CONFIDENTIAL UNTIL PUBLISHED

External Assessment Group Report commissioned by the NIHR Evidence Synthesis Programme on behalf of NICE

Linzagolix for treating moderate to severe symptoms of uterine fibroids

---

Produced by	**Southampton Health Technology Assessments Centre (SHTAC)**
Authors	Lois Woods, Senior Research Assistant, Evidence Synthesis and Information Specialist Neelam Kalita, Senior Research Fellow, Health Economics Asyl Hawa, Research Fellow, Health Economics David Alexander Scott, Principal Research Fellow, Statistics Geoff Frampton, Principal Research Fellow, Evidence Synthesis
Correspondence to	Dr Geoff Frampton Southampton Health Technology Assessments Centre (SHTAC) School of Healthcare Enterprise and Innovation University of Southampton Alpha House, Enterprise Road, University of Southampton Science Park, Southampton SO16 7NS www.southampton.ac.uk/shtac
Date completed	9th November 2023 (updated 28th November 2023 after factual accuracy check)
Source of Funding	This report was commissioned by the National Institute for Health and Care Research (NIHR) Evidence Synthesis Programme as project number 136078.

Acknowledgements

We thank the following for providing expert clinical advice to the project team and for reading and commenting on a draft of the report:

Mr Sameer Umranikar, Consultant Gynaecologist, Princess Anne Hospital, University Hospital Southampton NHS Foundation Trust.

We also thank Fay Chinnery and Karen Pickett, Southampton Health Technology Assessments Centre (SHTAC), for reading and commenting on a draft of this report for quality assurance.

Declared competing interests of the authors and advisors

The authors and clinical expert advisor have no conflicts to declare

Rider on responsibility for the report

The views expressed in this report are those of the authors and not necessarily those of the NIHR Evidence Synthesis Programme. Any errors are the responsibility of the authors.

This report should be referenced as follows:

Woods LE, Kalita N, Hawa A, Scott DA, Frampton G. Linzagolix for treating moderate to severe symptoms of uterine fibroids. A Single Technology Appraisal. Southampton Health Technology Assessments Centre, 2023

Contributions of authors

Lois Woods critically appraised the company searches and the clinical effectiveness systematic review, and drafted the report; Neelam Kalita critically appraised the health economic systematic review, critically appraised the economic evaluation, and drafted the report; Asyl Hawa critically appraised the health economic systematic review, critically appraised the economic evaluation, and drafted the report; David Alexander Scott critically appraised the clinical effectiveness systematic review and drafted the report; Geoff

Frampton critically appraised the clinical effectiveness systematic review, drafted the report and is the project co-ordinator and guarantor.

[redacted]

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

ABT	Add-back therapy
AE	Adverse event
AH	Alkaline haematin
AIC	Academic in confidence
BMD	Bone mineral density
BNF	British National Formulary
CI	Confidence interval
CIC	Commercial in confidence
CONSORT	Consolidated Standards of Reporting Trials
CRD	Centre for Reviews and Dissemination
CS	Company submission
CSR	Clinical study report
CT	Combination therapy
DSU	Decision Support Unit
DEXA or DXA	Dual energy X-ray absorptiometry
EAG	External Assessment Group
EMC	Electronic Medicines Compendium
EPAR	European Public Assessment Report
EQ-5D-3L	European Quality of Life Working Group Health Status Measure 3 Dimensions, 3 Levels
EQ-5D-5L	European Quality of Life Working Group Health Status Measure 5 Dimensions, 5 Levels
EQ-VAS	EuroQol Visual Analogue Scale
FAS	Full Analysis Set
FIGO	International Federation of Gynecology and Obstetrics
Hb	Haemoglobin
HMB	Heavy menstrual bleeding
HRG	Healthcare Resource Group
HRQoL	Health-related quality of life
HTA	Health technology assessment
ICER	Incremental cost-effectiveness ratio
IPD	Individual patient level data
ITC	Indirect treatment comparison
ITT	Intention to treat

MAIC	Matching-adjusted indirect comparison
MBL	Menstrual blood loss
mITT	Modified intention to treat
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NR	Not reported
NRS	Numeric Rating Scale
NSAIDs	Nonsteroidal anti-inflammatory drugs
PBAC	Pictorial blood-loss assessment chart
PGI-I	Patient Global Impression of Improvement scale
PSA	Probabilistic sensitivity analysis
PSS	Personal Social Services
QALY	Quality-adjusted life year
QoL	Quality of life
RCT	Randomised controlled trial
RR	Relative risk/risk ratio
SAE	Serious adverse event
SD	Standard deviation
SE	Standard error
SLR	Systematic literature review
SmPC	Summary of product characteristics
TA	Technology appraisal
TEAE	Treatment-emergent adverse event
TSD	Technical Support Document
UF	Uterine fibroid(s)
UFS-QoL	Uterine Fibroid Symptom-Quality of Life
UK	United Kingdom
US	United States
VAS	Visual analogue scale

1 EXECUTIVE SUMMARY

This summary provides a brief 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, health technology, evidence and information on the issues are in the main EAG report.

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

1.1 Overview of the EAG's key issues

The EAG's Key Issues refer to two different types of economic analysis since the company's submission (CS) includes both cost-comparison and cost-effectiveness analysis approaches. In the cost-comparison analysis the comparators are relugolix-CT and other gonadotrophin releasing hormone (GnRH) analogues. In the cost-effectiveness analysis the comparator is established clinical management, which the company refer to as best supportive care (BSC). An overview of how the economic analysis approaches map to the comparators and sub-populations of the CS is given in Table 3 of this report.

Table 1 Summary of Key Issues identified by the EAG

ID	Summary of issue	Report sections
1	Uncertain whether linzagolix has similar clinical effectiveness to relugolix CT and other GnRH analogues	2.3 (summary), 3.4, 3.5 (details)
2	Uncertain market share of relugolix CT	2.3
3	Issue 3 Uncertain relevance of the PRIMROSE pivotal trials to the three population subgroups that inform the company's economic analyses	2.2.3 (summary), 3.2.1.1.5.1 (details)
4	Uncertain whether patients can experience recurrence after undergoing surgery	4.2.5.2.4
5	Uncertainty surrounding the utility function	4.2.6

The key difference between the company's assumption and the EAG's conclusion for the cost-comparison analysis is that we are uncertain about the similarity in clinical efficacy between linzagolix and relugolix-CT for Populations #1 and #2.

The key differences between the company's preferred assumptions and the EAG's preferred assumptions for the cost-effectiveness analysis for Population #3 are:

- Inclusion of prophylactic regimens of calcium and vitamin D in the BSC arm.
- Distribution of surgery types.
- Use of healthcare resources.
- Unit costs of gynaecological consultation and MRI as identified by the EAG.
- Using EQ-5D-5L data from the PRIMROSE trial to estimate the health state utilities.

1.2 Overview of key model outcomes for the cost-effectiveness analysis

NICE technology appraisals compare how much a new technology improves length (overall survival) and quality of life in a quality-adjusted life year (QALY). An ICER is the ratio of the extra cost for every QALY gained.

Overall, the technology is modelled to affect QALYs by:

- Improving symptoms (based on menstrual blood loss) and affecting patients' transition through the health states (based on the response and recurrence rates).
- Reducing the overall probability of surgery, which is associated with a risk of mortality.
- Switching the surgery types, from open/abdominal to laparoscopic, which is associated with an improved quality of life.
- Utility associated with the controlled, uncontrolled, and post-surgery health states and disutility associated with adverse effects.

Overall, the technology is modelled to affect costs by:

- Increase in drug acquisition costs and health state resource use.
- Treatment discontinuation rates.
- Distribution of surgery types.

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

- Recurrence rate.
- Choice of the HRQoL data from the pivotal company trials (PRIMROSE) used to estimate EQ-5D values for the health states, and the source of utility (whether trial-based or published literature).

- Treatment withdrawal rates.
- Changing the probability associated with surgery and changing the distribution of surgery types.

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

The EAG have identified three key issues that are related both to the decision problem and the clinical efficacy evidence, summarised in the following tables.

Issue 1 Uncertain whether linzagolix has similar clinical efficacy to relugolix CT and other GnRH analogues

Report section	Sections 2.3 (summary), 3.4 and 3.5 (details)
Description of issue and why the EAG has identified it as important	According to the NICE Methods Guide[1] a cost comparison analysis is appropriate for technologies that are likely to provide similar or greater health benefits at similar or lower cost than the relevant comparator(s). Relugolix CT is considered by the company to be a relevant comparator to support their cost-comparison analyses. Cost comparisons were conducted for two population subgroups: those receiving linzagolix for short-term treatment (≤6 months) prior to surgical intervention (referred to in the CS as Population #1); and those receiving linzagolix for longer-term treatment (Population #2). No studies exist that directly compare linzagolix against relugolix CT in these populations so the company conducted network meta-analyses (NMAs) to make this comparison. Matching-adjusted indirect comparisons (MAICs) were also provided as a sensitivity analysis to help understand how sensitive the NMA results might be to heterogeneity in the trial characteristics. Results of the NMAs are generally highly uncertain and only convincingly show clinical similarity of linzagolix to relugolix CT for one outcome, the reduction in fibroid volume. Clinical similarity does not appear to be supported for key outcomes related to menstrual blood loss, including the company trials' primary outcome. However, interpreting clinical similarity from NMA results is challenging because a non-inferiority analysis should ideally have been pre-specified which, as far as we are aware, is uncommon in NMAs. Although not stated explicitly in the CS, the company appear to assume that statistical non-significance of NMA results implies similarity in clinical efficacy. Such an assumption would be very sensitive to statistical heterogeneity, and conclusions on clinical similarity may not be possible when NMA results have wide credible intervals that include the null, as was frequent in the NMAs provided by the company. The CS does not provide any

	explicit guidance on how the NMA results are expected to be interpreted.
What alternative approach has the EAG suggested?	The company had conducted the NMAs and MAICs using pooled data from the pivotal trials PRIMROSE 1 and PRIMROSE 2 but did not explore opportunities for reducing uncertainty in the NMA results. These trials, although of similar designs, differ in some aspects of their population baseline characteristics. The EAG requested the company to rerun the NMA analyses separately for the PRIMROSE 1 and PRIMROSE 2 trials. We also requested that the company explore alternative approaches for accounting for missing data, and for the company to provide posterior probabilities for the NMA and MAIC results to assist in judgements of the statistical similarity of the therapies. These analyses were provided by the company but do not reduce the uncertainty.
What is the expected effect on the cost-effectiveness estimates?	The existing cost-comparison analyses for Population #1 and Population #2 make the key assumption that linzagolix has similar clinical efficacy when compared to relugolix CT. If this assumption is not supported, then cost-comparison analyses might not be appropriate.
What additional evidence or analyses might help to resolve this key issue?	We requested extended NMAs from the company to include alternative comparators specified in the NICE scope. The company provided NMAs for one comparator, leuprolide acetate, but with almost no explanation of the methodology employed so the results are difficult to interpret. Results of these NMAs are highly uncertain and it is unclear whether other comparators could have been included in the evidence networks. A more thorough and transparent approach to the evidence synthesis, exploring ways to reduce uncertainty in the NMA results would be helpful. If uncertainty of the comparative clinical efficacy evidence for Populations #1 and #2 cannot adequately be resolved, then a cost-effectiveness modelling approach might be more appropriate for these population subgroups.

Issue 2 Uncertain market share of relugolix CT

Report section	Section 2.3
Description of issue and why the EAG has identified it as important	An assumption of NICE cost comparisons is that the selected comparator therapy, i.e. relugolix CT, should have an adequate market share. The EAG's clinical expert commented that most patients in his experience (around 90%) currently receive goserelin or leuprorelin, although relugolix CT is relatively new

	and its use would likely increase. The expert also commented that general practitioners are not yet aware of relugolix CT.
What alternative approach has the EAG suggested?	Consultation with further clinical experts to clarify the extent of relugolix CT use.
What is the expected effect on the cost-effectiveness estimates?	The cost comparison approach might not be appropriate if relugolix CT is not widely used in clinical practice.
What additional evidence or analyses might help to resolve this key issue?	Market share data if available.

Issue 3 Uncertain relevance of the PRIMROSE pivotal trials to the three population subgroups that inform the company's economic analyses

Report section	Sections 2.2.3 (summary) and 3.2.1.1.5.1 (details)
Description of issue and why the EAG has identified it as important	The company submission specifies three population subgroups are relevant to this technology appraisal, consistent with the NICE scope, and these influence the economic analysis approaches employed by the company: • Population #1: Patients having short-term treatment of 6 months or less whilst awaiting a surgical intervention (cost-comparison analysis); • Population #2: Patients having longer-term treatment with hormone-based therapy (cost-comparison analysis); • Population #3: Patients having longer-term treatment without hormone-based therapy (cost-utility analysis). The pivotal PRIMROSE trials do not explicitly and fully include any of these subgroups of patients, for the following reasons: • Population #1: Patients included in the trials were not eligible to receive surgery for their fibroids within 6 months regardless of the treatment provided. • Population #2: These patients, taking longer-term therapy, are not fully represented in the PRIMROSE trials since the trials had maximum duration 52 weeks, with most outcomes reported at 24 weeks. Patients who would take linzagolix for longer than 52 weeks are not specifically represented, although some limited efficacy outcomes data are available up to 64 weeks.

	• Population #3: People who were contraindicated to hormonal add-back therapy (ABT) were excluded from the PRIMROSE trials. The company assume that patients in the PRIMROSE trials randomised to receive linzagolix (100mg or 200mg) without ABT are suitable as a proxy for those contraindicated to ABT. The company do not provide a rationale for this assumption, and the EAG are uncertain whether the assumption is valid. Furthermore, there is uncertainty in the size of this sub-population in clinical practice. The EAG's clinical expert believed the sub-population unable to receive HRT to be very small, as he had not encountered this patient group in his clinical practice.
What alternative approach has the EAG suggested?	The CS does not discuss explicitly whether Population #1 and Population #2 are represented in the PRIMROSE trials and whether the trial outcomes can be applied to these population subgroups. The CS also does not discuss whether clinical efficacy or safety responses to linzagolix would differ according to patients' ability or willingness to receive ABT. It is therefore unclear whether the population in the PRIMROSE trials who could receive hormonal ABT is an appropriate proxy for those in Population #3. The EAG sought feedback from a clinical expert.
What is the expected effect on the cost-effectiveness estimates?	The company's approaches to economic analysis might not be appropriate if the clinical trial populations are not reflective of the modelled populations. In particular, the company's cost-effectiveness analysis for Population #3 might not be appropriate if: (1) patients who are unable or able to receive HRT differ in their response to linzagolix therapy; or (2) very few, or no, patients in clinical practice would be unable to receive ABT.
What additional evidence or analyses might help to resolve this key issue?	The EAG received advice from one clinical expert. Wider consultation with further clinical experts might help to understand whether the clinical trial populations can be extrapolated to the company's three sub-populations.

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

The EAG identified three key issues relating to both the decision problem and the clinical effectiveness evidence, summarised in section 1.3 above.

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

The EAG have identified two key issues relating to the cost-effectiveness evidence, summarised in the following tables.

Issue 4 Uncertain whether patients can experience recurrence after undergoing surgery

Report section	Section 4.2.5.2.4
Description of issue and why the EAG has identified it as important	The company's assumption that both linzagolix and BSC arms have similar distributions of surgery types is reasonable. With respect to patient distributions across the different surgery types, the EAG's clinical expert considered that some of the surgery types, e.g. laparoscopic hysterectomy and UAE, are more common than others. Furthermore, patients are also likely to undergo hysteroscopic myomectomy, which is not listed in the company's analyses. Lastly, our clinical expert suggested that recovery time after different types of surgery varies between 4 and 8 weeks. For example, the recovery time after laparoscopic surgery could be 4-6 weeks; open surgery: 6-8 weeks, UAE: 4-6 months. We have conducted scenario analyses changing the distributions across the different surgery types based on our expert's advice. While this impacts the total costs and total QALYs, the change is proportional as the distributions are similar for both the treatment arms and therefore there is no overall impact on the ICER (see Section 6) In their cost-utility analysis for Population #3, the company assume that after patients undergo surgery, they transition to the 'post-surgery' state until the onset of menopause. The EAG are uncertain if this is clinically plausible as patients undergoing different surgery types may have a different prognosis. While some may be completely cured (e.g., those undergoing hysterectomies), others may experience a recurrence of the symptoms post-surgery.
What alternative approach has the EAG suggested?	We suggest the company consider adding 'recurrence' from the post-surgery state to the cost-effectiveness model for Population #3. This would be appropriate if recurrence is found to be frequent, based on further discussion with clinicians and the NICE committee.
What is the expected effect on the cost-effectiveness estimates?	The direction and magnitude of the overall cost-effectiveness results are unclear as it depends on the recurrence rate(s) applied in both the treatment arms- Linzagolix and BSC. The EAG suspect the overall impact is unlikely to be significant if a similar recurrence rate is applied to both the treatment arms.
What additional evidence or analyses might	Further discussion and clarification of patients' prognosis after undergoing different surgery types in clinical practice, particularly with respect to the proportion of patients who may experience a recurrence of the symptoms, might enable a more accurate

help to resolve this key issue?	reflection of clinical practice. We suggest the company conduct scenario analyses by adding percentage(s) of recurrence in the post-surgery state for both arms.

Issue 5 Uncertainty surrounding the utility function

Report section	Section 4.2.6
Description of issue and why the EAG has identified it as important	In their base case, the company mapped UFS-QoL data from the PRIMROSE trials to EQ-5D-3L utility values using an unpublished algorithm that was applied in a previous NICE appraisal TA832. They also reported a scenario analysis using EQ-5D-5L data collected in the PRIMROSE trials, mapped to EQ-5D-3L utility values estimated using the NICE-preferred (Hernández-Alava) method. The EAG has some concerns over the use of utility estimates mapped from a disease-specific measure when EQ-5D data are available from the PRIMROSE trials. We also have some concerns about the lack of transparency of the UFS-QoL mapping algorithm. We note the use of the UFS-QoL mapped estimates in TA832, but question whether the TA832 committee's concerns about the availability of EQ-5D-5L data from the clinical trials apply in the current appraisal. The company applied a linear mixed model to analyse both mapped UFS-QoL and EQ-5D utility estimates from the PRIMROSE trials to estimate utilities for the 'controlled' and 'uncontrolled' health states. The same health state utilities were used in the economic model for both the treatment arms. The EAG have some concerns about the reporting of the utility analysis. The company did not define or justify the specification for the linear mixed model regression of utility data. It is not clear why they chose to include a single independent variable- reduction in menstrual blood loss (RMBL), or whether additional co-variates would have improved the model fit. Furthermore, no sensitivity or scenario analyses were reported for alternative specifications of the utility function, and uncertainty over the regression coefficients was not included in the probabilistic sensitivity analysis.
What alternative approach has the EAG suggested?	The EAG suggest that the base case should use utility estimates derived from EQ-5D data collected in the trial. Scenario analysis should also be reported to explore uncertainty over the coefficients from the linear mixed model analyses of trial utility data.
What is the expected effect on the cost-	Due to lack of information on the model specification, the EAG conducted a range of exploratory scenarios changing the coefficients of the utility function. The EAG's exploratory analyses

effectiveness estimates?	have an impact on the ICER, ranging between £13,968 per QALY and £34,376 per QALY.
What additional evidence or analyses might help to resolve this key issue?	Further information on the model specification for the linear mixed model utility function along with exploration of alternative specifications.

1.6 Other key issues: summary of the EAG's view

- The company's interpretation of BSC may not fully reflect clinical practice, as it does not include prophylactic doses of calcium and vitamin D, which are also given to the patients as well as NSAID and iron supplements, to protect against bone loss.

- There are no data to support the company's assumption that the treatment effect (i.e. response) of linzagolix is maintained beyond 1 year, although it may be biologically plausible.

- There is uncertainty about the recurrence rate in patients with uterine fibrosis.

- With respect to patient distributions across the different surgery types, advice from our clinical expert suggests that some of the surgeries (e.g., laparoscopic hysterectomy and umbilical artery embolization (UAE)) may be more common than others.

- The company's assumptions regarding healthcare resource use may not be reflective of the UK clinical practice. Their assumptions that patients would not have any GP visits, have full blood count and MRI scan once each, and people in the linzagolix arm receive one DEXA scan after 1 year may not be an appropriate representation of the clinical practice. We conduct scenario analysis on resource use, based on the advice of our clinical expert, see Section 6.

1.7 Summary of EAG's preferred assumptions and resulting ICER

The following changes were made to the company's base case to form the EAG preferred base case for Population #3:

- Inclusion of vitamin D and calcium in the BSC arm.
- Applying the distribution of surgery types based on the advice of the EAG's clinical expert.
- Using the health care resource use based on the EAG's clinical expert advice.
- Change in the unit costs for gynaecologist consultation and MRI scan.

- Using EQ-5D-5L data from the PRIMROSE trial to estimate the utilities for the controlled and uncontrolled health states.

Table 2 Company and EAG base case results for Population #3

Scenario	Incremental cost	Incremental QALYs	ICER (change from company base case)
Company's base case	[redacted]	[redacted]	£15,392
EAG's preferred base case	[redacted]	[redacted]	£28,973

For further details of the EAG's exploratory and sensitivity analyses see Section 6.

2 INTRODUCTION AND BACKGROUND

2.1 Introduction

This report is a critique of the company's submission (CS) to NICE from Theramex on the clinical effectiveness and cost effectiveness of linzagolix for treating moderate to severe symptoms of uterine fibroids. It identifies the strengths and weaknesses of the CS. Clinical experts were consulted to advise the external assessment group (EAG) and to help inform this report.

Clarification on some aspects of the CS was requested from the company by the EAG via NICE on 21st September 2023. A response from the company via NICE was received by the EAG on 11th October 2023 and this can be seen in the NICE committee papers for this appraisal.

2.2 Background

2.2.1 Background information on uterine fibroids

The CS provides an accurate overview of the disease in CS sections B.1.3.1 and B.1.3.2.

2.2.1.1 Overview of the condition

Uterine fibroids (also called myomas or leiomyomas) are non-malignant smooth muscle tumours of the uterus. The exact cause is not known but they have been linked to oestrogen and progesterone, occur in people of reproductive age, and can become smaller after menopause. Around 2 in 3 women develop at least one uterine fibroid. Incidence of fibroids increases with age until the menopause, with a peak in those aged in their 40s.

Uterine fibroids are classified according to their site of origin (CS Figure 2). Intramural fibroids (the most common type) develop within the uterine wall; subserosal fibroids develop on the outside of the uterus, projecting into the pelvis, where they can become very large; and submucosal fibroids develop from inside the uterus and protrude into the uterine cavity. Submucosal and subserosal fibroids may or may not have a stalk (pedunculate fibroids) and some fibroids may encompass more than one uterine location. Generally, fibroids in the uterus can cause bleeding symptoms, whereas fibroids outside the uterus can cause pressure symptoms. The EAG's clinical expert advisor confirmed that the International Federation of Gynecology and Obstetrics (FIGO) classification system for uterine fibroids (CS Figure 3) is used in clinical practice to guide treatment decisions.

2.2.1.2 Risk factors

Major risk factors for uterine fibroids, as confirmed by the EAG's clinical expert, are age up to menopause, family history, nulliparity and Black race. Specifically, Black women have an increased risk of developing uterine fibroids, are more likely to have large and multiple fibroids, develop these 5-6 years earlier, and experience higher rates of hospitalisation and surgical intervention compared to White women. The risk of developing uterine fibroids is also increased in women who have obesity, early menarche (first menstrual period), time since last birth more than 5 years, hypertension, and exposure to oestrogen-like chemicals (e.g. phytoestrogens in soy milk) (CS section B.1.3.1.3).

2.2.1.3 Symptoms and burden of disease

Most women with uterine fibroids do not experience symptoms, but for the 25% to 30% with fibroids who do, their symptoms can be moderate or severe [2] (this is the population specified in the NICE scope for this appraisal). The CS does not explicitly define severity of uterine fibroids. The EAG's clinical expert said severity of symptoms are judged according to their impact on a patient's quality of life and their work. If the patient needs to take time off work or their symptoms are causing disruption to their regular activity then these would be classed as moderate to severe. Symptoms that are often considered moderate or severe include heavy menstrual bleeding which can lead to anaemia, bladder or bowel pressure, pain, or infertility. For pain related to uterine fibroids, a numerical rating scale can be used for quantifying severity (CS Table 10) whilst symptom severity can also be assessed using a subscale of the UFS-QoL instrument (described in section 3.2.3.2). Black people typically present with more severe symptoms than White people.[3] The position, type, size, and number of fibroids present influences the type and severity of symptoms experienced.[4, 5] According to the recent (October 2022) NICE Technology Appraisal (TA) of relugolix CT for uterine fibroids (TA832), symptoms are broadly classed into heavy and prolonged menstrual bleeding, pelvic pain and pressure, and reproductive dysfunction. The CS (section B.1.3.2.1) notes that people with uterine fibroids can experience a wide range of symptoms including frequent menstrual cycles, bloating, increased urinary frequency, constipation, fatigue, anxiety or stress, and various types of pain (leg or back pain, menstrual pain or cramping, pelvic pain, or pain during intercourse).[6]

Iron deficiency anaemia (IDA) is an important complication in around two thirds of those who experience HMB caused by uterine fibroids and can lead to increased morbidity and mortality following surgery. Uterine fibroids can also cause infertility and pregnancy complications, including miscarriage, pre-term and caesarean delivery.

As reported in CS section B.1.3.2.1 and CS Table 3, the symptoms and sequelae of uterine fibroids can have a range of negative impacts on patients' wellbeing, including physical, social and emotional impacts which can interfere with sleep, relationships, social life and work or school. These can have a negative impact on patients' health-related quality of life (HRQoL) and productivity. The NICE Committee in TA832 concluded that uterine fibroids represent a significant burden for people who have them, affecting both physical and psychological aspects of quality of life.

2.2.2 Background information on linzagolix

Linzagolix (brand name Yselty®) is a gonadotrophin releasing hormone (GnRH) receptor antagonist which binds competitively to GnRH receptors in the pituitary gland. This alters GnRH signalling between the hypothalamus and pituitary, leading to a dose-dependent reduction in the production of serum luteinising hormone (LH) and follicle stimulating hormone (FSH) from the anterior pituitary gland. LH and FSH are key regulators of the production of estradiol and progesterone in the ovary. The effect of linzagolix on the production of LH and FSH causes immediate dose-dependent suppression of ovarian estradiol secretion and subsequent progesterone secretion, with the changes in hormone levels quickly reversible on stopping the therapy. The overall mode of action of linzagolix (as with all GnRH analogues) is therefore to reduce the levels of the hormones that are thought to be responsible for fibroid development, effectively inducing a controlled menopause.

GnRH analogues fall into two groups, agonists and antagonists, which differ in the way that they interact with pituitary GnRH receptors and modulate the secretion of LH and FSH. The GnRH agonists, such as leuprolide acetate and goserelin (which are potential comparators to linzagolix) cause an initial, transient, increase in sex hormone production before levels of estradiol and progesterone decrease, which can lead to a transient initial increase in symptoms such as heavy menstrual bleeding. In contrast, the more recently-developed GnRH antagonists, which include linzagolix and relugolix, do not cause a transient increase in oestradiol and progesterone levels or the associated initial symptom flare.

The use of GnRH analogues has the downside that patients may experience symptoms of early menopause (i.e. hot flashes, weight gain, fluid retention, among others) as well as potential adverse events related to early menopause, notably decreased bone mineral density (BMD) and increased risk of osteoporosis. Long-term use of GnRH analogues therefore requires a balancing act between management of uterine fibroids and management of menopausal sequelae. To achieve this, GnRH analogues are usually co-administered with hormonal therapy, except for short-term use (≤6 months). A linzagolix

tablet contains the GnRH antagonist without the hormonal therapy, and it is intended by the company that the hormonal therapy can be administered separately, referred to as "add-back therapy" (ABT). In contrast, relugolix CT is formulated as a combined therapy (CT) that includes both the GnRH antagonist and the hormonal therapy (estradiol-norethisterone acetate) in the same tablet.

As noted in the Summary of Product Characteristics (SmPC),[7] linzagolix is available as a daily oral therapy in two doses, 100mg and 200mg, each of which may be prescribed with or without ABT (where ABT comprises estradiol 1 mg and norethisterone acetate 0.5 mg).

The company have submitted a confidential Patient Access Scheme (PAS) discount of [redacted] to NHS England.

2.2.3 The position of linzagolix in the treatment pathway

According to CS Table 2, the four possible regimens of linzagolix (i.e. 2 doses, with or without ABT) allow for flexible dosing options to support the individualised treatment need of women with uterine fibroids. In summary:

• The 100mg and 100mg + ABT regimens enable "partial suppression" of estradiol, controlling uterine fibroids while minimising BMD loss. The CS states that this is suitable for either short-term (≤6 months) or long-term (>6 months) treatment.

• The 200mg dose can be used for "full suppression" of estradiol but for long-term use (> 6 months) concomitant ABT is required to control symptoms whilst minimising BMD loss.

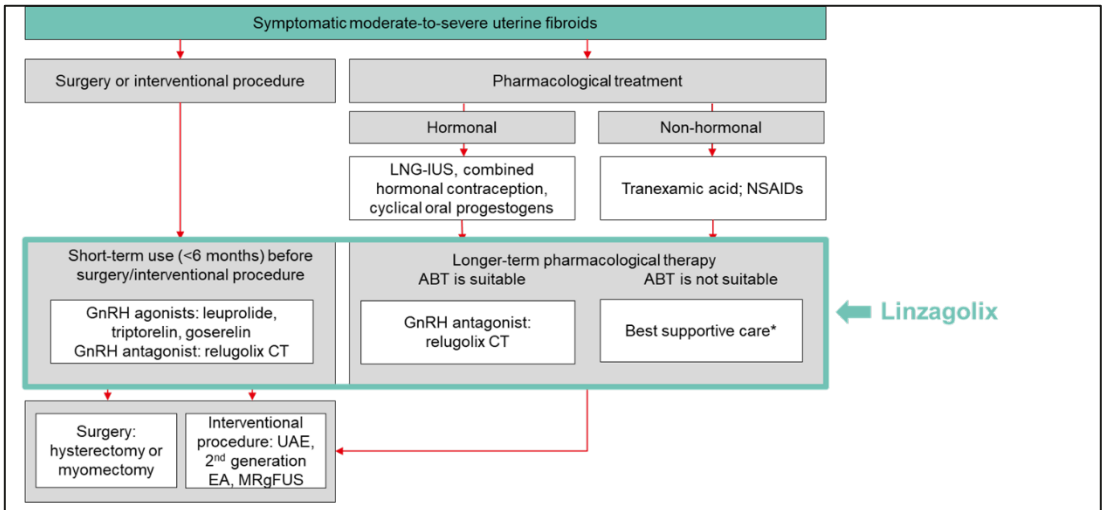
• The 200mg dose without ABT is suitable for short-term use when reduction of uterine and fibroid volume is desired, e.g. prior to surgery (such as myectomy or hysterectomy). NB this implies that whilst the 100mg dose without ABT can achieve symptom control, it does not provide the same magnitude of uterine or fibroid volume reduction as the 200mg dose.

According to the CS, the recommended dose of linzagolix is "100mg, or if needed 200mg, once daily with concomitant ABT". The CS does not provide any criteria for selecting whether the 100mg + ABT or 200mg + ABT dose is appropriate. The EAG assume that patients would likely be tried first on 100mg + ABT and if required for further symptom control the dose would be increased to 200mg + ABT, considering the patient's individual circumstances such as risk of osteoporosis. The SmPC recommends performing a dual X-ray absorptiometry (DXA) scan for patients with risk factors for osteoporosis, it does not recommend any particular dose of linzagolix.[7]

For patients in whom ABT therapy is not recommended, or those who prefer to avoid hormonal therapy, the recommended linzagolix dose is 100mg daily without ABT (CS Table 2). The CS specifies that contraindications to ABT include obesity, hypertension and dyslipidaemia; and women with an elevated risk of oestrogen- and progestogen-related side-effects (CS B.1.3.4.5) The size of this group of patients is uncertain and discussed further in section 3.2.1.1.5.1.

The care pathway is described in detail in CS section B.1.3.4, references the current NICE Guideline for heavy menstrual bleeding (NG88)[8], and is best summarised in CS Figure 5.

As suggested above, the CS has identified three relevant sub-populations for linzagolix therapy, i.e. patients having short-term treatment of 6 months or less (referred to by the company as Population #1); patients having longer-term treatment with hormone-based therapy (referred to as Population #2); and patients having longer-term treatment without hormone-based therapy (referred to as Population #3). The position of linzagolix in the treatment pathway for each of these sub-populations is summarised in CS Figure 6, which we have reproduced below in Figure 1.



* Best supportive care includes NSAIDs for pain management and iron supplements for blood loss
Abbreviations: ABT, add-back therapy; EA, endometrial ablation; LNG-IUS, levonorgestrel-releasing intrauterine system; MRgFUS: magnetic resonance-guided focused ultrasound; UAE, uterine artery embolization.

Source: Reproduction of CS Figure 6

Figure 1 The company's intended positioning of linzagolix in the treatment pathway

As shown in Figure 1, different comparators are relevant for each of these sub-populations. Furthermore, the company have used different economic analysis approaches for the sub-populations (cost comparisons for Populations #1 and #2 and cost-utility analysis for

Population #3). To help clarify the company's approach to the technology appraisal and assist interpretation of the company's Decision Problem (section 2.3 below), Table 3 provides an overview of the sub-populations, their relevant linzagolix dose regimens, comparators, and the company's economic analysis approaches.

Table 3 Summary of the company's approach to the technology appraisal

Sub-populations	Applicable linzagolix dose regimens	Comparators considered relevant by the company	Economic analysis approaches employed by the company
Population #1: Adults of reproductive age with moderate to severe symptoms associated with uterine fibroids having short-term treatment of 6 months or less	The company assume that patients receiving short-term treatment do not require ABT (CS section B.3.5.1) and the CS suggests 200mg would be used, without ABT, if shrinkage of uterine and fibroid volume is the primary aim (prior to surgery) (CS Table 2).	**Relugolix CT** (as explained later in this report, no direct in-trial comparisons exist so this comparison is made via network meta-analysis (NMA) comparing linzagolix in the pivotal placebo-controlled PRIMROSE trials against relugolix CT in the placebo-controlled LIBERTY trials using placebo as the common comparator) For detailed discussion of the PRIMROSE and LIBERTY trials see section 3.2.1; for explanation of the NMA approach see section 3.3.	**Cost comparison** Assumes linzagolix has similar clinical efficacy and safety to relugolix CT.
Population #2: Adults of reproductive age with moderate to severe symptoms associated with uterine fibroids having longer-term treatment, with hormone-based therapy	Either 100mg + ABT or 200mg + ABT		
Population #3: Adults of reproductive age	Either 100mg or 200mg, without ABT. The 200mg	**Best supportive care (BSC)[a]** (no alternative	**Cost-utility analysis** Follows the standard approach for Single

with moderate to severe symptoms associated with uterine fibroids having longer-term treatment, without hormone-based therapy	regimen is likely to be used short-term (the company base case assumes 200mg for 6 months followed by 100 mg, with a scenario of 100 mg all the way through) (CS B.3.5.1.1).	comparator without hormone-based therapy exists). This comparison is directly from the PRIMROSE trials assuming that the placebo arm represents BSC.	Technology Appraisals
a The company refer to established clinical management as 'best supportive care' (BSC) and we use this term in the current report for consistency. BSC is synonymous with established clinical management as stated in the NICE scope ABT: add-back therapy; BSC: best supportive care (=established clinical management); CT: combination therapy; NMA: network meta-analysis			

2.2.4 Characteristics of the appraisal populations

The CS does not discuss the size of these sub-populations in clinical practice. We note particular uncertainty regarding Population #3, i.e. those patients who are unable to or prefer not to receive hormonal ABT. The EAG's clinical expert thought Population #3 likely to be very small, since he had not seen many such patients with moderate or severe symptoms of uterine fibroids in his clinical practice. In contrast, the company's Market Research Survey[9] suggests that in the UK [redacted] of patients might be contraindicated and [redacted] unable to take ABT or would prefer to avoid hormone therapies, although it is unclear how relevant these data are to patients with moderate or severe symptoms, and methodological details of the company's survey are lacking. We have noted this uncertainty in the size of Population #3, together with uncertainty in how well Population #3 is supported with clinical evidence (discussed in detail section 3.2.1.1.5.1 below), as a key issue for further consideration (see Key Issue 3). The EAG are satisfied that Population #1 and Population #2 are relevant in clinical practice as they would cover most patients who would receive GnRH analogues,[9, 10] including those with moderate or severe symptoms of uterine fibroids, although the size of these groups in clinical practice is unclear. The EAG's clinical exert suggested that more patients would likely be in Population #1 than Population #2 but did not quantify this.

EAG conclusion on the condition and treatment pathway

The CS provides an accurate and thorough description of uterine fibroids and the associated symptoms. Details on the current treatment pathway are also accurate. The proposed position of linzagolix in the treatment pathway is either for short term use prior to surgery where the current treatments are GnRH agonists or relugolix CT, or for longer-term use where the current treatment is relugolix CT, or BSC if the

patient is contraindicated for hormonal therapy. Linzagolix has four dose regimens that enable this flexibility of use across the treatment pathway, however it is unclear what proportions of the indicated population correspond to each dose regimen.

2.3 Critique of the company's definition of the decision problem

For this appraisal, the company have submitted a cost-comparison analysis for Populations #1 and #2, and a cost-utility analysis for Population #3, as summarised in Table 3 above.

In a cost-comparison NICE appraisal, companies are not expected to provide a comparison of the intervention against all the comparators specified in the NICE scope.[10] Only one of the scoped comparators need be selected and should represent NICE recommended treatments as a whole in terms of costs and effects, and which has a significant market share. In the company's decision problem they have selected relugolix CT (relugolix combination therapy, i.e. includes hormonal add-back therapy; ABT) as the comparator in the cost-comparison analysis for Populations #1 and #2. CS section B.3.2.2 states that the selection of relugolix CT was based on the recommendations of NICE TA832 and clinical opinion.

The NICE guidance on cost-comparison appraisals[10] indicates that the intervention of interest (i.e. linzagolix regimens with or without ABT) should have similar clinical effectiveness and safety to the selected comparator(s) (i.e. relugolix CT). However, the EAG note that the clinical similarity of linzagolix to relugolix CT is uncertain, as explained in detail in section 3.5 of this report. This could have a bearing on whether cost-comparison analysis is an appropriate economic analysis approach for Population #1 and Population #2. We have therefore raised this as a key issue for further consideration (see Key Issue 1).

The company have not provided an estimate of the market share for relugolix CT when treating people with moderate or severe symptoms of uterine fibroids. The EAG's clinical expert estimated that relugolix CT currently has a low market share, with around 90% of patients in his practice receiving goserelin or leuprorelin, but he noted that, as relugolix CT is a relatively new therapy, its market share could increase. The expert also thought that general practitioners (GPs) are currently unfamiliar with relugolix CT. The EAG are uncertain whether the expert's observations are reflective of the use of GnRH analogues more widely in the NHS. As the market share of relugolix CT and other comparators could have a bearing on whether cost-comparison analysis is an appropriate economic analysis approach for Population #1 and Population #2 we have raised this as a key issue for further consideration (see Key Issue 2).

Table 4 below summarises the decision problem addressed by the company in the CS in relation to the final scope issued by NICE and the EAG's comments on this.

Table 4 Summary of the decision problem

	Final scope issued by NICE	Company's decision problem	Rationale if different from the final NICE scope	EAG comments
Population	People of reproductive age with moderate to severe symptoms associated with UFs	People of reproductive age with moderate to severe symptoms associated with UFs	Not applicable	The trial populations in the clinical evidence reported in the CS are mostly consistent with the NICE scope but exclude people at risk of BMD loss, and those with very large fibroids (see exclusion criteria CS Appendix M.1). It is also unclear whether they include people with pressure symptoms of uterine fibroids – see section 3.2.1 of this report.

	Final scope issued by NICE	Company's decision problem	Rationale if different from the final NICE scope	EAG comments
Subgroups	If the evidence allows the following subgroups will be considered: • People having short-term treatment of 6 months or less • People having longer-term treatment, with hormone-based therapy • People having longer-term treatment, without hormone-based therapy	• 1: People having short-term treatment of 6 months or less • 2: People having longer-term treatment, with hormone-based therapy • 3: People having longer-term treatment, without hormone-based therapy	Not applicable	The subgroups are consistent with the NICE scope. They are represented in the CS and in this report as Population #1, Population #2, and Population #3 – with respect to the numbering in the company's decision problem column to the left. The EAG are uncertain of the extent to which the trial populations represent populations #1, #2, and #3, see sections 3.2.1.1.4 and 3.2.1.1.5.
Intervention	Linzagolix (with or without hormone-based therapy)	Linzagolix (with or without hormone-based therapy)	Not applicable	This is consistent with the NICE scope.

	Final scope issued by NICE	Company's decision problem	Rationale if different from the final NICE scope	EAG comments
Comparators [a]	GnRH agonists (off-label for some GnRH agonists) [a] Relugolix-estradiol-norethisterone acetate Where hormone-based therapy is not suitable: established clinical management without linzagolix	GnRH agonists (off-label for some GnRH agonists) [a] Relugolix CT (relugolix-estradiol-norethisterone acetate) Where hormone-based therapy is not suitable: established clinical management without linzagolix (NSAIDs and iron supplements)	The company considers NSAIDs and iron supplements to be established clinical management for patients who cannot receive hormone-based therapy, based on guidelines and discussion with clinical experts	The CS includes relugolix-estradiol-norethisterone acetate (relugolix CT) as the main comparator for Population #1 (with GnRH agonists leuprorelin, goserelin and triptorelin in a supplementary comparison, CS Table 47) and relugolix CT as the only comparator for Population #2 which is appropriate for the NICE cost-comparisons if the intervention and comparator can be demonstrated to have similar clinical efficacy and safety. The company included other GnRH analogues as comparators for Population#1 and Population #2 in a clarification response. For Population #3 the company included best supportive care

	Final scope issued by NICE	Company's decision problem	Rationale if different from the final NICE scope	EAG comments
				(BSC) as the sole comparator which is appropriate for the cost-utility analysis for this population (which could not receive comparators containing hormone-based therapy). The EAG's clinical expert agreed with the company's definition of BSC except noting that tranexamic acid (an antifibrinolytic drug that reduces bleeding) would also be included in BSC in clinical practice, whilst iron supplements would be given specifically to anaemic patients. The EAG note that patients were prohibited from receiving tranexamic acid in the pivotal trials (PRIMROSE 1 CSR section 9.3.2).

	Final scope issued by NICE	Company's decision problem	Rationale if different from the final NICE scope	EAG comments
Outcomes	The outcome measures to be considered include: • Change in MBL volume • Time to MBL response • Pain • UF volume • Haemoglobin levels • Change in BMD • Rates and route of surgery • Impact on fertility and pregnancy and teratogenic effects • Mortality • AEs of treatment, including but not	• Change in MBL volume • Time to MBL response • Pain • UF volume • Haemoglobin levels • Change in BMD • Impact on pregnancy and teratogenic effects • Mortality • AEs of treatment, including but not limited to vasomotor symptoms and incontinence • HRQoL	Rates and route of surgery, impact on fertility, or pelvic organ prolapse were not specified endpoints in PRIMROSE 1 and PRIMROSE 2	The CS has only excluded outcomes which were not reported in the pivotal trials (impact on fertility and pregnancy and teratogenic effects; pelvic organ prolapse). Incontinence is reported in the CSRs, not the CS, as not enough adverse events occurred for the summary analysis. The EAG's clinical expert noted that pelvic organ prolapse may be a part of menopausal change and did not consider this to be important as an adverse effect of treatment.

	Final scope issued by NICE	Company's decision problem	Rationale if different from the final NICE scope	EAG comments
	limited to vasomotor symptoms, incontinence and pelvic organ prolapse • HRQoL			
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	• The most suitable type of economic evaluation varies between subgroups • For people having short-term treatment of 6 months or less and people having longer-term treatment with hormone-based therapy, where relugolix CT is the primary comparator	The blended approach to addressing the decision problem (an STA with cost-comparison methodology for a portion of the marketing authorisation population) was suggested by NICE and explored at the decision problem stage, and was considered	The company's economic approaches for analysing the three population subgroups are appropriate in principle. However, there is uncertainty whether linzagolix and relugolix CT have similar efficacy and safety (see Key Issue Issue 1 and report sections 3.4 and 3.5.) which is an assumption required for the cost-comparisons for Population #1 and Population #2.

	Final scope issued by NICE	Company's decision problem	Rationale if different from the final NICE scope	EAG comments
	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 availability and cost of biosimilar and generic products should be taken into account	of interest, cost-comparison methodology is used. This is based on population overlap between linzagolix and relugolix CT, findings from an indirect treatment comparison, clinical expert opinion, and guidance from NICE at the decision problem stage • For people having longer-term treatment without hormone-based therapy, where	appropriate by the company	

	Final scope issued by NICE	Company's decision problem	Rationale if different from the final NICE scope	EAG comments
		existing treatment options are limited, cost-effectiveness analysis is used, and expressed in terms of incremental cost per quality-adjusted life year		

Source: Reproduced from CS Table 1 with additional EAG comments.
Abbreviations: AEs: adverse effects; BMD: bone mineral density; GnRH: gonadotropin-releasing hormone; HRQoL: health-related quality of life; MBL: menstrual blood loss; NSAIDs: non-steroidal anti-inflammatory drugs; Relugolix CT: relugolix combined therapy (i.e. relugolix plus hormonal add-back therapy); UFs: uterine fibroids.
[a] There is a typographical error in the company's responses. The NICE scope refers to "GnRH analogues" which includes both GnRH agonists and GnRH antagonists; the company's description "agonists" should therefore read "analogues".

3 CLINICAL EFFECTIVENESS

3.1 Critique of the company's literature review methods

The company carried out three systematic literature reviews:

• one to identify randomised controlled trials (RCTs) for the clinical effectiveness evidence comparing linzagolix to placebo (CS section B.2.1 and CS Appendix D) and for a network meta-analysis (NMA) comparing linzagolix to relugolix CT (CS section B.2.9.1 and CS Appendix D),

• one for cost-effectiveness, cost, and healthcare resource use (CS section B.3.1 and CS Appendix G) which is discussed in section 4.1 of this report,

• and another for health-related quality of life (CS section B.3.4.3 and Appendix H which is discussed in section 4.2.6.1 of this report).

The EAG's full assessment of the methods of the clinical effectiveness review is summarised in Appendix 1. The review is generally comprehensive and appropriate for the decision problem. Searches were six months out-of-date when received by the EAG but we do not believe any relevant clinical efficacy studies were missed.

3.2 Critique of studies of the technology of interest and the company's analysis and interpretation of these

3.2.1 Included studies

The company systematic literature review identified six publications relating to the two company pivotal trials PRIMROSE 1 and PRIMROSE 2, and unpublished data relating to the company's extension trial PRIMROSE 3, that provide evidence on the efficacy and safety of linzagolix compared to placebo (CS section B.2.2).

The company additionally identified 19 publications relating to 14 studies for four GnRH analogue comparator therapies (relugolix CT, goserelin, leuprolide acetate, and ulipristal acetate) listed within CS Appendix D Table 7. These included the LIBERTY trials providing evidence for relugolix CT compared to placebo for use in the cost-comparison, and the PEARL trials providing evidence for ulipristal acetate compared to placebo and ulipristal acetate compared to leuprolide acetate for the company's indirect treatment comparison provided in clarification response A11. We believe it is likely that all relevant comparator studies were included.

As noted above (Table 3), the PRIMROSE trials provide evidence for the company's cost-utility analysis relevant to Population #3 (long-term treatment without ABT); whilst the PRIMROSE trials, the LIBERTY trials, and the PEARL trials provide evidence for the company's cost-comparison analyses relevant to Population #1 (short-term treatment with or without ABT) and Population #2 (long-term therapy with ABT). Characteristics of the PRIMROSE, LIBERTY, and PEARL trials are discussed below.

3.2.1.1 PRIMROSE trials and extension study

3.2.1.1.1 Role of the trials in the technology appraisal

The placebo-controlled PRIMROSE trials provide evidence of the clinical effectiveness and safety of linzagolix compared to BSC, with the placebo arms assumed to reflect BSC in clinical practice. This comparison is relevant to the company's cost-utility analysis for Population #3. The PRIMROSE trials are also used in indirect treatment comparisons of linzagolix against relugolix CT to support the company's cost-comparison analyses for Population #1 and Population #2, which assume that linzagolix and relugolix CT have similar clinical effectiveness and safety.

3.2.1.1.2 Study designs

PRIMROSE 1 and PRIMROSE 2 are large, completed, company-sponsored, phase III, multicentre, double-blind, RCTs; they had identical study designs but differed in location and patient baseline characteristics. PRIMROSE 3 is a completed company-sponsored off-treatment extension study for women who completed PRIMROSE 1 or PRIMROSE 2. Study characteristics are summarised in Table 5 below.

Table 5 Overview of the PRIMROSE trials

Study characteristics	PRIMROSE 1	PRIMROSE 2	PRIMROSE 3
Study ID	NCT03070899	NCT03070951	EudraCT 2021-000452-19
Study designs	Double-blind RCTs: 4 different dosing regimens of linzagolix vs placebo		Single arm, open-label, off-treatment extension study
Locations	USA (94 sites)	95 sites in USA and 8 European countries (no UK sites)	USA and Europe (no UK sites)

Study characteristics	PRIMROSE 1	PRIMROSE 2	PRIMROSE 3
Populations	Women aged ≥18 years with ultrasound-confirmed uterine fibroids (between 2-12 cm diameter) and heavy menstrual bleeding (≥80 mL MBL per cycle for at least 2 cycles)		Women completing either PRIMROSE 1 or 2 and who had a DXA scan within 35 days from the last treatment administration
Randomisation	1:1:1:1:1; stratified according to race (Black or African American vs other)		Not applicable: open-label, off-treatment
Regimens and participants	Randomised N=574; FAS N=511 Placebo N=103 Linzagolix 100 mg N=94 Linzagolix 100 mg + ABT N=107 Linzagolix 200 mg N=105 Linzagolix 200 mg + ABT N=102	Randomised N=535; FAS N=501 Placebo N=102 Linzagolix 100 mg N=97 Linzagolix 100 mg + ABT N=101 Linzagolix 200 mg N=103 Linzagolix 200 mg + ABT N=98	[redacted] Enrolled: [redacted] Completed: [redacted]
Primary outcome	Reduction in heavy menstrual bleeding at 24 weeks, reported as response (proportion of patients achieving the outcome); measured by the alkaline haematin method		Change in BMD at 12, 18 and 24 months from end of treatment in PRIMROSE 1 or 2
Duration and treatment switching	After 24 weeks (primary outcome), a second treatment period ran up to week 52 where treatment switching occurred for linzagolix 200 mg → linzagolix 200 mg + ABT in both trials, for 50% of the placebo group in PRIMROSE 1→ linzagolix 200 mg + ABT, and for all of the placebo group in PRIMROSE 2. There was no treatment switching in any of the other trial arms which		24 months. The study is complete.

Study characteristics	PRIMROSE 1	PRIMROSE 2	PRIMROSE 3
	also ran up to week 52. After this there was a follow-up period of no treatment for all study arms up to week 76. The trials are complete.		

Sources: CS sections B.2.3.1, B.2.3.4, CS Appendix M, and trial publication (Donnez et al. 2022;[11]) CS section B.2.11; PRIMROSE 3 study CSR.[12]
Abbreviations: ABT: add-back therapy; BMD: bone mineral density; DXA: dual-energy x-ray absorptiometry; FAS: full analysis set; MBL: menstrual blood loss; RCTs: randomised controlled trials.

Treatment switching was applied for practical and ethical reasons (Clarification Response A4b). Full suppression of serum estradiol (200 mg linzagolix without ABT) cannot be received for >6 months due to the impact on bone mineral density. Hence the treatment switching for this treatment group to 200 mg with ABT. The company explained that the treatment switch from placebo to 200 mg linzagolix without ABT was to support study participants to continue therapy and therefore avoid high discontinuation rates in that group (Clarification Response A4a). However, this switch was applied differently in the two trials as only 50% (selected at randomisation) of the placebo group in PRIMROSE 1 was switched. Nonetheless, there is no placebo comparator data, with sufficient statistical power, beyond 24 weeks for either of the trials and therefore for the cost-utility analysis.

Thus the available evidence appears to be relatively short-term considering that some of the linzagolix dosing regimens may be used for more than one year in clinical practice (subject to regular bone mineral density monitoring).[7] The company argue that because linzagolix leads to a dose-dependent reduction in serum estradiol and progesterone, and because it is well-known that fibroids are hormone-dependent, then as long as serum estradiol suppression is maintained clinical effectiveness is also expected to be durable (Clarification Response A7). The EAG's clinical expert agreed that this is likely. The company also point to the LIBERTY randomised withdrawal study for relugolix CT, which has a similar mechanism of action to linzagolix, and those trial results show durable effect in maintaining low MBL volume and amenorrhoea for over two years (Clarification Response A7). The EAG view the long-term similarity of linzagolix to relugolix CT to be plausible but speculative because the group of patients in the 52-week LIBERTY randomised withdrawal study who received relugolix CT for the whole 2-year period (beginning with LIBERTY 1 or 2), was relatively small (N=46), with limited results reported for this group.[13]

3.2.1.1.3 Pooled analysis of the trials

The CS primarily focuses on pooled data from PRIMROSE 1 and PRIMROSE 2 up to week 24 (prior to treatment switching), with results of the individual PRIMROSE trials reported in CS Appendix M. The strengths and limitations of the pooled analysis approach are discussed in section 3.2.4.2 below.

3.2.1.1.4 Relevance of the placebo arms to BSC in clinical practice

The placebo treatment group in the PRIMROSE 1 and 2 trials represents BSC (the comparator for Population #3 in the cost-utility analysis) which the company describe as consisting of NSAIDs and iron supplements (CS Table 1).

The EAG's clinical expert suggested that the company's interpretation of BSC does not fully reflect clinical practice. As well as NSAIDs and iron supplements, tranexamic acid would be prescribed according to the NG88 guidelines for management of heavy menstrual bleeding,[8] with prophylactic doses of calcium and vitamin D also given to protect against bone loss (doses as noted in section 4.2.4 of this report relating to the economic model). Expert opinion reported in the company's market research survey suggests that [redacted] patients (around [redacted]) on long-term pharmacological treatment in the UK would receive tranexamic acid. However, tranexamic acid is not part of the company description of BSC and patients were not permitted to receive tranexamic acid in the PRIMROSE trials. The EAG conclude that the effects of tranexamic acid on bleeding control that could be experienced by patients in clinical practice would not be reflected in placebo arms of the PRIMROSE trials, although the significance of this is unclear.

3.2.1.1.5 Trial population characteristics

3.2.1.1.5.1 Exclusion of patients unable to take hormonal therapy (Population #3)

Patients with moderate to severe symptoms of uterine fibroids who could not take hormonal ABT or preferred not to take hormonal ABT were excluded from the PRIMROSE trials (CS Document A, section A.9). As noted above, this group is referred to by the company as Population #3 and is the subject of their cost-utility analysis (Table 3) but is not represented by any patients in the trials. According to CS Document A (section A.9) an unspecified number of clinical experts agreed that the patients in the 100mg and 200mg arms of the PRIMROSE trials (i.e. those not randomised to receive ABT) are a suitable proxy population for those patients who cannot or prefer not to take hormonal ABT. The EAG are uncertain whether this is a reasonable proxy and the company do not provide a rationale for why it should be. We note that there may be several reasons why patients are unable to take

hormonal ABT, including having a history or risk of thrombosis, having diabetes, being a smoker, having a history of cancer, or personal preference. Given the uncertainty in whether patients in the PRIMROSE trials are appropriate as a proxy for those who cannot or prefer not to take ABT, as well as uncertainty in the size of this patient group in clinical practice (section 2.2.3), we have highlighted this uncertainty relating to Population #3 as a key issue for further consideration (see Key Issue 3).

3.2.1.1.5.2 Exclusion of patients eligible for surgery or with large fibroids indicative of surgery (Population #1)

Patients were excluded from participating in the PRIMROSE trials if their condition was so severe that they would require surgery within 6 months regardless of the treatment provided. This means that patients relevant to Population #1 were excluded. In addition, patients with fibroids over 12 cm in diameter were also ineligible to participate in the PRIMROSE trials. Such patients could be eligible for surgery or might benefit from fibroid reduction provided by linzagolix 200 mg without ABT. The linzagolix SmPC does not exclude these people. We highlight this uncertainty relating to how well the PRIMROSE trials represent Population #1 as an additional issue for further consideration (see Key Issue 3).

3.2.1.1.5.3 Exclusion of patients receiving long-term therapy (Population #2)

The PRIMROSE trials had a relatively short duration and patients receiving therapy for longer than 52 weeks are not represented.

3.2.1.1.5.4 Participants' baseline characteristics

Patient baseline characteristics for the individual trials are reported in CS Appendix M.3.2. In PRIMROSE 1 all treatment groups had similar baseline characteristics including similarity of prognostic characteristics, and likewise for PRIMROSE 2. Therefore, we agree with the CS that the within-trial baseline characteristics are generally comparable and that there is low risk of bias for any imbalance across treatment groups.

Most patients in PRIMROSE 1 were of Black race (61% to 65% across all trial arms) which differs from the PRIMROSE 2 population (4% to 6% were of Black race across all trial arms). Black race is a key risk factor for the development of uterine fibroids (section 2.2.1.2). The EAG's clinical expert commented that both PRIMROSE trials could reflect the proportion of Black people seen in clinical practice: PRIMROSE 1 would reflect those NHS Trusts who see a lot of Black patients, e.g. in London, whereas PRIMROSE 2 would be more representative of the population in the expert's clinical practice in Southampton. There do

not appear to have been any Asian women included in the PRIMROSE trials, although the EAG's clinical expert estimated that Asian patients make up 5-10% of the patients at his clinical practice and noted that Asian women not receiving HRT have greater risk of BMD loss if they continue on any therapy that does not include HRT so have narrower treatment options than lower-risk patients. Overall, the EAG's clinical expert agreed that, aside from not including Asian patients, both PRIMROSE trials are generally representative of the population likely to be seen in NHS clinical practice: the individual trial results (CS Appendix M) would reflect local population characteristics whilst the pooled trial results would reflect the overall population mix.

The EAG's clinical expert noted that BMD in the trials was in the normal range for most patients: this is consistent with the trials' exclusion criteria for BMD loss (which we assume reflect ethical considerations relating to the risk of worsening osteoporosis in susceptible patients).

Important fibroid characteristics (location, number, size; i.e. FIGO classification) are not reported in the patient baseline characteristics, although the FIGO classification appears to have been assessed in the trials, as the SmPC (section 5.1)[7] states "97.5% had FIGO classification from 1-6", and there was a pre-specified subgroup of participants with submucosal fibroids, FIGO 0, 1 or 2 at baseline, to evaluate impact on the primary outcome. The EAG's clinical expert confirmed that the FIGO classification is routinely used in clinical practice, and it is important because (as noted above in section 2.2.1.1), the type and location of a fibroid has a major influence on a patient's symptoms and prognosis. This view is supported by an independent academic paper which further states that lack of this clinical information restricts the predictability of expected effectiveness in a real-world population.[14]

Pressure symptoms are not part of the inclusion or exclusion criteria of the PRIMROSE trials. These can range from discomfort to pain, and influence mobility, urination, and constipation issues. These symptoms may be implied in the pain and quality of life assessments at baseline and in the results although this is not made clear. Patients with only subserosal fibroids were excluded from the trials and the EAG's clinical expert confirmed that these fibroids may be completely asymptomatic or have pressure symptoms. It is unclear if patients with pressure symptoms are included in the trial population or if a patient group is missing.

3.2.1.2 LIBERTY trials and extension studies

3.2.1.2.1 Role of the trials in the technology appraisal

The LIBERTY trials are used in indirect treatment comparisons of linzagolix against relugolix CT to support the company's cost-comparison analyses for Population #1 and Population #2, which assume that linzagolix and relugolix CT have similar clinical effectiveness and safety.

3.2.1.2.2 Study designs

LIBERTY 1 and LIBERTY 2 are completed, international, multicentre, phase III, double-blind, randomised controlled trials that investigated the efficacy and safety of relugolix CT for 24 weeks. A critique of LIBERTY 1 and 2, as applicable to the NMA for which they provide evidence, is in section 3.3.2 of this report. Study characteristics are summarised in Table 6 below, with further details on eligibility criteria in CS Appendix D.3.3.2.

Table 6 Overview of the LIBERTY trials

Study characteristic	LIBERTY 1	LIBERTY 2
Study ID	NCT03049735	NCT03103087
Study designs	Replicate double-blind RCTs: 3 arms: placebo, relugolix CT, and delayed relugolix CT (relugolix monotherapy followed by relugolix CT at 12 weeks)	
Locations	Africa, Europe, North America, South America	
	80 sites, 1 in the UK (number of UK patients not stated)	99 sites, none in the UK
Populations	Premenopausal women aged 18-50 years with ultrasound-confirmed fibroids and heavy menstrual bleeding (≥80 mL per cycle for 2 cycles or ≥160 mL for 1 cycle)	
Randomisation	1:1:1, unstratified	
Regimens and participants	Randomised N=388 Completed N=308	Randomised N=382 Completed N=302
	Placebo N=127 Relugolix CT N=128 Delayed relugolix CT[a] N=132	Placebo N=129 Relugolix CT N=125 Delayed relugolix CT[a] N=127

Study characteristic	LIBERTY 1	LIBERTY 2
Primary outcome	Reduction in MBL at 24 weeks reported as response (proportion of patients achieving the outcome); measured by alkaline haematin method	
Duration	24 weeks	

Sources: CS Appendix D.3.3.2, trial publications (Al-Hendy et al. 2021[15], Stewart et al. 2022[16]).
Abbreviations: LTE: long-term extension; MBL: menstrual blood loss; RCTs: randomised controlled trials.
[a] delayed relugolix CT: relugolix monotherapy for 12 weeks followed by relugolix CT for 12 weeks.

3.2.1.2.3 Relevance of the placebo arms to BSC in clinical practice

The therapies received by patients in the placebo arms of the LIBERTY trials are not well described in the trial publications.[15, 16] However, the trial protocol (available from clinicaltrials.gov) confirms that patients in the LIBERTY trials could not receive tranexamic acid. Therefore, as with the PRIMROSE trials described above (section 3.2.1.1.4), the placebo arms in the LIBERTY trials are not fully reflective of clinical practice.

3.2.1.2.4 Trial population characteristics

Baseline characteristics for participants in the individual LIBERTY 1 and LIBERTY 2 trials are provided in CS Appendix Table 9. Baseline characteristics for LIBERTY 1 were similar in both treatment groups except that actual menstrual blood loss was slightly lower in the placebo group at 218.8 mL compared to 239.4 mL in the relugolix CT group, although a similar proportion of patients had MBL<225 mL in both groups (67% and 66% respectively). Likewise for LIBERTY 2, baseline characteristics were similar in both treatment groups except that actual menstrual blood loss was slightly lower in the placebo group at 211.8 mL compared to 246.7 mL in the relugolix CT group, although a similar proportion of patients had MBL<225 mL in both groups (67% and 64% respectively). The study publication states that the demographic and clinical characteristics of the participants at baseline were similar across the trial groups.[15] The EAG's clinical expert confirmed that the baseline characteristics that are reported are consistent with what he would expect to see in clinical practice.

The characteristics are similar to those of the participants in the PRIMROSE trials, except that there were proportionally more Black people (almost half the population was Black compared to approximately one third of the population in the pooled analysis of the PRIMROSE trials). Similarly, FIGO classification, number and location of fibroids are not

reported. For a comparison of the baseline population characteristics of the LIBERTY and PRIMROSE trials see section 3.3.3.1.

3.2.1.3 PEARL trials

3.2.1.3.1 Role of the trials in the technology appraisal

The PEARL trials are used in NMA comparisons of linzagolix against leuprolide acetate to support the company's cost-comparison analyses for Population #1 and Population #2 (Clarification Response A11), which assume that linzagolix and other GnRH analogues have similar clinical effectiveness and safety.

3.2.1.3.2 Study designs

PEARL I and PEARL II are completed, randomised, multi-centre, phase III, double-blind, RCTs. PEARL I evaluated two different doses of ulipristal acetate versus placebo (i.e. three arms) in an anaemic population with symptomatic uterine fibroids. PEARL II evaluated two different doses of ulipristal acetate versus leuprolide acetate in a non-anaemic population with symptomatic uterine fibroids. Study characteristics are summarised in Table 7 below, with further details on eligibility criteria in CS Appendix D.3.3.3 and D.3.3.4.

Table 7 Overview of the PEARL trials

Study characteristics	PEARL I	PEARL II
Study ID	NCT00755755	NCT00740831
Study designs	Double-blind RCT	Double-blind RCT
Locations	Czech Republic, Hungary, India, Romania, Russia, Ukraine. No UK sites.	Austria, Belgium, Germany, Israel, Italy, Netherlands, Poland, Spain. No UK sites.
Populations	Premenopausal women aged 18-50 years and BMI 18-40 with excessive uterine bleeding (PBAC score >100) caused by fibroids (at least one fibroid ≥3 cm, none >10 cm) and anaemia (Hb ≤10.2 g/dL) for whom surgery is indicated. All participants were eligible for surgery after week 13.	Premenopausal women aged 18-50 years and BMI 18-40 with excessive uterine bleeding (PBAC score >100) caused by fibroids (at least one fibroid ≥3 cm, none >10 cm) for whom surgery is indicated. All participants were eligible for surgery after week 13.

Study characteristics	PEARL I	PEARL II
		(Anaemia was not an inclusion criterion)
Randomisation	2:2:1; stratified for haematocrit level (≤28% or >28%) and for race (Black or other)	1:1:1; stratified for race or ethnic group [race and ethnic groups not specified]
Regimens and participants	Ulipristal acetate 5 mg N=96 Ulipristal acetate 10 mg N=98 Placebo N=48 All study arms included 80 mg iron tablets (participants were anaemic).	Ulipristal acetate 5 mg N=98 Ulipristal acetate 10 mg N=104 Leuprolide acetate N=101
Primary outcome	Co-primary outcomes: Proportion of patients with reduction in uterine bleeding (PBAC score <75) at week 13. Change in total fibroid volume at week 13.	Proportion of patients with control of uterine bleeding at week 13 (PBAC score <75). (Change in fibroid volume was a secondary outcome)
Duration	13 weeks; plus off-treatment follow up to week 38	13 weeks; plus off-treatment follow up to week 38

Sources: CS Appendix D.3.3.3 and Donnez et al. 2012[17]; CS Appendix D.3.3.4 and Donnez et al. 2012[18]
Abbreviations: Hb: haemoglobin; PBAC: pictorial blood-loss assessment chart; RCT: randomised controlled trial.

3.2.1.3.3 Trial population characteristics

There is slight variation in baseline characteristics (where reported) between the arms within each of the PEARL trials.[17, 18] The EAG conclude that although there is within-trial heterogeneity in baseline population characteristics in each PEARL trial, there is no indication of systematic differences between arms in disease severity or prognostic factors.

Due to differences in reporting, only haemoglobin, age, BMI, uterine volume, and race can be directly compared between the two PEARL trials. Age and BMI were similar in both trials. PEARL had lower haemoglobin, reflecting the anaemic population, and higher total uterine volume, than PEARL II. Most patients (≥84%) in both trials were White. PEARL I had 10% to

15% Asian and no Black patients; whilst PEARL II had few if any Asian patients (only reported as “other”), and 9% to 11% were Black patients.

For a comparison of baseline characteristics between the PEARL, PRIMROSE and LIBERTY trials see section 3.3.3.2.

The EAG believe that there were no major outliers in baseline characteristics in comparison with the PRIMROSE and LIBERTY trials which were assessed by our clinical expert as representative of UK clinical practice, and therefore consider that the PEARL trials are likely to represent UK clinical practice with the caveat that PEARL I represents an anaemic population and in which the placebo group was prohibited from receiving tranexamic acid.

EAG conclusion on the included trials

All included trials are RCTs with comparative data up to 24 weeks for PRIMROSE and LIBERTY, and up to 13 weeks for PEARL, which is short-term relative to expected longer-term treatment of >6 months outlined in the proposed position in the treatment pathway. However, patient subgroups identified by the company as relevant to their decision problem and economic analyses are missing from the PRIMROSE trials, namely those eligible to receive surgical interventions (Population #1), those who would receive long-term treatment (Population #2), those unable to receive hormonal therapy (Population #3), and patients who would receive tranexamic acid. It is unclear how important these exclusions are, for example whether patients in the trials could be a suitable proxy for those not included. We have highlighted this for further consideration; see Key Issue 3.

3.2.2 Risk of bias assessment

3.2.2.1 Risk of bias assessment for the PRIMROSE trials

A critical appraisal of PRIMROSE 1 and PRIMROSE 2, where they were assessed together as ‘identical’ trials, using the NICE checklist for RCTs, is reported in CS section B.2.5. The EAG have critically appraised the trials in Appendix 2 of this report. The company included an additional assessment of the PRIMROSE trials in the relevant ITC section (CS Appendix D.3.7), but this was based on conference abstracts for the trials and the full publications are now available. In general, we believe the trials are at low risk of bias, with a few exceptions. There is unclear risk of selection bias due to lack of reporting of fibroid characteristics, which are important prognostic factors, although in the reported characteristics including other prognostic factors such as race and BMI there was balance across the treatment groups.

There is unclear risk of detection bias and reporting bias due to the way in which fibroid volume was measured (see section 3.2.3.1.3). There is also unclear risk of attrition bias due to the imbalance in the number of dropouts between groups which mostly affects the uterine and fibroid volume outcomes in PRIMROSE 1 where there was the most missing data.

3.2.2.2 Risk of bias assessment for the LIBERTY trials

Risk of bias assessments for LIBERTY 1 and 2 are reported in CS Appendix D.3.7. The company has not included justifications for their assessments, nor provided any overall statement of risk of bias. Therefore the EAG has assessed the LIBERTY trials for risk of bias using the NICE checklist, consistent with use of the NICE checklist for the PRIMROSE trials, and our completed checklist is in Appendix 2. Overall, our assessment of the trials is that they are at low risk of bias in all domains. There was an unclear risk of selection bias due to lack of reporting of the method of randomisation and lack of reporting of important prognostic fibroid characteristics, however, the reported participant baseline characteristics were generally similar across the treatment groups within each trial (section 3.2.1.2.4).

3.2.2.3 Risk of bias assessment for the PEARL trials

The PEARL trials were included by the company in this technology appraisal in response to Clarification Question A11 to extend the NMA network to include all GnRH analogues relevant to the NICE scope. The EAG have assessed these trials for risk of bias using the NICE checklist (see Appendix 2), consistent with use of this checklist for the PRIMROSE and LIBERTY trials. Overall, our assessment of both trials is that PEARL I is at low risk of bias and that PEARL II is at low risk of bias except for unclear risk of bias around non-reporting of fibroid characteristics.

EAG conclusion on risk of bias

All trials in this appraisal (PRIMROSE 1 and 2, LIBERTY 1 and 2, PEARL I and II) are mostly at low risk of bias across the various domains of bias. Where risk of bias is unclear this is mainly due to lack of reporting of FIGO classification for prognostic fibroid characteristics which is a consistent omission across all trials except PEARL I.

3.2.3 Outcomes assessment

3.2.3.1 Clinical effectiveness outcomes

3.2.3.1.1 Primary outcome

Response. The primary outcome in PRIMROSE trials (also primary in the LIBERTY trials) was the proportion of patients who achieved a reduction of HMB at week 24, defined as MBL

≤80mL and with a ≥50% reduction in MBL from baseline in the last 28 days before the week 24 visit (also assessed at week 52). This dichotomous outcome is referred to by the company as response. The EAG's clinical expert agreed that this represents a clinically meaningful reduction in HMB.

In the PRIMROSE and LIBERTY trials MBL was assessed using the alkaline haematin (AH) method which involves chemically measuring the blood content of used sanitary products and is considered the 'gold standard' approach.[19] All used sanitary products were sent to a central laboratory masked to the trial treatment for analysis and assessment of daily MBL (CS Table 10). In the PEARL trials MBL was assessed using a different approach: the pictorial blood assessment chart (PBAC) score, which records the number of tampons or towels used and the degree to which they are stained with blood. The PEARL trials therefore used a different definition of response, which was a PBAC score less than 75 (in the normal range), summed over the preceding 28-day period.

Estimates of MBL using the PBAC and AH approaches are generally correlated.[19] However, as noted by the company in TA832, calibration coefficients from the PBAC to the AH method differ between studies and publicly available data are required to enable translation of PBAC into AH measurement to enable comparison of the MBL estimates. Such information is available from the PEARL trial publications but not for other studies that might potentially be included in evidence networks for NMA such as the studies identified by the company that were conducted in Japan.[20-22] As noted in TA832, the specific conversion factor is not likely to be the same between the PEARL and Japanese trials which would make it incorrect to use the same calculations and translations between PBAC and AH across them. The Evidence Review Group in TA832 suggested that it might be possible to convert from PBAC score to MBL volume, using the approach adopted in Magnay et al. 2020,[19] but this depends on data availability in the trial publications. For the Japanese study by Osuga et al. 2019 this information could not be found even when the EAG had access to the study protocol and CSR.

3.2.3.1.2 Ranked secondary outcomes

These secondary outcomes were assessed in the following sequence (highest priority first) to protect against the risk of multiple testing inflating the overall type 1 error rate (CS Table 11). For statistical discussion please see section 3.2.4.4.

Time to reduced HMB. This was defined as the number of days from Day 1 of treatment to the first day the woman reached the definition of HMB (as defined for the primary outcome) and MBL was maintained up to week 24 and up to week 52.

Amenorrhoea (absence of bleeding). This was defined as having no sanitary material returned or the MBL volume below the lower limit of quantification within at least a 35-day interval maintained up to week 24 and up to week 52.

Time to amenorrhoea, defined as the number of days from Day 1 to the first day the woman reached the definition of amenorrhoea and without having bleeding after this time up to week 24 and up to week 52.

Number of days of uterine bleeding in the last 28-day interval before week 24 and before week 52.

Haemoglobin concentrations were assessed in a subgroup of patients who had anaemia (defined in CS Table 10 as Hb<12g/dL) whose baseline Hb was <10.5 g/dL (CS section B.2.9.6.5). For the comparison of linzagolix against placebo the CS reports differences in the change from baseline in Hb concentration (CS Table 21). However, for the comparison of linzagolix against relugolix CT the CS reports only the percentage change in Hb concentration from baseline (CS section B.2.9.6.5). This is less clinically informative but appears to reflect that only the percentage change is available from the LIBERTY trials.[15]

3.2.3.1.3 Additional outcomes

Pain related to uterine fibroids was assessed by patient self-report using an on-site eDiary with a numerical rating scale (NRS) from 0 (no pain) to 10 (worst possible pain) over the preceding 28 days. The NRS scores were categorised as 0 = none, 1 to 3 = mild, 4 to 6 = moderate, 7 to 10 = severe. Assessed at several timepoints including weeks 24 and 52. The CS does not report any validation or testing of this scale, although the EAG's clinical expert commented that a similar basic scale would be used in clinical practice.

In the PRIMROSE trials pain was reported as the mean and categorical changes in scores as well as the proportion who achieved a clinically meaningful pain response, defined as those who had a numerical rating scale score of at least 4 (indicating moderate or severe pain) at baseline and achieved a score of 1 or less at Week 24. The EAG's clinical expert agreed that achieving an NRS score of 1 or less would represent a meaningful clinical improvement. In the LIBERTY trials only the proportion achieving a clinically meaningful improvement in pan score was reported.

Primary fibroid volume was estimated using the same ultrasonography approach as for uterine volume. CS Table 10 states that “up to the three largest fibroids were included in the volume calculation” without explanation of why or when more than one fibroid would be included in the calculation. The EAG are uncertain whether this could be a source of bias, for instance if certain subgroups of the population (e.g. those with specific race, BMI or other anatomical characteristics) might have differed systematically in the way that their fibroid volume was calculated. The EAG also note that patients with large fibroids >12 cm diameter were excluded from the trials, imposing a ceiling effect on this outcome. Moreover, the CS does not report which types of fibroids were the largest in a given patient, e.g. whether they were situated inside or outside of the uterus. Due to these limitations the EAG believe that results for this outcome might not be consistently reliable when making comparisons between therapies or patient groups, so we have noted this as being a potential, although unclear, risk of bias (section 3.2.2.1). We also note that the total fibroid volume, potentially a more reliable measure, depending on how it was calculated, was available for the PRIMROSE trials when used in the NMA comparison of linzagolix against relugolix CT (section 3.3.3.1) but not reported by the CS for the within-trial comparison of linzagolix against placebo.

The reporting of primary fibroid volume is not consistent between the PRIMROSE within-trial comparisons of linzagolix against placebo (section 3.2.5.2.1) and the company’s indirect comparisons of linzagolix against GnRH analogues (section 3.5.1.4). In the PRIMROSE trials analyses, primary fibroid volume is reported as the least-squares mean ratio to placebo which has unclear clinical interpretation. It is unclear why the change in fibroid volume relative to the placebo arm, as used in the indirect treatment comparisons was not reported instead.

Uterine volume was estimated using ultrasonography (transvaginal, or abdominal if transvaginal was not available), done by the same operator at all visits where feasible. Due to the different ultrasound methods, operators and inherent imprecision of measurements the EAG believe some variation in this outcome is to be expected.

3.2.3.2 HRQoL outcomes

HRQoL outcomes are specified as a separate class of outcome (i.e. not “secondary” or “additional”). CS Table 11 does not mention any statistical interpretation criteria for these outcomes.

Disease-specific symptom severity and HRQoL were assessed using the 3-month recall version of the UFS-QoL (Uterine Fibroid Symptom-Quality of Life) questionnaire. Assessments were completed by participants on site using an eDiary set up by site staff. UFS-QoL is a validated measure comprising an 8-item symptom severity scale which measures a patient's objective symptoms (e.g. bleeding, cramping) and a 29-item HRQoL scale which measures a patient's subjective HRQoL experience based on six subdomains (concern, activities, energy/mood, control, self-consciousness, sexual function) (CS section B.2.6.2.7). Each symptom severity or HRQoL item is scored on a 5-point Likert scale and then scores are summed and transformed to give two 0-100 scales, one for symptom severity with improvement indicated by lower scores, and one for HRQoL with improvement indicated by higher scores. The CS does not state what a minimum clinically important difference or change in UFS-QoL score would be (results are presented in the CS as changes in mean scores without a numeric clinical interpretation). A recent (2023) study[23] states that a specific minimal clinically important difference for the UFS-QoL has not been established, but a difference of 10 points appears reasonable, whilst a previous study [24] suggested a change of 9-15 points could indicate a clinically meaningful improvement for the symptom severity or HRQoL scales. The EAG assume that the same minimum clinically important difference would apply to both the severity and HRQoL scales (this was not stated in the Anchan et al. study[23] but the results reported by Harding et al.[24] showed the severity and HRQoL scales to be within the 9-15 point range).

EQ-5D-5L index and visual analogue scale (VAS) scores were also assessed using on-site eDiaries. The company argue that "As the effects of fibroids are complex, and patients may report differently depending on exactly which timepoint in their menstrual cycle they complete the EQ-5D assessment, a singular measurement on a single day may not truly reflect patients' overall HRQoL. These issues raise questions as to the degree of validity and reliability of the EQ-5D scores from the PRIMROSE trials." As such the company favoured the UFS-QoL for assessing patients' HRQoL.

3.2.3.3 Safety outcomes

The CS reports the numbers of adverse events and the proportion of participants experiencing them that were reported by >2% in at least one active treatment group, for both treatment periods, up to Week 24 and Weeks 24-52 (CS B.2.10). This included vasomotor symptoms associated with hormonal treatment such as hot flushes and headache. Additionally, BMD was assessed as an adverse event of special interest due to the known effects of oestrogen suppression relating to osteoporosis. Assessments were made at three anatomic sites (lumbar spine, femoral neck, and total hip). The EAG's clinical expert

suggested that a 5% change in BMD would be clinically meaningful. Week 76 results from PRIMROSE 1 and 2, and unpublished data from the PRIMROSE 3 extension study, for which the primary outcome was change in BMD at 12, 18 and 24 months from end of treatment in PRIMROSE 1 or 2, were provided to support this safety outcome.

EAG conclusion on the outcomes assessment

The outcomes assessed are appropriate for the condition but the CS does not discuss the degree of change for each outcome that would be considered clinically meaningful. The EAG have particular concerns around the assessment and reporting of primary fibroid volume, which might be a source of bias. The company prefer the disease-specific UFS-QoL measure of HRQoL than the EQ-5D-5L, which the EAG agree is appropriate (based on precedent in TA832 and the opinion of the EAG's clinical expert).

3.2.4 Statistical methods of the included studies

3.2.4.1 Analysis populations

The intention to treat (ITT) population is not used in company analyses. Instead, the clinical effectiveness outcomes were analysed using the full analysis set (FAS) and safety outcomes were analysed using the safety analysis set (SAS).

The FAS was defined as all randomised patients in PRIMROSE 1 and PRIMROSE 2 who received at least one dose of double-blind study drug irrespective of the treatment received and who did not violate the following two exclusion criteria prior to first administration of double-blind study drug (based on the results of pre-treatment baseline assessments reported after Day 1): (1) significant risk, history of or known osteoporosis or other bone metabolic disease; (2) liver function test results ≥2 times the upper limit of normal. The difference in patient numbers between the randomised population (ITT) and the FAS ranged from 9 to 20 patients per trial arm (7% to 11%) in PRIMROSE 1 and from 4 to 10 patients per trial arm (4% to 9%) in PRIMROSE 2 (CS Appendix Table 51).

The EAG requested a sensitivity analysis using the ITT population to enable an assessment of the robustness of the FAS-based analyses to these missing data. Instead of providing this, the company provided a detailed explanation of why they believed an ITT analysis to be inappropriate - which was because the patients excluded from the FAS had failed the trial eligibility criteria after randomisation when they received delayed test results and therefore would be ineligible for inclusion. We note that the linzagolix EPAR does not provide an opinion on whether the FAS population is appropriate.[25] However, the EAG agree with the company's rationale for not employing an ITT analysis as reported in Clarification Response

A9. We note that the numbers of patients missing from the FAS were reasonably well-balanced across the trial arms and their baseline characteristics appear to be generally similar to those of the FAS population, with no clear signals of any systematic imbalances between the trial arms (Clarification Response Table 1). As noted above, the number of missing patients from the FAS was larger in PRIMROSE 1 than in PRIMROSE 2.

A per protocol (PP) population is also defined (CS Appendix M.3.1) and used in a sensitivity analysis (CS Appendix M.3.3.2) (see section 3.2.4.6 below).

The SAS was defined as all randomised patients in PRIMROSE 1 and 2 who received at least one dose of double-blind study drug irrespective of the treatment received. Patients were analysed according to treatment received.

3.2.4.2 Pooled analysis of PRIMROSE 1 and PRIMROSE 2

The analyses for all efficacy outcomes reported in the CS focus on a pooled FAS analysis of PRIMROSE 1 and PRIMROSE 2, which had identical study designs apart from their geographical locations and patient baseline characteristics. The EPAR states that the pooled analysis was done according to a statistical analysis plan written after the 24-week results from PRIMROSE 1 and 2 were available.[25] The CS primarily reports the pooled analysis, although individual analyses for PRIMROSE 1 and PRIMROSE 2 are provided in Appendix M. The CS is not explicit about how the two trials were combined in the pooled analysis but according to the statistical analysis plan it appears that the data from the individual trials were added together and statistical analyses were then run on the combined data.

Given that there are some differences between the individual PRIMROSE trials in the baseline characteristics, the EAG believe a meta-analysis of the PRIMROSE trials would be preferable to naïve pooling, to enable statistical heterogeneity of the trials and weighting of effects by sample size to be explored objectively. Meta-analysis would also clarify whether the naïve pooled analysis is appropriate. We note that the pooled analysis gives implausible results for two outcomes: effect estimates for the pooled analysis are outside the range of values in the individual trials for the response and UFS-QoL symptom severity score outcomes (Figure 2 and Figure 17 respectively in the results section of this report).

A pooled analysis of safety data was performed up to Week 52, with a supplemental post-hoc analysis including pooled data up to Week 76 for select BMD assessments. This was to provide a comprehensive overview and more precise estimates for the rates of adverse events and for potential bone BMD loss with linzagolix treatment (CS section B.2.4.1.1). The EAG agree that a pooled safety analysis is appropriate.

3.2.4.3 Sample size calculations

The sample size calculations ensured that the individual PRIMROSE trials had 90% statistical power to detect a difference between all four dose regimens of linzagolix and placebo for the primary outcome (MBL improvement response rate), assuming response rates of 30% for placebo and 70% for linzagolix (CS Table 11). After the exclusions noted above in section 3.2.4.1, sufficient patients remained in the FAS to achieve the intended statistical power.

3.2.4.4 Methods to account for multiplicity

For the individual PRIMROSE trials, secondary outcomes were analysed sequentially in ranked order within each linzagolix treatment group, to protect against an overall type I error. An outcome was only claimed to be statistically significant if the resulting p-value for that outcome and all outcomes higher up in the testing order (for a given treatment group) were less than 0·0125 (CS Table 11).

Additional efficacy outcomes in the individual PRIMROSE trials were tested using a p-value of less than alpha=0·0125 with no further adjustments for having multiple outcomes (CS Table 11). The CS does not clarify whether these outcomes were considered independently of the hierarchy for the preceding outcomes, or whether all outcomes in the preceding hierarchy would need to be statistically significant for the additional efficacy outcomes to be tested.

For the pooled analysis of week 24 outcomes, the company say that “as the analysis is to improve precision, statistical results are to be regarded from an exploratory perspective. No adjustment was made for multiplicity within the pooled analyses” (CS section B.2.4.1.1). This appears counterintuitive, as precision and exploration are opposite concepts. The EAG note that the need to protect against the effect of multiple testing resulting from analysing multiple linzagolix dose regimens and multiple outcomes is independent of whether the trials are pooled or not. Given that a specific statistical analysis plan was developed for the pooled analysis, the EAG are unclear why an objective approach to account for multiplicity was not included in the plan. The company’s statistical approach undermines their preferred (pooled trials) analyses by implying that these can only support “exploratory” inferences.

3.2.4.5 Outcome analyses

Overall, the statistical methods used for outcome analyses appear broadly appropriate. We note that the EPAR did not raise any concerns around the statistical methods used or their assumptions. [25] All outcome analyses appear to have used broadly consistent approaches

and included race (the randomisation stratification factor) as a covariate which is appropriate.

3.2.4.6 Handling of missing data

CS Table 11 states that in general missing data were not imputed. Patients who had less than 28 days of data were counted as non-responders. Patients who discontinued prematurely due to lack of efficacy or adverse events or who underwent operative or radiological interventions for uterine fibroids were considered as non-responders for the primary analysis and in a similar way for the secondary outcomes of amenorrhea and reduced MBL. These are appropriate assumptions.

Two sensitivity analyses were conducted for the primary outcome to assess the robustness of the primary clinical effectiveness analysis results under alternative assumptions for days on which there were no data from the AH method. The first was done by imputing daily bleeding data based on the eDiary responses for days when no sanitary products were returned but bleeding had been reported in the eDiary. The second was done by assigning patients who discontinued early or who did not return any sanitary protection tools and had missing bleeding information in the eDiary as non-responders. According to the EPAR and study CSRs these analyses confirmed the primary analysis finding of significantly reduced menstrual blood loss in each active treatment group compared to placebo.[25-27]

CS Appendix M.3.3.2 states, descriptively only, that results of sensitivity analyses imputing missing data and results in the per protocol set were consistent with those of the main analysis. Results of these analyses are not reported in the CS or trial publications. Missing values for continuous efficacy endpoints were handled within the analysis itself via mixed model repeated measures, with the assumption that the model specification was correct, and that the data were missing at random.

Outcomes with missing data are summarised in Table 8 (percentages missing are relative to the FAS population)

Table 8 Outcomes with missing data

Outcome	Missing data at week 24
Haemoglobin Pooled analysis (CS Table 21)	31 to 48 patients (24% to 33%) per arm
Pain score Pooled analysis (CS Table 22)	51-64 patients (25% to 31%) per trial arm

Primary fibroid volume (CS Appendix Table 63)	PRIMROSE 1: 33-52 patients (31% to 49%) per trial arm PRIMROSE 2: 13-20 patients (13% to 20%) per trial arm
Uterine volume (CS Appendix Table 63)	PRIMROSE 1: 32-47 patients (31% to 45%) per trial arm PRIMROSE 2: 13-20 patients (13% to 20%) pe trial arm
UFS-QoL symptom severity score	Same as for pain outcome
EQ-5D (CS Appendix Table 67)	Not reported NB for Week 52, around one third missing from placebo and 200mg + ABT arms in PRIMROSE 1 (as examples - treatment switching affected some other arms) (CS Appendix Table 68)

The EPAR noted a high rate of missing BMD measurements at week 24 and at week 52, for both PRIMROSE trials due to the dropout rate.[25] However, the clinical characteristics and baseline BMD of the patients missing week 24 BMD data and those with week 24 BMD data are very similar.

3.2.4.7 Sensitivity and post hoc analyses

Sensitivity analyses to assess the implications of missing data were conducted as noted in section 3.2.4.6 above. Results of pre-specified subgroup analyses on race, therapy cycle length, excessively heavy menstrual bleeding (defined by the third quartile of baseline MBL in the FAS), baseline FIGO classification, and fibroid size are reported in CS Appendix Figures 23 and 24 (CS Appendix E). Overall, the results show consistent effectiveness of the linzagolix compared to placebo for all the subgroups. There is a suggestion in CS Appendix Figure 23 that the strength of effect of linzagolix differs between the Black/Afro-American subgroup compared to the “Other” subgroup in PRIMROSE 1, as there appears to be less of a dose-response pattern evident in the Black/Afro-American group; this group did not achieve as high a response rate for the linzagolix 200mg plus ABT regimen as was achieved in the “Other” group.

EAG conclusion on study statistical methods

The company’s approach to trial statistics appears broadly appropriate for the individual PRIMROSE trials. However, the company’s approach to pooling the trials is not well explained and the company declared the results of the pooled analysis to be “exploratory” without requiring formal hypothesis testing or adjustment for multiple comparisons. The EAG believe that the same standard of statistical testing and adjustment for multiplicity used in the identical individual trials should have been applied

in the pooled analysis. We also suggest that a meta-analysis of the two individual PRIMROSE trials would be helpful to enable objective consideration of statistical heterogeneity and differences in sample sizes between the trials.

3.2.5 Efficacy results of the PRIMROSE studies

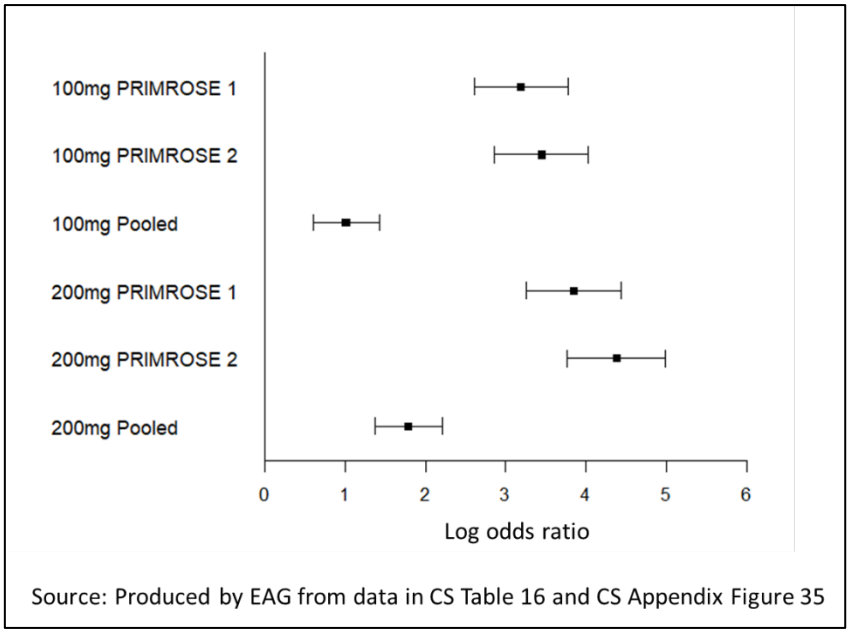
This section focuses on the evidence reported for the trial dosing regimen groups relevant to the cost-utility analysis of this appraisal, i.e. for Population #3, people receiving longer-term treatment without hormonal add-back therapy compared to placebo.

3.2.5.1 Primary outcome: clinically meaningful reduction in heavy menstrual bleeding (response) at Week 24

The primary outcome was the proportion of patients achieving a response (i.e. achieving a clinically meaningful reduction HMB). Response was defined as MBL $\leq$80 mL, with $\geq$50% reduction in MBL from baseline, at Week 24.

Odds ratios for the response outcome are reported in CS Table 16, but these are based on a skewed (non-normal) distribution. The log odds ratios, shown in Figure 2 below, are similar between the PRIMROSE 1 and PRIMROSE 2 trials, with no statistically significant differences between the regimens in the odds of achieving a response. However, there is a discrepancy in the data since the odds ratios provided by the company for the pooled trials analysis are outside the range of trial data, as shown for the log odds ratios in Figure 2. Nevertheless, overall, patients in the linzagolix groups have favourable odds of achieving a response compared to those in the placebo groups.

The response rates for the pooled analysis at Week 24 were 32.2% (66/205) in the placebo arm, 56.5% (108/191) in the linzagolix 100mg arm, and 74.5 (155/208) in the linzagolix 200mg arm (CS Figure 8 andCS Table 16). The results suggest a dose-response effect, as well as a placebo effect.



Source: Produced by EAG from data in CS Table 16 and CS Appendix Figure 35

Figure 2 Proportion achieving response: linzagolix versus placebo, Week 24

The placebo effect continued into the second treatment period increasing to 42% achieving response in the remainder of the placebo treatment group in PRIMROSE 1 at 52 weeks (CS Appendix M.3.3.1 Figure 36). The company argue that because of the high response rate in the placebo group the relative efficacy of linzagolix may be underestimated. They suggest that the high response rate in the placebo group may be due to patient non-compliance with the method of collecting used sanitary products, leading to overestimation of the number of days with no bleeding (CS sections B.2.12.2.2, and CS B.2.9.7, and clarification response A8). The EAG and our clinical expert considered that the reason for the observed placebo effect is uncertain, and the company's explanation is speculative. Regression to the mean might explain at least some of the placebo effect. The company conducted sensitivity analyses using two different methods of data imputation to check the robustness of the analysis to missing data and they found that the linzagolix versus placebo comparisons were not sensitive to missing data (CS section B.2.12.2.2).

Despite the response rate in the placebo group being 32.2%, the difference between placebo and each of the linzagolix treatment groups at Week 24 is statistically significant for all dose regimens in the pooled PRIMROSE trials ($p<0.001$) (CS Table 16) and in each of the individual PRIMROSE trials ($p<0.003$) (CS Appendix Figure 35).

At Week 52 The proportion of patients in the 100 mg treatment groups achieving a clinically meaningful reduction in HMB in the individual trials was maintained at 57% in PRIMROSE 1 and 53% in PRIMROSE 2 (compared with 56.5%, 56.4% and 56.7% in the pooled analysis, PRIMROSE 1, and PRIMROSE 2 respectively at Week 24). As noted above, there was an

unexplained increase in the proportion of responders in the PRIMROSE 1 placebo arm at Week 52.

3.2.5.2 Secondary and additional efficacy outcomes

3.2.5.2.1 Change in primary fibroid volume (additional outcome)

In the pooled analysis at Week 24, reductions in fibroid volume (limited to the largest three fibroids) were 25% and 48% (change from baseline) in the 100 mg and 200 mg linzagolix treatment groups (CS section B.2.6.2.1). Compared to placebo, the result for the 200mg linzagolix dose regimen is statistically significant (p<0.001) while the result for the 100mg dose is marginally significant (p=0.012) (the specified Bonferroni significance threshold was p≤0.0125) (CS Table 17).

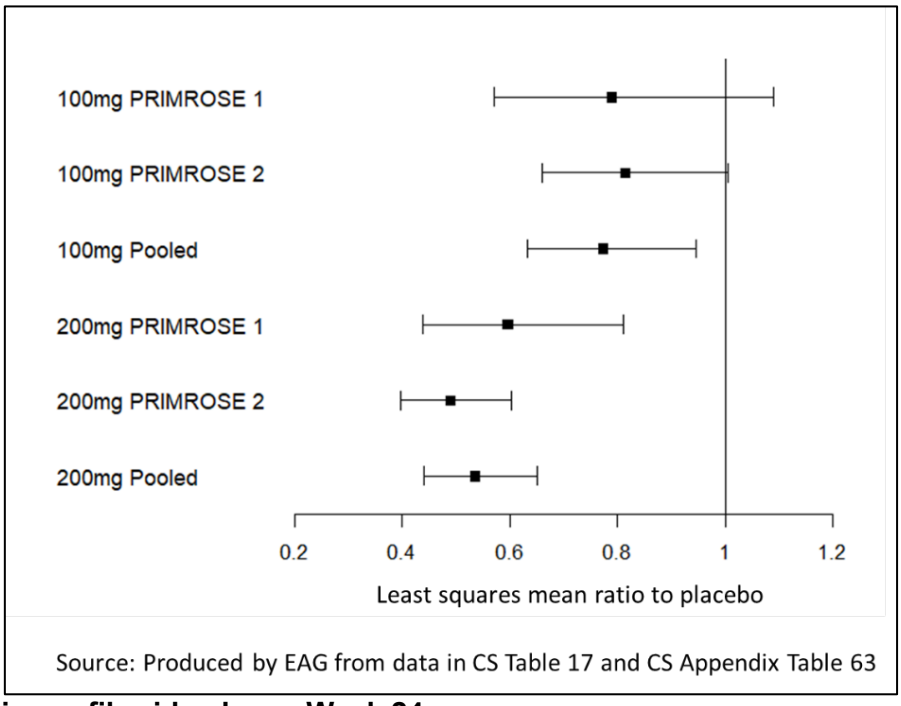


Figure 3 Primary fibroid volume, Week 24

Results for reduction in primary fibroid volume were similar at Week 24 for the individual PRIMROSE trials, with only the 200 mg group achieving a statistically significant result, although both doses achieved significant reduction in primary fibroid volume in the pooled analysis (Figure 3). There is a substantial amount of missing data, n/N: PRIMROSE 1 linzagolix 100 mg 58/94 and linzagolix 200 mg 72/105; and PRIMROSE 2 linzagolix 100 mg 79/97 and linzagolix 200 mg 85/103 (CS Appendix Table 63).

At Week 52, the change from baseline for primary fibroid volume (mean (SD)) was generally maintained at 9.42 mL (104.89) and -10.27 mL (55.53) in the linzagolix 100 mg groups of PRIMROSE 1 and PRIMROSE 2 respectively (CS Appendix Table 64). However, there was substantial missing data (19/61 missing in PRIMROSE 1 and 20/79 missing in PRIMROSE 2) (CS Appendix Table 64).

The EAG are uncertain how meaningful these results for the change in primary fibroid volume are, due to lack of clarity and potential inconsistency in how the primary fibroid(s) were selected and measured (section 3.2.3.1.3). We believe this outcome has potential for (but uncertain risk of) detection bias (section 3.2.2.1).

3.2.5.2.2 Change in uterine volume (additional outcome)

In the pooled analysis at Week 24, reductions in uterine volume (change from baseline) were 15% and 39% (change from baseline) in the 100 mg and 200 mg linzagolix treatment groups (CS section B.2.6.2.1). Compared to placebo, both results were statistically significant: p=<0.001 for both 100 mg and 200 mg linzagolix compared to placebo (CS Table 17). Results comparable across both trials and the pooled analysis were reported only as a ratio (Figure 4).

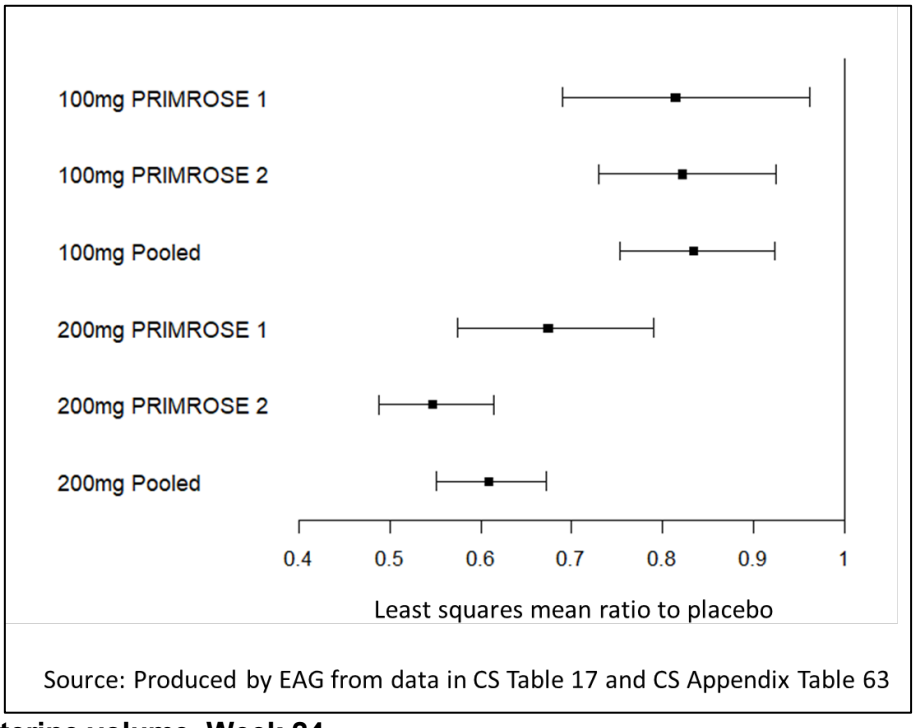


Figure 4 Uterine volume, Week 24

At Week 52, the change from baseline for uterine volume (mean (SD)) was generally maintained at 8.02 (229.05) mL and 94.98 (857.54) mL in the linzagolix 100 mg groups of PRIMROSE 1 and PRIMROSE 2 respectively (CS Appendix Table 64).

3.2.5.2.3 Secondary outcomes relating to reduction in heavy menstrual bleeding

Week 24 results for secondary outcomes related to reduction in HMB, are reported in CS sections B.2.6.2.2 to B.2.6.2.5. These are ranked secondary outcomes in the individual trials, in the following rank order:

- **Time to response (reduced HMB) (pooled analysis)** was significantly shorter in all linzagolix treatment groups compared to placebo. The 100 mg linzagolix group took a median 135.0 days (95% CI 119.0 to 146.0) and the 200 mg linzagolix group took a median 3.0 days (95% CI [redacted]). Median number of days to reduced HMB was non-evaluable in the placebo group (Pooled Analysis data on file Table 2.7.3.6.2.1).

- **Amenorrhoea (pooled analysis)** was achieved in a significantly larger proportion of patients in all linzagolix treatment groups compared to placebo. Amenorrhoea was achieved by 36.1% in the 100 mg linzagolix group and 65.4% in the 200 mg linzagolix group compared to 16.6% in the placebo group.

- **Time to amenorrhoea (pooled analysis)** was significantly shorter in all linzagolix treatment groups compared to placebo. KM estimates for the median number of days to achieve amenorrhoea was non-evaluable for the 100 mg linzagolix group and 33.0 days (95% CI 10.0 to 80.0) for the 200 mg linzagolix group (Pooled Analysis Data on file Table 2.7.3.6.4.1).

- **Number of days of uterine bleeding for the last 28 days (pooled analysis)** was significantly reduced in all linzagolix treatment groups compared to placebo. The percentage of patients with zero days of uterine bleeding was 76% in the 200mg group and 53% in the 100mg group. However, actual change from baseline in the mean number of days of uterine bleeding shows less difference between the dose regimens: 100 mg linzagolix -3.1 (3.2) days mean (SD), and 200 mg linzagolix -4.4 (2.6) days mean (SD) (Pooled Analysis Data on file Tables 2.7.3.6.5.1 and 2.7.3.6.5.2).

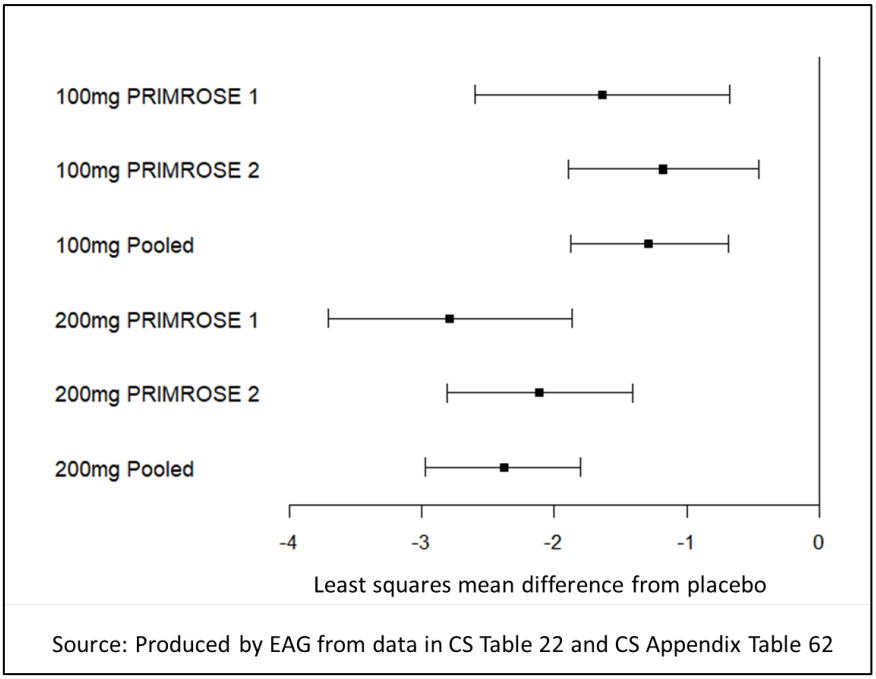
- **Haemoglobin concentrations in patients anaemic at baseline (Hb <12 g/dL) (pooled analysis)** were significantly improved in all linzagolix treatment groups at week 24. Summary values for Hb concentrations at 24 weeks show that the Hb levels only reach 12 g/dL in the linzagolix 100 mg plus ABT, 200 mg, and 200 mg plus ABT groups: mean g/dL (SD) 12.01 (1.57), 12.19 (1.46) and 12.25 (1.5) respectively (Tables 2.7.3.6.6.1 and 2.7.3.6.6.2 of Data on file PRIMROSE 1 and 2 Pooled

Analysis). Therefore, many patients remained anaemic or around the threshold for anaemia. Additionally, women with a haemoglobin level below 6 g/dL (severe anaemia) were excluded from the trials, so it is unclear if linzagolix is indicated for anaemia as a severe symptom of uterine fibroids.

These results support a meaningful reduction in HMB for all linzagolix treatment groups at Week 24 (CS Appendix M.3.3.3.1 to M.3.3.3.2). At week 52, CS Appendix M.3.3.3.1 to M.3.3.3.2 show that these results were sustained over a longer-term treatment period in the individual trials. Including the proportional differences between the PRIMROSE 1 and PRIMROSE 2 trials. However, the evidence is much weaker for this second treatment period as not all treatment groups had estimable results so there is less comparative evidence.

3.2.5.2.4 Uterine fibroid-associated pain (additional outcome)

In the pooled analysis, at Week 24 the uterine fibroid-associated pain score was statistically significantly reduced from baseline in all linzagolix treatment groups compared to the placebo group (CS section B.2.6.2.6). Results for the 100 mg and 200 mg linzagolix treatment groups also appear to be dose-dependent with linzagolix 200 mg effecting greater changes than linzagolix 100 mg compared to placebo (CS section B.2.6.2.6; CS Table 22). Results are broadly similar for the individual trials (Figure 5), and both 100 mg and 200 mg linzagolix treatment groups remained statistically significant in both PRIMROSE 1 and PRIMROSE 2 (CS Appendix Table 62).



Source: Produced by EAG from data in CS Table 22 and CS Appendix Table 62

Figure 5 Uterine fibroid-associated pain, Week 24

At week 52, these results were generally maintained in the linzagolix 100 mg treatment groups, although was a slight deterioration towards the end of the second treatment period in PRIMROSE 1 (CS Appendix Figure 41).

3.2.5.3 HRQoL outcomes

The company used UFS-QoL and EQ-5D-5L to measure HRQOL for the economic models. Choice of assessment tool, UFS-QoL or EQ-5D-5L, makes a difference as to whether any statistically significant changes were observed for the quality-of-life outcome: UFS-QoL scores show significant changes but EQ-5D-5L scores do not (see below). The explanation provided in the CS refers to the UFS-QoL score being based on a 3-month recall of overall pre-treatment and post-treatment experience, whereas the EQ-5D-5L score is based on a single measurement on a single day which may not adequately reflect a fluctuating menstrual cycle (CS Appendix M.3.3.3.6). However, the company base case uses the EQ-5D-5L data mapped to EQ-5D-3L which is according to NICE procedural preferences and is the most conservative option, with a scenario analysis using the UFS-QoL results.

3.2.5.3.1 UFS-QoL

In the pooled analysis, at Week 24 all the linzagolix treatment groups showed statistically significant improvements in both HRQoL scores (Figure 6) and symptom severity scores (Figure 7) compared to the placebo group (CS section B.2.6.2.7, CS Table 23). There is a discrepancy between the reported pooled and individual trial symptom severity scores; the pooled score is outside the range of the individual trial scores (Figure 7).

At Week 52, the linzagolix 100 mg group maintained the improved scores from baseline that were observed at Week 24, but they were not further improved during the second treatment period. This treatment group has the highest symptom severity score and lowest HRQoL score compared to all other treatment groups in both PRIMROSE 1 and PRIMROSE 2 (CS Appendix Table 66).

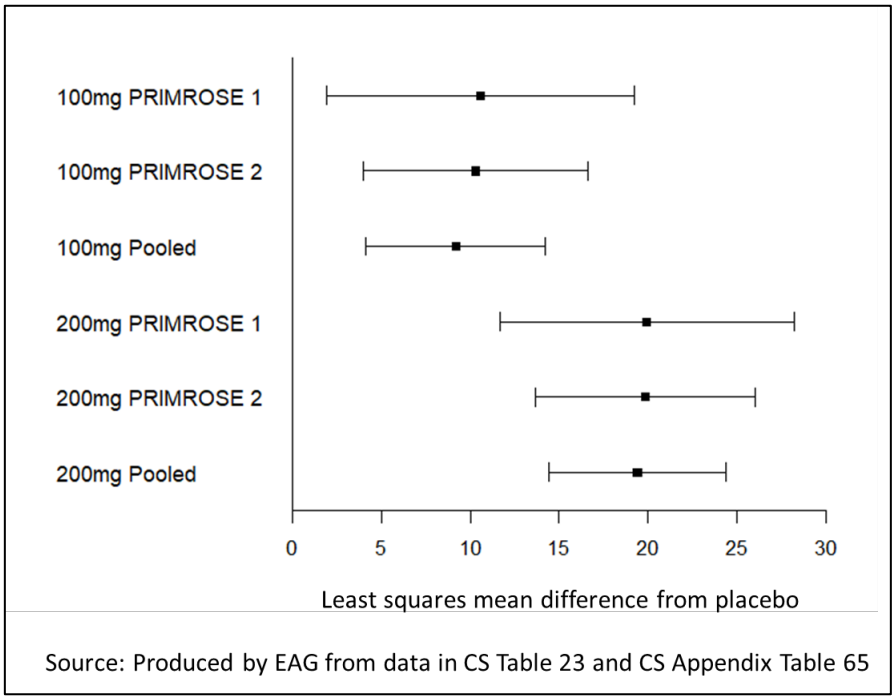


Figure 6 UFS-QoL HRQoL total score, Week 24

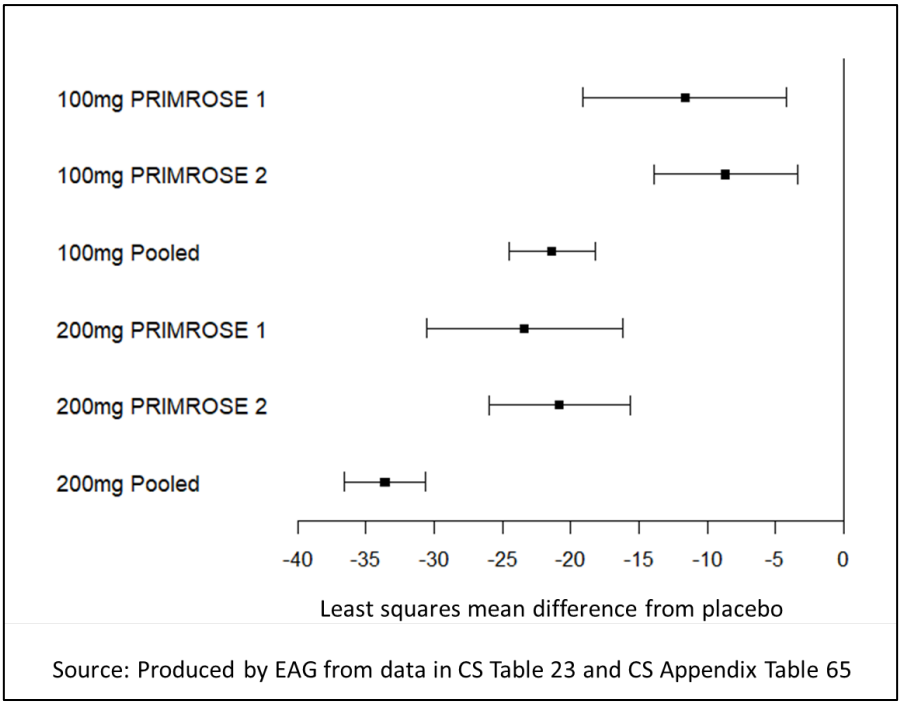


Figure 7 UFS-QoL symptom severity score, Week 24

3.2.5.3.2 EQ-5D-5L

In the pooled analysis at Week 24 all treatment groups showed similar small improvements in both index values and VAS scores, but there were no statistically significant differences for any of the linzagolix treatment groups, including the 100 mg and 200 mg linzagolix groups, compared to placebo (CS section B.2.6.2.7, CS Table 24).

Week 52 EQ-5D-5L results for the individual PRIMROSE trials appear consistent with the Week 24 results, i.e. there were no noticeable differences between treatment groups and placebo (CS Appendix Table 68).

3.2.5.4 Subgroup analyses

Subgroup analyses of the primary outcome (response, defined as a clinically meaningful reduction in HMB at Week 24) that were performed for the pooled analysis show responses across treatment groups were generally consistent for race (Black or African American; other), weight, BMI and age (CS section B.2.7; CS Appendix Figure 23); this is consistent with the individual PRIMROSE trials (CS Appendix Figure 24).

Subgroup analyses of the primary outcome for other planned subgroups were not performed for the pooled analysis, but some are reported for the individual PRIMROSE trials: excessively heavy menstrual bleeding, baseline FIGO 0, 1, or 2 in at least one fibroid (submucosal fibroids), and cycle length (≤28 days; <28 days). The results are generally

consistent but the EAG note that as the confidence intervals are wide (CS Appendix Figure 24).

A pre-specified subgroup of patients who had anaemia at baseline was included for the haemoglobin concentrations outcome which is reported above in section 3.2.5.2.3 as the subgroup is only relevant to that specific secondary ranked outcome.

EAG conclusion on the clinical efficacy outcomes

Overall, the pooled analyses for Week 24 outcomes show that linzagolix 200 mg without ABT is more effective than placebo for all reported outcomes and that linzagolix 100 mg without ABT is more effective than placebo for all reported outcomes except reduction in fibroid volume. The explanation for a placebo effect observed for the primary outcome is unclear. Caution should be exercised in interpreting results for fibroid volume (up to the 3 largest fibroids were measured, missing data) and for Hb concentrations in the subgroup who were anaemic at baseline (severely anaemic patients were excluded from the trials). The HRQoL results are ambiguous because linzagolix does not show any significant improvements compared to placebo according to the EQ-5D-5L results, yet linzagolix shows improvements in the UFS-QoL results (much greater than a 9-15 point change from baseline used in another study to indicate a clinically meaningful change).

3.2.5.5 Safety outcomes

CS section B.2.10 reports the results from the pooled Safety Analysis Set of PRIMROSE 1 and PRIMROSE 2 for Week 24 (Day 1 to Week 24) and for Week 52 (Week 24 to Week 52, i.e. not cumulative from Day 1). Safety results for the individual trials are reported in CS Appendix M.3.4. Safety results are relevant to all population groups, so this section of the report covers all trial regimens.

The pooled Safety Analysis Set included all randomised patients in the trials who received at least one dose of the study drug: Week 24 n=1,037; Week 52 n=757 (CS section B.2.4.2.1) and mean overall compliance was 98.7% and 99.3% at Weeks 24 and Week 52 respectively (CS section B.2.10.1.1).

Adverse event outcomes specifically noted in the NICE scope are: vasomotor symptoms (covered by hot flushes, below); incontinence, which was not reported for more than one patient in either treatment period (Confidential Pooled Analysis);[28] and pelvic organ prolapse which was not assessed and which the EAG's clinical expert considered inconsequential.

3.2.5.5.1 Treatment-emergent adverse events (TEAEs)

At Week 24, there were slightly more TEAEs in the linzagolix treatment groups compared to placebo, with the 200 mg group having the highest incidence at 63.3% compared to the placebo group at 49.3% (CS Table 34). However, during the second treatment period (weeks 24 to 52) fewer TEAEs were reported and there was very little difference in the incidence rate across all treatment groups (range 29.9% to 41.6%) with no apparent dose dependency (CS Table 35). Very few of these TEAEs were serious or severe, in either treatment period. There was low incidence of serious adverse events (CS section B.2.10.3). TEAEs leading to permanent treatment discontinuation were similar across all treatment groups: incidence was low at Week 24 (range 7.0% to 10.5%) with the most frequent reasons being headache (1.1%) and hot flushes (1.1%), and lower at Week 52 (range 1.3% to 8.1%) with the most frequent reason being related to bone mineral density loss (1.3%) (CS section B.2.10.2).

Incidence of hot flushes (vasomotor symptoms in the NICE scope) at Week 24 were considered dose dependent: incidence was higher in the groups treated without ABT (10.1% and 33.3% for the 100mg and 200mg groups respectively) compared to those who received ABT (5.2% and 9.6%) (CS Table 36). For the second treatment period, incidence of hot flushes decreased and was similar across all treatment groups (range 0.0% to 2.4%), with the highest incidence in the placebo/200 mg + ABT group experiencing their first 24-week exposure to linzagolix (CS Table 37). There may also have been a dose dependent response at Week 24 for the more common headache TEAE as incidence was higher in the groups without ABT (CS section B.2.10.1.3).

3.2.5.5.2 Mortality

None of the TEAEs were fatal and only one death occurred, which was accidental and unrelated to the trials.[11]

3.2.5.5.3 Bone mineral density

Decrease in BMD is a TEAE of special interest due to the mechanism of action of linzagolix, which suppresses the production of serum estradiol, and the known effects of low oestrogen levels which reduce bone mineral density. Changes in BMD from baseline are reported for the lumbar spine, femoral neck, and total hip (CS section B.2.10.6.1, and CS Tables 44 and 45).

At Week 24 dose dependent reductions in BMD were seen in all linzagolix treatment groups, although the changes are described as small and only the change for the 200 mg group is described as clinically meaningful. The EAG's clinical expert suggested that a $\geq$5% change

in BMD would be clinically meaningful, therefore, as the highest percentage change from baseline was -3.697 (2.859) (mean (SD), in the 200 mg without ABT group) we concur with the company.

At Week 52 the greatest reduction in BMD was seen in the 200 mg/200 mg + ABT group, although the addition of ABT limited the risk, and for the other treatment groups the BMD decrease appeared to stabilise because it was less rapid. There was no clinically meaningful change in BMD at Week 52. However, the EPAR noted the (slight) increase in BMD-related adverse events of musculoskeletal and connective tissue disorders (osteopenia, osteoporosis and bone loss) during this second treatment period, [25] therefore the EAG view these results cautiously.

Week 76 data, reported in Table 26 of the EPAR, show that the rate of BMD loss had generally slowed or reversed, and more so in the treatment groups that included ABT. [25] Focussing on the lumbar spine (considered to be the most sensitive), the slowest to recover group was that which did not receive ABT at all i.e. the 100mg group. Furthermore, the results from PRIMROSE 3, an off-treatment extension study reported in CS section B.2.11, appear (the evidence has several limitations) to show a continued trend for partial or complete BMD recovery, and suggests that the overall bone health of the participants is [redacted], thus implying [redacted]. The EAG consider the longer-term evidence around BMD uncertain due to small patient numbers remaining in the treatment arms in the off-treatment periods (up to Week 76 in PRIMROSE 1 and 2; PRIMROSE 3), and the increase (though small in number) in BMD-related adverse events in later treatment periods.

EAG conclusion on safety results

Linzagolix appears to be well-tolerated, with very few serious or severe adverse events, few adverse events leading to treatment discontinuation, and no associated mortality. Hot flushes were common and appeared to be dose-dependent during the first 24 weeks of treatment but were much reduced and not dose dependent afterwards. It appears that reduction in BMD was dose-dependent during the first few months of treatment, but not a clinically meaningful change. During continued treatment BMD loss was less rapid, although it is uncertain whether this pattern would be sustained in the longer term.

3.3 Critique of studies included in the indirect comparisons

3.3.1 Rationale for the network meta-analysis

Placebo-controlled trials of linzagolix (PRIMROSE) and relugolix CT (LIBERTY) are available, but no direct comparisons of linzagolix against relugolix CT exist. The company therefore conducted NMAs for each outcome, where data were available, to compare linzagolix against relugolix CT (CS section B.2.9).

3.3.2 Identification, selection and feasibility assessment of studies for ITC

3.3.2.1 Comparison of linzagolix against relugolix CT

The aim of the indirect treatment comparisons was to investigate whether linzagolix has similar clinical effectiveness and safety to relugolix CT, to provide supporting information for the cost-comparisons for Population #1 and Population #2.

The company's systematic literature review of clinical efficacy studies (CS Appendix Table 7) identified five studies of relugolix CT. These were the two LIBERTY pivotal placebo-controlled trials which had informed the relugolix CT technology appraisal (T832) and three trials conducted in Japan that compared relugolix CT against placebo (Osuga et al. 2019[21]), against leuprorelin acetate (also called leuprolide acetate) (Osuga et al. 2019[20]), or compared three doses of relugolix CT (Hoshhiai et al. 2021[22]). The company excluded non-US and non-EU trials, meaning that these Japanese trials were excluded, and they also excluded trials published more than 20 years ago.

The company did not explore potential relevance of the three excluded Japanese studies but the EAG believe that these studies were excluded appropriately. Two of them included comparisons unlikely to be connected or useful in an evidence network (relugolix CT versus placebo;[21] relugolix CT dose-ranging [22] whilst the third (relugolix CT versus leuprolide[20]) used the PBAC method for measuring MBL which is not directly comparable with the AH method used in the PRIMROSE and LIBERTY trials. Although an adjustment can sometimes be made for comparing PBAC scores against AH-derived estimates of MBL (discussed in section 3.2.3.1.1), the necessary data to derive the required coeffcients for such an adjustment were not available in the trial publications and CSR for the Osuga et al. study[20] when it was scrutinised by the company in TA832.[29]

The CS reports that NMAs were conducted on the following outcomes according to data availability: response (reduced HMB, defined as a menstrual blood loss ≤80 mL and ≥50% reduction from baseline), percentage change in MBL, improvement in pain (defined as a NRS score ≤1 for participants with an NRS score ≥4 at baseline), percentage change in

primary fibroid volume, percentage change in haemoglobin for participants with haemoglobin ≤10.5 g/dL at baseline, and improvement in HRQoL (defined as the change UFS-QoL total score). The EAG believe it unlikely that other outcomes could be included, as not all outcomes of interest were reported in the trials.

3.3.2.2 Comparison of linzagolix against other GnRH analogues

The EAG requested further evidence from the company on the relative clinical effectiveness of linzagolix compared to the other GnRH analogues that could be used in clinical practice. In their response the company provided additional NMAs, using a fixed-effects model only, based on an extended network (Clarification Response A11). The company's clarification response states that the network could include the PRIMROSE, LIBERTY and PEARL trials for the response outcome but only the PRIMROSE and PEARL trials for the changes in total fibroid volume and haemoglobin, with no other outcomes being available from more than one trial. The Clarification Response does not describe the evidence network, but we assume it was as shown in Figure 8. PEARL I is an RCT that compared ulipristal acetate (a selective progesterone receptor modulator) against placebo,[17] and PEARL II is an RCT that compared ulipristal acetate against leuprolide acetate (a GnRH agonist).[18] Both PEARL I and PEARL II had been identified in the company's systematic literature search (CS Appendix Table 7) and their study designs are summarised in section 3.2.1.3 of this report. Ulipristal acetate is only indicated for intermittent treatment when uterine fibroid embolisation or surgery are unsuitable or unsuccessful, and as noted in TA832 it is unlikely that many people with uterine fibroids needing treatment would agree to have ulipristal acetate, given the level of monitoring needed and potential risks of liver damage. According to Clarification Response A11 leuprolide acetate was the only additional relevant comparator that could be included in the extended NMA network. However, the company have not reported their study selection process for this extended evidence network. It is unclear whether a more thorough search and study selection process would identify further studies relevant for inclusion in the evidence network.

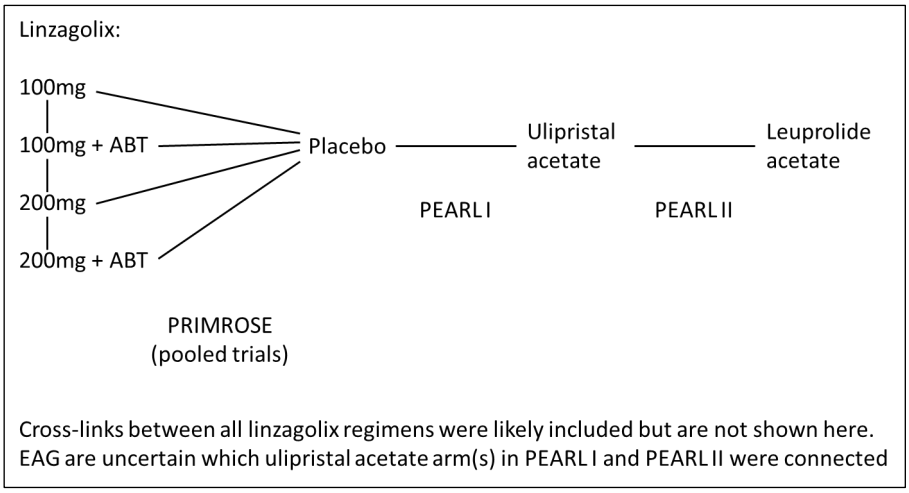


Figure 8 EAG presumed evidence network for the comparison of linzagolix against leuprolide acetate

According to the company's list of studies included in their clinical effectiveness systematic literature search (CS Appendix Table 7) and their list of studies that are available but were excluded from the systematic literature search (CS Appendix D.3.2) we believe it unlikely that other relevant studies and comparators could have been included in the NMAs.

3.3.3 Clinical heterogeneity assessment

3.3.3.1 Comparison of linzagolix against relugolix CT

Heterogeneity assessment is discussed in CS section B.2.9.4. The CS states that in general, there was good alignment between the trials; the inclusion and exclusion criteria were identical between PRIMROSE 1 and 2, as well as between LIBERTY 1 and 2; the outcomes in the PRIMROSE and LIBERTY trials were defined similarly: and pooled PRIMROSE patient-level data allowed additional comparisons to be made.

However, the CS acknowledges differences in population characteristics between the PRIMROSE and LIBERTY trials. The trials differed notably in the proportion of Black patients (approximately 63% versus 5% across arms in PRIMROSE 1 and PRIMROSE 2, respectively; and 47% and 42% across arms in LIBERTY 1 and LIBERTY 2, respectively). Other differences in baseline characteristics were in the proportion of Hispanic or Latino patients (11.8% across arms in the pooled PRIMROSE trials versus 20.7% across arms in LIBERTY 1 and 2; p<0.001), mean baseline MBL (207.6 across arms in the pooled PRIMROSE data versus 229.2 across arms in LIBERTY 1 and 2; p=0.007), uterine volume (328.2 across arms in the pooled PRIMROSE data versus 393.2 across arms in LIBERTY 1 and 2; p<0.001), uterine fibroid volume (98.9 across arms in the pooled PRIMROSE data versus 72.9 across arms in LIBERTY 1 and 2; p<0.001), and the proportion of patients with a

pain score ≥4 (77.1% across arms in the pooled PRIMROSE data versus 71.8% across arms in LIBERTY 1 and 2; p=0.029).

The company note in Clarification Response A8 that the method of accounting for missing MBL data differed between the PRIMROSE and LIBERTY trials. Missing return of menstrual products was considered 'no bleeding' in the PRIMROSE trials whereas missing data were imputed in the LIBERTY trials. The company argue that this might explain why a placebo effect exists in the PRIMROSE placebo arms but is less clear in the LIBERTY trials. The company conducted sensitivity analyses using two different methods of data imputation in the pooled PRIMROSE trials to check the robustness of the analysis to missing data and they found that the linzagolix versus placebo comparisons were not sensitive to missing data (CS section B.2.12.2.2). The company discuss how these trial differences could influence the results of NMAs that compare linzagolix against relugolix CT in CS section B.2.9.7.

The company concluded that overall, the trials appeared to be broadly comparable and an NMA was an appropriate method of indirect comparison. The company also conducted a matching adjusted indirect comparison (MAIC) as a scenario analysis to explore whether differences in baseline characteristics may have impacted comparative results from the NMA. The EAG agree with this overall approach for exploring heterogeneity, although there are caveats around the methodology of the MAIC analyses (see section 3.4.4 below).

The company elected to use the pooled PRIMROSE trials and separate LIBERTY trials for their NMAs. We assume (section 3.2.4.2) that pooling means that the trial data were added together, and then statistical analyses were run on the combined data set. Given the differences in the baseline characteristics of PRIMROSE 1 and PRIMROSE 2 noted above, the EAG requested the company to provide a NMA that included the two PRIMROSE trials separately in the network (Clarification Question A10). We compare the company's NMA results for the pooled and separate analyses of the PRIMROSE trials in section 3.5.1 below.

3.3.3.2 Comparison of linzagolix against other GnRH analogues

For their extended NMAs comparing linzagolix against leuprolide acetate (Clarification Response A11) the company did not provide a systematic comparison of the baseline characteristics of the PEARL, PRIMROSE, and LIBERTY trials. However, the clarification response highlights that the method for estimating MBL differed between the PRIMROSE and LIBERTY trials which used the alkaline haematin method (the 'gold standard'), and the PEARL trials which used the pictorial blood loss assessment chart (PBAC) method (see section 3.2.3.1.1 for a discussion of these methods and their comparability).

The EAG note there are a number of differences in the baseline characteristics of the PEARL I and PEARL II trials[17, 18] when compared to the PRIMROSE and LIBERTY trials (CS Appendix 8):

- All patients in the PEARL trials were eligible to undergo fibroid surgery at the end of the treatment period. In the PRIMROSE trials patients who would require surgery within 6 months were excluded (CS Appendix M.1).

- In PEARL I the inclusion criteria specified patients should meet a specified PBAC score and should have fibroid-related anaemia whereas these criteria were not used in the PRIMROSE and LIBERTY trials.

- PRIMROSE and LIBERTY excluded patients with fibroids of 12cm diameter or larger, whereas the PEARL trials excluded patients with fibroids of 10cm or larger.

- PEARL I did not include any Black patients but had approximately 10-15% Asian patients whereas PEARL II had approximately 9-11% Black patients and few if any Asian patients (not separated from “other”). As noted previously, the population of PRIMROSE 1 was different, with approximately 61-65% Black patients. The balance of Black and White patients in the PEARL trials aligns most closely with PRIMROSE 2.

- Total fibroid volume in PEARL I (ranging from 61.9 cm3 in the placebo arm to 100.7 cm3 in the 5mg ulipristal acetate arm) was notably larger than in the PRIMROSE trials (approximately 43-72cm3).

- BMI in both PEARL trials is similar to PRIMROSE 2 but lower than in PRIMROSE 1.

- Total fibroid volume (only measured in PEARL I) is more similar to PRIMROSE 1, being slightly higher than in PRIMROSE 2

- Uterine volume measurements in PEARL I are similar to those in PRIMROSE 1. whilst the measurements in PEARL II are similar to those in PRIMROSE 2

- UFS-QoL scores (only measured in PEARL II) correspond with PRIMROSE 2 for the symptom severity score and are better than both PRIMROSE trials for the total HRQoL score.

The company did not conduct any MAIC analyses to further explore heterogeneity in the trial population characteristics in the comparison of linzagolix against leuprolide acetate.

3.3.4 Risk of bias assessment for studies included in the NMAs

Risks of bias in the PRIMROSE, LIBERTY, and PEARL II trials are discussed in section 3.2.2 and summarised in Appendix 2. No high risks of bias were identified. We noted uncertainty in how reliable measures of fibroid volume are in the PRIMROSE trials; and uncertainty in why some outcomes had extensive missing data (summarised in Table 8) which was more frequent in PRIMROSE 1 than in PRIMROSE 2. However, there was no indication of any substantive differences in the amount of missing data between trial arms.

3.4 Critique of the indirect comparison

3.4.1 Data inputs to the NMA and MAIC analyses

The company provided the statistical code for the NMA and MAIC analyses in CS Appendix D.3.6.1 and D.3.6.2 and in Clarification Response Document Appendices 2, 4 and 5. No input data for the NMAs were provided with the code, and the NMA code does not contain sufficient detail of the data sets analysed (e.g. sample size for the input data used) for the EAG to check whether the NMAs were conducted appropriately. The EAG were not provided with the individual participant data so we could not verify the MAIC analyses.

3.4.2 Statistical methods for the NMA and MAIC analyses

The statistical methods of the NMAs are described in CS Appendix D.3.5.2, including a network diagram (CS Appendix Figure 5). The NMA was conducted in a Bayesian framework using Monte Carlo Markov Chain (MCMC). Overall, the EAG believe the methods of the NMA are appropriate, aside from the caveat noted above that we were unable to validate the analysis.

The statistical methods of the MAIC analysis are described in CS Appendix D.3.8. The MAIC statistical methods appear to have been correctly applied, with the caveat that the EAG could not verify this. The target population for weighting in the matching was the pooled LIBERTY trials, which the EAG's clinical expert agreed are broadly representative the UK clinical practice population (section 3.2.1.2.4). Matching was conducted on the proportion of Black patients, uterine volume, total fibroid volume, MBL, and haemoglobin; these were the most important variables for matching according to two internal company experts. However, the CS does not explain what the key treatment effect modifiers are (all should be included in anchored population matching) and no scenarios were conducted to investigate the influence on outcomes of different matching variables.

A comparison of trial baseline characteristics before and after weighting shows that the matching was partially successful but differences in the proportion of Hispanic / Latino patients worsened post-matching and some differences in mean pain score remained (CS Appendix Table 19). The distribution of weights (CS Appendix Figure 6) is reasonable with no very large weights used.

3.4.3 Selection of random and fixed-effects models

Fixed and random effects models were conducted in the NMAs. With only small differences observed in the deviance information criterion (DIC), the company preferred fixed effects models across all outcomes. Given the heterogeneity present and the uncertainty as to whether these differences are treatment effect modifiers, the EAG prefer the results of the random effects models (section D.3.5.3).

3.4.4 Summary of EAG critique of the NMA and MAIC analyses

In NICE technology appraisals cost-comparison analyses assume that the intervention and comparator have similar clinical efficacy and safety. Results of the company's NMA analyses are used to support inferences about the clinical similarity of linzagolix and relugolix CT but do not directly inform the company's economic models.

A challenge with interpreting NMA results to infer the similarity of linzagolix and GnRH analogue comparators is that a robust conclusion on the similarity of the treatments would require an inference of non-inferiority. However, no non-inferiority trials are available for linzagolix and so the NMA results are based on analysis of treatment differences (superiority) rather than similarity (non-inferiority). When heterogeneity is present, random-effects models produce wider credible intervals than fixed-effects models, potentially increasing the risk of falsely concluding that treatments are similar. That is, the more heterogeneity that is present, the greater the risk of falsely concluding that there is no treatment difference. The company state that "the outcomes of the NMA from the available evidence does not generally indicate any expected differences in treatment efficacy for linzagolix when compared with relugolix CT" (CS section B.2.9.8). The EAG are uncertain what is meant by "any expected differences" and we note that for most outcomes the company do not discuss whether clinical similarity (i.e. non-inferiority) of linzagolix compared to relugolix CT can be inferred from the NMA results.

The EAG were unable to verify the NMA and MAIC analyses and so the possibility of errors cannot be excluded. Results of NMAs on the company's extended evidence network, provided in Clarification Response A11 are very difficult to interpret because of lack of clarity in how the company conducted the analysis. Important limitations are:

- The PRIMROSE and PEARL II trials used in the evidence network employed different methods for estimating MBL (alkaline haematin and PBAC respectively) which in TA832 were considered incompatible without adjustment (section 3.2.3.1.1). It is unclear whether an adjustment was made to allow the different approaches to be compared in the network and, if so, how.
- The PEARL trials reported outcomes at 13 weeks whereas the timepoint of interest in the PRIMROSE trials is 24 weeks, although the PRIMROSE trials also reported some outcomes at 12 weeks. The company do not explain what timepoint their NMA results refer to (i.e. whether 12, 13 or 24 weeks) and what assumptions were applied to allow the trials' different assessment timepoints to be compared.
- The PEARL I and PEARL II trials include two ulipristal arms (5mg and 10mg) but it is unclear which of these the company included in the evidence network.
- The company do not explain in Clarification Response A11 how the trials were selected for the extended evidence network and so we are uncertain whether any relevant studies of GnRH analogues might have been missed.

3.5 Results from the NMA and MAIC analyses

Results from the NMA and MAIC analyses comparing linzagolix against relugolix CT are presented below in section 3.5.1. A summary of the results across all the outcomes and linzagolix regimens is provided in section 3.5.2.

Given the extensive limitations of the NMAs comparing linzagolix against leuprolide acetate noted above, results of these analyses are provided in Appendix 3 for reference only.

3.5.1 NMA results comparing linzagolix against relugolix CT

The results reported here are relevant to the company's cost comparison analyses for Population #1 and Population #2. For results relevant to the company's cost utility analysis (Population #3), for which the comparator is best supportive care, see section 3.2.5.

3.5.1.1 Response

3.5.1.1.1 Fixed-effects model results

Forest plots produced by the EAG from the data in company Clarification Response Table 3 (Figure 9) show [redacted] compared to those receiving r
not seen in PRIMROSE 1 it is difficult to explain. In contrast, relugolix CT is [redacted] the 100
100mg + ABT regimen. The remaining comparisons have wide credible intervals that include the null and ar
and MAIC results for PRIMROSE 2, particularly for the linzagolix 200mg dose regimen, but broad agreemen

sample size (ESS) values in the MAIC analysis for PRIMROSE 2 are very low (range [redacted] to [redacted] across the linza
ESS [redacted] for all regimens (Clarification Response Table 12), suggesting that the pooled analysis achieved r
reliable. The EAG therefore base our inferences for this outcome on the pooled trials analysis.

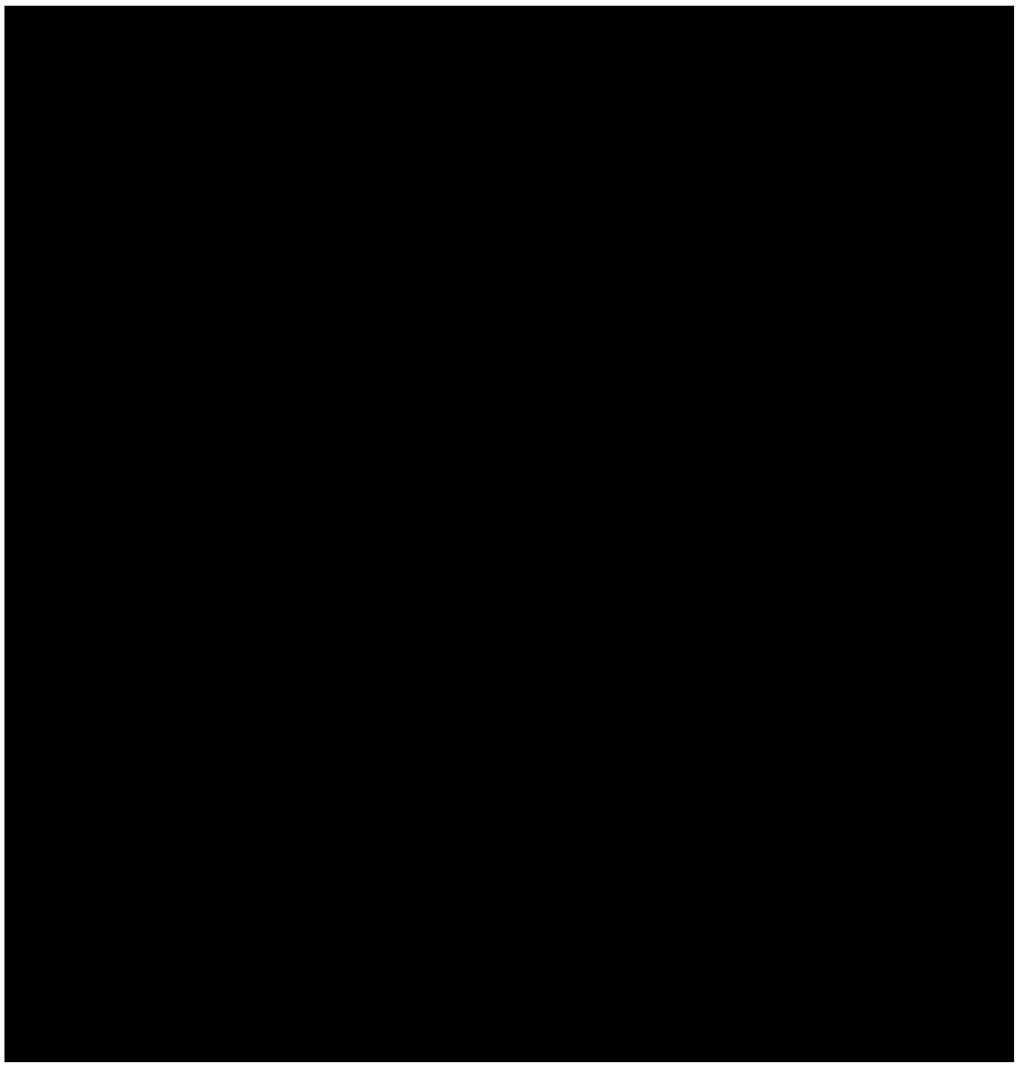


Figure 9 Linzagolix vs relugolix CT: NMA and MAIC results for response

The NMA forest plot and the posterior rank distribution probabilities for the fixed-effects pooled analysis (Cla
200mg + ABT is [redacted] relugolix CT at eliciting a response (with [redacted]% probability of being [redacted]
be driven mainly by the large effect in PRIMROSE 2 which was not seen in PRIMROSE 1. The forest plot sh
analysis the point estimates have credible intervals that lie [redacted], favouring [redacted], with a
favourable than [redacted]In conclusion, similarity of linzagolix and relugolix
[redacted]

3.5.1.1.2 Comparison with random-effects model results

The company provided random-effects model results only for the pooled analysis. As shown in Figure 10, th
is not possible to say with any certainty where the true point estimates lie. [redacted]
provide any guidance on interpretation. To be confident that one therapy is similar (i.e. non-inferior) to the ot
reasonable confidence that the point estimate lies close to the null, or that the log odds of achieving a mean
zero for that therapy. Posterior rank probabilities for the relative effectiveness of the intervention and compa
the company did not provide these for the random-effects analyses.

Figure 10 Response, pooled analysis, fixed & random effects models compared

3.5.1.2 Menstrual blood loss

3.5.1.2.1 Fixed-effects model results

Forest plots for the change in menstrual blood loss (Figure 11) show disagreement between the NMA and MAIC results for PRIMROSE 2 but broad agreement for PRIMROSE 1 and the pooled PRIMROSE trials. ESS values for MAIC analyses for PRIMROSE 2 are very low (range [redacted] to [redacted] across the linzagolix regimens) whilst the MAIC pooled analysis has the

highest ESS [redacted] for all linzagolix regimens (Clarification Response Table 13), suggesting that the pooled analysis achieved reasonably good matching of the trial populations. The EAG therefore base our inferences for this outcome on the pooled trials analysis.

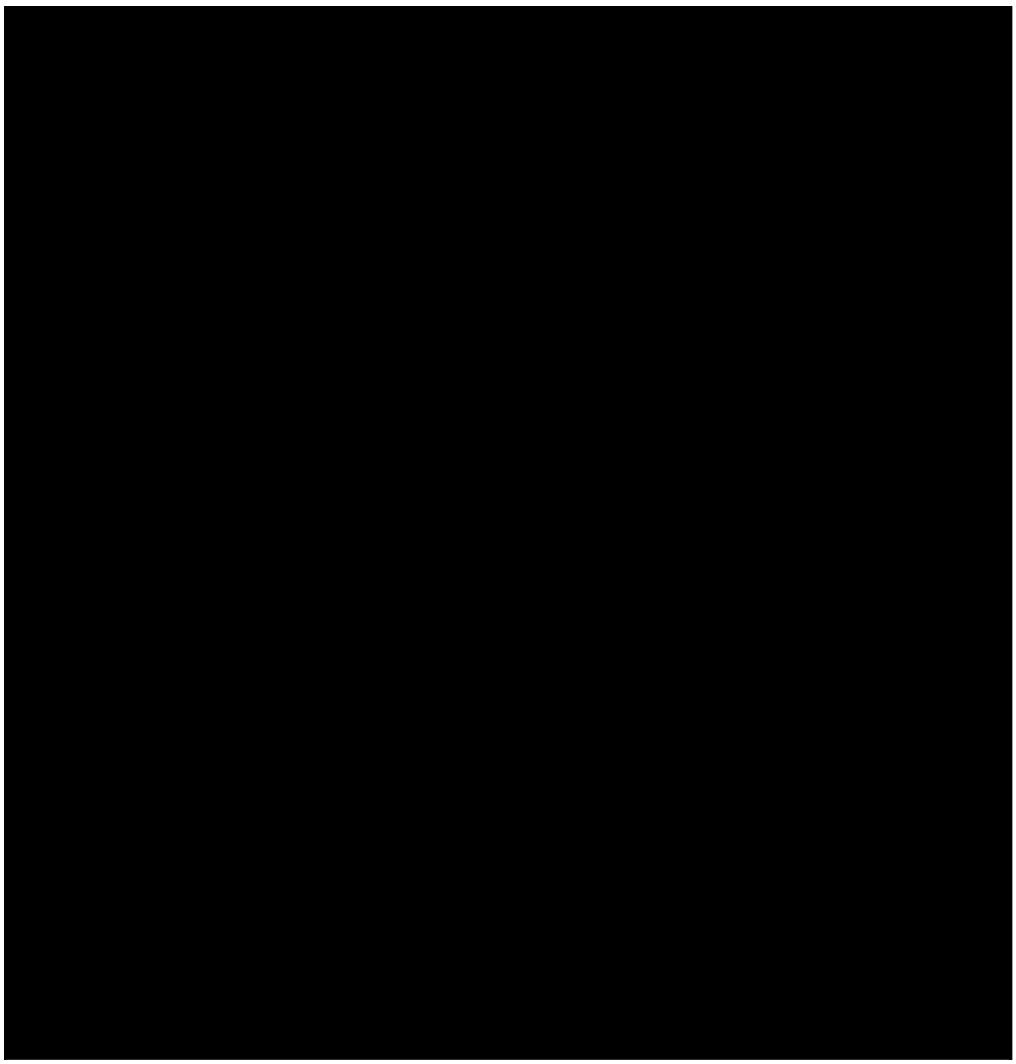


Figure 11 Linzagolix vs relugolix CT: NMA and MAIC results for menstrual blood loss

For the pooled analysis, relugolix CT is statistically superior to the 100mg, 100mg + ABT and 200mg regimens of linzagolix, so a conclusion of clinical similarity of linzagolix to relugolix CT is not supported for these regimens. The difference is non-significant for the 200mg + ABT regimen but the credible interval is wide and lies mostly above the null, suggesting relugolix CT is more likely to be superior. The posterior rank probability that relugolix CT is superior to all regimens of linzagolix and placebo is [redacted] (Clarification Response Document Appendix 1).

3.5.1.2.2 Comparison with random-effects model results

There is [redacted] between the fixed-effects and random-effects model analyses for the change in menstrual blood volume (Figure 12), with the random-effects analysis having [redacted] credible intervals. The statistical [redacted] of relugolix CT [redacted] the 100mg, 100mg + ABT and 200mg regimens of linzagolix at reducing MBL is therefore supported. Although statistically non-significant, the credible interval for the 200mg +ABT point estimate [redacted], suggesting that relugolix CT is [redacted] this linzagolix regimen. These findings [redacted] a conclusion that linzagolix is statistically similar to relugolix CT at reducing menstrual blood loss.

Figure 12 Change in MBL, pooled analysis, fixed & random effects models compared

3.5.1.3 Proportion experiencing improvement in fibroid-related pain

3.5.1.3.1 Fixed-effects model results

Forest plots for the log odds of achieving a meaningful improvement in the pain score (Figure 13) show disagreement between the NMA and MAIC results for PRIMROSE 2 but broad agreement for PRIMROSE 1 and the pooled PRIMROSE trials. The ESS values for the MAIC analysis for PRIMROSE 2 are very low (range [redacted] to [redacted] across the linzagolix regimens) whilst the MAIC pooled analysis has the highest ESS range [redacted] to [redacted] across the

linzagolix regimens, suggestive of relatively poor matching of the trial populations (Clarification Response Table 14).

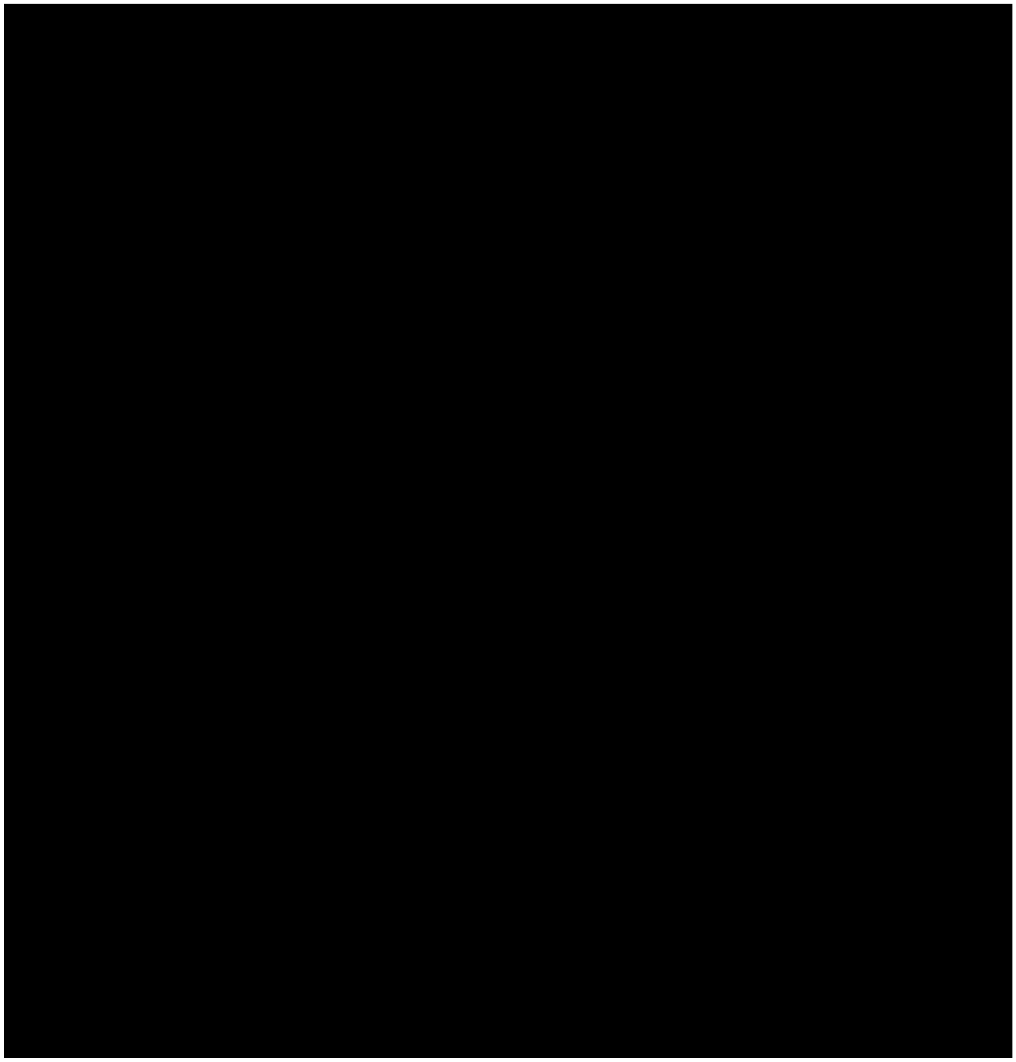


Figure 13 Linzagolix vs relugolix CT: NMA and MAIC results for the proportion achieving an improvement in pain total score

The NMA forest plot for the pooled analysis indicates [redacted] difference for the odds of improvement in the pain score between [redacted] of the linzagolix regimens and relugolix CT (all credible intervals include zero). With wide credible intervals either side of zero for the 100mg and 100mg + ABT regimens it is not possible to say with any certainty where the true point estimates lie. We are therefore unable to conclude that 100mg and 100mg + ABT linzagolix regimens are [redacted] as relugolix CT. For the linzagolix 200mg and 200mg + ABT regimens most of the credible intervals lie above zero. The posterior rank probabilities for each of these treatments (Clarification Response Document Appendix 1) suggest that there

is a [redacted] probability that the linzagolix 200mg regimen and a [redacted] probability that the linzagolix 200mg + ABT regimen are rank 1 or 2, which supports the suggestion that these linzagolix regimens may be [redacted] as relugolix CT at reducing fibroid-related pain.

3.5.1.3.2 Comparison with random-effects model results

As shown in Figure 14, the fixed-effects pooled trials analysis underestimates the heterogeneity present, with much narrower credible intervals than the random-effects analysis. The credible intervals in the random-effects analysis are so wide that is not possible to say with any certainty where the true point estimates lie.

[redacted].

Figure 14 Pain improvement, pooled analysis, fixed & random effects models compared

3.5.1.4 Primary fibroid volume

3.5.1.4.1 Fixed-effects model results

Forest plots based on fixed-effects models for the change in primary fibroid volume (Figure 15) show disagreement between the NMA and MAIC results for PRIMROSE 2 but broad agreement for PRIMROSE 1 and the pooled PRIMROSE trials. The MAIC pooled trials analysis has the highest ESS (range [redacted] to [redacted] across the linzagolix regimens) (Clarification Response Table 15) but there is also a moderate amount of missing data ([redacted]% to [redacted]%)

compared to the sample size for the response outcome. Although the pooled analysis has better matching and sample size compared to the individual trials there are uncertainties around the approach used for measuring fibroid volume (section 3.2.3.1.3), reducing confidence in the findings.

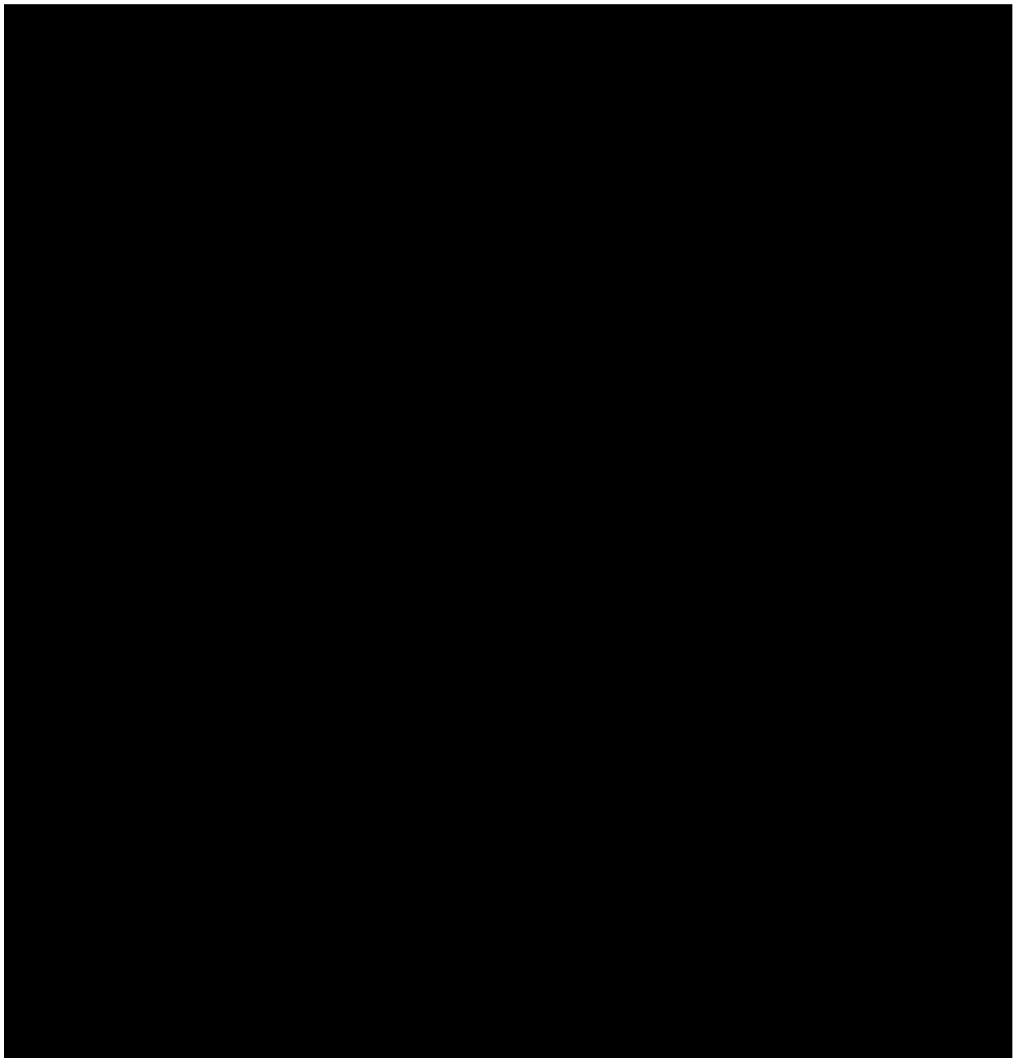


Figure 15 Linzagolix vs relugolix CT: NMA and MAIC results for primary fibroid volume

The NMA forest plot for the fixed-effects pooled analysis indicates that linzagolix 200mg is statistically [redacted] for reducing the primary fibroid volume, with [redacted]% posterior rank probability of being the most effective therapy at reducing primary fibroid volume, with the 100mg regimen having [redacted]% probability of being rank 2 (i.e. the second most effective therapy), followed by the 200mg + ABT regimen with [redacted]% probability of being rank 3 (i.e.

the third most effective therapy) (Clarification Response Document Appendix 1). However, for the linzagolix 100mg + ABT regimen the credible interval is [redacted], hindering a clear inference regarding similarity of linzagolix and relugolix CT. This regimen has a [redacted] posterior rank probability of being among the most effective therapies at reducing fibroid volume.

In summary, we conclude that, the 200mg regimen of linzagolix is [redacted] compared to relugolix CT for reducing primary fibroid volume whilst the linzagolix 100mg and 200mg + ABT regimens appear likely to be [redacted] for this outcome.

3.5.1.4.2 Comparison with random-effects model results

The random-effects analyses for the pooled trial populations generally confirm the findings of the fixed-effects analyses, demonstrating [redacted] of the linzagolix 200mg regimen compared to relugolix CT at reducing the primary fibroid volume. The credible intervals for the 100mg and 200mg + ABT linzagolix regimens remain [redacted] and largely [redacted], suggestive that linzagolix would be [redacted] relugolix CT at reducing the primary fibroid volume. A caveat, however, is that the EAG have concerns about whether the approach for selecting and measuring the primary fibroids was in the PRIMROSE trials was appropriate (section 3.2.2.1).

[redacted]

Figure 16 Change in primary fibroid volume, pooled analysis, fixed & random effects models compared

3.5.1.5 Haemoglobin, % change from baseline in patients with anaemia at baseline

3.5.1.5.1 Fixed-effects model results

Forest plots for the change in haemoglobin (Figure 17) show disagreement between the NMA and MAIC results, especially for PRIMROSE 2. However, sample sizes in the MAIC analyses are very low, with no ESS value greater than [redacted] for any of the linzagolix regimen groups, and ESS only [redacted] to [redacted] for the PRIMROSE 2 trial analyses (Clarification Response Table 16). Unsurprisingly, the credible intervals are very wide, making it difficult to determine with any certainty where the true effect estimates would lie.

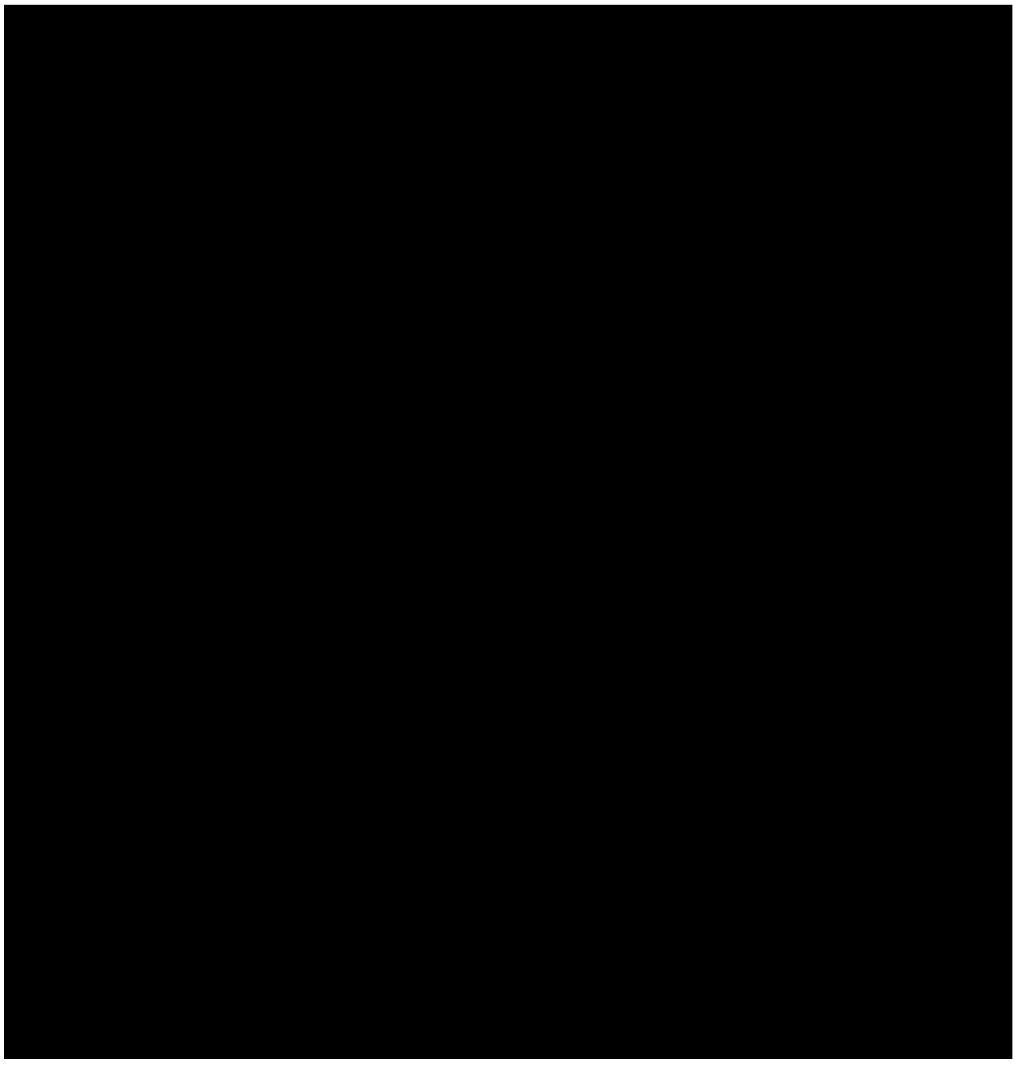


Figure 17 Linzagolix vs relugolix CT: NMA and MAIC results for % haemoglobin change (baseline anaemia subgroup)

3.5.1.5.2 Comparison with random-effects model results

The random-effects model results (Figure 18) have wider credible intervals, indicating that the fixed-effects analyses do not fully account for the statistical heterogeneity. Both the fixed- and random-effects analyses show similar distributions of the effect estimates in the forest plots, but do not resolve the uncertainty in where the true effect estimates lie. In conclusion, [redacted] at improving haemoglobin levels in patients who were anaemic at baseline.

Figure 18 Change in haemoglobin (baseline anaemia subgroup), pooled analysis, fixed & random effects models compared

3.5.1.6 UFS-QoL total HRQoL score

3.5.1.6.1 Fixed-effects model results

Forest plots for the change in UFS-QoL total score (Figure 19) show disagreement between the NMA and MAIC results for PRIMROSE 2, especially for the linzagolix 200mg and 200mg + ABT regimens, but broad agreement for PRIMROSE 1 and the pooled trials analysis. The MAIC pooled trial analysis has the highest ESS (range [redacted] to [redacted] across the linzagolix regimens (Clarification Response Table 17), suggesting moderate matching of the trial populations for this outcome, albeit with [redacted]% to [redacted]% fewer data than were available for analysis of the

response outcome. The EAG therefore base our inferences for this outcome on the pooled trials analysis, acknowledging that incomplete population matching introduces uncertainty.

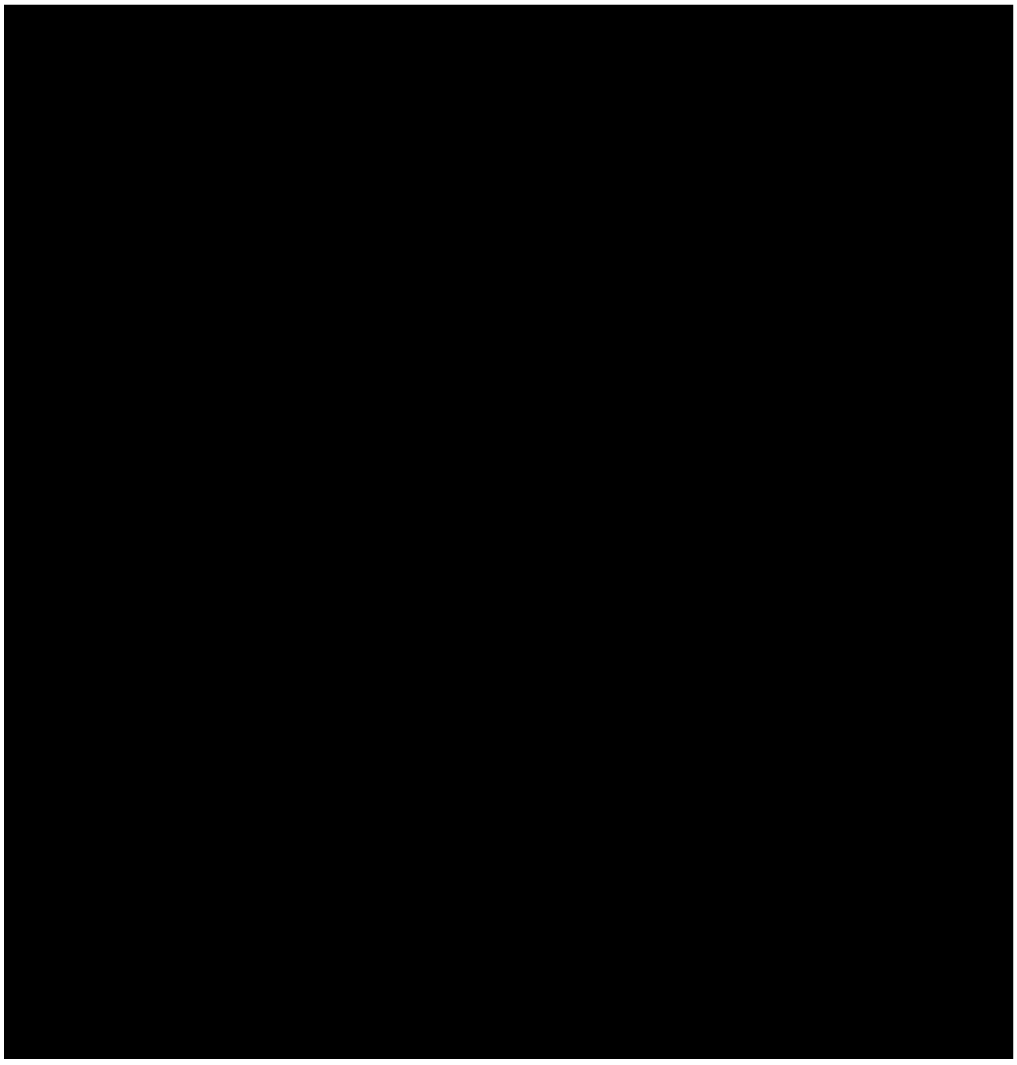


Figure 19 Linzagolix vs relugolix CT: NMA and MAIC results for UFS-QoL total score

The NMA forest plot for the pooled analysis indicates that relugolix CT is [redacted] to the 100mg and 100mg + ABT regimens of linzagolix for improving the UFS-QoL total score. Credible intervals for the 200mg and 200mg + ABT linzagolix regimens [redacted], suggesting that relugolix CT is [redacted] these linzagolix regimens. The posterior rank probability that relugolix CT is the most effective of the treatments for this outcome is [redacted]% (Clarification Response Document Appendix 1).

3.5.1.6.2 In summary, we conclude from the fixed-effects analysis that [redacted] the linzagolix regimens are conclusively [redacted] compared to relugolix CT, with relugolix CT being [redacted] to the linzagolix 100mg and 100mg + ABT regimens. Comparison with random-effects model results

The random-effects model results (Figure 20) have wider credible intervals, indicating that the fixed-effects analyses do not fully account for the statistical heterogeneity. The random-effects analyses still support the [redacted] of relugolix CT compared to 100mg linzagolix for improving fibroid-related quality of life, but the [redacted] of relugolix CT compared to linzagolix 100mg + ABT is less certain. In summary, we still conclude that [redacted] the linzagolix regimens are conclusively [redacted] compared to relugolix CT.

[redacted]

Figure 20 Change in UFS-QoL total HRQoL score, pooled analysis, fixed & random effects models compared

3.5.2 Summary of NMA results for Populations #1 and #2

Table 9 summarises the EAG's interpretation of the NMA results for the comparisons of linzagolix against relugolix CT. Despite the similar mechanisms of action of these therapies there appears to be very little evidence for the similarity or superiority of linzagolix compared to relugolix CT for most of the outcomes and dose regimens tested. This might to some extent reflect the weaknesses in the evidence synthesis methods discussed above, which

appear to have failed to cope well with heterogeneity in the clinical trials designs and populations.

Table 9 Overview of NMA results, linzagolix versus relugolix CT at Week 24

Outcome/model	Linzagolix regimen			
	100mg	100mg + ABT	200mg	200mg + ABT
Response – odds of achieving a response (meaningful reduction in MBL)				
Fixed effects Section 3.5.1.1.1	[redacted]	[redacted]	[redacted]	[redacted]
Random effects Section 3.5.1.1.2	[redacted]	[redacted]	[redacted]	[redacted]
% change in menstrual blood loss				
Fixed effects Section 3.5.1.2.1	[redacted]	[redacted]	[redacted]	[redacted]
Random eff	[redacted]	[redacted]	[redacted]	[redacted]

ects Section 3.5.1.2.2				
Pain – odds of achieving a meaningful improvement in NRS pain score				
Fixed effects Section 3.5.1.3.1	[redacted]	[redacted]	[redacted]	[redacted]
Random effects Section 3.5.1.3.2	[redacted]	[redacted]	[redacted]	[redacted]
% change in primary fibroid volume				
Fixed effects Section 3.5.1.4.1	[redacted]	[redacted]	[redacted]	[redacted]
Random effects	[redacted]	[redacted]	[redacted]	[redacted]

Section 3.5.1.4.2	[redacted]		[redacted]	[redacted]
% change in haemoglobin in patients anaemic at baseline				
Fixed effects Section 3.5.1.5.1	[redacted]	[redacted]	[redacted]	[redacted]
Random effects Section 3.5.1.5.2	[redacted]	[redacted]	[redacted]	[redacted]
Change in UFS-QoL HRQoL score				
Fixed effects Section 3.5.1.6.1	[redacted]	[redacted]	[redacted]	[redacted]
Random effects Section 3.5.	[redacted]	[redacted]	[redacted]	[redacted]

1.6.2				

3.6 Clinical efficacy conclusions

3.6.1 Population #1

As noted in Table 3, Population #1 (those receiving short-term therapy prior to surgical intervention) could in theory receive any of the linzagolix regimens but the 200mg regimen without ABT is suggested by the company be the optimal regimen for achieving the fastest fibroid shrinkage prior to surgery.

The NMA results suggest that the 200mg dose of linzagolix without ABT, which we assume to be the most relevant regimen for Population #1, is [redacted] to relugolix CT for the reduction in primary fibroid volume (Table 9) but not for any of the other outcomes assessed.

Insufficient reliable evidence has been provided by the company to determine whether linzagolix 200mg would have comparable effectiveness to other GnRH analogues.

3.6.2 Population #2

The most appropriate linzagolix regimens for Population #2, who receive longer-term therapy, would be 100mg + ABT or 200mg + ABT. As shown in Table 9, the 200mg + ABT regimen of linzagolix appears to have [redacted] compared to relugolix CT for reducing primary fibroid volume. However, there is no conclusive evidence that the 100mg + ABT regimen has comparable effectiveness to relugolix CT for any of the outcomes tested.

Insufficient reliable evidence has been provided by the company to determine whether linzagolix 100mg + ABT or 200mg + ABT would have comparable effectiveness to other GnRH analogues.

3.6.3 Population #3

Evidence relevant to Population #3 (longer-term treatment without ABT, where the comparator is BSC) is summarised in section 3.2.5 of this report. Overall, the pooled analyses for Week 24 outcomes show that linzagolix 100mg without ABT and 200 mg without ABT are more effective than placebo for all reported outcomes. However, in the PRIMROSE 1 trial population, 100 mg linzagolix without ABT was not more effective than placebo for some outcomes.

Caution should be exercised in interpreting results for fibroid volume (only 3 largest fibroids were measured, missing data) and for Hb concentrations in the subgroup who were anaemic at baseline (severely anaemic patients were excluded from the trials), none of the trial participants were contraindicated for ABT, and there is a placebo effect observed for the primary outcome.

The HRQoL results are ambiguous because linzagolix does not show any significant improvements compared to placebo according to the EQ-5D-5L results, yet linzagolix shows improvements in the UFS-QoL results (much greater than a 9-15 point change from baseline used in another study to indicate a clinically meaningful change).

Safety results show linzagolix is well-tolerated, but there is some uncertainty around long-term effects on BMD.

3.7 Uncertainties in the clinical efficacy evidence

There are numerous uncertainties in the evidence base for this technology appraisal. Major uncertainties have been raised as Key Issues with the aim that further consultation might enable some of the uncertainty to be resolved (see section 1.1 of this report for details).

Key Issue 1: Uncertain clinical similarity of linzagolix to other GnRH analogues. NMAs were conducted for six outcomes but linzagolix was only clinically similar (at least as good as) relugolix CT for one of these - reducing the volume of the primary fibroids. However, results are uncertain due to challenges in interpreting clinical similarity when NMA effect estimates have wide credible intervals. Lack of methodological details about how the NMAs were conducted precludes any interpretation of whether linzagolix might have clinical similarity to leuprolide acetate.

Key Issue 2: Uncertain relevance of the PRIMROSE trials to the three sub-populations in the company Decision Problem and NICE scope. Due to the trials' eligibility criteria and short duration, few patients if any from the three population sub-groups #1 to #3 are included in the trials. We are uncertain whether those who were included could serve as a proxy for those not included, for instance whether the efficacy and safety of linzagolix without ABT would differ between people who can or cannot receive hormone therapy.

Uncertainties in the company's analysis methods. The EAG were unable to validate the company's NMA or MAIC analyses as we did not have access to the individual participant data, and the statistical code provided did not specify sufficient details of the data format. There are inconsistencies between the individual PRIMROSE trials and the pooled trials analysis that raise uncertainty in how the pooled analysis was conducted and whether it was

quality-checked. These affect the odds ratio for response, and the UFS-QoL symptom severity score, where the pooled outcome effect estimates lie outside the range of the individual trial effects. The NMAs provided for the comparison of linzagolix against leuprolide cannot be usefully interpreted because of the lack of methodological clarity; even the outcome assessment timepoint was not provided.

In the MAIC analyses the PRIMROSE 2 trial was poorly matched. Simulated treatment comparison (STC) could be explored as an alternative indirect treatment comparison approach which may be more suitable than MAIC when there is less overlap of the population characteristics. Other opportunities to further explore and perhaps reduce uncertainty could be to conduct sensitivity analyses on the NMA and MAIC analyses adjusting for different sets of covariates and to test robustness of the analyses to missing data.

4 COST EFFECTIVENESS

4.1 Company review of cost-effectiveness evidence

The company report their economic search strategy in CS section B.3.1 and CS Appendix G. They conducted searches for published economic evaluations for GnRH antagonists. Five cost-effectiveness or cost-minimization studies that assessed pharmacological treatments for uterine fibrosis and one prior NICE appraisal in moderate to severe symptoms of uterine fibrosis (TA832) were identified and summarised (CS Table 46). In TA832, a cohort-level Markov model was developed to assess the cost-effectiveness of relugolix CT compared to GnRH agonists (goserelin, triptorelin, leuprorelin) in pre-menopausal women with moderate to severe UF symptoms who have failed or are unsuitable for conventional hormonal therapy.

EAG conclusion on cost-effectiveness searches

The original searches were conducted on 21 July 2021, and update searches in March 2022 and February 2023. No grey literature sources were reported in CS Appendix G, although the company reported hand searching of published literature across several conferences. Of the identified and reported studies in the company's search, the NICE appraisal TA832 is the most pertinent to the current appraisal.

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

4.2.1 NICE reference case checklist

The company's economic evaluation is discussed in relation to the NICE reference case in Table 10.

Table 10 NICE reference case checklist

Element of health technology assessment	Reference case	EAG comment on company's submission
Perspective on outcomes	All direct health effects, whether for patients or, when relevant, carers	Yes
Perspective on costs	NHS and PSS	Yes

Element of health technology assessment	Reference case	EAG comment on company's submission
Type of economic evaluation	Cost–utility analysis with fully incremental analysis	Yes, for Population #3; cost-comparison for Populations #1 & #2
Time horizon	Long enough to reflect all important differences in costs or outcomes between the technologies being compared	Yes. Six months for Populations #1, 10 years for Populations #2 and #3. In the cost-effectiveness analysis for Population #3, scenarios were conducted with 30 years and 60 years (lifetime horizon). Changing the time horizon has no impact on the ICER as mortality is not affected.
Synthesis of evidence on health effects	Based on systematic review	Yes
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.	HRQoL not applied in the CCA for Populations #1 and #2. In the CEA for Population #3, trial-based disease-specific measure UFS-QoL was mapped to EQ-5D-3L in the base case and scenario was conducted using the estimates obtained from mapping the trial-based EQ-5D-5L to EQ-5D-3L.
Source of data for measurement of health-related quality of life	Reported directly by patients and/or carers	Yes, HRQoL data using UFS-QoL and EQ-5D-5L were collected in the PRIMROSE trials.

Element of health technology assessment	Reference case	EAG comment on company's submission
Source of preference data for valuation of changes in health-related quality of life	Representative sample of the UK population	Yes
Equity considerations	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	Yes (severity modifier does not apply, CS B.3.6)
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	Yes
Discounting	The same annual rate for both costs and health effects (currently 3.5%)	No discounting in the CCA for Populations #1 and 2; same discount rate of 3.5% for costs and health effects applied in the CEA for Population #3.
Source: EAG assessment based on the company submission CCA: Cost Comparison Analysis; CEA: Cost Effectiveness Analysis		

4.2.2 Model structure

The company presented a blended approach for this appraisal, submitting two economic models for three subgroups of patients:

- A cost-comparison model for patients having short-term treatment of 6 months or less (Population #1); and those having longer-term treatment, with hormone-based therapy (Population #2)
- A cost-effectiveness model for patients having longer-term treatment, without hormone-based therapy (Population #3).

4.2.2.1 Overview of the model structure

The key features of the cost-comparison model for Population #1 and Population #2 and the cost-effectiveness model for Population #3 are presented in the following sub-sections.

4.2.2.1.1 Cost-comparison model for Populations #1 and Population #2

Key features of the model are:

- Proportion of patient estimated in four states: on-treatment, off-treatment, menopause, and death.
- For Population #2, patients are assumed to undergo no further pharmacological treatment or surgery after menopause.
- No clinical efficacy parameters were included for Population #1. For Population #2, treatment discontinuation was incorporated: a discontinuation rate of [redacted] was obtained from the PRIMROSE trials and converted to per cycle probability, and applied across all the treatment arms (discussed in Section 4.2.5.1).
- Time horizon: 6 months (Population #1); 10 years (Population #2)
- No discounting
- Perspective: National Health Service (NHS)/Personal Social Services (PSS)
- Cycle length: 28 days
- Costs included: Drug costs, administration costs, healthcare resource use costs and costs associate with surgery.

4.2.2.1.2 Cost-effectiveness model for Population #3

Key features of the model are:

- Markov model with six health states comprising:
  - Uncontrolled (defined by heavy menstrual bleeding (HMB) >80 mL MBL per cycle)
  - Controlled (defined as achieving MBL ≤ 80 mL and ≥50% reduction from baseline)
  - Surgery
  - Post surgery
  - Menopause
  - Death
- Time horizon: 10 years
- Discounting: 3.5% p.a.
- Perspective: NHS/PSS
- Cycle length: 28 days
- Half cycle correction applied.

Patients enter the model in the ‘uncontrolled’ state where they receive treatment. After this, their symptoms can remain ‘uncontrolled’ or can be ‘controlled’. Patients transition to surgery, menopause, or the death state from the ‘uncontrolled’ state. Those in the

'controlled' state may remain there or may lose response and transition to uncontrolled, surgery, menopause, or death. Surgery is assumed to last for only one cycle after which patients transition to the 'post-surgery' state and remain there till the onset of menopause. A schema of the model structure is presented in CS Figure 25.

EAG conclusion on the model structure

The company's simple modelling approach for the cost-comparison analysis is reasonable. We view the model structure for the cost-effectiveness analysis, based on symptom control (controlled, uncontrolled) as appropriate, based on clinical expert advice to the EAG and committee discussions in NICE TA832. The company explored the impact of varying model features in their scenario analyses. For further details, see section 5.2.2.

4.2.3 Population

The company specify the target population for linzagolix in CS Section B.3.2.1. Adults of reproductive age with moderate to severe symptoms of UF are divided into three subgroups:

- Population #1: People with short-term treatment of ≤6 months
- Population #2: People with longer-term treatment, with hormone-based therapy.
- Population #3: People with longer-term treatment, without hormone-based therapy.

The baseline characteristics used in the economic analyses were mean age (42.25 years) and average age of menopause (51 years). The CS assumed that all patients transitioned to the menopause state on reaching the age of 51 years, after which they did not experience any disease-related symptoms.

EAG conclusion on the model population

The patient subgroups included in the company analyses align with the final NICE scope for this appraisal. Patient characteristics in the company's analyses, based on the PRIMROSE 1 & 2 trial populations and TA832, are reflective of UK clinical practice. We note that Population #3, which includes patients who are unable to receive hormone-based therapy (shown in CS Figure 1), is inconsistent with the population in the PRIMROSE trials. The clinical trials included people who could be randomised to any trial arm, with or without hormone-based therapy, and therefore do not include a population unable to receive hormone-based therapy (see the eligibility criteria in CS Table 9). The EAG's clinical expert suggested that, in practice, very few patients would be unfit/prefer not to receive hormone-based therapy, so the unique population who could benefit from linzagolix without hormone-based therapy in clinical practice could be very small. However, this is uncertain (see section 2.2.4 above).

4.2.4 Interventions and comparators

The economic models evaluate the intervention (linzagolix) against specific comparators for each sub-population (GnRH analogues for Population #1 and Population #2; best standard of care [BSC] for Population #3). The company describe the intervention in CS section B.1.2 and we discuss the intervention and its intended use in practice earlier in Section 2.2.2 of this report. The dosing regimen for linzagolix (see CS Section B.3.2.2), that received marketing authorization, is consistent with the PRIMROSE trials.

In their cost-comparison model, the company assumed a dosage of 200mg for Population #1; and 200mg + ABT for Population #2. For Population #3 (in the cost-effectiveness model), a dosage of 200mg for 6 months followed by 100mg was assumed for their base case. Scenario analyses were conducted using different dosing regimens in both the cost-comparison analysis and cost-effectiveness analysis. These assumptions had no significant impact on the incremental costs in the cost-comparison analysis for Population #1 and Population #2. But changing the regimen in the cost-effectiveness analysis for Population #3 had significant impacts on the overall ICERs. For instance, changing to 200mg for 6 months followed by BSC or using 100mg continuously from the start, increased the company's ICER by £1,443 and £1,973, respectively.

The comparators included across the three subgroups are:

- Population #1: relugolix CT and GnRH agonists (goserelin, leuprorelin and triptorelin)
- Population #2: relugolix CT
- Population #3: BSC comprising concomitant medications for pain management (NSAIDs) and iron supplements.

EAG conclusion on the intervention and comparators

The intervention included in the economic models is consistent with the NICE scope. However, there is uncertainty with respect to the comparators used across the three populations. To elaborate, in the cost-comparison analysis for Population #1 and Population #2, relugolix CT may not be most appropriate comparator as it appears to have low market share; clinical expert advice to the EAG suggests that most of the patients (circa 90%) currently receive goserelin or leuprorelin, although the use of relugolix would likely increase over time. With respect to Population #3, our expert advice is that these patients would require protection against bone loss, which is not included within the modelled BSC. This would include prophylactic regimens of calcium

and vitamin D. We conducted scenario analyses including these regimens as part of BSC, for further details see Section 6.

4.2.5 Treatment effectiveness and extrapolation

4.2.5.1 Clinical parameters used in the cost-comparison analyses

Population #1:

- Across both the treatment arms, a similar proportion of patients (45.10%) was assumed to receive surgery.
- The distribution of types of surgery for linzagolix was assumed to be same as for relugolix CT and sourced from the TA832 company submission.
- Background mortality was not incorporated as treatment only lasts for 6 months.

Population #2:

- Treatment discontinuation for linzagolix doses were obtained from the pooled PRIMROSE trials. It was assumed to be the same as linzagolix 200mg + ABT for relugolix CT ([redacted]).
- Like Population #1, 45.10% of patients were assumed to experience surgery across both the treatment arms.
- The distribution of the types of surgery was obtained from TA832.
- Mortality due to surgery was not incorporated.

Across both the populations, the costs of surgery were applied as a one-off cost. No data from company indirect treatment comparisons was used to inform clinical efficacy. The company justified their approach citing the ITC findings (discussed above in Section 3.4), their clinical expert opinion and NICE TA832 for similar efficacy of relugolix CT with GnRH agonists.

EAG conclusion on treatment effectiveness and extrapolation

The EAG have two concerns for the cost-comparison analysis.

1. We are uncertain whether:
    a. linzagolix has similar clinical efficacy as relugolix CT and other GnRH analogues (see Key Issue 1) and
    b. relugolix-CT has an adequate market share to qualify as the selected comparator for the cost-comparison analysis (see Key Issue 2).

If the company's assumption is not supported, then cost-utility analyses would be more appropriate.

2. The company did not apply any additional risk of mortality from surgery in Population #2. Although the mortality is very low, and particularly with the short time horizon is unlikely to impact on costs, we view that it ought to be included for completeness.

4.2.5.2 Clinical parameters used in the cost-effectiveness analysis.

The company applied the following transition probabilities in their economic model for Population #3, reproduced from CS Table 54 and shown in Table 11 below. We noted an inconsistency in the transition probability from surgery to death between the CS and the economic model (Clarification Response B2) which has been corrected in the table below. We discuss and critique the derivation of these estimates in the following subsections.

Table 11 Transition probabilities used in the cost-effectiveness model.

FROM / TO	Controlled	Uncontrolled	Surgery	Post-surgery	Procedural death
Linzagolix					
Controlled	[redacted]	[redacted]	[redacted]	0.000%	0.000%
Uncontrolled	[redacted]	[redacted]	[redacted]	0.000%	0.000%
Surgery	0.000%	0.000%	0.000%	99.997%	0.003%
Post-surgery	0.000%	0.000%	0.000%	100.000%	0.000%
Procedural death	0.000%	0.000%	0.000%	0.000%	100.000%
Best Supportive Care					
Controlled	[redacted]	[redacted]	[redacted]	0.000%	0.000%
Uncontrolled	[redacted]	[redacted]	[redacted]	0.000%	0.000%
Surgery	0.000%	0.000%	0.000%	99.997%	0.003%
Post-surgery	0.000%	0.000%	0.000%	100.000%	0.000%
Procedural death	0.000%	0.000%	0.000%	0.000%	100.000%

Note: Transition matrix does not include background mortality which is applied separately within the model calculations
Source: CS Table 54 and company economic model

4.2.5.2.1 Response rate

Response rates for linzagolix and BSC arms were informed by the pooled data from the PRIMROSE 1 and 2 trials at 24 weeks (which we discuss earlier in Section 3.2.5.1). While the trials included four treatment dosing regimens for linzagolix (100mg, 100mg + ABT, 200mg, and 200mg + ABT), the cost effectiveness model included clinical effectiveness estimates for only 100mg and 200mg doses, as these are the doses indicated for Population #3. The placebo-arm of the PRIMROSE trials was assumed to be representative of the

clinical effectiveness of BSC, due to lack of active treatment options for this subgroup. This is consistent with the approach adopted in TA832.

The pooled 24-week response rate from the PRIMROSE trials, as reported in CS Table 16, determined the proportion of patients entering the 'controlled' health state (defined as MBL ≤80 mL and ≥50% reduction from baseline at 24 weeks). Those who do not have a response or achieve but subsequently lose their response enter the 'uncontrolled' health state. The company applied a standard equation (see CS Equation 1) to convert the 24-week response rate to a per cycle probability of moving from the uncontrolled to the controlled health state. These probabilities are reproduced below in column 3 of Table 12.

Table 12 Response probabilities included in the model.

Treatment	24-week response rate	Response probability per cycle	Source for response rate
Linzagolix 200mg	[redacted]	[redacted]	Pooled PRIMROSE 1 & 2 (CS Table 16)
Linzagolix 100mg	[redacted]	[redacted]	
BSC	[redacted]	[redacted]	
Source: Company model, CS Table 50			

EAG conclusion on response

We note that the PRIMROSE trials provide a maximum duration of efficacy outcome assessments of 52 weeks, whereas the economic model used the efficacy outcomes at 24 weeks. Furthermore, the SmPC suggests linzagolix may be used for more than one year in clinical practice (subject to regular bone mineral density monitoring). In their response to Clarification Question A7, the company justified using 24-week data based on the argument that treatment effect of linzagolix is expected to be maintained over 52 weeks. They cited 2-year data from the LIBERTY randomised withdrawal study[30] for relugolix CT, where the treatment effect was maintained. They argued that given linzagolix has a similar mechanism of action to relugolix CT, its efficacy is likely to be maintained beyond 24 weeks. We are uncertain about this assumption; while the company's response is biologically plausible, there are no data to support this.

We did not identify any inconsistency in the response rate between CS Table 16 and the economic model. No error was identified in the conversion from rate to per cycle probabilities.

4.2.5.2.2 Recurrence rate

Recurrence rates of uterine fibrosis, converted to per cycle probabilities using the same formula as for response rate (i.e., CS Equation 1), were used to inform the probability of losing response and transitioning from the 'controlled' to the 'uncontrolled' state. The company state that these estimates (shown in Table 13 below) were obtained from a market research survey with UK gynaecologists (n=50). In a scenario analysis conducted by the company, these estimates are shown to have a significant impact on the cost-effectiveness results: increasing the recurrence rate during treatment with linzagolix to that used for BSC increased the base case ICER by £5,315, to £20,707.

Table 13 Recurrence rates used in the model for Population #3

Treatment	Recurrence rate	Estimated recurrence probability per cycle	Source
Linzagolix	[redacted]	[redacted]	Clinical opinion
BSC	[redacted]	[redacted]	
Source: CS Table 51 and the company's cost-effectiveness model			

EAG conclusion on recurrence rate

We conducted a targeted search of recurrence rate in patients with uterine fibrosis and found recurrence rates of 11.7% and 15.3% at 1-year post-laparoscopic myomectomy, [31, 32] and 23% at 40 months post-abdominal myomectomy[33] respectively. The EAG's clinical expert suggested that a recurrence rate of circa 23% in patients with UF may be reasonable for the linzagolix arm. For patients receiving BSC, the recurrence rate is likely to be higher as these patients are unable to receive any hormones including ABT. For completeness, we conducted a range of exploratory scenarios assuming a similar recurrence rate for linzagolix and BSC, thereby varying the rates between 10% and 25%, in Section 6.1.

4.2.5.2.3 Surgery

In the economic model, surgery is assumed to last for one model cycle. The following types of surgery were included:

- Urinary artery embolization (UAE)
- Magnetic resonance-guided focused ultrasound surgery (MRgFUS)
- Abdominal myomectomy
- Laparoscopic myomectomy
- Abdominal hysterectomy
- Laparoscopic hysterectomy

In the model, patients transition from the 'controlled' and 'uncontrolled' health states to the 'surgery' state. The surgery rate was informed by the PEARL II trial, which compared ulipristal acetate with leuprorelin acetate for the pre-operative treatment of symptomatic fibroids. Using the surgery rate from PEARL II (45.10%) and a waiting time of 18 months on average, the per cycle probability of surgery from the two health states was estimated to be 3.02%. In the base case, the probability of experiencing surgery from the two health states are assumed to be equivalent. Furthermore, the probability was assumed to be the same for linzagolix and BSC. Scenario analyses assuming lower surgery probabilities of 1% and 2% for patients in the 'controlled' state reduced the overall ICER by £4,251 and £863, respectively.

The company assume similar distributions of patients across the surgery types for both the treatment arms. The distribution for their base case, reproduced below in Table 14, were obtained from the CS of NICE TA832. Scenario analyses were conducted using the EAG estimates from the same appraisal (which had no impact on the overall cost-effectiveness results) and assuming a 10% switch from open/abdominal to laparoscopic surgery for patients in the linzagolix arm (which reduced the overall base case ICER by £2,873).

Table 14 Distribution of surgery types for the base case model

Surgery types	Patient distribution
UAE	4.8%
MRgFUS	3.0%
Open/abdominal myomectomy	25.7%
Laparoscopic myomectomy	8.2%
Open/abdominal hysterectomy	51.8%
Laparoscopic hysterectomy	6.4%

Source: CS Table 52
UAE: Urinary artery embolization; MRgFUS: Magnetic resonance-guided focused ultrasound surgery

EAG conclusion on surgery

The company's assumption that both linzagolix and BSC arms have similar distributions of surgery types is reasonable. With respect to patient distributions across the different surgery types, the EAG's clinical expert considered that some of the surgery types (e.g., laparoscopic hysterectomy and UAE are more common than others. Furthermore, patients are also likely to undergo hysteroscopic myomectomy, which is not listed in the company's analyses. Lastly, our clinical expert suggested that recovery time after

different types of surgery varies between 4 and 8 weeks. For example, the recovery time after laparoscopic surgery could be 4-6 weeks; open surgery: 6-8 weeks; UAE: 4-6 months. We have conducted scenario analyses changing the distributions across the different surgery types based on our expert's advice. While this impacts the total costs and total QALYs, the change is proportional as the distributions are similar for both the treatment arms and therefore there is no overall impact on the ICER (see Section 6).

4.2.5.2.4 Post-surgery health state

In the model, patients enter the 'post-surgery' state after undergoing surgery and remain there until the onset of menopause. We note that the choice of surgery type is dependent on a range of factors including disease characteristics and patient preferences, and therefore, the prognosis of the patients may vary depending on the type of surgery undergone. While some patients may be completely cured (e.g., those who underwent a hysterectomy), others may experience a recurrence of the symptoms post-surgery. In NICE TA832, the post-surgery state was sub-divided into two: patients who received hysterectomies and those who did not, with a proportion of the latter cohort transitioning to a second surgery state, based on re-surgery rates. Furthermore, evidence from published literature[34] indicates that although surgery has a high impact on the symptoms of the fibroids, these may recur (except after hysterectomies). Overall, we view the company's assumption that patients stay in the 'post-surgery' state until they experience menopause as simplistic and are unclear whether this is representative of the disease pathway.

EAG conclusion on the post-surgery state

The EAG are uncertain about: i) how the prognosis of different surgery types will vary; and ii) whether patients undergoing surgeries other than hysterectomies may experience a recurrence. We suggest the company could conduct scenario analyses by adding a percentage of recurrence within the post-surgery state for both the arms. Due to the uncertainties discussed above, we view further discussion with clinical experts is warranted.

4.2.5.2.5 Treatment discontinuation

In the economic model, patients could discontinue treatment either for any reasons (named as 'trial-based' treatment discontinuation by the company) or due to adverse events (named as 'modified' treatment discontinuation by the company). The discontinuation rates were obtained from the pooled PRIMROSE analysis at 24 weeks (shown below in Table 15) and converted to per cycle probabilities. These were applied throughout the time horizon for all

patients in the 'controlled' and 'uncontrolled' health states on both the treatment arms. In the base case, a constant risk of trial-based discontinuation was assumed.

The cited reasons for trial discontinuation were participant request, loss to follow-up, adverse event, lack of efficacy, pregnancy and other. Scenario analysis was conducted using the modified rates, i.e., discontinuation due to AEs only. The trial-based withdrawal rates are much higher than the modified rates, implying that fewer patients remain on treatment when the trial-based rates are applied. This has a significant impact on the ICER as it impacts the drug costs. For example, using modified rates (treatment discontinuation due to AEs) increases the ICER by £10,436, to £25,828. This is driven by a significant increase in the drug costs of linzagolix, particularly that of linzagolix 100mg as the proportion of patients receiving linzagolix 100mg post the initial 6 months is much higher due to modified discontinuation rates, compared to that of trial-based rates.

Table 15 Treatment discontinuation

Treatment	Base case (Trial based)		Scenario analysis (modified)		Source
	Withdrawal for any reason	Per cycle prob. of TTD (estimated)	Withdrawal due to AE	Per cycle prob. of TTD (estimated)	
Linzagolix 100mg	[redacted]	[redacted]	[redacted]	[redacted]	Pooled PRIMROSE analysis
Linzagolix 200mg	[redacted]	[redacted]	[redacted]	[redacted]	
BSC	[redacted]	[redacted]	[redacted]	[redacted]	

Source: Company's cost-effectiveness analysis model
TTD: Time to Treatment Discontinuation

The company do not discuss treatment discontinuation in their CS. Like the response rates, they applied the discontinuation rates available at 24 weeks in the economic model. Data available for 24-52 weeks were not used. We view that it is appropriate to use long-term data where available. In previous NICE TA832, the company applied the discontinuation rates for relugolix-CT from the LIBERTY trials, based on clinical opinion, and from PEARL II for GnRH agonists. The TA832 committee concluded the rates used in the model as highly uncertain and that the company's model did not accurately capture the uncertainty. In the current appraisal, the discontinuation rate, whether 'trial-based' or 'modified', is important because it impacts the acquisition costs, and hence the ICER. We conducted a range of exploratory

scenarios assuming similar discontinuation rates for linzagolix 200mg, 100mg and BSC, see Section 6.

EAG conclusion on treatment discontinuation

Overall, we view the company's approach to modelling treatment discontinuation to be reasonable as the estimates are based on the PRIMROSE trials and reflect the expectation that in practice patients may discontinue due to many reasons, other than AE. However, we suggest that analysis based on 52-week data would be appropriate.

4.2.5.2.6 Adverse event rates

The economic model included treatment emergent adverse events occurring in 5% or more of patients across the treatment arms of the pooled PRIMROSE trial. These included four adverse events: anaemia, headache, hot flash, and nausea, see CS Table 55. The average duration of these adverse events was assumed to be one model cycle and the rates (in CS Table 55) were multiplied with the associated disutilities (in CS Table 64) to obtain QALY loss due to AEs. This is applied in the first model cycle.

EAG conclusion on adverse events

Overall, we agree with the company's approach. Clinical advice to the EAG suggested that anaemia should be expected to improve on treatment in patients with UF. Furthermore, allergies and intolerances to medications are common adverse events witnessed among these patients, which are not included. However, inclusion of these events is unlikely to make any significant impact on the overall cost-effectiveness results. Lastly, we are unclear if the duration of the PRIMROSE trial is sufficient to capture any detrimental effects of reduction in bone density in these patients.

4.2.5.2.7 Mortality

Age-adjusted background mortality rates obtained from the ONS data for England were incorporated in the economic model. Mortality associated with surgery related complications were also incorporated. These estimates, reproduced below in Table 16, were obtained from NICE TA832.

Table 16 Risk of procedural death

Treatment arm	Risk of death	Source
UAE	0.0200%	TA832/Zowall et al., 2008
MRgFUS	0.0000%	TA832/Gorny et al., 2011
Open/abdominal myomectomy	0.0028%	TA832/Assumption
Laparoscopic myomectomy	0.0000%	TA832/Assumption

Open/abdominal hysterectomy	0.0028%	TA832/Settnes et al 2020
Laparoscopic hysterectomy	0.0020%	TA832/Settnes et al 2020
Source: CS Table 53		

EAG conclusion on mortality

The company's approach to modelling background mortality is appropriate. We also agree with the assumption of excess mortality associated with surgical procedures. Clinical advice to the EAG indicates that there is likely to be a higher risk of mortality associated with hysterectomies and myomectomies as these are major surgeries, whereas UAE is less risky than open surgery. Considering this, we conducted a scenario analysis with a decreased mortality rate (0.0002%) associated with UAE. For further details, see Section 6.

4.2.6 Health related quality of life

The discussion and critique of the health-related quality of life (HRQoL) data in the following sub sections relate to the cost-effectiveness model developed for the Population #3, comparing linzagolix with BSC. An overview of the utility values used in the cost-effectiveness analysis is presented in Table 17. The cost-comparison analyses for Population #1 and Population #2 do not incorporate HRQoL data.

Table 17 Summary of utility values for cost-effectiveness analysis

Health state	Treatment arm	Utility value	Source	EAG discussion
Controlled	Linzagolix (200 mg)	[redacted]	PRIMROSE 1 and 2 (UFS-QoL mapped to EQ-5D)	Section 4.2.6.2.1
	BSC	[redacted]		
Uncontrolled	Linzagolix (200 mg)	[redacted]		
	BSC	[redacted]		
Surgery	Surgery and post-surgery utility values are non-treatment specific in the base case	0.677	Literature	Section 4.2.6.2.2
Post-surgery		0.846	Literature	
Source: CS Table 67				

4.2.6.1 Systematic literature review for utilities

The company conducted a systematic literature review of existing HRQoL studies in patients with uterine fibrosis and report the search and findings in CS Appendix H. In total 47 studies met their inclusion criteria. Of these, seven included patients from the UK. Most of these studies used Uterine Fibroid Symptom and Quality of Life questionnaire (UFS-QOL) (n=6) to measure HRQoL, either as the singe instrument (n=4) or in combination with other instruments (n=2). Other instruments used included SF-36, EQ-5D-3L, and the EQ-5D visual analogue scale. Of the 47 included studies, the company applied the utility values obtained

from EQ-5D-5L for two health states, controlled (0.73) and uncontrolled (0.55), from a Canadian study by Hux et al. 2015[35] in their scenario analysis (see Section 5.2.2).

In the previous NICE appraisal TA832, the company included treatment-specific utility values, which were informed by MBL from the treatment arms of the LIBERTY trials. The UFS-QoL data, obtained from the LIBERTY trials, were mapped to EQ-5D using an unpublished mapping algorithm. An ordinary least squares (OLS) model, adjusted for age and MBL, was used to predict the impact of MBL on mapped EQ-5D utilities to generate time-varying utilities. The Evidence Review Group in TA832 expressed concerns over this approach due to lack of sufficient justification for the choice of regression model. They preferred a repeated measures model to allow for exploring uncertainties and generating utility estimates closer to that of general population averages when MBL was low. The TA832 NICE committee concluded that the model was likely to underestimate the utility values to inform the QALY gains with relugolix-CT.

4.2.6.2 Study-based health related quality of life

4.2.6.2.1 Controlled and uncontrolled health states

HRQoL data from the PRIMROSE trials were used to estimate utilities for the 'controlled' and 'uncontrolled' health states in the model. PRIMROSE 1 and 2 used patient self-reported, disease specific UFS-QoL scores and the EQ-5D-5L to collect HRQoL data at baseline and at weeks 12, 24, 36 and 52. The UFS-QoL and EQ-5D-5L data from the trials were mapped onto EQ-5D-3L utility values. The company then used a linear mixed model (LMM) to predict the utility values for 'controlled' and 'uncontrolled' health states based on reduced menstrual blood loss (RMBL). The utility function is not reported in the CS: we derived this as below, based on the information in the economic model. The coefficient of the intercept (α) is the estimated utility for the 'uncontrolled' state as this reflects patients without RMBL; whereas (α + β*RMBL) gives the estimated utility for the 'controlled' state, reflecting patients with RMBL.

$$EQ\text{-}5D_{mapped} = \alpha + \beta\ RMBL + \varepsilon$$

We reproduced the coefficients of the intercept and the RMBL from the utility functions from the economic model as shown in Table 18 below.

Table 18 Coefficients from the utility function obtained from the company's model.

	Base case (EQ-5D-3L utility mapped from UFS-QoL data)	Scenario (EQ-5D-3L utility mapped from EQ-5D-5L data)
Intercept	[redacted]	[redacted]
RMBL	[redacted]	[redacted]
Source: company cost-effectiveness model		

For their base case, the company mapped UFS-QoL to EQ-5D-3L and conducted a scenario analysis with EQ-5D-5L mapped to EQ-5D-3L.

- **Base case: Mapping of UFS-QoL to EQ-5D-3L**

Following the approach adopted in TA832, the company used an unpublished algorithm reported in a paper by Rowen and Brazier 2011 and shown in CS Equation 2, to map UFS-QoL to EQ-5D-3L.[29] To account for within-patient repeated measures (a critique raised by the Evidence Review Group in TA832), a linear mixed model (LMM) was chosen to estimate the health-state utilities. It is noteworthy the UFS-QoL measure includes two scales: symptom severity and HRQoL. The directions of these scales are opposite (e.g., decrease in symptom severity indicates improvement whereas increase in HRQoL scale indicates improvement). The CS does not explicitly define which of these scales was used. CS Equation 2 is based on individual UFS-QoL questions that appear to include parts of both the symptom severity and HRQoL scales rather than focusing on HRQoL questions. The rationale for this approach is unclear.

- **Scenario analysis: Mapping of EQ-5D-5L to EQ-5D-3L**

The company state that a linear mixed model was used to map EQ-5D-5L responses in the PRIMROSE trials to EQ-5D-3L using the algorithm developed by Hernandez-Alava et al. 2017,[36] as recommended by NICE.[1] The equation was not reported in the CS. Like in their base case, LMM was used to estimate the utility values.

Table 19 Mapped utilities used in the cost-effectiveness model.

Health state	Base case	Scenario analysis
Controlled	[redacted]	[redacted]
Uncontrolled	[redacted]	[redacted]
Source: CS Tables 58 and 61		

4.2.6.2.2 Surgery and post-surgery health states

The PRIMROSE trials did not collect HRQoL data for surgery or post-surgery health states. Therefore, utility estimates for these health states were informed by published literature (CS Table 66, reproduced below in Table 20). As stated earlier in Section 4.1, the company reported a literature search of HRQoL data associated with surgical/interventional procedures in patients with uterine fibrosis. Weighted average utility values for surgery and post-surgery were obtained based on the distribution of surgery types (discussed earlier in Section 4.2.5.2.3)

Table 20 Health state utilities for surgery and post-surgery

Surgery	Health state	Value
UAE	Surgery	0.620
	Post-surgery	0.800
MRgFUS	Surgery	0.783
	Post-surgery	0.802
Open/abdominal myomectomy	Surgery	0.628
	Post-surgery	0.878
Laparoscopic Myomectomy	Surgery	0.630
	Post-surgery	0.880
Open/abdominal Hysterectomy	Surgery	0.705
	Post-surgery	0.834
Laparoscopic hysterectomy	Surgery	0.707
	Post-surgery	0.836
Source: Reproduced from CS Table 66 and the company's economic model		

4.2.6.3 Adverse events

Disutilities associated with the adverse events (see Table 21) were informed by published literature. Further details are in CS section B.3.4.4. These estimates were multiplied with the frequency of the AE of the two treatment arms obtained from PRIMROSE trials, shown in CS Table 55, and discussed earlier in Section 4.2.5.2.6, to obtain the AE QALY decrement, shown in Table 22 below. These decrements were applied as a one-off in the first model cycle.

Table 21 Adverse event disutilities

AE	Disutility

Anaemia	-0.0209
Headache	-0.0297
Hot flush/flash	-0.0600
Nausea	-0.0480
Source: CS Table 64 and the economic model	

Table 22 AE QALY decrement

Treatment arm	AE disutility
Linzagolix (200 mg, base case)	-0.002
BSC	-0.001
Source: CS Table 65 and the economic model	

4.2.6.4 Age-adjusted utilities

To account for an age-related decrease in quality of life, the company applied age-related utility decrements using a widely used algorithm published by Ara and Brazier.[37] The utility multiplier was applied in each model cycle throughout the time horizon.

EAG conclusions on HRQoL

The company adopted a similar approach as in TA832 to estimate utilities for the controlled and uncontrolled health states. An unpublished algorithm for mapping UFS-QoL to EQ-5D-3L was used, based on the previous appraisal TA832. The company did not report the mapping algorithm used for EQ-5D-5L to EQ-5D-3L. The coefficients used in the mapping algorithm were provided in the economic model; therefore, we are unable to verify if the estimates in the algorithms were implemented appropriately. Using the estimates obtained from mapping EQ-5D-5L to EQ-5D-3L has a significant impact on the ICER: the base case ICER doubles, to £30,803. This is driven by a higher increment in the total QALY gain for the BSC arm than that for linzagolix, thereby decreasing the incremental QALY gain.

We acknowledge the company's mapping in the base case is consistent with TA832. We have also noted their rationale for using a disease-specific measure in the base case (in CS Section B.3.4.5.3). Advice from our clinical expert concurs with the company's rationale for preferring disease-specific measures over generic measures. However, we note NICE's preference for trial-based EQ-5D data when it is available,[1] and question whether the TA832 committee's concerns about the availability of EQ-5D-5L data from the clinical trials apply in the current appraisal.

The company addressed one of the key uncertainties raised in TA832 associated with the choice of regression model for the utility function, by using a linear mixed model to account for repeated measures. However, they did not provide any information on the utility function used in their LMM within the CS. The intercept and the coefficient of the only covariate (RMBL) for the two LMMs (one for the base case and the other for scenario analysis) are hard coded within the model. No clarity or rationale was provided for their choice of the covariate included therefore we are unclear as to why other covariates (such as age) were excluded from the utility function. Furthermore, no sensitivity or scenario analyses were conducted to explore the impact of varying assumptions around the utility function (e.g., using alternative utility functions exploring non-linear impacts of RMBL on utility). Due to this lack of information, the EAG are uncertain about the robustness of the utility function for the LMM that was used to estimate utilities for the 'controlled' and 'uncontrolled' health states. We conducted a range of exploratory scenarios changing the coefficients by +/-10% in EAG analyses, see Section 6.

In conclusion, we view that there are uncertainties with respect to the company's approach for estimating utilities for the 'controlled' and 'uncontrolled' health states that warrant further investigation.

4.2.7 Resources and costs

We outline below the costs included in the cost-comparison and the cost-effectiveness analyses.

- Drug acquisition costs
- Administration costs
- Healthcare resource use costs
- Costs associated with surgery.

4.2.7.1 Drug acquisition and administration

CS Tables 68 and 69 report the cost of drugs used in the model. The cost reported in the model for the two doses of linzagolix (100mg and 200mg) is [redacted] using the Patient Access Scheme (PAS) discount. Within the cost-comparison model, no additional costs of ABT were applied in the comparator arms for relugolix CT and the GnRH agonists. All patients were assumed to receive concomitant medications (including ibuprofen and iron supplements) in the base case. Linzagolix is administered orally; there are no associated administration costs. However, as the GnRH agonists are administered via subcutaneous injection, the

company model assumes that a patient requires 10 minutes of a nurse's time in GP clinic, at a cost of £7.67.

4.2.7.2 Resource use

4.2.7.2.1 Health care

The health care resource use and costs used in the cost comparison and cost-effectiveness analyses are reported in CS Table 74 and CS Table 75 and summarised in Table 23 below. The costs are obtained from PSSRU 2022, and the NHS reference costs 2021/22.

Table 23 Health care resource use and costs

Resource	GnRH analogues	BSC	Cost
Gynaecologist consultation	Once only	Once only	£185.51
GP visits	None	None	£42.00
DEXA scans	One after 1 year [a]	None	£95.45
Ultrasound	Once (67% of patients)	Once (67% of patients)	£235.60
Full blood count	Once	Once	£2.96
Hysteroscopy	Once (17% of patients)	Once (17% of patients)	£286.41
MRI	Once (17% of patients)	Once (17% of patients)	£197.34

Source: Reproduced from CS Table 74 and CS Table 75
[a] Applied to 100% of patients in the first model cycle as a conservative assumption

The healthcare resource use costs for Population #3 are assumed to be equivalent to those used in TA832. The company aggregated the costs, which are applied as a one-off in the first cycle of the economic model.

4.2.7.2.2 Adverse events

Four adverse events are captured in the company model for Population #3, the costs of which are reported in CS Table 76. Headache and hot flush are assumed to incur no cost, while treatment for anaemia is assumed to be the cost of a GP surgery consultation, £42.00, as in TA832. For nausea, a cost of £0.96 is applied to cover treatment with metclopramadine, as used in TA832. CS Table 77 reports the adverse event costs which are applied in the first model cycle, obtained by combining the unit costs with the adverse event probabilities in Section 4.2.6.3 above. In the base case, a cost of £1.25 is applied for adverse events related to linzagolix 200mg treatment in the company model. The corresponding cost for BSC is £2.82. The company also implemented a scenario using linzagolix 100mg, which incurs an adverse event related cost of £4.24.

4.2.7.2.3 Surgery

For Populations #1 and #2, total costs associated with surgery are aggregated based on the proportion of patients undergoing surgery (see CS Tables 79 and 80).

For Population #3, the proportion of patients moving to the surgery state in each cycle is used to estimate the cost of surgery. The total cost of surgery is based upon a weighted average of different types of surgery. These surgery types can vary by treatment. CS Table 78 presents the costs of each type of surgery, obtained from the NHS schedule of NHS costs 2021/22, and reproduced in Table 24 below. The company assume that the distributions of surgery type are equivalent across both treatment arms in the base case, with a surgery cost of £5,278. CS Table 81 reports the surgery costs for Population #3, including the base case costs for linzagolix and BSC, and costs used in two scenario analyses with treatment-independent surgery distributions used in the TA832 Evidence Review Group report, and treatment-specific surgery distributions.

Table 24 Costs by surgery type

Surgery type	Cost
UAE	£2,786
MRgFUS	£1,131
Open/abdominal myomectomy	£4,670
Laparoscopic myomectomy	£3,496
Open/abdominal hysterectomy	£6,336
Laparoscopic hysterectomy	£5,273
Source: Reproduced from CS Table 78	

EAG conclusions on health care resource use and unit costs

The EAG have some concerns over the company's assumptions for healthcare resource use and associated unit costs applied to the economic model. Consultation with our clinical expert suggests that: i) patients receive two GP visits on average, unlike the company's assumption of no GP visits; ii) patients are likely to have two full blood count tests; and iii) they are unlikely to undergo DEXA scans or an MRI scan. With respect to unit costs for the resource use, we noted an inconsistency in the cost of gynaecologist consultation. Lastly, we view that the company may have underestimated the costs associated with MRI. We outline the EAG's assumptions (based on our expert advice) for these parameters in Table 25 and Table 26 and we conducted scenarios in the cost-effectiveness model for Population #3 (see Section 6).

Table 25 Health care resource usage

Resource	Company's assumption		EAG assumptions	
	GnRH antagonists	BSC	GnRH antagonists	BSC
GP visits	None	None	Twice	Twice
DEXA scans	One after 1 year	None	None	Same as company
Full blood count	Once	Once	Twice	Twice
MRI	Once (17% of patients)	Once (17% of patients)	None	None
Source: Company assumptions are obtained from CS Table 74				

Table 26 Unit costs for healthcare resource use

Resource	Company	EAG (source)
Gynaecologist consultation	£185.51	£181.26 (NHS Reference costs 2021/22, Gynaecologist consultation non-admitted face-to-face attendance, WF01A)
MRI	£197.34	£255.41 (NHS Reference costs 2021/22, Magnetic Resonance Imaging Scan of One Area, with Post-Contrast Only, 19 years and over, RD02A)
Source: Company assumptions are obtained from CS Table 75		

5 COST EFFECTIVENESS RESULTS

5.1 Company's cost effectiveness results

The company report the deterministic base case cost-comparison results for Population #1 and Population #2 in CS Tables 86 and 87 respectively, and the cost-effectiveness results for Population #3 in CS Table 88. These results are summarised in Table 27 and Table 28 below. Note that all results use the PAS price for linzagolix.

Table 27 Base case cost-comparison results for Populations #1 and #2 using the PAS discount

Treatment	Population #1		Population #2	
	Total costs	Incremental costs	Total costs	Incremental costs
Linzagolix	[redacted]	-	[redacted]	-
Relugolix CT	£3,411	[redacted]	£4,752	[redacted]
Leuprorelin	£3,441	[redacted]	-	-
Goserelin	£3,407	[redacted]	-	-
Triptorelin	£3,482	[redacted]	-	-

Source: Reproduced from CS Tables 86 and 87
PAS: patient access scheme

Table 28 Base case cost-effectiveness results for Population #3 using the PAS discount.

Treatment	Total			Incremental			ICER (£/QALY)
	Costs	LYG	QALYs	Costs	LYG	QALYs	
BSC	£5,107	[redacted]	[redacted]				
Linzagolix	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Source: Reproduced from CS Table 88.
BSC, best supportive care; LYG, life-years gained; QALY, quality-adjusted life year; ICER, incremental cost-effectiveness ratio; PAS: patient access scheme

5.2 Company sensitivity analyses

5.2.1 Deterministic sensitivity analyses

For the Population #3 cost-effectiveness analysis the company report deterministic sensitivity analysis results for the ten most influential parameters in CS Table 90 and CS Figure 28. The ranges of variation for the input parameters were based on 95% confidence

intervals where available, or an assumption that the standard error was 10% of the mean. The company's results indicate that the assumptions regarding the BSC response rates are the main drivers of the model results, increasing the ICER to £19,035 per QALY. The proportion of patients receiving surgery and the recurrence rates for patients on BSC also have a high impact on the ICER. All ICERs remained below £20,000. No deterministic sensitivity analyses were performed in the cost-comparison analysis for Population #1 and Population #2.

5.2.2 Scenario analyses

The company report the results of 19 scenarios for Population #3 in CS Table 93. The scenarios explored included: time horizons, discount rates, transition probabilities, utility values, and dosing regimens for linzagolix. Changing the source of utility values from UFS-QoL to EQ-5D-5L mapped to 3L utility values had the largest effect on the results, increasing the ICER by £15,411 to £30,803 per QALY. Implementing the modified trial percentages (including adverse events as the reason for discontinuation) for treatment withdrawal rates produced the next-highest ICER of £25,838 per QALY, an increase of £10,436 from the base case. The greatest reduction in the ICER was obtained by using utility values from Hux et al.[35] as opposed to those from the PRIMROSE trials (£10,098 per QALY). Modifying the time horizons, source of surgery distribution to values reported in the Evidence Review Group report from TA832, and post-surgery utilities had no effect on the ICER. The results for scenario analyses performed for Population #1 and Population #2 are in CS Table 91 and CS Table 92 respectively.

5.2.3 Probabilistic sensitivity analysis

The company conducted a probabilistic sensitivity analysis (PSA) with input parameter distributions as presented in CS Appendix N. The PSA was run for 1,000 iterations, and mean results reported in CS Table 89. The cost-effectiveness plane and cost-effectiveness acceptability curve are presented in CS Figure 26 and CS Figure 27 respectively. The probabilistic results were in line with the deterministic results when run by the EAG. No probabilistic sensitivity analyses were performed in cost-comparison analyses for Population #1 or Population #2.

EAG conclusion on the company's sensitivity and scenario analyses

The EAG did not find any errors in any of the company's analyses. The company included all necessary parameters in the PSA, with appropriate corresponding distributions. The EAG note a minor inconsistency in CS Table 93 reporting scenario analyses for Population #3: for concomitant medicine distribution, the company appear to have transposed the base case and scenario columns. The base case should be to

assume 100% distribution, whilst the scenario should be to assume treatment specific distributions. The results in the table are correct for the scenario.

5.3 Model validation and face validity checks

We conducted a range of checks on both the company's models (i.e., the cost-comparison model and the cost-effectiveness model) using an EAG checklist:

- Input checks: comparison of all parameter values in the model against the values stated in the company submission and cited sources.
- Output checks: replication of results reported in the company submission using the company model.
- 'White box' checks: manual checking of formulae working from the cohort-level Markov model, which includes reviewing the calculations across each cycle and working backwards to trace links to input parameters and forwards to the results.
- 'Black box' checks: working through a list of tests to assess whether changes to key model inputs or assumptions have the expected effects on the model results.
- The model is well-implemented, and no coding errors were identified.

We noted a few minor inconsistencies, such as incorrect NHS reference cost codes for laparoscopic myomectomy (the model stated MA08A and MA08B, but the correct codes are MA09A and MA09B; the corresponding unit costs applied in the model are correct) which did not impact the model results.

5.3.1 EAG corrections to the company model

Except one minor correction, the EAG did not identify any that needed to be made to either of the company models. As stated earlier in Section 4.2.7, we noted an inconsistency in the unit cost of a gynaecologist consultation. The company used an estimate of £185.51; we view the correct estimate is £181.26. This change in unit cost does not impact the overall cost effectiveness result. We incorporate this correction the EAG scenarios as well as in our preferred base case (see Section 6) within the cost-effectiveness model for Population #3. We also re-ran the cost-comparison analyses for Populations #1 and #2 with the correct unit cost (see Section 6).

5.3.2 EAG summary of key issues and additional analyses

We present a summary of the issues identified by the EAG and our additional analyses for the cost-effectiveness analysis for Population #3 in Table 29.

Table 29 Additional EAG scenarios conducted in the CEA for Population #3

Parameter	Company base case	EAG scenarios	EAG preferred	Reason for analysis
Baseline patient characteristics: mean age and age of menopause	Mean age 42.25 years, average age of menopause 51 years	+/- 10%	-	To explore the variation around the mean age.
Model time horizon	10 years	20 years Lifetime	-	Based on NICE Reference case
Medications used for BSC	NSAIDs and iron supplements	Addition of vitamin D and calcium supplements. Vit D: 10mg/day Calcium: 1500mg/day	Same as EAG scenario	Based on EAG expert opinion.
Recurrence rate	Data collected from survey completed by gynaecologists	Assume same recurrence rate of 23.8% for linzagolix and BSc Assuming 10% and 25% recurrence rate for both treatment arms	-	Exploratory analyses to illustrate the effect of using similar recurrence rates for linzagolix and BSC
Treatment discontinuation rate	Trial-based withdrawal rates: Linzagolix 100mg: [redacted] Linzagolix 200mg: [redacted] BSc: [redacted]	Assume same discontinuation rates for linzagolix 100mg, linzagolix 200mg and BSc at: [redacted]	-	Exploratory analysis

Parameter	Company base case	EAG scenarios	EAG preferred	Reason for analysis
		Assume same discontinuation rates for linzagolix 100mg, linzagolix 200mg and BSc at: [redacted]		
Surgery: procedural death	Mortality rates of UAE: 0.02%	Mortality rates: of UAE: 0.0002%	-	Based on clinical opinion
Surgery: Distribution of surgery types	UAE: 4.8% Endometrial ablation: 0.0% MRgFUS: 3.0% Abdominal myomectomy: 25.7% Laparoscopic myomectomy:8.2% Abdominal hysterectomy: 51.8% Laparoscopic hysterectomy:6.4%	UAE: 20% Endometrial ablation: 0.0% MRgFUS: 0% Abdominal myomectomy: 0% Laparoscopic myomectomy:0% Hysteroscopic myomectomy: 20% Abdominal hysterectomy: 6% Laparoscopic hysterectomy:54%	Same as EAG scenario	Based on clinical opinion
Healthcare resource use	GP visits: GnRH antagonists: none BSC: None DEXA scans GnRH antagonists: one after 1 year	GP visits: GnRH antagonists: Twice BSC: Twice DEXA scans GnRH antagonists: None BSC: None	Same as EAG scenario	Based on EAG expert opinion

Parameter	Company base case	EAG scenarios	EAG preferred	Reason for analysis
	BSC: None Full Blood Count GnRH antagonists: Once BSC: Once MRI GnRH antagonists: Once BSC: Once	 Full Blood Count GnRH antagonists: Twice BSC: Twice MRI GnRH antagonists: None BSC: None		
Unit costs	Gynaecologist: £185.51 MRI: £197.34	Gynaecologist: £181.26 MRI: £255.41	Same as EAG scenario	
Utilities	Estimates used for the EQ-5D utility function for UFS-QoL Intercept: [redacted] RMBL: [redacted]	Exploratory scenario reducing the estimates used for the EQ-5D utility function by 10%. Intercept: [redacted] RMBL: [redacted]	-	Exploratory analysis
	Estimates used for the EQ-5D utility function for UFS-QoL Intercept: [redacted] RMBL: [redacted]	Increasing the estimates used for the EQ-5D utility function for EQ-5D-5L by 10%. Intercept: [redacted]		

Parameter	Company base case	EAG scenarios	EAG preferred	Reason for analysis
		RMBL: [redacted]		
	Estimates used for the EQ-5D utility function for EQ-5D-5L Intercept: [redacted] RMBL: [redacted]	Exploratory scenario reducing the estimates used for the EQ-5D utility function by 10%. Intercept: [redacted] RMBL: [redacted]		
	Estimates used for the EQ-5D utility function for EQ-5D-5L Intercept: [redacted] RMBL: [redacted]	Increasing the estimates used for the EQ-5D utility function by 10%. Intercept: [redacted] RMBL: [redacted]		

6 EAG'S ADDITIONAL ANALYSES

6.1 Exploratory and sensitivity analyses undertaken by the EAG.

6.1.1 Cost-comparison analysis for Population #1 and Population #2

For the cost-comparison analyses, we corrected the unit cost of a gynaecologist consultation as reported in Section 5.3.1, and also applied our preferred cost of an MRI at £255.41 (see Table 29) for Population #1 and Population #2. Results are reported below in Table 30 and Table 31. Changing the unit costs do not impact the incremental costs in Population #1 and Population #2 because of the proportional change in the total costs across the treatment arms. No other scenario analyses were conducted for the cost-comparison analyses on Population #1 and Population #2, due to the uncertainties surrounding the assumption of similar clinical efficacy between linzagolix and relugolix-CT and secondly, whether relugolix-CT has an adequate market share to qualify as the selected comparator for the cost-comparison analysis (discussed earlier in Section 3 and Section 4.2.5.1).

Table 30 Cost-comparison results using EAG unit cost for a gynaecologist consultation and MRI for Population #1

Treatment	Total costs	Incremental costs, linzagolix versus
Linzagolix	[redacted]	
Relugolix CT	£3,417	[redacted]
Leuprorelin	£3,446	[redacted]
Goserelin	£3,413	[redacted]
Triptorelin	£3,488	[redacted]

Table 31 Cost-comparison results using EAG unit cost for a gynaecologist consultation and MRI for Population #2

Treatment	Total costs	Incremental costs, linzagolix versus
Linzagolix	[redacted]	
Relugolix CT	£4,757	[redacted]

6.1.2 Cost-effectiveness analysis for Population #3

Results from the EAG scenario analyses (outlined in Table 29) conducted on the company's base case analysis for Population #3 are shown in Table 32 below. The ICERs vary between the range of £13,968 per QALY (Scenario: increasing the coefficients used for the EQ-5D utility function for UFS-QoL by 10%) and £34,376 per QALY (Scenario: reducing the coefficients used for the EQ-5D utility function for EQ-5D-5L by 10%).

Table 32 EAG additional scenarios applied to the company's base case for the Population #3 cost-effectiveness analysis.

Scenario	Treatment	Total Costs	Total QALYs	ICER (£/QALY)
Company base case	BSC	[redacted]	[redacted]	£15,392
	Linzagolix	[redacted]	[redacted]	
Baseline characteristics				
Patient mean age + 10%	BSC	[redacted]	[redacted]	£17,017
	Linzagolix	[redacted]	[redacted]	
Patient mean age -10%	BSC	[redacted]	[redacted]	£15,252
	Linzagolix	[redacted]	[redacted]	
Average menopause age +10%	BSC	[redacted]	[redacted]	£15,271
	Linzagolix	[redacted]	[redacted]	
Average menopause age – 10%	BSC	[redacted]	[redacted]	£17,999
	Linzagolix	[redacted]	[redacted]	
Concomitant medication				
Addition of vitamin D and calcium	BSC	[redacted]	[redacted]	£15,705
	Linzagolix	[redacted]	[redacted]	
Recurrence rate				
10% for both treatment arms	BSC	[redacted]	[redacted]	£22,137
	Linzagolix	[redacted]	[redacted]	
25% for both treatment arms	BSC	[redacted]	[redacted]	£21,108
	Linzagolix	[redacted]	[redacted]	
Treatment discontinuation				

Scenario	Treatment	Total Costs	Total QALYs	ICER (£/QALY)
[redacted] for linzagolix 100mg, linzagolix 200mg, and BSC	BSC	[redacted]	[redacted]	£14,864
	Linzagolix	[redacted]	[redacted]	
[redacted] for linzagolix 100mg, linzagolix 200mg, and BSC	BSC	[redacted]	[redacted]	£16,384
	Linzagolix	[redacted]	[redacted]	
[redacted] for linzagolix 100mg, linzagolix 200mg, and BSC	BSC	[redacted]	[redacted]	£14,930
	Linzagolix	[redacted]	[redacted]	
[redacted] for linzagolix 100mg, linzagolix 200mg, and BSC	BSC	[redacted]	[redacted]	£26,509
	Linzagolix	[redacted]	[redacted]	
[redacted] for linzagolix 100mg, linzagolix 200mg, and BSC	BSC	[redacted]	[redacted]	£23,806
	Linzagolix	[redacted]	[redacted]	
[redacted] for linzagolix 100mg, linzagolix 200mg, and BSC	BSC	[redacted]	[redacted]	£26,345
	Linzagolix	[redacted]	[redacted]	
Procedural death				
0.0002% mortality rate for UAE	BSC	[redacted]	[redacted]	£15,392
	Linzagolix	[redacted]	[redacted]	
Distribution of surgery types				
UAE: 20% Hysteroscopic myomectomy: 20% Abdominal hysterectomy: 6% Laparoscopic hysterectomy: 54%	BSC	[redacted]	[redacted]	£15,392
	Linzagolix	[redacted]	[redacted]	
Health care resource use				

Scenario	Treatment	Total Costs	Total QALYs	ICER (£/QALY)
GP visits: twice for both treatment arms DEXA scans: none for both treatment arms Full blood count: twice for both treatment arms MRI: none for both treatment arms	BSC	[redacted]	[redacted]	£14,165
	Linzagolix	[redacted]	[redacted]	
Unit costs				
Gynaecologist: £181.26 MRI: £255.41	BSC	[redacted]	[redacted]	£15,392
	Linzagolix	[redacted]	[redacted]	
Utilities				
Reducing the coefficients used for the EQ-5D utility function for EQ-5D-5L by 10%. Intercept: [redacted]; RMBL: [redacted]	BSC	[redacted]	[redacted]	£34,376
	Linzagolix	[redacted]	[redacted]	
Increasing the coefficients used for the EQ-5D utility function for EQ-5D-5L by 10%. Intercept: [redacted]; RMBL: [redacted]	BSC	[redacted]	[redacted]	£27,903
	Linzagolix	[redacted]	[redacted]	
Reducing the coefficients used for the EQ-5D utility function for UFS-QoL by 10%. Intercept: [redacted]; RMBL: [redacted]	BSC	[redacted]	[redacted]	£17,140
	Linzagolix	[redacted]	[redacted]	
Increasing the coefficients for the EQ-5D utility function for UFS-QoL by 10%. Intercept: [redacted]; RMBL: [redacted]	BSC	[redacted]	[redacted]	£13,968
	Linzagolix	[redacted]	[redacted]	

BSC, best supportive care; QALY, quality-adjusted life year; ICER, incremental cost-effectiveness ratio; UAE, uterine artery embolization; RMBL, reduced menstrual blood loss.

6.2 EAG's preferred assumptions for Population #3

The EAG's preferred base case assumptions for Population #3 are:

- Inclusion of prophylactic regimens of calcium and vitamin D in the BSC arm.
- Distribution of surgery types based on advice from our clinical expert.
- Use of healthcare resources based on advice from our clinical expert.
- Including the unit costs of gynaecological consultation and MRI as identified by the EAG.
- Using EQ-5D-5L data from the PRIMROSE trials to estimate the health state utilities.

The results are presented in Table 33 in a cumulative manner. The ICER for the EAG base case is £28,973 [redacted], an [redacted] £13,581 [redacted] from the company's base case. The ICER remains below the £30,000 [redacted] threshold.

Table 33 Cumulative EAG preferred assumptions for the Population #3 cost-effectiveness analysis

Assumption	Treatment	Total Costs	Total QALYs	ICER (£/QALY)
Company base case	BSC	[redacted]	[redacted]	£15,392
	Linzagolix	[redacted]	[redacted]	
+ Include vitamin D and calcium in BSC	BSC	[redacted]	[redacted]	£15,705
	Linzagolix	[redacted]	[redacted]	
+ EAG preferred surgery type distribution	BSC	[redacted]	[redacted]	£15,705
	Linzagolix	[redacted]	[redacted]	
+ EAG preferred health care resource use	BSC	[redacted]	[redacted]	£14,478
	Linzagolix	[redacted]	[redacted]	
+ EAG preferred unit costs (EAG preferred base case)	BSC	[redacted]	[redacted]	£14,478
	Linzagolix	[redacted]	[redacted]	
+ Utilities obtained from mapping EQ-5D-5L to EQ-5D-3L	BSC	[redacted]	[redacted]	£28,973
	Linzagolix	[redacted]	[redacted]	
EAG base case	BSC	[redacted]	[redacted]	£28,973
	Linzagolix	[redacted]	[redacted]	

BSC, best supportive care; QALY, quality-adjusted life year; ICER, incremental cost-effectiveness ratio

6.3 Scenarios conducted on the EAG base case for Population #3 cost-effectiveness analysis.

The EAG conducted further scenarios on the EAG base case economic model. Results from these scenarios are reported in Table 34 below. The ICERs ranges between £9,498 per QALY (scenario: using utility values for the 'controlled' and 'uncontrolled' health states from Hux et al.) and £49,857 per QALY (scenario: treatment discontinuation due to AEs only).

Table 34 EAG scenarios on the EAG base case model for the Population #3 cost-effectiveness analysis

Scenario	Treatment	Total Costs	Total QALYs	ICER (£/QALY)
EAG base case	BSC	[redacted]	[redacted]	£28,973
	Linzagolix	[redacted]	[redacted]	
Linzagolix dosing				
200 mg for 6 months followed by BSC	BSC	[redacted]	[redacted]	£29,325
	Linzagolix	[redacted]	[redacted]	
100 mg	BSC	[redacted]	[redacted]	£32,023
	Linzagolix	[redacted]	[redacted]	
Surgery probability from the 'controlled' health state				
1%	BSC	[redacted]	[redacted]	£17,102
	Linzagolix	[redacted]	[redacted]	
2%	BSC	[redacted]	[redacted]	£24,907
	Linzagolix	[redacted]	[redacted]	
Concomitant medication distribution				
Treatment specific	BSC	[redacted]	[redacted]	£28,711
	Linzagolix	[redacted]	[redacted]	
Treatment withdrawal rates				
Modified trial % (withdrawal due to AEs)	BSC	[redacted]	[redacted]	£49,857
	Linzagolix	[redacted]	[redacted]	
Utility for 'controlled' and 'uncontrolled' health states				
Utility mapping from UFS-QoL to EQ-5D-3L	BSC	[redacted]	[redacted]	£14,478
	Linzagolix	[redacted]	[redacted]	
Utility source 'controlled' and 'uncontrolled' health states				
Hux et al.	BSC	[redacted]	[redacted]	£9,498
	Linzagolix	[redacted]	[redacted]	
Post-surgery utility				

Scenario	Treatment	Total Costs	Total QALYs	ICER (£/QALY)
General population	BSC	[redacted]	[redacted]	£28,973
	Linzagolix	[redacted]	[redacted]	
Equal to controlled	BSC	[redacted]	[redacted]	£28,973
	Linzagolix	[redacted]	[redacted]	
Adverse event disutility				
Exclude	BSC	[redacted]	[redacted]	£27,877
	Linzagolix	[redacted]	[redacted]	
Utility age adjustment				
Exclude	BSC	[redacted]	[redacted]	£28,763
	Linzagolix	[redacted]	[redacted]	

BSC, best supportive care; QALY, quality-adjusted life year; ICER, incremental cost-effectiveness ratio.

6.4 Conclusions on the cost-comparison evidence for Population #1 and Population #2

The company developed a cost-comparison model for linzagolix compared to relugolix CT and GnRH agonists for Population #1 (people having short-term treatment of 6 months of less) and for linzagolix compared to relugolix CT for Population #2 (people having longer-term treatment with hormone-base therapy). The EAG performed validation checks on the cost-comparison model as discussed in section 5.3. No errors or inconsistencies were found, except the one discussed in section 5.3.1. We corrected this error as well as updating the company's model with the unit cost for MRI as identified by the EAG (see Table 29 Additional EAG scenarios conducted in the CEA for Population #3Table 29). These did not change the overall results (incremental costs) as shown in Table 30 and Table 31 due to the proportional change in the total costs across the treatment arms. Overall, we view the company's simple modelling approach for the cost-comparison analysis is reasonable.

The EAG did not conduct any scenario analyses for Population #1 or Population #2, as we are uncertain whether:

- linzagolix has similar clinical efficacy as relugolix CT and other GnRH analogues (see Key Issue 1) and
- relugolix-CT has an adequate market share to qualify as the selected comparator for the cost-comparison analysis (see Key Issue 2).

6.5 Conclusions on the cost-effectiveness evidence for Population #3

The company developed a model to estimate the cost-effectiveness of linzagolix compared to BSC for Population #3 (people with longer-term treatment, without hormone-based therapy). The EAG consider the overall model structure to be appropriate. The model uses clinical efficacy data from the PRIMROSE trials. The company base case produced an ICER of £15,392 per QALY gained for linzagolix compared to BSC. This ICER was obtained by applying a confidential PAS discount for linzagolix. The EAG did not identify any technical errors on checking the economic model, except a few minor inconsistencies in reporting which did not have any impact on the overall results.

The EAG disagree with some of the assumptions in the company's model. Our preferred assumptions include:

- Inclusion of prophylactic regimens of calcium and vitamin D in the BSC arm.
- Distribution of different surgery types based on clinical expert opinion.
- Inclusion of healthcare resource use based on clinical expert opinion.
- Using a unit cost of £181.26 for gynaecologist consultation (WF01A) and £255.41 for an MRI scan (RD02A), respectively, obtained from NHS Reference Costs 2021/22.
- Using the utility estimates derived from EQ-5D-5L data collected in the PRIMROSE trials.

The EAG preferred assumptions increase the ICER to £28,973 per QALY gained for linzagolix compared to BSC. In addition to the above issues addressed by the EAG, there are other key uncertainties in the company's assumptions. These include:

- Population #3, which includes patients unable to receive hormone-based therapy, is inconsistent with the population in the PRIMROSE trials where people who could receive hormone-based therapy were randomised. Our expert advice received was that very few patients would be unfit/prefer not to receive hormone-based therapy, so the unique population that linzagolix without hormone-based therapy could benefit in clinical practice could be very small. However, as we noted in section 2.2.4 above, the size of this sub-population is uncertain, and we have suggested further clarification would be helpful (see Key Issue 3).

- There is uncertainty regarding the company's assumption that treatment efficacy (response rate) of linzagolix will be maintained in the long run, over 52 weeks. This assumption is based on the evidence from 2-year data from the LIBERTY randomised withdrawal study for relugolix CT. While maintaining the response rate may be biologically plausible, there are no data to confirm this.

- We have concerns whether patients staying in the 'post-surgery' state until they experience menopause is reflective of the disease prognosis. While some patients may be cured, others may experience a recurrence of the symptoms post-surgery. This would benefit from further clarification (see Key Issue 4).

- There are also uncertainties about the reporting of the utility function. The company did not define or justify the specification for the linear mixed model regression of utility data. It is not clear why they chose to include a single independent variable (RMBL), or whether additional co-variates would have improved the model fit. Furthermore, no sensitivity or scenario analyses were reported for alternative specifications of the utility function (see Key Issue 5).

- There are uncertainties in the company's assumptions regarding healthcare resource. Their assumptions that patients would not have any GP visits, have full blood count and MRI scan once each, and people in the linzagolix arm receive one DEXA scan after 1 year may not reflect clinical practice. Based on the advice of our clinical expert, we conducted scenario analysis assuming that on average, patients visit GPs and have the full blood count, twice each. They are unlikely to undergo DEXA scans or MRI scans. We note that wider consultation with further clinical experts might help to better understand the heterogeneity in resource use across hospitals and resolve the uncertainty.

7 REFERENCES

1. National Institute for Health and Care Excellence. NICE health technology evaluations: the manual.
2. Theramex. Data on File: Uterine Fibroids Market Research Survey for Yselty (Linzagolix) Full Report (Confidential). 2022 2022.
3. Marsh EE, Al-Hendy A, Kappus D, Galitsky A, Stewart EA, Kerolous M. Burden, Prevalence, and Treatment of Uterine Fibroids: A Survey of U.S. Women. J Womens Health. 2018;27(11):1359-67.
4. Giuliani E, As-Sanie S, Marsh EE. Epidemiology and management of uterine fibroids. International Journal of Gynecology & Obstetrics. 2020;149(1):3-9.
5. Hunsche E, Rakov V, Scippa K, Witherspoon B, McKain L. The Burden of Uterine Fibroids from the Perspective of US Women Participating in Open-Ended Interviews. Women's Health Reports. 2022;3(1):286-96.
6. Soliman AM, Margolis MK, Castelli-Haley J, Fuldeore MJ, Owens CD, Coyne KS. Impact of uterine fibroid symptoms on health-related quality of life of US women: evidence from a cross-sectional survey. Current Medical Research and Opinion. 2017;33(11):1971-8.
7. European Medicines Agency (EMA). Yselty (linzagolix) Summary of Product Characteristics. EMA; 2023.
8. National Institute for Health and Care Excellence. NG88. Heavy menstrual bleeding: assessment and management. 2018.
9. Theramex. Data on File: Linzagolix for treating moderate to severe symptoms of uterine fibroids: an indirect treatment comparison. 2023 2023/06/09/.
10. National Institute for Health and Care Excellence. Fast track appraisals - guiding notes for ERGs. NICE; 2017.
11. Donnez J, Taylor HS, Stewart EA, Bradley L, Marsh E, Archer D, et al. Linzagolix with and without hormonal add-back therapy for the treatment of symptomatic uterine fibroids: two randomised, placebo-controlled, phase 3 trials. The Lancet. 2022;400(10356):896-907.
12. Kissei Pharmaceutical Co. Ltd. Data on File: PRIMROSE 3 CSR. 2023 2023/03/09/.
13. Al-Hendy A, Venturella R, Arjona Ferreira JC, Li Y, Soulban G, Wagman RB, et al. LIBERTY randomized withdrawal study: relugolix combination therapy for heavy menstrual bleeding associated with uterine fibroids. American Journal of Obstetrics and Gynecology. 2023.
14. de Lange ME, Semmler A, Clark TJ, Mol BWJ, Bet PM, Huirne JAF, et al. Considerations on implementation of the newest treatment for symptomatic uterine

fibroids: Oral GnRH antagonists. British Journal of Clinical Pharmacology. 2023;n/a(n/a).

15. Al-Hendy A, Lukes AS, Poindexter AN, Venturella R, Villarroel C, Critchley HOD, et al. Treatment of Uterine Fibroid Symptoms with Relugolix Combination Therapy. The New England Journal of Medicine. 2021;384(7):630-42.
16. Stewart EA, Lukes AS, Venturella R, Arjona Ferreira J-C, Li Y, Hunsche E, et al. Relugolix Combination Therapy for Uterine Leiomyoma–Associated Pain in the LIBERTY Randomized Trials. Obstetrics & Gynecology. 2022;139(6):1070-81.
17. Donnez J, Tatarchuk TF, Bouchard P, Puscasiu L, Zakharenko NF, Ivanova T, et al. Ulipristal acetate versus placebo for fibroid treatment before surgery. The New England Journal of Medicine. 2012;366(5):409-20.
18. Donnez J, Tomaszewski J, Vázquez F, Bouchard P, Lemieszczuk B, Baró F, et al. Ulipristal acetate versus leuprolide acetate for uterine fibroids. The New England Journal of Medicine. 2012;366(5):421-32.
19. Magnay JL OBS, Gerlinger C, Seitz C. A systematic review of methods to measure menstrual blood loss. BMC Womens Health. 2018;18(1):1-13.
20. Osuga Y KE, Kudou K, Tanimoto M, Hoshiai H. Oral gonadotropin-releasing hormone antagonist relugolix compared with leuprorelin injections for uterine leiomyomas: a randomized controlled trial. Obstetrics and Gynecology. 2019;133(3):423-33.
21. Osuga Y KE, Kudou K, Hoshiai H. Relugolix, a novel oral gonadotropin-releasing hormone antagonist, in the treatment of pain symptoms associated with uterine fibroids: a randomized, placebo-controlled, phase 3 study in Japanese women. Fertility and Sterility 2019;112(5):922-9.e2.
22. Hoshhiai H, Seki Y, Kusumoto T, Kudou K, Tanimoto M. Relugolix for oral treatment of uterine leiomyomas: a dose-finding, randomized, controlled trial. BMC Women's Health. 2021;21:375 (11pp).
23. Anchan RM SJ, Zhang S, Wojdyla D, Bortoletto P, Terry K, et al. Long-term health-related quality of life and symptom severity following hysterectomy, myomectomy, or uterine artery embolization for the treatment of symptomatic uterine fibroids. American Journal of Obstetrics and Gynecology. 2023;275:275.e1-.e17.
24. Harding G, Coyne KS, Thompson CL, Spies JB. The responsiveness of the uterine fibroid symptom and health-related quality of life questionnaire (UFS-QOL). Health and Quality of Life Outcomes. 2008;6(1):99.
25. European Medicines Agency (EMA). Yselty (linzagolix) European Public Assessment Report (EPAR). 2022 2022/07/01/.
26. ObsEva S.A. Data on File: PRIMROSE 1 CSR Week 24, v2.0. 2021 2021/06/22/.

27. ObsEva S.A. Data on File: PRIMROSE 2 CSR Week 52, v2.0. 2021 2021/05/26/.
28. ObsEva S.A. Data on file: PRIMROSE 1 and 2 Pooled Analysis, 16-OBE2109-008_and_16-OBE2109-009_Combined. 21-09-08. 2021 2021.
29. National Institute for Health and Care Excellence. Relugolix-estradiol-norethisterone acetate for treating moderate to severe symptoms of uterine fibroids (TA832). Available at: https://www.nice.org.uk/guidance/ta832 (Last accessed: 8 Nov 2023). 2022.
30. Al-Hendy A, Venturella R, Ferreira JCA, Li Y, Wagman R, Lukes AS. LIBERTY RANDOMIZED WITHDRAWAL STUDY: 2-YEAR EFFICACY AND SAFETY OF RELUGOLIX COMBINATION THERAPY IN WOMEN WITH HEAVY MENSTRUAL BLEEDING ASSOCIATED WITH UTERINE FIBROIDS. Fertility and Sterility. 2021;116(3):e2.
31. Yoo E-H, Lee PI, Huh C-Y, Kim D-H, Lee B-S, Lee J-K, et al. Predictors of leiomyoma recurrence after laparoscopic myomectomy. Journal of Minimally Invasive Gynecology. 2007;14(6):690-7.
32. Shiota M, Kotani Y, Umemoto M, Tobiume T, Hoshiai H. Recurrence of uterine myoma after laparoscopic myomectomy: What are the risk factors? Gynecology and Minimally Invasive Therapy. 2012;1(1):34-6.
33. Rossetti A, Sizzi O, Soranna L, Cucinelli F, Mancuso S, Lanzone A. Long-term results of laparoscopic myomectomy: recurrence rate in comparison with abdominal myomectomy. Human Reproduction. 2001;16(4):770-4.
34. Radosa MP, Owsianowski Z, Mothes A, Weisheit A, Vorwergk J, Asskaryar FA, et al. Long-term risk of fibroid recurrence after laparoscopic myomectomy. European Journal of Obstetrics & Gynecology and Reproductive Biology. 2014;180:35-9.
35. Hux M, Ng C, Ortega GL, Ferrazzi S, Goeree R. Utility values for pre-menopausal women suffering from symptomatic uterine fibroids. Expert Review of Pharmacoeconomics & Outcomes Research. 2015;15(1):181-9.
36. Hernández-Alava M, Pudney S. Econometric modelling of multiple self-reports of health states: The switch from EQ-5D-3L to EQ-5D-5L in evaluating drug therapies for rheumatoid arthritis. Journal of Health Economics. 2017;55:139-52.
37. Ara R, Brazier JE. Populating an economic model with health state utility values: moving toward better practice. Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research. 2010;13(5):509-18.

8 APPENDICES

Appendix 1 Systematic review critique

Table 35 – APPENDIX 1 – EAG appraisal of systematic review methods of the clinical effectiveness review

Systematic review components and processes	EAG response (Yes, No, Unclear)	EAG comments
Was the review question clearly defined using the PICO framework or an alternative?	Yes	PICO criteria are in CS Appendix Tables 5 and 6, and the searches were structured accordingly.
Were appropriate sources of literature searched?	Yes	Core healthcare and medical databases were searched alongside handsearching of multiple conferences (CS Appendix D.1.1).
What time period did the searches span and was this appropriate?	Partly	An initial search and two update searches covered the period from database inception to 7th February 2023 (conferences since 2019). Although searches were 6 months old when the CS was received by the EAG, no studies are thought to have been missed because the EAG's background search on 4th September 2023 only found six further conference abstracts for results from the same studies already identified by the company (i.e. not listed in the complete reference lists for included studies in CS Appendix Table 7). We did not find additional studies.
Were appropriate search terms used and combined correctly?	Yes	Search strategies are reported in CS Appendix D.1.1.
Were inclusion and exclusion criteria specified? If so, were these criteria appropriate and	Probably	Eligibility criteria are reported in CS Appendix Tables 5 and 6. Eligibility criteria for the first update search differs

Systematic review components and processes	EAG response (Yes, No, Unclear)	EAG comments
relevant to the decision problem?		from the original search in that it omits treatments such as radiofrequency ablation, hysterectomy, and watchful waiting, which are not in scope. It is not clear how watchful waiting might be different from BSC in this appraisal without looking at the studies it applies to. Watchful waiting does not involve any therapy whereas BSC in this appraisal includes NSAIDs and/or iron supplements (and tranexamic acid according to NG88). The eligibility criteria for the second update search reverted to that of the original search. Further eligibility criteria were applied for suitability for the NMA, e.g. excluding non-USA and non-EU studies. Whilst this would exclude Asian participants, it was appropriate to exclude those trials due to other limitations.
Were study selection criteria applied by two or more reviewers independently?	Yes	Both citation screening and full-text screening were done by two independent reviewers, with discrepancies reconciled by a third independent reviewer (CS Appendix D.1.2).
Was data extraction performed by two or more reviewers independently?	Yes	Data was extracted by two independent reviewers and any discrepancies were reconciled by a third independent reviewer (CS Appendix D.1.4).
Was a risk of bias assessment or a quality assessment of the included studies undertaken? If so, which tool was used?	Yes	A critical appraisal of PRIMROSE 1 and PRIMROSE 2 combined, using the NICE checklist for RCTs, is reported in CS section B.2.5. An earlier assessment was

Systematic review components and processes	EAG response (Yes, No, Unclear)	EAG comments
		carried out for the NMA in CS Appendix D.3.7 for PRIMROSE 1 and 2 and for LIBERTY 1 and 2, but the critical appraisal tool and sources are not reported. .
Was risk of bias assessment (or other study quality assessment) conducted by two or more reviewers independently?	Unclear	The methods for conducting the risk of bias assessments are not reported.
Is sufficient detail on the individual studies presented?	Yes	Study details for PRIMROSE 1 and PRIMROSE 2 are reported in CS sections B.2.3 to B.2.4, and in CS Appendices D and M. Additionally the CSRs, SAPs, and study publications provided necessary details. Study details for LIBERTY 1 and LIBERTY 2 are reported in CS Appendix D.3.3.2 and various study publications. Study details for PEARL I and PEARL II are reported in CS Appendix D.3.3.3 and D.3.3.4 and the study publications.
If statistical evidence synthesis was undertaken (e.g. pairwise meta-analysis, ITC, NMA) was undertaken, were appropriate methods used?	Yes	An NMA was undertaken to demonstrate equivalence of efficacy and safety of linzagolix with relugolix CT. Further analyses were provided by the company in Clarification Responses A10 and A12. The methods were generally appropriate although the EAG was unable to validate the results. The methods are critiqued in section 3.4 of this report.

Systematic review components and processes	EAG response (Yes, No, Unclear)	EAG comments
Abbreviations: CSR: clinical study report; ITC: indirect treatment comparison; NMA: network meta-analysis; PICO population, intervention, comparator, outcome; SAP: statistical analysis plan.		

Appendix 2 Risk of bias assessments

Table 36 – APPENDIX 2.1 – Risk of bias assessment for PRIMROSE 1 and PRIMROSE 2

Questions	PRIMROSE 1	PRIMROSE 2	EAG Comments
Was randomisation carried out appropriately?	Yes. Patients were randomised using a computer-generated randomisation list using the random allocation of treatment according to a permuted block randomisation stratified by race (Black or African American vs. Other)		Agree. Low risk of bias.
Was the concealment of treatment allocation adequate?	Yes. The patients were randomised to treatment groups by IWRS		Agree, the interactive web response system should ensure concealment of treatment allocation. Low risk of bias.
Were the groups similar at the outset of the trial in terms of prognostic factors?	Yes. As the prevalence of fibroids is higher and symptoms are more severe in Black women, randomisation was stratified to ensure equal distribution of Black patients among treatment groups. In the PRIMROSE 1 trial, patients had a higher mean BMI, a higher number of Black patients and a higher percentage of patients who were anaemic at baseline (Hb <12 g/dL) compared the PRIMROSE 2 trial		Stratification of randomisation by race is important and there is low risk of bias in that respect. The company consider race, BMI, and anaemia; however, fibroid characteristics such as FIGO type, size, number, and location are not reported or accounted for. Fibroid characteristics are important as they can influence which symptoms are experienced and to what extent. There is lack of clarity around whether the total fibroid volume could be measured consistently across all

Questions	PRIMROSE 1	PRIMROSE 2	EAG Comments
			groups (report section 3.2.3.1.3). Unclear risk of bias.
Were the care providers, patients and outcome assessors blind to treatment allocation?	Yes. Masked treatment kits were sent to each site and kept in controlled conditions. Masking was achieved by using tablets with an identical appearance between the linzagolix treatments and corresponding placebo and over-encapsulation of the ABT and corresponding placebo. All patients took two tablets and one capsule daily. The operational teams were masked to group allocation until unmasking after the database was locked; patients and investigation teams at each site remained blinded.		Agree. Clarification Response A5 states that investigators and subjects in both trials were blinded until the trials were complete, and the unmasking of PRIMROSE 1 at week 24 described in Donnez et al. 2022 only refers to the study Sponsor. Key endpoints were assessed by central laboratory. Low risk of bias
Were there any unexpected imbalances in dropouts between groups?	No. In order to consider all randomised and treated patients in the analysis, the assessment of the primary endpoint for patients who discontinued prior to Week 24 for a reason other than lack of efficacy, AEs, or operative or radiological interventions for UF was based on the results from the 28 days prior to the last eDiary entry in order to use as many data as possible up to Week 24 after the start of treatment, irrespective of actual treatment taken. Patients who had less than 28 days of data were considered as non-responders. The secondary endpoint of amenorrhea was assessed in a similar way		Patient flow, including discontinuations, are reported in the CONSORT diagrams in CS Appendix M.2. The CSRs report the randomised set numbers. For each individual trial the numbers were similar across groups except for weeks 24-52 (the second treatment period) where more patients discontinued in the 200mg/200mg+ABT group in both trials. Comparing the two trials, about twice as many patients discontinued during weeks 1-24 in

Questions	PRIMROSE 1	PRIMROSE 2	EAG Comments
			PRIMROSE 1 than in PRIMROSE 2 in each group. The amount of missing data varied by outcome and was almost 50% for some outcomes (Table 8). Unclear risk of bias.
Is there any evidence to suggest that the authors measured more outcomes than they reported?	No		All outcomes are reported in CS section B and CS Appendix M.3.3.3. FIGO classification of fibroid type was assessed (reported in summary in the SmPC and as evidenced by the existence of the FIGO 0, 1, 2 at baseline subgroup) but not reported in the patient baseline characteristics. Unclear risk of bias.
Did the analysis include an intention-to-treat analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	Yes. A sensitivity analysis was conducted on the primary efficacy endpoint to check the robustness of the analysis results under alternative assumptions with regards to missing data. Results of the sensitivity analyses imputing missing data and results in the PP Set were consistent with those of the main analysis. Missing values for continuous efficacy endpoints were handled within the analysis itself via mixed model repeated measures, with the assumption that the model specification was correct, and that the data were missing at random. All data recorded in the eCRF were included in data listings.		The CS reports a modified ITT analysis, the Full Analysis Set (FAS) which includes all randomised patients who received at least one dose of the study drug. The number of discontinuations varies between trials (n=48 PRIMROSE 1; n=21 PRIMROSE 2) and between treatment groups (randomized sets for each treatment group are reported in the CSRs) the

Questions	PRIMROSE 1	PRIMROSE 2	EAG Comments
			linzagolix 100 mg group had the most discontinuations. However the baseline characteristics do not show imbalances, nor did the study lose statistical power. The FAS also excludes patients who met exclusion criteria for liver function or BMD based on results of pre-treatment baseline assessments reported after Day 1 (CS Appendix M.3.1) which is appropriate (Clarification Response A9). Probably low risk of bias.
Was there good quality assurance for this trial?	Yes the trial was conducted in accordance with ICH GCP guidelines and regulatory requirements. The study monitor reviewed eCRFs and other study documents, and conducted source data verification, to verify that these and the trial protocol were followed.		No impact on risk of bias.

Source: reproduced from CS Table 15 with added EAG comments.
Abbreviations: ABT: add-back therapy; AE: adverse event; BMD: bone mineral density; BMI: body mass index; CONSORT: CONsolidated Standards Of Reporting Trials; eCRF: electronic case report form; FIGO: International Federation of Gynecology and Obstetrics; ICH GCP: International Council for Harmonisation Good Clinical Practice; ITT: intention to treat; IWRS: interactive web response system; PP per protocol.

37 – APPENDIX 2.2 – Risk of bias assessment for LIBERTY 1 and LIBERTY 2

The name of the critical appraisal tool used is not reported although it covers the main areas of bias, except reporting bias. The company sometimes made different assessments for LIBERTY 1 and LIBERTY 2 in answer to the same question for which the Al-Hendy et al. 2021 paper could answer as it covers both trials – described as “replicate”.[15] We have used the NICE checklist below for our own assessment.

Questions	EAG comments based on Al-Hendy et al. 2021[15]		EAG assessment
	LIBERTY 1	LIBERTY 2	
Was randomisation carried out appropriately?	The randomisation method is not reported. Stratification is not reported.		Unclear risk of bias
Was the concealment of treatment allocation adequate?	Yes. Allocation was performed using an interactive website.		Low risk of bias
Were the groups similar at the outset of the trial in terms of prognostic factors?	"Within each trial, the demographic and clinical characteristics of the participants as baseline were similar across the trial groups (Tables 1 and S4)" - Al-Hendy et al. 2021. However, FIGO type, location, and number are not reported, and these are important prognostic factors.		Unclear risk of bias
Were the care providers, patients and outcome assessors blind to treatment allocation?	The trials are described as "double-blind", however, blinding/unmasking methods and policies are not reported.		Probably low risk of bias
Were there any unexpected imbalances in dropouts between groups?	Figure S4, Al-Hendy et al. 2021 reports patient disposition. Dropouts ranged between 17.2% and 21.9% for each group, fairly high, but similar numbers. With regard to reasons for discontinuation, proportions were similar across groups.		Low risk of bias
Is there any evidence to suggest that the authors measured more	The reported outcomes in the study publication, Al-Hendy et al. 2021, match the primary and key secondary outcomes in the statistical analysis plan (Table S2 of study publication) which although		Low risk of bias

Questions	EAG comments based on Al-Hendy et al. 2021[15]		EAG assessment
	LIBERTY 1	LIBERTY 2	
outcomes than they reported?	different from the protocol was finalised prior to database lock and unblinding.		
Did the analysis include an intention-to-treat analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	A modified ITT analysis was performed on all participants who underwent randomisation and received at least one dose of relugolix or placebo. Any missing data was imputed and used in a mixed methods model (Al-Hendy et al. 2021 supplement).		Low risk of bias
Was there good quality assurance for this trial?	"The trials were conducted in accordance with the guidelines of the International Council for Harmonisation and the principles of the Declaration of Helsinki. All the participants provided written informed consent." – Al-Hendy et al. 2021.		No impact on risk of bias.

Abbreviations: FIGO: International Federation of Obstetrics and Gynecology; ITT: intention-to-treat.

38 – APPENDIX 2.3 – Risk of bias assessment for PEARL I

Questions	EAG comments based on Donnez et al. 2012[17]	EAG assessment
	PEARL I	
Was randomisation carried out appropriately?	Yes. Patients were randomly assigned in a 2:2:1 ratio which was stratified according to the haematocrit level at screening (≤28% or >28%) and race (black or other).	Low risk of bias
Was the concealment of treatment allocation adequate?	Yes. Patients were assigned to study group using a Web-integrated interactive voice-response system.	Low risk of bias

Questions	EAG comments based on Donnez et al. 2012[17]	EAG assessment
	PEARL I	
Were the groups similar at the outset of the trial in terms of prognostic factors?	Patients were stratified by race: black or other; but baseline characteristics for race are reported for White and Asian – with similar proportions across groups. Patients in the ulipristal acetate 10 mg arm had slightly more subserosal fibroids than patients in the other study arms (study publication Table 1). Study publication states there were no significant differences between the ulipristal acetate and placebo groups.	Low risk of bias
Were the care providers, patients and outcome assessors blind to treatment allocation?	Yes. Double-blind trial. Study materials and medication packaging were identical for all three groups. MRI results were assessed centrally by a radiologist unaware of study-group assignments.	Low risk of bias
Were there any unexpected imbalances in dropouts between groups?	There were very few dropouts: 1 in the placebo group, 5 in the 5mg group and 6 in the 10 mg group.	Low risk of bias
Is there any evidence to suggest that the authors measured more outcomes than they reported?	All assessments for outcomes in the protocol (available online with the study publication) are reported in the study publication and/or its supplement, except for ferritin. Ferritin levels help understand iron deficiency, but anaemia is defined by haemoglobin levels which are reported, therefore the EAG have no concern.	Low risk of bias
Did the analysis include an intention-to-treat analysis?	Yes. A modified ITT analysis was carried out. Only 1 patient in the 5 mg group (withdrawn prior to receiving study drug) and 4 patients in the 10 mg group (no efficacy data available) were excluded	Low risk of bias

Questions	EAG comments based on Donnez et al. 2012[17]	EAG assessment
	PEARL I	
If so, was this appropriate and were appropriate methods used to account for missing data?	from the primary analysis. Due to comparing two doses of ulipristal acetate with placebo, a Bonferroni correction was used (all p-values doubled). Missing values were imputed using the last available post-baseline value. A sensitivity analysis included the 4 patients in the 10 mg group who were excluded due to having no efficacy data by using baseline data carried forward.	
Was there good quality assurance for this trial?	The study was approved by the independent ethics committee at each study site and conducted in accordance with the principles of the International Conference on Harmonization – Good Clinical Practice guidelines. The original protocol, amendments, and statistical analysis plan are available with the full text article.	No impact on risk of bias

Abbreviations: ITT: intention-to-treat; MRI: magnetic resonance imaging.

39– APPENDIX 2.4 – Risk of bias assessment for PEARL II

Questions	EAG comments based on Donnez et al. 2012[18]	EAG assessment
	PEARL II	
Was randomisation carried out appropriately?	Yes. “The randomization list followed a stratification process for avoiding imbalance with respect to race or ethnic group among the three study groups”	Low risk of bias
Was the concealment of treatment allocation adequate?	Yes. “A Web-integrated voice-response system transmitted the randomization to the packaging organization, which delivered the medications to the treatment centers”	Low risk of bias

Questions	EAG comments based on Donnez et al. 2012[18] PEARL II	EAG assessment
Were the groups similar at the outset of the trial in terms of prognostic factors?	Patient baseline characteristics are reported in study publication Table 1. They are similar, except that the 5 mg ulipristal treatment group had a much larger median uterine fibroid volume than the other two treatment groups, although the ranges were similar, and they had a pain score of 9 whereas the other two groups had a pain score of 7. Similar to other trials of uterine fibroids, the FIGO classification for the characteristics of the fibroids, which can indicate type and severity of symptoms, is not reported.	Unclear risk of bias
Were the care providers, patients and outcome assessors blind to treatment allocation?	The study was double-blind. “Data were collected by an independent contract research organization (ICON Clinical Research) and handled and analyzed by an independent data-management organization (MDSL International).”	Low risk of bias
Were there any unexpected imbalances in dropouts between groups?	Study publication Figure 1 describes patient flow. There were very few dropouts: 4 excluded between randomisation and receiving treatment; and a further 2, 3 and 6 patients were withdrawn from each treatment group; 2 patients from the ulipristal treatment groups (1 each) did not receive the study drug.	Low risk of bias
Is there any evidence to suggest that the authors measured more	No. Outcomes reported in the study publication match the primary and secondary endpoints outlined in the study protocol.	Low risk of bias

Questions	EAG comments based on Donnez et al. 2012[18] PEARL II	EAG assessment
outcomes than they reported?		
Did the analysis include an intention-to-treat analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	There were two analyses: a modified ITT analysis that excluded 5 patients – two who never received the study drug and three with missing efficacy data after baseline – and a per-protocol population – the ITT population with the exclusion of patients with major protocol deviations and a compliance rate of <80%. The per-protocol population was favoured as the most conservative analysis, with the modified ITT population used for sensitivity analysis of missing data by using the baseline data carried forward.	Low risk of bias
Was there good quality assurance for this trial?	The study was approved by the independent ethics committee at each study site and conducted in accordance with the principles of the International Conference on Harmonization – Good Clinical Practice guidelines. The original protocol, amendments, and statistical analysis plan are available with the full text article.	No impact on risk of bias

Abbreviations: FIGO: International Federation of Obstetrics and Gynecology; ITT: intention-to-treat.

Appendix 3 NMA results: linzagolix versus leuprolide acetate

As noted above, the NMAs for the comparison of linzagolix against GnRH analogues suffer from serious methodological limitations and were based only on fixed-effects models. According to the company, only leuprolide acetate could be included in the evidence network and only three outcomes could be assessed: odds of response, % change in total fibroid volume, and % change in haemoglobin (Clarification Response A11). Results of these analyses are provided here for illustrative purposes and should be interpreted with caution. The assessment timepoint for the outcomes was not reported, so it is unclear whether these results refer to Week 12 or Week 24 assessments.

Figure 21 – APPENDIX 3.1 – Linzagolix vs leuprolide acetate: NMA results for odds of achieving a response

Figure 22 – APPENDIX 3.2 – Linzagolix vs leuprolide acetate: NMA results for the percentage change in total fibroid volume (fixed-effects model)

Figure 23 – APPENDIX 3.3 – Linzagolix vs leuprolide acetate: NMA results for % change in haemoglobin in patients who were anaemic at baseline (fixed-effects model)