.4/pp); 74.1%	32 (2.4/pp); 59%	24 (2.1/pp); 44.4%	37
7-18 months ABR	1.51 (0.81, 2.82)* Δ-64% (95%CI 36, 80)	0.51 (0.23, 1.12)	0.44 (0.17, 1.12)	0.84 (0.41, 1.73)* Δ77% (95%CI 54, 88)
Total bleeds	54	19	14	30

	All bleeds	Joint bleeds	Spontaneous	FIX-treated
People who had bleeds	20 (2.7/pp); 37%	11 (1.7); 20.3%	9 (1.6/pp); 16.7%	15
7-24 months ABR	1.51 (0.83, 2.76)* Δ-64% (95%CI 37, 79)	0.46 (0.24, 0.89)*	0.38 (0.16, 0.89)	0.99 (0.48, 2.03)*
Total bleeds	74	26	18	43
People who had bleeds	27 (2.7/pp); 50%	15 (1.7/pp); 27.8%	11 (1.6); 20.4%	19

*adjusted ABR: generalized estimating equations negative binomial regression model accounting for the paired design of the study with an offset parameter to account for the differential collection periods. Treatment period was included as a categorical covariate.

Source: Company submission Document B; HOPE-B CSR[13]

FIX levels

During the lead-in phase, the company reported that 79.6% of participants had FIX levels <12% of normal, which changed to 7.8% after 3 months, 8.0% at 12 months, 6% at 18 months, and 10% following 24-months of treatment.

The company also reported change in mean FIX levels; however, the EAG noted that the baseline FIX levels reported in the CS (i.e. Table 12, p.73) and used to calculate change from baseline represented an estimate of FIX levels as if participants were not receiving FIX replacement therapy, rather than FIX levels assessed during the lead-in phase. This estimate was based on the conventional FIX threshhold for each of the participants' diagnosed disease severity, i.e. a participant with severe disease (FIX levels <1%) was awarded a baseline FIX level of 1%. FIX levels during the lead-in phase were not reported in the CS and the EAG was unable to identify these during its appraisal. The EAG considered the company's approach to be unusual and one that could be potentially misleading.

Clinical advice to the EAG was that the target with prophylactic FIX replacement therapy was to keep trough (i.e. minimum) levels of circulating FIX between 3 – 5% of normal, though following each treatment FIX levels may initially be much greater. Studies evaluating the efficacy of prophylactic FIX replcement therapies reported FIX levels in the normal range following treatment, which then returned to the trough level over hours or days (depending on whether the treatment is a short- or extended half-life product). [22, 23, 24, 25, 26] Without knowing true baseline FIX levels during the lead in phase, the EAG was only able to comment on absolute FIX levels

after treatment with ED, without comment on whether the level of FIX after treatment was meaningfully different from treatment with prophylactic FIX replacement.

On the basis of the data provided, the EAG was able to make the following observations:

- Mean FIX levels reported following administration of ED were >35%, which was above the threshold for mild haemophilia B.
- FIX levels varied widely across participants, ranging from a minimum of 4.5%, which was within the threshold for moderate disease severity, to above 100% of normal.
- FIX levels appeared to remain consistent between 6-months and 24-months following treatment.

FIX levels in AMT-061-01 appeared relatively stable until final follow-up (N=3). In the study of AMT-060-01, FIX levels also appeared [redacted] (see Figure 1). Overall, acknolwedging the small samples, the EAG considered these data to support a plausible maintenance of treatment effect up to five years' following treatment.

Figure 1: FIX Levels up to 5-years following treatment with AMT-060

Source: Study CSR [16]. Data is from participants in Cohort B.

Consumption of FIX replacement therapy

During the lead-in phase, 100% of participants in HOPE-B were receiving FIX replacements at a mean of 44.1 infusions per participant. Throughout the post-treatment period, including the 0-6 months immediately following treatment with ED, FIX replacement reduced and stayed reasonably stable to 24 months. By the time of the final follow-up (months 19-24), 24.5% of participants were receiving FIX replacement, each receiving a mean of 3.2 infusions. The EAG considered that this reduction represented a potential major reduction in healthcare resource use and treatment burden. However, as discussed in Section 3.2.2.3, the EAG considered it plausible that rates of FIX replacement following ED would be somewhat higher in practice.

Joint health

There was a very small, statistically significant improvement in Haemophilia Joint Health Score (HJHS) following treatment with ED. The EAG was unable to identify a minimally clinical importance difference (MCID) for the HJHS, though a LS mean change of [redacted] on a scale of 0-124 may be unlikely to demonstrate a major change in joint health. The EAG understood that joint damage occurs following years of joint bleeds, and that this damage would not be reversible. Within the short timeframe of the available data from HOPE-B, the EAG therefore did not consider it surprising that there was no clear difference in joint health as measured by the HJHS. If ED was found to lead to improvements in joint health, this may be evident at the latest follow-up timepoints of the study not yet collected, though may be better represented by long-term follow-up data from the HOPE-B study in comparison with naturalistic studies of joint health in people with the target condition. The company did not report data for the prevalence and resolution of target joints in the CS, and the tables containing these data were not supplied with the HOPE-B CSR. From the text in the amended CSR provided[20], it appeared as if [redacted], though the wording was somewhat unclear.

Health-related quality of life and funcion

A numerical benefit in EQ-5D-5L scores was reported at the 24-month follow-up, though the difference in score was under the threshold considered to be a meaningful change in HRQoL for people with haemophilia[27].

The EAG was unable to identify an established MCID threshold for the Haem-A-QoL, which would determine what change in scores would be clinically meaningful for people with haemophilia B. One paper[28] reported a MCID threshold of 7 points for the total score and 10 points for two domains (physical health sports and leisure). Using these thresholds for those subscales and an arbitrary threshold of 7 points for the other domains, a benefit of ED was demonstrated for the 'feelings' domain at 12- and 24-months, the 'treatment' domain at 12- and 24-months, and the 'dealing with haemophilia' domain at 24-months (though the latter was not statistically significant). These domains would suggest that study participants felt less emotional burden from their haemophilia, had reduced treatment burden and may also feel more able to manage their condition. The company stated that a statistically significant change in scores was noted for the 'work and school' and 'future' domains. The EAG was unclear if these changes were clinically meaningful, but if so, it would suggest that those treated with ED felt more able to go to school/work, and were less concerned about the impact of haemophilia on their future lives. There was no change in total Haem-A-QoL scores or in the other domains: participants' physical health, engagement in sports/leisure, view of themselves (including impact of the disease on their current lives), family planning, or personal relationships.

There were no differences in scores on the WPAI (work productivity), BPI (pain), and HAL (functional ability) following treatment with ED.

3.2.3.2. Subgroup analyses

Results of the company's planned subgroup analyses for ABR were shown in Figure 13 of the CS (p.86). Across subgroups, ED was associated with a benefit for ABR as compared to the lead-in phase with the exception of a subgroup of participants aged ≥60 years, in whom ABRs were shown to increase following treatment with ED. The company did not discuss the potential interpretation of this finding except to note the small sample size of this group (N=8). The EAG acknowledged that the small sample for the subgroup meant that there was uncertainty in the finding, as evidenced by the wide 95%Cis around the effect, though noted that the effect was large (RR 1.90, 95%CI 0.38, 9.57). Conversely however, absolute ABR rates reported for older participants reported in the study CSR[20] appeared to show an overall reduction in bleeding between the lead-in phase and follow-up after treatment with ED, and subgroup analyses for older participants reported in the study CSR showed that FIX levels increased and FIX replacement reduced following treatment with ED. The EAG therefore considered it possible that the increase in ABR shown in Figure 13 of the CS for participants aged ≥60 years could be a data inputting error. Nevertheless, as there also appeared to be some numerical difference in

the size of effect for ABR between age groups generally, this may suggest that the age of participants affects treatment outcome. Given the uncertainty surrounding the effect for older participants, and the pattern of effect across groups which could not be plausibly explained, the EAG did not feel able to conclude about the presence or lack of a difference in effect according to age.

ED showed a beneficial effect for ABR compared to the lead-in phase for participants both with and without neutralising antibodies to AAV, though the effect for those with antibodies was somewhat smaller in magnitude. There was also a reduced effect for those with ≥grade 2 liver steatosis (i.e. moderate or severe). In response to a request from the EMA, the company submitted a subgroup analysis comparing ABR between those who did and did not receive corticosteroid treatment due to elevated transaminase[29]. The results were comparable across groups. However, those who received corticosteroid treatment during the study follow-up reported lower mean FIX levels: 14.30 (SD 7.65) at 24-months compared to 40.12 (SD 17.55)[29].

3.2.3.3. Safety

Administration of ED

Six people in HOPE-B (6/54, 11.1%) experienced an infusion reaction to ED, of whom three (3/54, 5.6%) required a dose interruption. One participant (1/54, 1.9%) did not receive the full dose of ED due to the reaction (10% of dose received only).

Deaths

There was one death (1/54, 1.9%) following administration of ED, which the company stated was due to a bacterial urinary infection followed by cardiogenic shock. The study investigator did not consider this death to be related to treatment with ED, and the EMA assessment was that there was no evidence to refute this conclusion[29]. There was one death in AMT-060-01 that occurred following the end of the five-year follow-up. This death was also considered by the study investigator to be unrelated to treatment with ED.

Serious adverse events

During the lead-in phase, four participants (4/54, 7.4%) experienced a serious adverse event compared to fourteen participants (14/54, 25.9%) following treatment with ED. A total of 17 serious adverse events occurred during the follow-up period, compared to five in the lead-in phase. Despite the increase in events, the company stated in the CS that none of these events

were related to treatment. The CS did not list the types of severe event experienced by participants, though the study CSR[13] provided more detail. Of the events reports, , [redacted] were considered severe and [redacted] were considered mild or moderate. Events included two cases of blood loss anaemia, hepatocellular carcinoma, acute myocardial infarction, COVID-19, jaw fracture, haemophilic arthropathy, cardiogenic shock, upper gastrointestinal haemorrhage, muscle haemorrhage and cellulitis. The EMA assessment of these events was that there was no evidence that these events were caused by ED, though it could not rule out the possibility of instances where ED had exacerbated a condition, thus leading to the event.

All adverse events

All study participants experienced at least one adverse event following treatment with ED, compared to 68.5% (37/54) of people during the lead-in phase. Events showing a marked increase included nasopharyngitis, arthralgia, back pain, extremity pain, fatigue, toothache, diarrhoea, ALS and ALT increases, creatinine increase, and headaches.

Treatment-related adverse events

As the potential implications of gene therapies for broader processes in the body were yet unknown, understanding the potential for adverse effects following these types of therapies was more uncertain. Of the 557 adverse events reported following treatment with ED, 93 (16.7%) were considered by the investigator to be related to ED. Treatment-related AEs affected 38/54 (70.4%) of study participants. A full breakdown of the treatment-related adverse events was not provided in the CS and the data table accompanying the study CSR was also not provided. The CS stated that the majority were mild or moderate in severity, with only [redacted] experiencing [redacted]. The most commonly reported treatment-related event was an increase in ALT (experienced by 9/54, 16.7% of participants).

Adverse events of special interest

No participants exhibited raised ALT/AST levels during the lead-in phase. Following treatment with ED, 11/54 (20.4%) and 8/54 (14.8%) participants experienced increases in ALT and AST, respectively. Of these, 8/54 (14.8%) and 5/54 (9.3%) ALT and AST increases were more than twice baseline levels, and of these almost all were considered by the company to be treatment-related. The company reported that only one event was considered to be severe, though based on discussion in the EMA report of ED (p.110, [29]), there was some uncertainty about the severity classifications of AST and ALT increases, with similar increases described as severe,

moderate and mild across different participants. Overall, 9/54 (16.7%) participants received treatment with corticosteroids, including prednisone, prednisolone, and methylprednisolone, for ALT/AST increases, over a mean duration of 79.8 days (SD 26.6; range 51 - 130)[20 29]. The company stated that all participants discontinued corticosteroids between 85 – 170 days following ED, and that all ALT/AST increases were resolved within 3 to 127 days. However, the EAG was unable to resolve that against information in the SmPC for ED where it was reported that "onset of ALT elevations [in the clinical studies] ranged from day 22 to 787 post-dose" (p. 13). All events were described as non-serious and resolved with treatment; however, clinical advice to the EAG was that late-onset increases in ALT/AST were more concerning as these could result in repeat events and further need for immunosuppression therapy, potentially throughout participants' lives.

No serious adverse events related to the use of corticosteroids were reported, though a list of adverse events of other severity levels associated with corticosteroid use were not provided in the CS or identified by the EAG elsewhere.

Eligibility criteria for the study required participants not to exhibit inhibitors to FIX at screening. Following treatment with ED, inhibitors to FIX were not detected in any participants, suggesting that treatment did not result in the development of inhibitors during the study follow-up. Anti-AAV5 NAbs were identified in 38.9% of participants at baseline, and in 100% of participants from week three onwards following treatment with ED.

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

To establish comparisons against ED, the company located four studies including relevant comparators: PROLONG-9FP for Idelvion (relevant n=40); B-LONG for Alprolix (relevant n=63); Paradigm-2 for Refixia (relevant n=29); and NCT00093171 for BeneFIX (relevant n=34). Study-level details are presented in CS Appendix D, Table 22, and bleeding outcomes from these studies are reported in Table 11 of this report.

The included studies differed from HOPE-B in several important ways, principally relating to analysis populations, outcome definitions and background care. First, comparator studies often included different analysis populations. Most notably, B-LONG and Paradigm-2 included significant numbers of patients who would not have been classed as having prior prophylaxis, which significantly limited the number of patients relevant for each group. Thus, the patients available in the most relevant subgroups for B-LONG and Paradigm-2 were 33 and 17,

respectively. These small sample sizes meant that comparisons were more imprecise, and subject to imbalance in subsequent matched adjusted indirect comparisons, than with larger sample sizes.

Second, outcome definitions and time at risk varied between studies. This was especially important with respect to bleeds and was a common issue across all studies. For example, in B-LONG, the time at risk for a bleed was defined with respect to all follow-up, whereas the company noted that HOPE-B limited follow-up time to exclude a period following FIX use. The inclusion of bleeds also differed between studies. In PROLONG-9FP and B-LONG, the validation of individual bleeds used different clinical algorithms than in HOPE-B, where bleeds were investigator assessed. In Paradigm-2, included bleeding events were counted only if they were identified as spontaneous or traumatic, whereas in HOPE-B all bleeding events were counted. The combined impacts of these differences are hard to quantify. While the company presented a range of sensitivity analyses using different definitions of outcomes, these were naïve in nature and thus it was not clear what the impacts would be on a 'target randomised study'.

Third, it is difficult to understand what the standard of care was for the different analyses presented, as the time range of these studies and the range of populations included may have meant different standards of care were in place. Again, it was difficult to quantify the totality of these impacts on estimates of effectiveness.

The EAG identified a further study, NCT01335061 (Kavakli et al., 2015), that was excluded by the company in their SLR. The reason for exclusion given by the company was that the population was not relevant, but the EAG considered the population eligibility criteria to be comparable with other included studies and the inclusion criteria for review question 2 outlined in the CS appendix D1.1.1.

3.4. Critique of the indirect comparison and/or multiple treatment comparison

The methods used for the indirect comparison depended on the study used. As a rule, because all comparisons were non-randomised, a range of matching and adjusting approaches were taken. The EAG noted that while the statistical methods used to undertake these were of an appropriate standard, the inconsistency between comparisons in variables available for adjustment creates significant variability in the credibility of different analyses used, as does the

differing interpretations of estimates from indirect comparisons. The EAG also regarded that comparisons with BeneFIX were especially uncertain for reasons detailed by the company.

3.4.1. Statistical methods

Methods used to undertake comparisons differed by the availability of participant-level data. Of four indirect comparisons undertaken, only comparisons with Idelvion included participant-level data both for ED and Idelvion. Comparisons with Alprolix, Refixia and BeneFIX relied on summary data for comparator treatments. A strength of the company's approach was the transparency of sensitivity analyses presented, including sensitivity analyses where relevant on different outcome definitions.

3.4.1.1. Comparisons with Idelvion

Methods used to create comparisons with Idelvion relied on first excluding patients from PROLONG-9FP that were 'unique' to that study, i.e. adolescent patients and patients with different thresholds for ALT/AST values, and then estimating inverse probability of treatment weights (IPTW). IPTW relies on considering how patient characteristics 'predict' membership to either treatment group, and then reweighting patients to balance characteristics between groups. In this analysis, patients were reweighted from PROLONG-9FP to be similar to patients in HOPE-B. This is an important point of incommensurability between the different indirect comparisons undertaken.

Because of the availability of patient-level data for both studies, this comparison included the richest set of factors for adjustment, specifically severity of haemophilia B, prior ABR, and age, though prior FIX product class, BMI, weight, ALT/AST thresholds, HIV status, total bilirubin threshold, family with FIX inhibitor antibodies and duration of diagnosed haemophilia B were also considered. Based on estimates provided in Table 5.1 of the report of indirect comparisons[30 31], it was clear that IPTW analysis generated improvements in many, but not all factors; notable differences between groups in BMI, prior FIX product class, ALT/AST thresholds, and HIV status (among other characteristics) remained significantly imbalanced. The optimal combination of covariates for adjustment was selected after ranking covariates and considering trade-off between improvement in balance, effective sample size and overall balance of groups. After estimation of IPTW, differences between groups were estimated using standard regression models.

3.4.1.2. Comparisons with Alprolix, Refixia and BeneFIX

In contrast, methods used to create comparisons with Alprolix, Refixia and BeneFIX relied on a combination of matching populations by inclusion criteria and then adjusting using weights estimated on the HOPE-B patient-level data to create a HOPE-B group similar in mean and distribution of key variables as in the summary data available from comparator studies. This is an important point of incommensurability between these comparisons and the comparisons estimated against Idelvion; in the present comparisons, the interpretation of the effect is the average treatment effect on the comparator population.

Matching and adjusting HOPE-B data to B-LONG, which was the study for Alprolix, relied on first selecting a subset of patients in B-LONG with prior prophylaxis. The EAG regarded that this was appropriate to ensure balance on this key moderator, though this limited the number of additional variables used for comparison; in particular, primary analyses relied on adjusting only for prior ABR. Secondary analyses using the full B-LONG dataset included additional variables, but the EAG regarded that these analyses were not likely to be probative given major differences in populations by prior prophylaxis. Importantly, very few data were available to compare balance of covariates between groups in the primary analysis, which the EAG regards as a significant threat to the credibility of the analysis.

Matching and adjusting HOPE-B data to Paradigm-2, which was the study for Refixia, relied on a subset of patients with prior prophylaxis as primary analysis. As above, the EAG noted that this was appropriate (and relatedly that full-population secondary analyses were not likely to be reliable), but acknowledged that these primary analyses were only inconsistently able to adjust for both prior ABR and prior FIX product class. Correspondingly, it was not possible to ascertain covariate balance in the primary analysis.

Finally, comparison of data from HOPE-B and NCT00093171, which was the source of clinical data for BeneFIX, was limited by a lack of baseline data and ambiguities in outcome definitions, in addition to a lack of precision estimates (i.e. standard errors) for outcomes. Analyses thus required imputation of standard errors. Only age and prior FIX product class were available for adjustment, and it was not possible to ascertain covariate balance. The EAG thus regarded these analyses as especially tenuous. The company also reported pre-post analyses for patients from HOPE-B who were previously on BeneFIX but the EAG regarded these as being even less probative for decision-making than the MAICs given the lack of a comparator group.

3.4.2. Results of indirect comparisons

We report primary analyses for indirect treatment comparisons below, focusing on ABR, AsBR and AjBR as key outcomes, and reporting 'final' multivariable adjusted comparisons (Table 9). Additional analyses (not reported here) were undertaken for percentage with no ABR, with no AsBR, and with no AjBR; consumption of FIX; and for HRQoL estimates. All estimates suggest superiority of etranacogene against comparators in reducing the rate of key outcomes.

Table 9: Primary analyses of indirect comparisons of etranacogene vs key comparators

	ABR	AsBR	AjBR
Idelvion	[redacted]	[redacted]	[redacted]
Alprolix	[redacted]	[redacted]	[redacted]
Refixia	[redacted]	[redacted]	[redacted]
BeneFIX	[redacted]	[redacted]	[redacted]

Note: All estimates are expressed as rate ratios (95% CI).

Of note is that in Document B, the company reports secondary analyses for some comparisons instead of primary analyses; and these secondary analyses often include outcomes for which primary analyses are not available (e.g. AsBR for etranacogene vs Alprolix). No secondary analyses were reported for comparisons with Idelvion. However, as noted above, the EAG regards that the dissimilarity in populations to be too large for these analyses to be meaningful. These are nevertheless provided in summary form below (Table 12). These reflect a similar pattern of effects as for primary analyses above.

Table 10: Secondary analyses of indirect comparisons of etranacogene vs key comparators

	ABR	AsBR	AjBR
Alprolix	[redacted]	[redacted]	[redacted]
Refixia	[redacted]	[redacted]	[redacted]

Note: All estimates are expressed as rate ratios (95% CI).

3.5. Additional work conducted by the EAG

Clinical outcomes of prophylactic FIX replacement therapies as identified by the company and the EAG are shown in Table 11. Note that, for simplicity, the EAG has reported a small selection of effect estimates from the cited studies; a broader range of estimates (e.g. for different groups/regimens) were reported in CS appendix D and the cited publications. In some cases,

alternative estimates were used by the company in their ITC, which may have been as part of an effort to select populations that were most comparable. All studies were conducted with people with moderately severe and severe haemophilia B.

Table 11: Annualised bleeding rates following treatment with prophylactic FIX replacement vs etranacogene dezaparvovec

Study and product	ABR	ASBR	AJBR	Number of participants with bleeds requiring treatment (follow-up)
PROLONG-9FP[22] Extended-action FIX replacement; 7-day regimen N=40	1.58 (95%CI 1.02, 2.44)*	0.65 (95%CI 0.37, 1.13)*	NR	75% (26 weeks)
B-LONG[23] Short-acting FIX replacement; 7-day regimen N=63	3.12 (95%CI 2.46, 3.95)*	Median 1.0 (IQR 0.0, 2.22)	Median 1.1 (IQR 0.0, 4.0)	77% (median 12 months)
Paradigm-2[24] Extended-action FIX replacement; 7-day regimen N=29	2.51 (95% CI 1.42, 4.43)*	1.22 (0.48, 3.10)	NR	55% (12 months)
NCT00093171[25] Short-acting FIX replacement; 1->3 times weekly N=17	Mean 3.11 (SD 3.76)	Mean 0.72 (SD NR)	NR	64.7% (median 32 weeks)
NCT01335061[26] Short-acting FIX replacement; 7-day regimen N=25	Median 2.0 (range 0, 13.8)	Median 1.0 (range 0, 13.8)	Median 0.0 (range 0, 9.8	64.0% (12 months)
HOPE-B ED[13] N=52	1.51 (0.83, 2.76)*	0.38 (0.16, 0.89)	0.46 (0.24, 0.89)*	50% (7-24 months)

Abbreviations: ABR, annualised bleed rate; AJBR, annualised joint bleed rate; ASBR, annualised spontaneous bleed rate; CI, confidence interval; ED, etranacogene dezaparvovec; FIX, factor IX; IQR, interquartile range; NR, not reported; SD, standard deviation

Note: *estimated rate based on author's choice of statistical model

3.6. Conclusions of the clinical effectiveness section

3.6.1. Evidence quality

There was a small, low-quality evidence base for ED and the EAG had significant concerns about the reliability of the findings. The best available evidence was from the HOPE-B study, though the EAG identified a number of serious risk of bias concerns that could favour ED in the results. A major cause of concern in the HOPE-B study was the potential impact that the onset of the COVID-19 pandemic had on the risk of bleeding in the study, and the EAG considered that forthcoming follow-up data from the study (i.e. once participants daily activities had begun to return to levels comparable with the lead-in phase) may be more reliable. The phase IIb and phase I/II studies of ED were of limited value for decision-making, and the EAG was sceptical about their value in supporting the durability of the ED treatment effect given the difference in the treatment formulation (the phase I/II study) and the small sample size (phase IIb study[32]). The company's approach to the ITC were the best available to them, but the evidence base for the efficacy of prophylactic FIX replacement was limited and variations in the methods used by included studies resulted in unreliable relative effects.

3.6.2. Clinical benefits of ED

The EAG considered that the uncertainty in the findings due to evidence quality concerned the magnitude of the treatment effect, rather than its presence *per se*. Despite the lack of baseline data for FIX levels in the lead-in phase, FIX levels after treatment with ED were at a level considered to offer a meaningful benefit to people with severe and moderately severe haemophilia B. These FIX levels appeared to be stable over the study follow-up and would likely have benefits for people from having safer and more stable FIX activity. Bleeding rates were reduced for those receiving ED as compared to the lead-in phase of HOPE-B and in naïve comparisons with the best available evidence for routine prophylactic FIX replacement. While absolute ABRs were not drastically lower following ED than for its comparators, there appeared to be a major increase in the number of people who were without bleeds. Moreover, while acknowledging the limitations in the data about rates of FIX replacement, the EAG considered it likely that treatment with ED would result in a reduction in FIX replacement treatments than they would have in usual practice. Some minimal benefits in PROs were reported in HOPE-B, and the EAG expected that, with time, these benefits may increase (e.g. as people adjust to the new normal with their condition).

3.6.3. Reduced treatment benefit for some populations

Following treatment with ED, cells in the liver are directed to produce FIX. Subgroup analysis in the HOPE-B study suggested that participants with moderate or severe liver steatosis at baseline had reduced benefit of ED, as did those who required corticosteroid treatment (e.g. to treat elevated transaminase elevations). People with serious liver conditions were ineligible to participate in HOPE-B, though it may be that people affected by less severe liver conditions may also experience reduced benefits. The EAG considered that the onset of liver conditions at any time following treatment with ED could also affect the durability of the treatment response. Subgroup analysis also suggested that there may be reduced benefit for people with pre-existing neutralising antibodies to AAV. Given the immaturity of the evidence base and the broader uncertainties about the mechanisms involved in gene therapy, the EAG considered it plausible that other as yet unknown factors may moderate the treatment effect.

3.6.4. Potential for long-term clinical benefits

Evidence for the durability of the ED treatment effect was limited to the 2-year follow-up of the HOPE-B study since the EAG did not consider the longer follow-up evidence from the Phase I/II and Phase IIb studies to be useful in this regard, nor did it consider the durability model presented by the company (Shah 2022)[17] to be informative given the lack of available data. The potential for gene therapies to deliver long-term, even lifelong, clinical benefits was an area of significant clinical interest. To date, the evidence for long-term effects of gene therapies was lacking across indications, and the EAG understood that the presence of a long-term effect in one gene therapy would not necessarily confer benefit in another. Researches have posited that various factors may influence the potential for long-term gene expression, including the rate of cell turnover, patient demographics, and immune-response[33]. Illnesses experienced by people who have received a gene therapy, and any treatments that they receive, may also affect the durability of a gene therapy treatment response. On the basis of the evidence available and current thinking about gene therapies, the EAG considered it both plausble that ED could have a lifelong effect or that the treatment effect of ED could last only a few years until (for example) liver cell turnover has progressed and/or people experience conditions that affect liver function or the body's immune response. Clinical advice to the EAG was that this was an area of great uncertainty, considered that a 6 – 8 year duration of effect was plausible based on the current evidence. The EAG noted that the uncertainty in this issue was unlikely to be resolved without further data collection, though the EAG explored the potential impact that variation in the durability of treatment response has on the cost effectiveness of ED in Section 6.2.2.

3.6.5. Safety

Overall, ED was not associated with significant safety concerns during the studies; those adverse events reported appeared to be mild or were considered to be unrelated to treatment. An exception to this was the risk of transaminase elevations, which affected a significant minority of study participants. The EMA highlighted inconsistencies in the decisions made by the study investigators about whether these elevations were caused by ED, and the EAG considered it reasonable to assume that treatment with ED did result in an increased risk of ALT/AST increases that require immunosuppressive treatment. The data presented by the company suggested that all these events responded to treatment and were resolved within the first year after ED, however evidence reported by the EMA suggested that this was not the case and further elevations occurred more than one year following treatment. Pending further clarification from the company on this point, this would suggest that people receiving ED are at an increased risk of repeated transaminase elevations that require corticosteroid treatment, and which may therefore have long-term impacts on their health.

4. COST-EFFECTIVENESS

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

The company conducted a SLR to identify existing evidence to support this appraisal, including published cost effectiveness analyses and studies reporting cost, resource use, and HRQoL data. Overall, the methods used by the company were appropriate. A summary of the EAG's assessment is provided in Table 12.

Table 12: Summary of EAG's critique of the methods implemented by the company to identify cost-effectiveness, cost and resource, and HRQoL evidence

Systematic review step	Section of CS in which methods are reported	EAG assessment of robustness of methods
Searches	Appendix D1.1 and G	Searches were well conducted with a variety of keywords and subject headings used in a range of databases. A variety of grey literature sources were also searched. There was some discrepancy in the sources searched in the original searches of August 2021 and then in the update searches of October 2022. One search strategy was used to search for economic, cost and HRQoL evidence simultaneously.
Inclusion criteria	Appendix D1.1.1	The inclusion criteria were appropriately and appeared sufficiently broad to capture all relevant evidence
Screening	Appendix D1.1.2	The methods used were consistent with best practice.
Data extraction	Appendix D1.1.2	The methods used were consistent with best practice.
QA of included studies	NA	Quality assessment was not conducted

Abbreviations: CS, Company Submission; EAG, External Assessment Group; HRQoL, health-related quality of life; NA, not applicable; QA, quality assessment

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

4.2.1. NICE reference case checklist

Table 13: NICE reference case checklist

Attribute	Reference case	EAG comment on company's submission
Perspective on outcomes	All direct health effects, whether for patients or, when relevant, carers	QALYs were used as appropriate and captured the health benefit to patients. Adverse events disutility was corporated

Attribute	Reference case	EAG comment on company's submission
		into the company's model.
Perspective on costs	NHS and PSS	NHS and PSS, as appropriate.
Type of economic evaluation	Cost–utility analysis with fully incremental analysis	The company submitted a cost utility analysis. The company made multiple pairwise comparisons rather than a fully incremental analysis in its deterministic base case, but presented a fully incremental analysis in the PSA. Comparisons were complicated by changing the IV FIX treatment used post-failuire of ED for each scenario.
Time horizon	Long enough to reflect all important differences in costs or outcomes between the technologies being compared	The model incorporated a time horizon of 59 years until the cohort of patients reached an age of 100 years. The EAG considered this to be sufficiently long enough to capture important differences in costs and benefits between the intervention and comparators.
Synthesis of evidence on health effects	Based on systematic review	Bleeding rates and consequently transition probabilities in the economic analysis for ED were estimated from the HOPE-B study, and comparator arms via ITC[30 31].
Measuring and valuing health effects	Health effects should be expressed in QALYs. The EQ-5D is the preferred measure of health-related quality of life in adults.	Outcomes were reported in QALYs as per the reference case.
Source of data for measurement of health-related quality of life	Reported directly by patients and/or carers	Utility associated with ED treatment was taken from EQ-5D-5L data at 24 months from HOPE-B. A lower utility was assigned to IV FIX based on expert opinion. Disutilities for bleed events were taken from US-ICER 2022[34].
Source of preference data for valuation of changes in health-related quality of life	Representative sample of the UK population	The population of the HOPE-B study were generally representative of the target haemophilia B population in the UK..
Equity considerations	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	Women were excluded from the study, though clinical experts to the EAG advised that evidence from the studies could nevertheless be generalised to the small minority of females with severe disease.

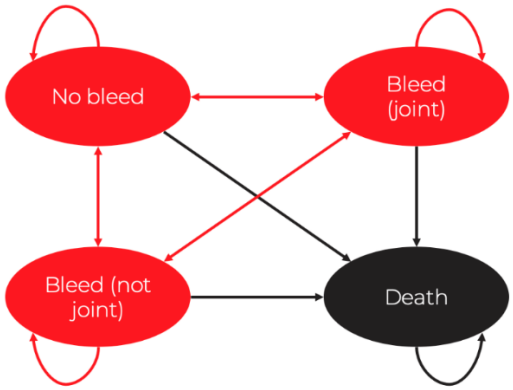
Attribute	Reference case	EAG comment on company's submission
Evidence on resource use and costs	Costs should relate to NHS and PSS resources and should be valued using the prices relevant to the NHS and PSS	Resource use and costs were based on NHS Costs 2019/2020 and PSSRU 2021. It was unclear whether prices were adjusted to a common price year, but the EAG considered this unlikely to be of consequence to the results in this case.
Discounting	The same annual rate for both costs and health effects (currently 3.5%)	Costs and benefits were discounted at 3.5%.

Abbreviations: EQ-5D, EuroQol 5 dimension; HRQoL: health-related quality of life; NHS, National Health Service; PSS, Pseronal Social Services; QALY: quality-adjusted life year; TA: technology appraisal

4.2.2. Model structure

The company presented a cohort-based Markov model whereby patients moved through four health states (Figure 2). Health states were based on bleeding events, and each health state was associated with specific costs and utilitites. The company defined health states as 'no bleed', 'non-joint bleed', 'joint bleed' and 'death'. All patients started from the 'no bleed' health state and the cycle length was seven days.

Figure 2: Model structure



Source: CS Document B, page 147

Rates of bleeding were used to calculate transition probabilities between the health states. Utilities were attached to each of the four health states. In addition, a treatment-specific decrease in health utility was applied to patients receiving IV FIX to account for the inconvenience of regular (i.e. once or twice weekly) IV injections versus a once only administration of ED.

The cycle length of one week was in line with the adminstration of routine prophylactic IV FIX. The EAG's clinical expert also noted that duration of bleeding events and their management was highly unlikely to be more than a week. Hence the EAG believed that the cycle length in the model was appropriate.

Overall the EAG considered that the company's approach to defining health states according to bleeding events was likely to be appropriate.

4.2.3. Population

Modelled baseline characteristics for participants in the HOPE-B study are outlined in Table 14. The EAG noted that the mean age of participants in HOPE-B was 41.5 years old whilst the expected indication for ED was for people over the age of 18 (CS B p.51).

Table 14: Patient baseline characteristics

Patient characteristics	
Male	100% (N=54)
Age mean (SD, min-max), years	41.5 (15.8, 19-75)
Severity of haemophilia B at time of diagnosis, n (%)	
• Severe (Factor IX <1%)	44 (81.5)
• Moderately severe (Factor IX ≥1% and ≤2%)	10 (18.5)
• Positive HIV status, n (%)	3 (5.6)
• Prior hepatitis B infection, n (%)	9 (16.7)
• Prior or ongoing hepatitis C infection, n (%)	31 (57.4)
Pre-screening Factor IX prophylaxis therapy n (%)	
• Extended half-life	31 (57.4)
• Standard half-life	23 (42.6)

Abbreviations: SD, standard deviation; HIV, human immunodeficiency virus

Source: CS Document B, Table 8

4.2.4. Interventions and comparators

The company compared treatment with ED followed by IV FIX on ED failure versus four IV FIX products available in the NHS (BeneFIX [standard half-life] and Alprolix, Idelvion and Refixia [all

extended half-life]). The EAG noted that people who receive ED may receive supplementary IV FIX on demand, and therefore the comparison may be considered ED+IV FIX followed by IV FIX on ED failure as compared with four IV FIXes (see Section 2.4). The decision model excluded on demand FIX as this would be expected to be equal between arms, which the EAG agreed with, albeit noting that the incremental cost may be affected where a different FIX treatment was used as an on-demand therapy than was considered in the comparator arm because of variability in the price of different FIX treatments (e.g. ED+Refixia should be compared with Refixia). For simplicity, we refer to the variations of the ED arm as ED+BeneFIX, ED+Alprolix etc.

In its base case, rather than presenting a comparison of all five strategies, the company presented a series of four pairwise comparisons. In each case, ED was compared with a specific FIX treatment, thus the pairwise comparisons were ED+BeneFIX vs BeneFIX, ED+Alprolix vs Alprolix etc. In each case, the FIX treatment used was also the treatment administered to all patients after ED failure. The EAG understood the logic of this approach, however, multiple pairwise comparisons in the presence of multiple comparators can lead to misleading conclusions. The correct way to analyse this decision problem would have been to compare all options simultaneously against each other, excluding dominated and extended dominated strategies. This is the approach NICE refers to as a 'fully incremental' analysis. The EAG noted that a fully incremental analysis was presented in the company's probabilistic sensitivity analysis (PSA), where the ED arm was defined as ED+Refixia. This was the most expensive IV FIX treatment and thus represented a conservative (least favourable) estimate of the cost-effectiveness of ED.

Clinical experts advised the EAG that a large number of people with haemophilia B were indeed receiving Refixia in the NHS, due to its longer half-life compared with BeneFIX resulting in reduced treatment burden. The EAG's preferred approach was based on a five-way comparison with ED+Refixia as the ED arm, but with some additional sensitivity analysis around the choice of IV FIX on failure of ED (see Section 6).

4.2.5. Perspective, time horizon and discounting

The economic analysis was conducted from an NHS and PSS perspective, as was consistent with the NICE reference case.

The time horizon used in the economic analysis was patient life-time (up to age 100). The EAG considered this appropriate to capture all relevant differences in cost and outcomes between arms. However, the EAG noted that the starting age applied in model was 41.5 years' of age, whilst the anticipated indication was adults aged ≥18 years. In response to clarification, the company provided incremental cost effectiveness results for a cohort of patients aged 18-years with a life-time horizon. The results showed that ED was still dominant over all comparators. However, to investigate the effect of age in combination with other scenarios, the EAG also included it in EAG base case (see Section 6.3).

Costs and benefits were discounted at the NICE reference case rate of 3.5%.

4.2.6. Treatment effectiveness and extrapolation

4.2.6.1. Durability of ED

The decision model included a predicted failure rate of ED (durability function) based on extrapolations of observed data, with failure in the company base case defined as FIX activity <2%. Once ED failed, the model assumed that patients resumed prophylactic treatment with one of the IV FIX treatments. This was modelled by calculation of a weighted average of the ED and relevant IV FIX costs and health state utilities in each cycle.

The company stated that the median durability of ED was [redacted] years (CS B p150) on the basis of modelled projections (Shah 2022a & 2022b[17 35]), although the EAG noted this is a presumed typographical error as the median stated in Shah et al. was [redacted] years.

Shah et al.[17] combined observed data from the HOPE-B (n=52) and AMT-061-01 (n=3) studies (total n = 55) and modelled extrapolations using Bayesian and frequentist linear mixed models. The 52 participants from the HOPE-B data were those who received the full dose of ED, excluding one person who received only 10% of the dose due to a reaction to treatment, and one participant who responded poorly to treatment and continued to require routine prophylactic FIX treatments. Baseline was defined as FIX activity levels at 6 months post-treatment, with failure defined as a predicted FIX activity of <2%. Predicted failure rates were extrapolated to 25.5 years, with a supplementary analysis extending to 60 years (Shah 2022b).

The EAG considered the source data for the extrapolation to be appropriate, drawing on the two studies using the hFIXco-Padua gene variant and excluding AMT-060 (wild-type Factor IX transgene). Whilst the statistical modelling technique was considered reasonable, the EAG was concerned with the low participant numbers available to inform the model and the short follow-

up of the source data: 24 months follow-up data were available for only 6/55 (10.9%) participants in the analysis, and 30 months for 3/55 (5.45%), which were then extrapolated out to 60 years. The EAG also did not consider it appropriate to exclude the participant who did not have a satisfactory response to ED and continued to require prophylactic therapy; while Shah stated that this person had high levels of neutralising antibodies to AAV at baseline, the EAG understood that the presence of such antibodies was common in the general population and people with high levels were not currently expected to be ineligible for treatment with ED. Overall, the EAG considered that there was a high degree of uncertainty in the extrapolation.

There were limited long-term data available to determine the potential durability of gene therapy treatment effects, both within haemophilia and in other conditions, due to their relatively recent development. The EAG understood that expectations of durability would be specific to the treatment and its indication, and that evidence of durability in other gene therapies (even those using a similar vector) may not be indicative. Long-term extrapolations of treatment effect beyond the study follow-up period were therefore highly uncertain. Clinical advice to the EAG acknowledged this limitation but considered a durability of 6-8 years to be plausible on the basis of current thinking.

Within haemophilia, the EAG understood there to be several reasons why gene therapies using an AAV vector may experience reduced durability[33]. Evidence from the HOPE-B study suggested that specific subgroups of people treated with ED may experience a reduced treatment effect and the EAG considered it plausible that they may be more susceptible to reductions in treatment efficacy over time. This included people who received corticosteroids to treat transaminase increases, people who develop AAV antibodies, and those with moderate or severe liver steatosis at baseline(Section 3.2.3.2).[29] It had also been posited that the rate of cell turnover in the areas of the body that receive the treatment, and subsequent illnesses and treatments that interfere with that area of the body or the broader mechanisms of treatment, may lead to reduced efficacy over time[33]. Following ED, cells in the liver becomes responsible for producing FIX, and study participants with liver conditions were either excluded from the study or else showed reduced treatment efficacy. The EAG also understood that the liver was known to have a higher rate of cell turnover compared to other areas of the body. Overall, the EAG understood that there are reasons why the ED treatment effect may not sustain over time, and that further evidence would be needed to demonstrate its durability.

Given their impact on the cost-effectiveness estimates, the EAG explored a number of assumptions around durability, including threshold analyses to identify the minimum duration of effect for ED to be considered cost-effective (sections 6.2.3 and 6.2.10.1).

4.2.6.2. Definition of ED failure

The company base case set a threshold for the re-introduction of prophylactic IV FIX when patients' levels of circulating FIX were ≤2% of normal, which was highlighted by the company as a typical FIX activity level considered for prophylactic treatment (CS B p150). However, clinical advice to the EAG stated that a 2% - 5% FIX activity level would be considered as a 'trough' (i.e. minimum level of FIX activity when people are routinely receiving FIX therapy), and that this may be too low to engage safely in some routine activities (e.g. certain sports). Furthermore, following administration of ED, participants in HOPE-B only discontinued IV FIX once FIX levels were >5% of normal activity (Section 3.2.2.3). Therefore, the EAG considered that the 2% threshold may underestimate the proportion of patients returning to prophylactic IV FIX. Shah et al.[17] calculated the durability function (described in Section 4.2.6.1 above) at a 5% FIX activity level threshold, resulting in a median durability of [redacted] years. The EAG considered a 5% FIX activity level to be a more plausible threshold at which prophylactic IV FIX would be recommenced.

4.2.6.3. Response to ED in the first six months post administration

The EAG noted that participants in HOPE-B continued to receive IV FIX post administration of ED until their FIX levels had stabilised at >5% of normal activity and that the clinical outcomes reported in the CS excluded data measured during the initial 6-month period post-administration, which the company stated was to allow participants' circulating FIX levels to stabilise (see Section 3.2.2.5). In contrast, the company base case assumed 100% durability of ED from the first model cycle in terms of risk of bleeds but included cost of 3 weeks of IV FIX immediately following ED administration.

As there may be greater need for people receiving ED to receive prophylactic FIX treatments in the initial 6-months after treatment, the EAG considered that the company base case was optimistic and so the EAG explored the impact of a longer 'induction period'.

4.2.6.4. Transition probabilities for bleeds

The clinical data to derive transition probabilities for the ED arm were based on those observed in the HOPE-B study, with comparisons with IV FIX based on rate ratios estimated via several

ITCs [30 31] (Section 3.4). Given the limitations of the data (Section 3.4), and the risk of overestimation of the treatment effect of ED (Key Issue 2), the EAG explored the impact on the ICER of reduced treatment efficacy of ED comparative to prophylactic FIX replacement.

4.2.7. Health-related quality of life

The company model assumed that patients receiving ED had a higher utility than those receiving IV FIX. Bleed health states are associated with a disutility penalty and adverse events also incur a disutility. These are discussed in turn below.

4.2.7.1. Treatment-specific health utility

Health utility for patients receiving ED was based on EQ-5D-5L data from HOPE-B at 24 month follow-up. Health profiles were converted to utilities using the Van Hout cross-walk algorithm[36], yielding a utility of [redacted] for patients in the 'no bleed' health state. The EAG considered this an appropriate algorithm.

The EAG noted that the utility estimate was applied from the first model cycle and therefore may have overestimated utility during the first cycles post-administration while patients were still receiving IV FIX.

Health utility for patients receiving IV FIX was defined as the utility for ED less [redacted] (i.e. [redacted]) to represent "patients living a precautionary life, as they fear bleeding events and lack of freedom to enjoy usual activities, as described in the dimensions of EQ-5D" (CS p156). The source of the quoted disutility was expert opinion, described as "conservative and a minimum, but reasonable" (CS p156). The methods for deriving the clinical expert opinion were not reported and it was unclear how the SE was estimated.

The EAG noted that the utility decrement associated with 'some problems with performing usual activities' in the MVH algorithm (NICE's preferred preference weightings for the EQ-5D-3L) was 0.036, rising to 0.363 for 'unable to perform usual activities' (calculated as $2\beta_3 + \beta_8 + \beta_{11}$ as per Equation 1, Dolan 1997[37]). The availability of FIX replacement therapies, particularly regular prophylactic treatment, was associated with a major benefit for survival and HRQoL in people with haemophilia B. However, ED may have further benefits for HRQoL over and above this, for example by reducing treatment burden and benefits for functioning and psychological wellbeing of a higher and more stable circulating FIX level. This implied that an appropriate upper estimate for the disutility associated IV FIX compared with ED would be 0.036 (per the MVH algorithm for a 'some problems performing usual activities'), somewhat below the company's

base case assumption. However, allowing for additional impact above and beyond this from being free of injections whilst ED remains active, the EAG preferred an estimate of 0.042 for the disutility associated with IV FIX compared with ED.

4.2.7.2. Health state utility (bleed events)

In the model, patients experiencing a bleed event experienced a reduction in health state utility, incorporated in the model as the treatment-associated utility less the disutility of a joint or non-joint bleed, as appropriate. Based on the company's expert opinion, the non-joint bleeds lasted up to two days and joint bleeds up to four days. The EAG considered the company's reasoning for how to model the bleeding events disutility was appropriate. The data relevant to disutilities came from US-ICER 2022[34]. Table 15 provides the summary of disutility values of bleeding events.

Table 15: Summary of disutility values of bleeding events

Bleed type	Disutility
Disutility of non-joint bleed per cycle	0.05 (-)
Disutility of joint bleed per cycle	0.16 (-)

Abbreviations: ED, etranacogene dezaparvovec

Source: Table 32 from Document B of CS

4.2.7.3. Adverse events disutility

The company reported disutilities associated with ED and IV FIX treatment, sourced from various literature (CS p154-5). Weekly probabilities for ED were taken from the HOBE-B study and for IV FIX from a study relating to BeneFIX (source not stated), which was then assumed equal across all IV FIXes. The company stated that the impact of AEs from all treatments was "captured in the model via the application of disutility values and estimated AE duration, where necessary" (CS p153), but the time over which patients were at risk of AEs was not stated. The EAG noted that the company's initial model included a life-time duration for adverse events in ED, but that the version submitted post clarification altered this to the first year only with no rationale provided (see Section 6.1).

The EAG considered that the disutility estimates were appropriate but the lack of clarity over the source of IV FIX disutilities increased uncertainty. The EAG explored the impact of altering the assumption over duration of AEs in Section 6.2.9.

4.2.8. Resources and costs

The company conducted a SLR to identify cost and healthcare resource use for ED and its comparators. Model costs were separated into the following types:

- Drug acquisition costs
- Administration costs
- Follow-up cost for etranacogene dezaparvovec
- Monitoring costs
- Bleed-related management costs
- Adverse event costs

Follow-up costs were presented in Table 35 of the CS, with weekly/monthly follow-up sessions assumed to be with a nurse in the hospital and a liver function test carried out twice weekly, presumably at home. The EAG considered there to be uncertainty around whether follow-up sessions would be with a nurse or consultant haematologist. The EAG explored the impact of this in a scenario analysis (see Section 6.2.8).

The company presented the cost per bleed for the intervention and comparators in Table 38 of the CS (note that these costs did not consider any confidential discounts to the NHS for these treatments). This reported that the cost for Refixia (an extended half-life treatment increasingly used in the NHS) was £8,247.89 per bleed.

Section 3.5.4.1 of the company CS highlighted societal costs associated with treatments, including estimates of the workdays lost due to bleeding events (CS Table 42). The company included these costs in a scenario analysis presented in CS Section 3.10.3, concluding that as ED is associated with fewer bleeds, the incremental costs are even lower compared with any of the IV FIXes, thus reinforcing the conclusions of the NHS+PSS (NICE reference case) analysis.

5. COST-EFFECTIVENESS RESULTS

5.1. Company's cost-effectiveness results

5.1.1.1. Base case results

The company's base case comprised pairwise comparison of ED with each of the four IV FIXes (Table 1). In each case, the IV FIX used alongside ED and post ED failure was changed to be the relevant comparator. As stated in Section 4.2.4, where there were more than two treatment strategies being compared a fully incremental analysis was required, taking into account dominance and extended dominance. Changing the IV FIX used alongside ED complicated this but the company did present a fully incremental analysis for its PSA, assuming Refixia was used alongside ED.

The figures reported included the company's PAS discount but list prices for IV FIXes. Results including confidential discounts for all drugs are reported in the confidential appendix to this report. In the company's base case, ED dominated all IV FIXes. The PSA showed that ED had a [redacted] probability of being the most cost-effective of all five comparators at a willingness to pay of £30,000/QALY. The EAG's critique of the PSA is in Section 5.2.2.

Table 16: Company base case results (including PAS price for ED)

	Discounted costs	Discounted QALYs	Incremental discounted costs	Incremental discounted QALYs	Cost per QALY gained
Company deterministic base case					
ED followed by Alprolix	[redacted]	[redacted]	-	-	-
Alprolix	[redacted]	[redacted]	[redacted]	[redacted]	Dominated
ED followed by BeneFIX	[redacted]	[redacted]			
BeneFIX	[redacted]	[redacted]	[redacted]	[redacted]	Dominated
ED followed by Idelvion	[redacted]	[redacted]			
Idelvion	[redacted]	[redacted]	[redacted]	[redacted]	Dominated
ED followed by Refixia	[redacted]	[redacted]			

Refixia	[redacted]	[redacted]	[redacted]	[redacted]	Dominated
Company probabilistic base case					
ED followed by Refixia	[redacted]	[redacted]	-		-
BeneFIX	[redacted]	[redacted]	[redacted]	[redacted]	Dominated
Alprolix	[redacted]	[redacted]	[redacted]	[redacted]	Dominated
Idelvion	[redacted]	[redacted]	[redacted]	[redacted]	Dominated
Refixia	[redacted]	[redacted]	[redacted]	[redacted]	Dominated

Abbreviations: QALYs, quality adjusted life years

5.2. Company's sensitivity analyses

5.2.1. One-way sensitivity analysis (OWSA)

The company conducted pairwise OWSA varying a number of model parameters by +/-20% including annual bleeding rates, treatment costs, utilities, disease monitoring cost, disease follow up cost, disease management cost, disutility of adverse events and non-joint and joint bleeding events. Result were presented as tornado diagrams of NMB rather than ICERs for greater clarity when ICERs are negative (CS p197-200). The results were most sensitive to variation in IV FIX annual bleed rates and subsequent costs, except for Refixia where the two most important parameters were cost of treatment of bleeds and total disease monitoring cost (rather than bleed rates).

The EAG preferred to see five-way comparisons of ED and IV FIXES (i.e. comparing net monetary benefit with each treatment option). However, in all cases the model conclusions were insensitive to changes in the input parameters, and the EAG believed this was true for multi-way comparisons.

5.2.2. Probabilistic sensitivity analysis

Parametric distributions were assigned to all model inputs except for HOPE-B demographics (e.g age), list prices, dosing regimens, durability and mortality. The EAG noted measures of variability were reported as a mix of standard deviations and standard errors (CS B p186-90). In its response to clarification question B19 the company confirmed that all variability measures

were standard errors. The PSA was run for 10,000 simulations. Results are reported in section 5.1.1.1 above.

The company reported the PSA results as a fully incremental analysis, which was more appropriate for decision making than the pairwise analysis presented in the deterministic base case (NICE manual section 4.10.8[38]). The EAG agreed with the company's decision to include the assumption of a 100% market share of Refixia following treatment with ED. Whilst this may not be a full representation of real world practice, clinical advice to the EAG was that individuals were more commonly switching to Refixia due to the need for less frequent treatments.

The EAG considered the parameterisations broadly appropriate, although it was unclear how standard errors and hyperparameters were defined for costs and adverse event incidences, utilities and costs. Given the relatively low incidence of AEs and low influence on the cost-effectiveness results the EAG considered this to be of minor consequence.

The EAG agreed that HOPE-B demographics, list prices and mortality were appropriately entered in the model as constants, and noted that durability was handled in a separate scenario analysis. However, not allowing dosing regimens to vary may have underestimated uncertainty.

In summary, the EAG considered that the PSA was appropriately performed but that assuming fixed dosing regimens may have underestimated uncertainty.

5.2.3. Scenario analyses

Scenario analyses explored a number of assumptions over durability, utilities, model time horizon and societal costs, each of which is discussed below.

5.2.3.1. Durability

The durability scenario analyses compared ED with BeneFIX only, assuming (1) 100% lifetime durability, (2) reintroduction of IV FIX at a 5% FIX activity threshold, (3) 100% durability for five years followed by a 20% drop in durability over five years and (4) 100% durability for 24 months followed by a 20% drop in durability over five years. Results showed that a 5 year durability (followed by 20% decline over the next 5) yielded an ICER of [redacted] (with PAS discount) compared with BeneFIX, substantially in excess of the £20,000 to £30,000 wilingness to pay threshold used by NICE in the STA programme. Whilst useful, the EAG preferred to see simultaneous comparison of ED versus all four IV FIXes, and so conducted additional threshold

analyses to identify the minimum durability for ED to yield an ICER below £20,000 and below £30,000 per QALY (Section 6.2.10.1).

5.2.3.2. Utilities

The company presented a scenario analysis without the treatment-specific health utility difference (see Section 4.2.7.1 and CS p.206), although it was unclear whether the ED utility (of [redacted]) was applied to IV FIX, or the IV FIX utility (of [redacted]) applied to ED. Nevertheless, the results were insensitive to this scenario: ED remained dominant. As expected, the scenario did not affect incremental cost, but reduced the incremental QALYs. The EAG noted presentation of the results as pairwise comparisons rather than fully incremental and explored alternative estimates of the utility difference between ED and IV FIX (see Sections 4.2.7.1 and 6.2.5).

5.2.3.3. Time Horizon

The company explored the impact of 5-, 10- and 20-year time horizons of ED compared with BeneFIX. Over 5 years, the ICER was [redacted] (including PAS discount), reducing to [redacted] at 10 and 20 years.

The EAG considered the shorter time horizons to be inappropriate as they were not considered long enough to capture all differences in cost and outcomes. The EAG noted the company's preferred base case had a life-time horizon.

5.2.3.4. Societal costs

The company did not make an argument for the inclusion of a broader cost perspective including societal costs, therefore the EAG provides no comment on this scenario.

5.3. Model validation and face validity check

The model structure and key inputs were assessed for face validity by the company's clinical experts and reviewed by an external agency, following which minor adjustments were made.

6. EVIDENCE ASSESSMENT GROUP'S ADDITIONAL ANALYSES

The EAG identified a number of limitations within the company's base case and explored the impact of parameter values and assumptions that the EAG believed were more plausible.

This section is organised as follows: Section 6.1 details the impact of errors identified in the EAG's validation of the executable model. Section 6.2 details a series of scenario analyses exploring the robustness of the cost-effectiveness results to specific assumptions and additional uncertainties identified by the EAG. These analyses were conducted within the company corrected base-case analysis. Section 6.3 identifies the EAG's preferred base case based on a combination of the exploratory analyses presented in Section 6.2. A summary of the decision modelling results is then in Section 6.4.

6.1. EAG corrections and adjustments to the company's base case model

Due to a number of mechanical errors in the company model the EAG was unable to fully explore the submitted version of the model. The company supplied a revised model file following clarification, however, the EAG noticed several undocumented changes in the calculations in this model leading to small changes in the base case. These were:

- Reduction in the unit cost of Refixia from £1221.50 to £1211.50
- Increase in the administration cost of ED from £635.55 to £808.64
- Cessation of costs and quality of life impact associated with adverse events from ED after one year

The EAG noted that the change in unit cost of Refixia appeared erroneous: the list price supplied to the EAG by NICE was £1221.50 and so the EAG reverted the price to the original (this adds approximately £3000 to the annual cost of treatment with Refixia).

In its clarification response, the company stated that the cost of administration was omitted from Table 34 of the company submission (CS Table 34, p159) and it presented a revised version (clarification response, Table 5, p15) stating a revised administration cost for ED (£808.64, reported as £808.62). However, the EAG noted that this was also omitted from the original model submission but were added by the company post clarification. The additional costs of £133.92 and £39.17 represented an outpatient procedure and one hour of nursing time

(covering time required to handle biomarker test results), yielding the increased figure of £808.64.

The EAG noted the change to the model assumptions regarding duration of adverse events post administration of ED, but also noted that there was no explanation or justification for this in the company's clarification response. The company submission (CS p153) stated that the impact of AEs from all treatments was "captured in the model via the application of disutility values and estimated AE duration, where necessary", but without further elaboration. The EAG explored the reintroduction of adverse events cost and disutility after the first year of treatment as a scenario (Sections 6.2.5 and 6.2.9).

Therefore, the EAG corrected base case (Table 17) comprised the company base case results reported in the revised model file (including the increased administration cost of ED, duration of AEs associated with ED lasting one year only but with a reverted unit cost for Refixia). The EAG reproduced the pairwise comparisons as per the deterministic company base case, but also presented deterministic and probabilistic fully incremental analyses, with Refixia used alongside ED. Due to resource constraints within the timeline of the appraisal, PSA results are presented based on 1,000 iterations rather than the 10,000 used in the CS.

Table 17: EAG-corrected company base case results (including ED PAS price)

	Discounted costs	**Discounted QALYs**	**Incremental discounted costs**	**Incremental discounted QALYs**	**Cost per QALY gained**
*EAG-corrected company **deterministic** base case*					
ED followed by Alprolix	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Alprolix	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
ED followed by BeneFIX	[redacted]	[redacted]			
BeneFIX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
ED followed by Idelvion	[redacted]	[redacted]			
Idelvion	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
ED followed by Refixia	[redacted]	[redacted]			
Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
*EAG-corrected company **deterministic** base case (fully incremental)*					

ED followed by Refixia	[redacted]	[redacted]	[redacted]		[redacted]
BeneFIX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Alprolix	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Idelvion	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
*EAG-corrected company **probabilistic** base case (run by EAG)*					
ED followed by Refixia	[redacted]	[redacted]			
BeneFIX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Alprolix	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Idelvion	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Abbreviations: QALYs, quality adjusted life years

6.2. Exploratory and sensitivity analyses undertaken by the EAG

The EAG conducted a number of scenario analyses to test the impact of alternative model assumptions on the ICER. These are discussed in the Sections below. The first scenario analysis explored sensitivity to the IV FIX used alongside ED / post ED failure. The EAG's preferred IV FIX was Refixia, so all subsequent analyses show ED+Refixia for the intervention arm.

6.2.1. IV FIX taken alongside ED and post ED failure

In its base case, the company provided four pairwise comparisons rather than a fully incremental analysis, each time changing the IV FIX used alongside ED to the relevant comparator. The EAG preferred a fully incremental analysis but this required a decision on the IV FIX (see Section 4.2.4). Based on discussions with its clinical expert, the EAG preferred Refixia. However, the company submitted market share data of IV FIX for 2020, and so the EAG explored the impact of assuming a weighted average by market share in a scenario analysis. This was included as an additional comparator (ED+mkt share), alongside ED+Refixia and the four IV FIXes in Table 18 and Table 19.

6.2.2. Definition of ED failure

The company base case assumed prophylactic IV FIX was reinstated once FIX activity levels fall below 2%. As discussed in Section 3.2.2.4, this may be lower than would be used in clinical practice. The EAG therefore calculated a scenario where prophylactic IV FIX was resumed at 5% FIX activity level.

6.2.3. Durability of ED

Due to the small sample size and limited follow-up of the available evidence, the durability of ED treatment effect was a key area of uncertainty. The company's base case assumed a median durability of [redacted] years (reported in the CS as [redacted] years), based on a definition of failure of 2% FIX activity levels. At a 5% FIX activity definition of failure, median durability using the company analysis was [redacted] years (see Sections 4.2.6.1 and 4.2.6.2). The EAG's clinical expert acknowledged the uncertainty in this area, though anticipated that a much lower duration of effect of around 6 to 8 years may be plausible. As discussed in Section 3.6.4, the EAG was aware of reasons why treatment with ED may not have lifetime durability, and therefore considered the assumption in the company model to be uncertain. The EAG conducted a threshold analysis calculating the minimum durability required to achieve an ICER below £20,000 and £30,000 per QALY.

6.2.4. Time to steady state

The company base case allowed for a three-week period during which IV FIX was maintained post administration of ED, adjusting costs and bleed rates accordingly. However, the company did not supply data pertaining to bleed rates for months 0-6 post administration. The EAG therefore explored a scenario where all patients maintained IV FIX for six months rather than three weeks.

6.2.5. Utility assumptions

The EAG conducted additional scenario analyses comprising (a) a disutility for IV FIX of 0.042 in place of the company's base case of [redacted], and (b) equal utility associated with treatment with ED and all IV FIXes (set to [redacted]). An additional scenario assuming disutility for adverse events continues beyond the first year is described in Section 6.2.9 below.

At clarification, the company stated that the benefits of the intervention gradually improved to get to its maximum value at month 24. Hence, the EAG calculated an additional scenario (c)

setting utility equal to [redacted] for the first 24 months[redacted]. This approximated a gradual improvement in utility over the first two years post administration.

6.2.6. Estimation of transition probabilities for bleeds

Given the uncertainty in the company's ITCs (Section 3.4) and the risk of overestimation of the treatment effect of ED (Key Issue 2), the EAG conducted two scenario analyses assuming equal probabilities of bleeds across all five treatments, one ('low bleed rate scenario') setting ABR and AjBR bleed rates to the treatment with the lowest ABR (ED) and one ('moderate bleed rate scenario') setting rates to those of Refixia.

To explore the impact of a gradual increase in ED effectiveness, the EAG also presented a scenario analysis with a gradual reduction in bleed rates from that associated with Refixia to that associated with ED over a period of 24 months. Note that this scenario overlapped with the 'time to steady state' scenario (Section 6.2.4) which explored the impact on cost and bleed rates of continuation of IV FIX for a period of six months. This scenario assumed the company base case continuation of prophylactic IV FIX (3 weeks) and focussed on the impact of assuming a gradual 24 month time period for bleed rates to reach those estimated for ED. The cost and QALY impact of bleeds were as per the company base case (including IV FIX).

6.2.7. Age at administration of the intervention

The company base case assumed an age of 41.5 years at ED administration, whilst it was anticipated that the licence would specify a minimum age of 18 years. The EAG was aware of qualitative evidence that suggested that the decision to receive a gene therapy was complex, and that people weigh up a number of considerations before taking a decision to receive a treatment[39 40]. The decision to receive a gene therapy may also be influenced by the extent to which treatment precludes any future gene therapy treatment; in which case, people may choose to wait until evidence was available for several gene therapies (given that there are more gene therapies for haemophilia B in the pipeline) before making a choice. Those who find their disease difficult to manage with prophylactic FIX replacement (for example, where their lifestyle means that regular treatments are challenging) may be more likely to opt for a gene therapy earlier in their lives. The EAG therefore considered it plausible that some people may choose to receive treatment at aged 18 years while others may wait several years before deciding to do so. In its clarification response (question B22), the company provided an analysis with age at administration of 18. This did not affect the conclusions of the model but increased

the magnitude of costs and QALYs in each arm. The EAG repeated this with its corrected base case.

6.2.8. Follow-up visit with haematologist rather than nurse

The company base case assumed follow-up care post administration of ED would be provided by a nurse. The EAG explored a scenario where this follow-up care was provided by a haematologist.

6.2.9. Adverse Events continue whilst ED durability persists

The EAG explored a scenario that reverts to the company base case assuming adverse events continued (and so imposed costs and disutility penalties) whilst ED durability continued.

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

The EAG made the changes described in Sections 6.2.1 to 6.2.9. Each change was made individually. The results of the EAG's exploratory analyses are provided in Table 18 (deterministic) and

Table 19 (probabilistic). In each case, to facilitate fully incremental analysis, the comparators are listed in order of increasing cost, which may lead to a change in the ordering of interventions listed. The highest NMB at £20,000 and £30,000 thresholds are highlighted in bold. Deterministic and probabilistic results of the durability threshold analysis are presented in the text below

Table 19. Results presented here include PAS prices for ED and list prices for all IV FIXes. Results including confidential discounts for all treatments are reported in the confidential appendix to this report.

Table 18: Deterministic EAG scenario analyses

Scenario	Section in EAG report	Comparators	Costs	QALYs	ICER	NMB @ £20k	NMB @ £30k
EAG corrected company base case	6.1	ED+mkt share	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		ED+Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		BeneFIX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Alprolix	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Idelvion	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Scenario	Section in EAG report	Comparators	Costs	QALYs	ICER	NMB @ £20k	NMB @ £30k
		Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
5% FIX activity definition of failure	6.2.2	ED+Refixia	[redacted]	[redacted]		[redacted]	[redacted]
		BeneFIX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Alprolix	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Idelvion	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
6 month time to steady state	6.2.4	ED+Refixia	[redacted]	[redacted]		[redacted]	[redacted]
		BeneFIX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Alprolix	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Idelvion	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Utility assumptions							
a. Disutility of IV FIX treatment of 0.042	6.2.5	ED+Refixia	[redacted]	[redacted]		[redacted]	[redacted]
		BeneFIX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Alprolix	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Idelvion	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
b. Equal utility in all arms	6.2.5	ED+Refixia	[redacted]	[redacted]		[redacted]	[redacted]
		BeneFIX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Alprolix	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Idelvion	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
c. ED utility 0.815 for the first 24 months	6.2.5	ED+Refixia	[redacted]	[redacted]		[redacted]	[redacted]
		BeneFIX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Alprolix	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Idelvion	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Transition probabilities for bleed							
a. Low bleed rates scenario	6.2.6	ED+Refixia	[redacted]	[redacted]		[redacted]	[redacted]
		BeneFIX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Alprolix	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Idelvion	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
b. Moderate bleed rate scenario	6.2.6	ED+Refixia	[redacted]	[redacted]		[redacted]	[redacted]
		BeneFIX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Alprolix	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Scenario	Section in EAG report	Comparators	Costs	QALYs	ICER	NMB @ £20k	NMB @ £30k
		Idelvion	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
c. Gradual improvement with ED over 24m	6.2.6	ED+Refixia	[redacted]	[redacted]		[redacted]	[redacted]
		BeneFIX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Alprolix	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Idelvion	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
18 at admin	6.2.7	ED+Refixia	[redacted]	[redacted]		[redacted]	[redacted]
		BeneFIX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Alprolix	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Idelvion	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Follow up visit with haematologist	6.2.8	ED+Refixia	[redacted]	[redacted]		[redacted]	[redacted]
		BeneFIX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Alprolix	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Idelvion	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Adding AE cost and disutility to ED after first year	6.2.9	ED+Refixia	[redacted]	[redacted]		[redacted]	[redacted]
		BeneFIX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Alprolix	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Idelvion	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Abbreviations: AE, adverse events; EAG, External Assessment Group; ED, etranacogene dezaparvovec; FIX, factor IX; ICER, incremental cost-effectiveness ratio; mkt, market; NMB, net monetary benefit; QALY, quality adjusted life year

Table 19: Probabilistic EAG scenario analyses

Scenario	Section in EAG report	Comparators	Costs	QALYs	ICER	NMB @ £20k	NMB @ £30k
EAG corrected company base case	6.1	ED+mkt share	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		ED+Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		BeneFIX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Alprolix	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Idelvion	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Scenario	Section in EAG report	Comparators	Costs	QALYs	ICER	NMB @ £20k	NMB @ £30k
		Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
5% FIX activity definition of failure	6.2.2	ED+Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		BeneFIX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Alprolix	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Idelvion	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
6 month time to steady state	6.2.4	ED+Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		BeneFIX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Alprolix	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Idelvion	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Utility assumptions							
a. Disutility of IV FIX treatment of 0.042	6.2.5	ED+Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		BeneFIX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Alprolix	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Idelvion	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
b. Equal utility in all arms	6.2.5	ED+Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		BeneFIX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Alprolix	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Idelvion	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
c. ED utility 0.815 for the first 24 months	6.2.5	ED+Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		BeneFIX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Alprolix	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Idelvion	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Transition probabilities for bleed							
a. Low bleed rates scenario	6.2.6	ED+Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		BeneFIX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Alprolix	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Idelvion	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
b. Moderate bleed rate scenario	6.2.6	ED+Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		BeneFIX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Alprolix	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Scenario	Section in EAG report	Comparators	Costs	QALYs	ICER	NMB @ £20k	NMB @ £30k
		Idelvion	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
c. Gradual improvement with ED over 24m	6.2.6	ED+Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		BeneFIX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Alprolix	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Idelvion	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
18 at admin	6.2.7	ED+Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		BeneFIX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Alprolix	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Idelvion	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Follow up visit with haematologist	6.2.8	ED+Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		BeneFIX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Alprolix	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Idelvion	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Adding AE cost and disutility to ED after first year	6.2.9	ED+Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		BeneFIX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Alprolix	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Idelvion	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Abbreviations: AE, adverse events; EAG, External Assessment Group; ED, etranacogene dezaparvovec; FIX, factor IX; ICER, incremental cost-effectiveness ratio; mkt, market; NMB, net monetary benefit; QALY, quality adjusted life year

6.2.10.1. Durability threshold analysis results

To perform the threshold analysis, rather than adjusting the durability model used in the company's base case (Shah et al. 2022a & 2022b[17 35]), the EAG assumed a simple 'cliff edge' whereby durability was assumed to persist at 100% (i.e. where no ED patients require prophylactic IV FIX) until n years had elapsed, after which durability dropped to 0% (where all ED patients require prophylactic IV FIX). This represented an optimistic scenario for ED as a gradual decline over a number of years was considered to be more plausible, starting within a few years post administration (as per the Shah et al. extrapolation in the company's base case).

The identified minimum durability should therefore be seen as absolute minima required for ED to be cost-effective. Durability that exceeded this length of time would be regarded as increasing confidence that ED was cost-effective. Full results are reported in Appendix A in this report.

Figure 3 and Figure 4 below show that under the company's base case assumptions, ED durability needs to be maintained for a minimum of 17 to 18 years to yield an ICER below £20,000 and £30,000 (the results are mostly insensitive to varying the WTP threshold between £20,000 and £30,000).

Figure 3: NMB as a function of ED durability (EAG corrected company base case, NMB at £20,000 / QALY)

Figure shows net monetary benefit of each of the five comparators as a function of durability of ED. The vertical line identifies the point where ED yields the highest net monetary benefit at a WTP threshold of £20,000 per QALY.

Figure 4: NMB as a function of ED durability (EAG corrected company base case NMB at £30,000 / QALY)

Figure shows net monetary benefit of each of the five comparators as a function of durability of ED. The vertical line identifies the point where ED yields the highest net monetary benefit at a WTP threshold of £30,000 per QALY.

6.3. EAG's preferred assumptions

The EAG preferred deterministic and probabilistic base case ICERs are provided in Table 20 and Table 21. Incremental costs including comparator PAS prices are provided in the confidential appendix to this report.

Table 20: EAG's deterministic preferred model assumptions

Preferred assumption	Section in EAG report	Comparators	Costs	QALYs	ICERs	ICER change from base case	NMB @ £20k	NMB @ £30k
EAG corrected company base case (excl. ED+mkt share)	6.1	ED+Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		BeneFIX	[redacted]	[redacted]	[redacted]		[redacted]	[redacted]
		Alprolix	[redacted]	[redacted]	[redacted]		[redacted]	[redacted]
		Idelvion	[redacted]	[redacted]	[redacted]		[redacted]	[redacted]
		Refixia	[redacted]	[redacted]	[redacted]		[redacted]	[redacted]
EAG preferred base case assumptions								
5% FIX activity definition of failure	6.2.2	ED+Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		BeneFIX	[redacted]	[redacted]	[redacted]		[redacted]	[redacted]
		Alprolix	[redacted]	[redacted]	[redacted]		[redacted]	[redacted]
		Idelvion	[redacted]	[redacted]	[redacted]		[redacted]	[redacted]
		Refixia	[redacted]	[redacted]	[redacted]		[redacted]	[redacted]
6 month time to steady state	6.2.4	ED+Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		BeneFIX	[redacted]	[redacted]	[redacted]		[redacted]	[redacted]
		Alprolix	[redacted]	[redacted]	[redacted]		[redacted]	[redacted]
		Idelvion	[redacted]	[redacted]	[redacted]		[redacted]	[redacted]
		Refixia	[redacted]	[redacted]	[redacted]		[redacted]	[redacted]
Disutility of IV FIX	6.2.5	ED+Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		BeneFIX	[redacted]	[redacted]	[redacted]		[redacted]	[redacted]

treatment of 0.042		Alprolix	[redacted]					
		Idelvion	[redacted]					
		Refixia	[redacted]					
Adding AE cost and disutility to ED after first year	6.2.9	ED+Refixia	[redacted]					
		BeneFIX	[redacted]					
		Alprolix	[redacted]					
		Idelvion	[redacted]					
		Refixia	[redacted]					
Cumulative		ED+Refixia	[redacted]					
		BeneFIX	[redacted]					
		Alprolix	[redacted]					
		Idelvion	[redacted]					
		Refixia	[redacted]					

Abbreviations: AE, adverse events; EAG, External Assessment Group; ED, etranacogene dezaparvovec; ICER, incremental cost-effectiveness ratio; mkt, market; NMB, net monetary benefit; QALY, quality adjusted life year

Table 21: EAG's probabilistic preferred model assumptions

Preferred assumption	Section in EAG report	Comparators	Costs	QALYs	ICERs	ICER change from base case	NMB @ £20k	NMB @ £30k
EAG corrected company base case (excl. ED+mkt share)	6.1	ED+Refixia	[redacted]					
		BeneFIX	[redacted]					
		Alprolix	[redacted]					
		Idelvion	[redacted]					
		Refixia	[redacted]					
EAG preferred base case assumptions								
5% FIX activity definition of failure	6.2.2	ED+Refixia	[redacted]					
		BeneFIX	[redacted]					
		Alprolix	[redacted]					
		Idelvion	[redacted]					
		Refixia	[redacted]					
6m time to steady state	6.2.4	ED+Refixia	[redacted]					
		BeneFIX	[redacted]					
		Alprolix	[redacted]					
		Idelvion	[redacted]					

		Refixia	[redacted]	[redacted]			[redacted]	[redacted]
Disutility of IV FIX treatment of 0.042	6.2.5	ED+Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		BeneFIX	[redacted]	[redacted]	[redacted]		[redacted]	[redacted]
		Alprolix	[redacted]	[redacted]	[redacted]		[redacted]	[redacted]
		Idelvion	[redacted]	[redacted]	[redacted]		[redacted]	[redacted]
		Refixia	[redacted]	[redacted]	[redacted]		[redacted]	[redacted]
Adding AE cost and disutility to ED after first year	6.2.9	ED+Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		BeneFIX	[redacted]	[redacted]	[redacted]		[redacted]	[redacted]
		Alprolix	[redacted]	[redacted]	[redacted]		[redacted]	[redacted]
		Idelvion	[redacted]	[redacted]	[redacted]		[redacted]	[redacted]
		Refixia	[redacted]	[redacted]	[redacted]		[redacted]	[redacted]
Cumulative		ED+Refixia	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
		BeneFIX	[redacted]	[redacted]	[redacted]		[redacted]	[redacted]
		Alprolix	[redacted]	[redacted]	[redacted]		[redacted]	[redacted]
		Idelvion	[redacted]	[redacted]	[redacted]		[redacted]	[redacted]
		Refixia	[redacted]	[redacted]	[redacted]		[redacted]	[redacted]

Abbreviations: AE, adverse events; EAG, External Assessment Group; ED, etranacogene dezaparvovec; ICER, incremental cost-effectiveness ratio; mkt, market; NMB, net monetary benefit; QALY, quality adjusted life year

6.3.1. Durability threshold analysis around EAG's preferred base case

Figure 5 and Figure 6 below show that under the EAG's preferred base case assumptions, ED durability needs to be maintained at 100% for a minimum of 18-19 years to yield an ICER below £20,000 and £30,000 (the results are mostly insensitive to varying the WTP threshold between £20,000 and £30,000).

Figure 5: NMB as a function of ED durability (EAG base case, NMB at £20,000 / QALY)

Figure shows net monetary benefit of each of the five comparators as a function of durability of ED. The vertical line identifies the point where ED yields the highest net monetary benefit at a WTP threshold of £20,000 per QALY.

Figure 6: NMB as a function of ED durability (EAG base case NMB at £30,000 / QALY)

Figure shows net monetary benefit of each of the five comparators as a function of durability of ED. The vertical line identifies the point where ED yields the highest net monetary benefit at a WTP threshold of £30,000 per QALY.

6.4. Conclusions of the cost-effectiveness section

The EAG considered that the overall methodological approach used by the company in its analysis was mostly sound. However, several of its base case assumptions were unduly optimistic, and the analysis was severely limited by the quantity and quality of data available (in part, a consequence of the rarity of the disease) and the short follow-up period (which will be resolved by time). Individually the EAG's exploratory analyses did not alter the conclusions of the model and ED dominated all FIX replacement treatments. However, the EAG analyses demonstrated that the cost effectiveness of ED depended largely on assumptions concerning the durability of its effect. The EAG did not consider that the Shah analysis provided by the company was a reliable source of evidence for durability, given the lack of available data, and therefore the cost effectiveness of ED depended almost entirely upon conjecture about long-term durability. Relatedly, the definition of treatment failure in the model, which informed assumptions concerning treatment costs and utilities, also influenced cost effectiveness. Results varied substantially with the results of analyses including the confidential comparator prices (included in the confidential appendix to this report), though durability assumptions remained a significant issue.

7. QALY MODIFIER

The company stated that there was no excess mortality for people with haemophilia B and therefore the technology did not meet the criteria for the severity modifier (CS B.3.6, p166).

References

1. Adivo. Data on file: Patient Prophylactic Market Share for all Factor IX Drugs in U.S. and EU5 from 3Q 2020 Advivo Source. Unpublished. 2021
2. Berntorp E, LeBeau P, Ragni MV, et al. Quality of life in a large multinational haemophilia B cohort (The B-Natural study) - Unmet needs remain. *Haemophilia* 2022;28(3):453-61. doi: 10.1111/hae.14525 [published Online First: 2022/03/10]
3. Burke T, Asghar S, O'Hara J, et al. Clinical, humanistic, and economic burden of severe hemophilia B in the United States: Results from the CHESS US and CHESS US+ population surveys. *Orphanet J Rare Dis* 2021;16(1):143. doi: 10.1186/s13023-021-01774-9 [published Online First: 20210320]
4. O'Hara J, Hughes D, Camp C, et al. The cost of severe haemophilia in Europe: the CHESS study. *Orphanet J Rare Dis* 2017;12(1):106. doi: 10.1186/s13023-017-0660-y [published Online First: 2017/06/02]
5. Schwartz CE, Stark RB, Michael W, et al. Understanding haemophilia caregiver burden: does appraisal buffer the impact of haemophilia on caregivers over time? *Psychol Health* 2020;35(12):1516-30. doi: 10.1080/08870446.2020.1766042 [published Online First: 2020/06/02]
6. Cutter S, Molter D, Dunn S, et al. Impact of mild to severe hemophilia on education and work by US men, women, and caregivers of children with hemophilia B: The Bridging Hemophilia B Experiences, Results and Opportunities into Solutions (B-HERO-S) study. *Eur J Haematol* 2017;98 Suppl 86:18-24. doi: 10.1111/ejh.12851 10.1111/ejh.12851.
7. Khair K, Mackensen SV. Caregiver burden in haemophilia: results from a single UK centre. *The Journal of Haemophilia Practice* 2017;4(1):40-48. doi: doi:10.17225/jhp00094
8. Buckner TW, Witkop M, Guelcher C, et al. Impact of hemophilia B on quality of life in affected men, women, and caregivers-Assessment of patient-reported outcomes in the B-HERO-S study. *Eur J Haematol* 2018;100(6):592-602. doi: 10.1111/ejh.13055 [published Online First: 20180411]
9. Berger K, Schopohl D, Hilger A, et al. Research in haemophilia B--approaching the request for high evidence levels in a rare disease. *Haemophilia* 2015;21(1):4-20. doi: 10.1111/hae.12603 [published Online First: 2014/12/17]
10. Thorat T, Neumann PJ, Chambers JD. Hemophilia Burden of Disease: A Systematic Review of the Cost-Utility Literature for Hemophilia. *J Manag Care Spec Pharm* 2018;24(7):632-42. doi: 10.18553/jmcp.2018.24.7.632 [published Online First: 2018/06/29]
11. National Institute for Health and Care Excellence. Single technology appraisal and highly specialised technologies evaluation: User guide for company evidence submission template. Chapter 2 Clinical effectiveness. Process and methods [PMG24], 2022.
12. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health* 1998;52(6):377-84. doi: 10.1136/jech.52.6.377
13. Behring CSL. HOPE-B Study Results Overview: 24-month data [data on file], 2022.
14. Von Drygalski A, Gomez E, Giermasz A, et al. Stable and durable factor IX levels in hemophilia B patients over 3 years post etranacogene dezaparvovec gene therapy. *Blood Adv* 2022 doi: 10.1182/bloodadvances.2022008886 [published Online First: 2022/12/10]
15. Von Drygalski A, Giermasz A, Castaman G, et al. Etranacogene dezaparvovec (AMT-061 phase 2b): normal/near normal FIX activity and bleed cessation in hemophilia B. *Blood Adv* 2019;3(21):3241-47. doi: 10.1182/bloodadvances.2019000811 [published Online First: 2019/11/08]

16. Miesbach W, Meijer K, Coppens M, et al. Gene therapy with adeno-associated virus vector 5-human factor IX in adults with hemophilia B. *Blood* 2018;131(9):1022-31. doi: 10.1182/blood-2017-09-804419 [published Online First: 2017/12/17]
17. Shah J, Kim H, Sivamurthy K, et al. Comprehensive analysis and prediction of long-term durability of factor IX activity following etranacogene dezaparvovec gene therapy in the treatment of hemophilia B. *Curr Med Res Opin* 2022:1-11. doi: 10.1080/03007995.2022.2133492 [published Online First: 2022/10/27]
18. Pierce GF, Haffar A, Ampartzidis G, et al. First-year results of an expanded humanitarian aid programme for haemophilia in resource-constrained countries. *Haemophilia* 2018;24(2):229-35. doi: 10.1111/hae.13409 [published Online First: 20180314]
19. High KA. The gene therapy journey for hemophilia: are we there yet? *Blood* 2012;120(23):4482-7. doi: 10.1182/blood-2012-05-423210 [published Online First: 20120724]
20. Behring CSL. CT-AMT-060-01 CSR: A phase I/II, open-label, uncontrolled, single-dose, dose-ascending, multi-centre trial investigating an adeno-associated viral vector containing a codon-optimized human Factor IX gene (AAV5-hFIX) administered to adult patients with severe or moderately severe haemophilia B [data on file]. 2022
21. Agency for Healthcare Research and Quality. Global Health Evidence Evaluation Framework: Research White Paper 2013 [Available from: https://effectivehealthcare.ahrq.gov/sites/default/files/pdf/global-health-evidence-evaluation-framework_white-paper.pdf accessed March 2023.
22. Santagostino E, Martinowitz U, Lissitchkov T, et al. Long-acting recombinant coagulation factor IX albumin fusion protein (rIX-FP) in hemophilia B: results of a phase 3 trial. *Blood* 2016;127(14):1761-9. doi: 10.1182/blood-2015-09-669234 [published Online First: 2016/01/13]
23. Powell JS, Pasi KJ, Ragni MV, et al. Phase 3 study of recombinant factor IX Fc fusion protein in hemophilia B. *N Engl J Med* 2013;369(24):2313-23. doi: 10.1056/NEJMoa1305074 [published Online First: 2013/12/07]
24. Collins PW, Young G, Knobe K, et al. Recombinant long-acting glycoPEGylated factor IX in hemophilia B: a multinational randomized phase 3 trial. *Blood* 2014;124(26):3880-6. doi: 10.1182/blood-2014-05-573055 [published Online First: 2014/09/28]
25. Lambert T, Recht M, Valentino LA, et al. Reformulated BeneFix: efficacy and safety in previously treated patients with moderately severe to severe haemophilia B. *Haemophilia* 2007;13(3):233-43. doi: 10.1111/j.1365-2516.2007.01458.x [published Online First: 2007/05/15]
26. Kavakli K, Smith L, Kuliczkowski K, et al. Once-weekly prophylactic treatment vs. on-demand treatment with nonacog alfa in patients with moderately severe to severe haemophilia B. *Haemophilia* 2016;22(3):381-8. doi: 10.1111/hae.12878 [published Online First: 20160129]
27. O'Hara J, Martin AP, Nugent D, et al. Evidence of a disability paradox in patient-reported outcomes in haemophilia. *Haemophilia* 2021;27(2):245-52. doi: 10.1111/hae.14278 [published Online First: 2021/02/18]
28. Wyrwich KW, Krishnan S, Poon JL, et al. Interpreting important health-related quality of life change using the Haem-A-QoL. *Haemophilia* 2015;21(5):578-84. doi: 10.1111/hae.12642 [published Online First: 20150331]
29. European Medicines A. Hemgenix (etranacogene dezaparvovec) [Available from: https://www.ema.europa.eu/en/medicines/human/summaries-opinion/hemgenix accessed 10 February 2023.
30. Eversana™. EtranaDez indirect treatment comparisons versus recombinant Factor IX products for hemophilia B. Technical report 24 month data – FINAL, 2022.

31. Eversana™. EtranaDez indirect treatment comparisons versus BeneFIX. Addendum of technical report – FINAL, 2022.
32. Behring CSL. CT-AMT-061-01 CSR: Phase IIb, open-label, single-dose, single-arm, multi-center trial to confirm the factor IX activity level of the serotype 5 adeno-associated viral vector containing the Padua variant of a codon-optimized human factor IX gene (AAV5-hFIXco-Padua, AMT-061) administered to adult subjects with severe or moderately severe hemophilia B [data on file]. 2022
33. Muhuri M, Levy DI, Schulz M, et al. Durability of transgene expression after rAAV gene therapy. *Mol Ther* 2022;30(4):1364-80. doi: 10.1016/j.ymthe.2022.03.004 [published Online First: 2022/03/15]
34. Tice Ja WSH-HBFSMMASJCJAF, Pearson Sd RDM. Gene Therapy for Hemophilia B and An Update on Gene Therapy for Hemophilia A: Effectiveness and Value. In: Economic, ed. Draft Evidence Report: ICER, 2022.
35. Shah J. CSL222 Durability estimatin update (60 years) (data on file), 2022.
36. Van Hout B, Janssen MF, Feng YS, et al. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value Health* 2012;15(5):708-15. doi: 10.1016/j.jval.2012.02.008 [published Online First: 2012/08/08]
37. Dolan P. Modeling valuations for EuroQol health states. *Med Care* 1997;35(11):1095-108. doi: 10.1097/00005650-199711000-00002
38. National Institute for Health and Care Excellence. NICE health technology evaluations: the manual (PMG36), 2022.
39. Fletcher S, Jenner K, Pembroke L, et al. The experiences of people with haemophilia and their families of gene therapy in a clinical trial setting: regaining control, the Exigency study. *Orphanet J Rare Dis* 2022;17(1):155. doi: 10.1186/s13023-022-02256-2 10.1186/s13023-022-02256-2. [published Online First: 20220404]
40. Aradom E, Gomez K. The patient gene therapy journey: Findings from qualitative interviews with trial participants at one UK haemophilia centre. *The Journal of Haemophilia Practice* 2021;8(1):32-44. doi: doi:10.17225/jhp00174

Appendix A: Threshold analysis on durability of ED

This appendix shows the full results of the threshold analysis presented in Figure 3: NMB as a function of ED durability (EAG corrected company base case, NMB at £20,000 / QALY)Figure 3 to Figure 6, showing how the cost-effectiveness of ED and its comparators varies with assumptions over the durability of ED (presented as net monetary benefit, NMB). The durability function assumed for this analysis was a 'cliff-edge' function, whereby 100% durability was assumed until year n, dropping instantly to 0% the following year. It is thus an approximation of a more plausible gradual tailing off of durability. The EAG adopted this approach to avoid assuming a specific parametric form for the durability function.

Table 22 shows the NMB of the four IV FIXes at £20,000 and £30,000 per QALY gained. These do not change as durability of ED changes. Table 23 shows how the cost and QALYs gained with ED change as the durability of ED increased. The option with the highest NMB ('winner') is mathematically identical to identifying the most cost-effective option with an ICER below the threshold taking account of dominance and extended dominance.

The same data are shown in Table 24 and Table 25 for the EAG-preferred base case.

Table 22: Net Monetary benefit for IV FIXes at £20,000 and £30,000/QALY thresholds (deterministic, EAG corrected company base case, ED PAS discount)

IV FIX	NMB @ £20,000 / QALY	NMB @ £30,000 / QALY
Benefix	[redacted]	[redacted]
Alprolix	[redacted]	[redacted]
Idelvion	[redacted]	[redacted]
Refixia	[redacted]	[redacted]

Table 23: Threshold analysis varying durability of ED (deterministic, EAG corrected company base case, ED PAS discount)

	ED+Refixia		NMB (£20,000/QALY)			NMB (£30,000/QALY)		
Years	£	QALY	ED	Max	Winner	ED	Max	Winner
1	[redacted]	[redacted]	[redacted]	[redacted]	Benefix	[redacted]	[redacted]	Benefix
2					Benefix			Benefix
3					Benefix			Benefix

4	[redacted]	[redacted]	[redacted]	[redacted]	Benefix	[redacted]	Benefix
5					Benefix		Benefix
6					Benefix		Benefix
7					Benefix		Benefix
8					Benefix		Benefix
9					Benefix		Benefix
10					Benefix		Benefix
11					Benefix		Benefix
12					Benefix		Benefix
13					Benefix		Benefix
14					Benefix		Benefix
15					Benefix		Benefix
16					Benefix		Benefix
17					Benefix		Benefix
18					ED		ED
19					ED		ED
20					ED		ED
21					ED		ED
22					ED		ED
23					ED		ED
24					ED		ED
25					ED		ED
26					ED		ED
27					ED		ED
28					ED		ED
29					ED		ED
30					ED		ED
31					ED		ED
32					ED		ED
33					ED		ED
34					ED		ED
35					ED		ED
36					ED		ED
37					ED		ED
38					ED		ED
39					ED		ED
40					ED		ED
41					ED		ED
42					ED		ED
43					ED		ED
44					ED		ED
45					ED		ED
46					ED		ED
47					ED		ED
48					ED		ED
49					ED		ED
50					ED		ED

Table 24: Net Monetary benefit for IV FIXes at £20,000 and £30,000/QALY thresholds (deterministic, EAG preferred base case, ED PAS discount)

IV FIX	NMB @ £20,000 / QALY	NMB @ £30,000 / QALY
Benefix	[redacted]	[redacted]
Alprolix	[redacted]	[redacted]
Idelvion	[redacted]	[redacted]
Refixia	[redacted]	[redacted]

Table 25: Threshold analysis varying durability of ED (deterministic, EAG preferred base case, ED PAS discount)

	ED+Refixia		NMB (£20,000/QALY)			NMB (£30,000/QALY)		
Years	£	QALY	ED	Max	Winner	ED	Max	Winner
1	[redacted]	[redacted]	[redacted]	[redacted]	Benefix	[redacted]	[redacted]	Benefix
2					Benefix			Benefix
3					Benefix			Benefix
4					Benefix			Benefix
5					Benefix			Benefix
6					Benefix			Benefix
7					Benefix			Benefix
8					Benefix			Benefix
9					Benefix			Benefix
10					Benefix			Benefix
11					Benefix			Benefix
12					Benefix			Benefix
13					Benefix			Benefix
14					Benefix			Benefix
15					Benefix			Benefix
16					Benefix			Benefix
17					Benefix			Benefix
18					Benefix			Benefix
19					ED			ED
20					ED			ED
21					ED			ED
22					ED			ED
23					ED			ED
24					ED			ED
25					ED			ED
26					ED			ED
27					ED			ED
28					ED			ED