Single Technology Appraisal

Zuranolone for treating postnatal depression [ID6431]
Committee Papers

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

SINGLE TECHNOLOGY APPRAISAL

Zuranolone for treating postnatal depression [ID6431]

Contents:

The following documents are made available to stakeholders:

The **final scope** and **final stakeholder list** are available on the NICE website.

1. **Company submission** from Biogen
2. **Company summary of information for patients (SIP)** from Biogen
3. **Clarification questions and company responses**
4. **Patient group, professional group and NHS organisation submissions** from:
   a. College of Mental Health Pharmacy
   b. NHS England
5. **Expert personal perspectives** from:
   a. Dr Livia Martucci, Chair of the Perinatal Faculty, consultant in perinatal psychiatry – clinical expert, nominated by Biogen Idec Ltd
   b. Professor Heather O'Mahen, Professor of Perinatal and Clinical Psychology: National Clinical Advisor to NHS England Perinatal Mental Health – clinical expert, nominated by NHS England
6. **External Assessment Report** prepared by Bristol Technology Assessment Group
7. **External Assessment Report – factual accuracy check**

Any information supplied to NICE which has been marked as confidential, has been redacted. All personal information has also been redacted.

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Single technology appraisal

Zuranolone for treating postnatal depression [ID6431]

Company evidence submission

April 17, 2025

File name	Version	Contains confidential information	Date
ID6431_Zuranolone for the treatment of PND_Biogen_CS_Final_v2.0 [Redacted]	Version 2	YES	April 17, 2025

Contents

List of tables

List of figures

List of abbreviations

Abbreviation	Definition
ACOG	American College of Obstetricians and Gynecologists
ADHD	Attention Deficit Hyperactivity Disorder
ADT	Antidepressant Treatment
AE	Adverse Event
AIC	Akaike Information Criterion
ALSPAC	Avon Longitudinal Study of Parents and Children
APA	American Psychiatric Association
BA	Behavioural Activation
BIC	Bayesian Information Criterion
BIMF	Barkin Index of Maternal Functioning
BL	Baseline
BMI	Body mass Index
BP	Bodily Pain
CACE	Complier-Average Causal Effect Analysis
CBT	Cognitive Behavioural Therapy
CE	Cost-Effectiveness
CEA	Cost-Effectiveness Analysis
CFB	Change From Baseline
CGI-I	Clinical Global Impression - Improvement

CGI-S	Clinical Global Impression - Severity
CI	Confidence Interval
CPI	Consumer Price Inflation
C-SSRS	Colombia Suicide Severity Rating Scale
CYP	Cytochrome P450
DSA	Deterministic Sensitivity Analysis
DSM	Diagnostic and Statistical Manual of Mental Disorders
DSU	Decision Support Unit
ECG	Echocardiogram
ECM	Established Clinical Management
EHIS	European Health Interview Survey
EPDS	Edinburgh Postnatal Depression Scale
EQ-5D	EuroQoL 5 Dimension
FAS	Full Analysis Set
FDA	Food and Drug Administration
GABA	Gamma Amino Butyric Acid
$GABA_A$	Gamma Amino Butyric Acid Type A
GAD-2/7	General Anxiety Disorder-2/7 Items
GBP	Great British Pound
GH	General Health
GHQ	General Health Questionnaire
GI	Gastrointestinal
GP	General Practitioner
HAMD-17	17-item Hamilton Rating Scale for Depression
HR	Hazard Ratio
HRQoL	Health-related Quality of Life
HRU	Healthcare Resource Utilisation
HTA	Health Technology Assessment
HUI	Health Utilities Index
ICD	International Classification of Diseases
ICER	Incremental Cost-effectiveness Ratio
IFSAC	Inventory of Functional Status After Childbirth
ILAP	Innovative Licensing Access Pathway
INMB	Incremental Net Monetary Benefit
IP	Investigational Product
IPT	Interpersonal Therapy
ITC	Indirect Treatment Comparison
ITQOL SF-47	Infant and Toddler Quality of Life Questionnaire Short Form-47
LOCF	Last Observation Carried Forward

LSM	Least-Square Means
LYG	Life Years Gained
MADRS	Montgomery-Åsberg Depression Rating Scale
MAE	Mean Absolute Error
MBU	Mother and Baby Unit
MCS	Mental Component Score
MCT	Meaningful Change Threshold
MDD	Major Depressive Disorder
MDE	Major Depressive Episode
ME	Mean Error
MH	Mental Health
MHP	Mental Health Problem
MHRA	Medicines and Healthcare products Regulatory Agency
MID	Minimal Important Difference
MMHA	Maternal Mental Health Alliance
MMRM	Mixed Model for Repeated Measures
MoA	Mechanism of Action
MRHD	Maximum Recommended Human Dose
NAS	Neuroactive Steroid
NCDR	National Commissioning Data Repository
NHB	Net Health Benefit
NHS	National Health Service
NHSCII	National Health Service Cost Inflation Index
NICE	National Institute for Health and Care Excellence
NIMH	National Institute of Mental Health
NR	Not Reported
OLS	Ordinary Least Square
ONS	Office for National Statistics
OR	Odds Ratio
PCS	Physical Component Score
PF	Physical Functioning
PHQ	Patient Health Questionnaire
PMH	Perinatal Mental Health
PMP S-E	Perceived Maternal Parenting Self-Efficacy
PND	Postnatal Depression
PNMH	Perinatal Mental Health
PQN	Perinatal Quality Network
PSA	Probabilistic Sensitivity Analysis
PSS	Personal Social Services

PSSRU	Personal Social Services Research Unit
PWC-20	Physician Withdrawal Checklist
QALY	Quality-Adjusted Life Year
RCT	Randomised Controlled Trial
RE	Role Emotional
RMSE	Root Square Mean Error
RP	Role Physical
RR	Relative Risk
SAE	Serious Adverse Event
SAS	Statistical Analysis System
SD	Standard Deviation
SE	Standard Error
SEM	Standard Error of Measurement
SF	Social Functioning
SF-36	36-item Short Form Health Survey
SLR	Systematic Literature Review
SNRI	Serotonin Noradrenaline Reuptake Inhibitor
SSRI	Selective Serotonin Reuptake Inhibitor
TCA	Tricyclic Antidepressant
TEAE	Treatment-emergent Adverse Event
THIN	The Health Improvement Network
TSD	Technical Support Document
UK	United Kingdom
US	United States
VBA	Visual Basic for Applications
WT	Weight
WTP	Willingess-to-pay

1 Decision problem, description of the technology and clinical care pathway

1.1 Decision problem

Table 1 presents the final National Institute for Health and Care Excellence (NICE) scope and the decision problem addressed in this submission.

The target population for zuranolone in this submission – adults with postnatal depression (PND) – is aligned with the expected Medicines and Healthcare Products Regulatory Agency (MHRA) label and the final NICE scope (Biogen 2025e). Efficacy of zuranolone in PND is supported by the largest clinical study programme in PND to date, consisting of the SKYLARK and ROBIN studies (see Sections 2.6.1 and 2.6.2). In line with the guideline on clinical investigation of medicinal products in the treatment of depression from the European Medicines Agency (European Medicines Agency 2023), the SKYLARK and ROBIN studies did not recruit mild PND patients as it is difficult to demonstrate an effect in mildly ill patients. Accordingly, the SKYLARK and ROBIN studies included patients with moderate to severe PND (as defined by the baseline 17-item Hamilton Rating Scale for Depression [HAMD-17] score [Section 2.6] and Montgomery-Åsberg Depression Rating Scale [MADRS] score [see Section 2.8.2]). While Biogen is aware of the limitations of the data package in providing evidence in mild PND patients, it is believed that given the unique mechanism of action (MoA) of zuranolone, the data package broadly supports its efficacy in all PND patients. Additionally, zuranolone, will be the only licensed treatment in PND, and is expected to be initially prescribed mainly by perinatal psychiatrists within the community perinatal mental health services, and therefore, most likely be used for the treatment of patients with moderate to severe PND.

Table 1: The decision problem

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope
Population	• Adults with PND	• As per scope	N/A
Intervention	• Zuranolone with ECM	• As per scope	N/A

Comparator(s)	ECM without zuranolone which may include:
	• Psychological therapies (e.g. CBT; facilitated self-help)
	• Antidepressant treatments:
	  ○ TCAs
	  ○ SSRIs
	  ○ SNRIs
	  ○ Atypical antidepressants (e.g. mirtazapine)
	• High-intensity psychological intervention combined with antidepressant treatments
	• Augmentation with additional antidepressants, antipsychotics, or electroconvulsive therapy
	• Best supportive care
		• As per scope	N/A

Outcomes	The outcome measures to be considered include: • depressive symptoms including relapse and remission rates • severity of depression • cognitive function • anxiety • sleep quality • hospitalisation • mortality • child health-related outcomes • adverse effects of treatment (including adverse effects of treatment discontinuation) • HRQoL	• As per scope	N/A
Subgroups to be considered	If the evidence allows, the following subgroups may be considered: • Previous history of depression • Severity of PND	• Subgroup analysis for severity of PND is provided • Subgroup analysis for previous history of depression is not provided	• In SKYLARK, 12% of patients were reported to have a history of depression. The subgroup sample size was deemed too small to produce stable assessment on efficacy; therefore, the analyses were not conducted.
Economic analysis	• The reference case stipulates that the cost effectiveness of treatments should be expressed in terms of incremental cost per quality-adjusted life year. • The reference case stipulates that the time horizon for estimating clinical and cost effectiveness should be sufficiently long to reflect any differences in costs	• As per scope	N/A

	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		
Special considerations including issues related to equity or equality	N/A	• Women's health and mental health have historically been underserved areas of healthcare. Zuranolone is anticipated to be the first and only approved treatment indicated for PND, and ensuring access to zuranolone has the potential to address the longstanding unmet needs in care for this vulnerable patient population.	• The first-ever government-led Women's Health Strategy for England programme was initiated in 2022. Ensuring equitable access to zuranolone may support implementation of this strategy by addressing several key commitments to the improvement of perinatal mental health services.

Source: (National Institute for Health and Care Excellence 2025)
Abbreviations: CBT: Cognitive Behavioural Therapy; ECM: Established Clinical Management; HRQoL: Health-related Quality of Life; MDD: Major Depressive Disorder; N/A: Not Applicable; NHS: National Health Service; PND: Postnatal depression; SNRI: Serotonin Noradrenaline Reuptake Inhibitors; SSRI: Selective Serotonin Reuptake Inhibitors; TCA: Tricyclic Antidepressant.

1.2 Description of the technology being evaluated

A summary of the technology is presented in Table 2.

Table 2: Technology being evaluated

UK approved name and brand name	Zuranolone (ZURZUVAE)
Mechanism of action	• Zuranolone is a synthetic neuroactive steroid and a selective positive allosteric modulator of synaptic and extrasynaptic $GABA_A$ receptors, designed to selectively retain the pharmacological properties of allopregnanolone on $GABA_A$ receptors and to have improved oral bioavailability compared with allopregnanolone. Similar to allopregnanolone, zuranolone is hypothesised to bind to the neuroactive steroid binding site on synaptic and extrasynaptic $GABA_A$ receptors (Martinez Botella 2017, Althaus 2020, Hoffmann 2020) • Activation of synaptic $GABA_A$ receptors results in short-term (millisecond time frame) GABAergic inhibition while activation of extrasynaptic $GABA_A$ receptors results in sustained GABAergic inhibition also known as tonic inhibition (seconds to minutes time frame). Upon binding of zuranolone to $GABA_A$ receptors, zuranolone enhances expression of various transmembrane $GABA_A$ receptors and signalling through these receptors (Lee 2014, Reddy 2016) Zuranolone is hypothesised to exert its antidepressant effects by enhancing GABAergic inhibition, in particular tonic inhibition, and may provide a mechanism to normalise function in brain networks dysregulated during PND (Northoff 2014, Fee 2017, Althaus 2020, Li 2020, Antonoudiou 2022)
Marketing authorisation/ CE mark status	Zuranolone does not currently have a marketing authorisation in the UK. An application for marketing authorisation was submitted to the MHRA in July 2024, and approval is expected in [redacted] • Zuranolone is expected to be the first and only licensed treatment for PND in the UK Zuranolone was granted an Innovation Passport by the MHRA as part of the ILAP, demonstrating zuranolone to be an innovative product that addresses area of high unmet need and offering potential benefits to patients in this vulnerable patient population (GOV.UK 2021)
Indications and any restriction(s) as described in the SmPC	Biogen anticipates zuranolone to be indicated for the treatment of PND in adults, following childbirth
Method of administration and dosage	• The recommended dose of zuranolone is 50 mg (two 25 mg capsules) taken orally once daily with fat-containing food in the evening or at bedtime for 14 days, as a single course of treatment (Biogen 2025e) • Dose reductions (Biogen 2025e): o The dose may be reduced to 40 mg (two 20 mg capsules) if the patient does not tolerate the 50 mg dose o In special populations such as patients with renal impairment, hepatic impairment, and patients with concomitant use of strong CYP3A inhibitors, the recommended dose of zuranolone is 30 mg
Additional tests or investigations	The introduction of zuranolone for the treatment of adults with PND would not require additional tests, investigations or administration beyond those that are currently required for all patients with PND in the UK
List price and average cost of a course of treatment	• Proposed list price (2025): [redacted] • Estimated cost per treatment course: o At list price: [redacted]
Patient access scheme (if applicable)	[redacted] o Estimated cost per treatment course after net discount: [redacted]

Source: See table
Abbreviations: CE: Cost-effectiveness; CYP: Cytochrome P450; GABA: γ-aminobutyric Acid; ILAP: Innovative Licensing and Access Pathway; MHRA: Medicines and Healthcare Products Regulatory Agency; PND: Postnatal Depression; SmPC: Summary of Product Characteristics; UK: United Kingdom.

1.3 Health condition and position of the technology in the treatment pathway

1.3.1 Disease overview

PND is one of the most common complications of pregnancy and childbirth. In the United Kingdom (UK), it is estimated that up to 16% of patients develop depressive symptoms within 1 year of giving birth (Harrison 2023). PND is defined as a depressive episode characterised by onset in peripartum period and driven by pathophysiological mechanisms that make PND a distinct disease from Major Depressive Disorder (MDD) (Slomian 2019). While the exact definitions can vary, it is widely accepted that the symptoms can manifest during pregnancy or within the first 4 weeks and up to 12 months following childbirth (American College of Obstetricians and Gynecologists (ACOG) , National Institute for Health and Care Excellence 2023a). PND can be a debilitating condition that leads to severe short- and long-term consequences for affected individuals, their children, partners and their families (Moore Simas 2019).

Patients with PND may present with multiple symptoms, including persistent sadness, anxiety, or depressed mood, which can cause significant functional impairment, a loss of interest in daily activities, difficulty sleeping and concentrating, changes in eating habits, fatigue and decreased energy, and feelings of worthlessness, shame, or guilt (Section 1.3.3) (National Institute of Mental Health (NIMH) 2022). Patients with PND are also three times more likely to experience suicidal ideation compared to those who are unaffected (Yu 2024). In 2020, maternal suicide was the leading cause of direct deaths in the UK, in the first year postnatal (Knight 2022).

PND can also increase the risk of poor mother-child bonding, resulting in worse short- and long-term outcomes for the child, including delayed cognitive development and a higher degree of emotional disorders (Kerstis 2012, Letourneau 2012, Netsi 2018, Slomian 2019). Furthermore, the negative long-term impact of PND on patients, children, partners, and families, leads to substantial economic impact (health and social care costs, health-related quality of life [HRQoL] and productivity losses) and healthcare resource utilisation (Bauer 2014, Bauer 2016b). The overall burden of PND is reflective of the National Health Service' (NHS) commitment to investing in perinatal mental health services. Mental health was specifically named as a substantial healthcare challenge in the NHS long term plan published in 2019 (National Health Service 2019), the first-ever government-led Women's Health Strategy for England programme initiated in 2022 (Department of Health & Social Care 2022),

and 3-year delivery plan for maternity and neonatal services in 2023 (National Health Service England 2023).

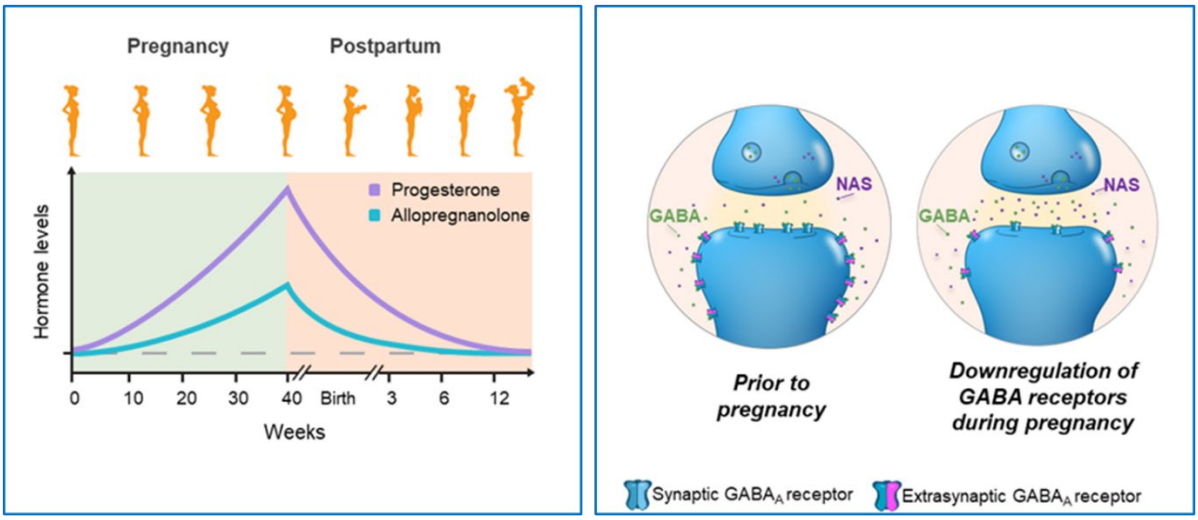
1.3.1.1 Pathophysiology

There are distinct physiological mechanisms associated with PND, which differentiate it from MDD. While the underlying pathophysiology of PND is not completely understood, it is thought that genetic, metabolic, endocrine, environmental, and neurobiological factors are associated with the manifestation of depressive symptoms in PND (Duman 2019, Batt 2020). Specifically, the perinatal period is characterised by rapid physiological changes in levels of endocrine hormones, peptides and neuroactive steroids, which may contribute to the vulnerability to depressive symptoms in some patients (Maguire 2009, Zorumski 2019).

Normal brain functioning relies on a balance of inhibitory (e.g., GABAergic) and excitatory (e.g., glutamatergic) signalling inputs. Gamma amino butyric acid (GABA) is the primary inhibitory neurotransmitter in the central nervous system and provides the inhibitory input into signalling pathways (Fogaça 2019), making it a key mediator in a variety of complex emotional and cognitive behavioural functions, such as mood, regulation, sensory processing, stress response, memory formation, and attention (Sarawagi 2021). Gamma amino butyric acid type A ($GABA_A$) receptors have been shown to regulate brain networks in regions of the brain relevant to PND (Deligiannidis 2019, Bryson 2020, Horáková 2022, Feng 2024).

Neuroactive steroids are a class of endogenous or synthetically produced steroids that can act as positive allosteric modulators of $GABA_A$ receptor (Zorumski 2019) – allopregnanolone, a metabolite of progesterone, is an example of an endogenous neuroactive steroid and a positive allosteric modulators of $GABA_A$ receptors (Zorumski 2019, Hoffmann 2020). During pregnancy and the postnatal period, neuroactive steroid levels fluctuate; allopregnanolone levels increase during pregnancy and fall after childbirth (Chrousos 1998, Meltzer-Brody 2020). As allopregnanolone levels increase during pregnancy, the body adapts by downregulating $GABA_A$ receptor expression to avoid excessive neuronal inhibition and maintain an ideal level of inhibition throughout pregnancy (Figure 1) (Maguire 2009, Maguire 2019, Meltzer-Brody 2020). Disruption in the ability of the GABA system to adapt to changes in allopregnanolone levels during the peripartum period may be one factor that mediates the onset of PND symptoms (Maguire 2008, Maguire 2019, Meltzer-Brody 2020).

Figure 1: Levels of neuroactive steroid (left) and $GABA_A$ receptors (right) during the perinatal period



Source: (Chrousos 1998, Maguire 2009, Meltzer-Brody 2020)
Abbreviations: GABA: γ-aminobutyric acid; NAS: Neuroactive steroid.

Given the role of $GABA_A$ receptors in the pathophysiology of neurologic and neuropsychiatric disorders, $GABA_A$ receptor positive allosteric modulators, such as zuranolone are being explored for the treatment of mood disorders (Luscher 2011).

Similarly to the endogenous allopregnanolone, zuranolone is hypothesised to bind to the neuroactive steroid binding site on synaptic and extrasynaptic $GABA_A$ receptors (Martinez Botella 2017, Althaus 2020). Activation of synaptic $GABA_A$ receptors results in short-term (millisecond time frame) GABAergic inhibition while activation of extrasynaptic $GABA_A$ receptors results in sustained GABAergic inhibition, also known as tonic inhibition (seconds to minutes time frame) (Lee 2014, Reddy 2016). Upon binding of zuranolone to $GABA_A$ receptors, zuranolone enhances signalling and expression of various transmembrane $GABA_A$ receptors. Zuranolone has also been shown to increase cell surface expression of $GABA_A$ receptors in *in vitro* studies (Althaus 2020). Therefore, zuranolone is hypothesised to exert its antidepressant effects by enhancing GABAergic inhibition, in particular tonic inhibition, and may provide a mechanism to normalise function in brain networks dysregulated during PND (Northoff 2014, Fee 2017, Althaus 2020, Li 2020, Antonoudiou 2022).

1.3.1.2 Risk factors

Some factors may increase the likelihood of developing PND symptoms. As demonstrated by a recent study based in England, risk factors associated with a significantly increased risk of developing PND symptoms included the following: being aged <25 years (adjusted risk ratio [aRR]: 1.37; p=0.017), living at a different address to the father of the baby (aRR: 1.44; p=0.006), having physical health problem (aRR: 1.34; p=0.027), having mental health problem (aRR: 1.94; p<0.001) or antenatal anxiety (aRR: 2.12; p<0.001), and experiencing antenatal depression (aRR: 1.44; p=0.007).

Satisfaction with birth was associated with a decreased risk of PND (aRR: 0.92; p<0.001) (Harrison 2023). These findings are also supported by evidence from outside of the UK (Agrawal 2022, American Psychological Association 2022).

However, it is important to stress that any woman, irrespective of the history of mental health problems and/or situational factors can experience symptoms of PND (American Psychological Association 2022), and for many patients, the onset of PND can be the first time they experience any mental health problems.

1.3.1.3 Signs and symptoms

Due to the heterogeneous nature of PND, patients may present with varying signs and symptoms. Common symptoms include: depressed mood; loss of interest or pleasure in activities; changes in sleep, appetite, or weight; psychomotor disturbances such agitation (e.g., restlessness) or retardation (e.g., slowed movement); difficulty with speech or thinking; decreased energy; feelings of guilt or worthlessness and trouble concentrating (National Institute of Mental Health (NIMH) 2022). In addition, patients with PND may experience persistent sad, anxious, or “empty” mood, irritability, aches, pains, headaches, cramps, or digestive problems that do not have a clear physical cause or do not ease with treatment (National Institute of Mental Health (NIMH) 2022). In particular, symptoms such as anxiety (experienced by up to 70% of PND patients) and sleep problems have been associated with more severe depression and frequent self-harm ideation, underscoring the importance of addressing all symptoms to improve the function and psychological health for patients suffering from PND (Deligiannidis 2023a). In the UK, deaths from mental health-related causes account for approximately 38% of deaths between 6 weeks and 1 year after the end of pregnancy, with maternal suicide being the leading cause of death during this period (Knight 2022).

A troubled relationship with the newborn baby can also be a characteristic feature of PND, for example, difficulty bonding or forming an emotional attachment with the baby; persistent doubts about one’s ability to care for the baby; and thoughts about harming the baby (National Institute of Mental Health (NIMH) 2022).

The intensity and number of symptoms experienced by patients with PND defines the severity of the PND episode as well as the degree of functional disability (American Psychiatric Association (APA) 2022). In research settings, tools such as MADRS and HAMD are commonly used to determine disease severity. However, in clinical practice, distinguishing between severities of PND remains challenging due to a variety of diagnostic criteria, screening tools, and their cut-offs (Kroenke 2001, Hanusa 2008, Brodey 2016). Therefore, in the real-world clinical practice, the severity of PND is primarily based on clinical judgement. Mild PND is defined as having few, if any, symptoms which are in excess of those required to make a diagnosis, with symptoms being distressing yet manageable

and resulting in minor impairment in social or occupational functioning. In moderate PND patients, the quantity and severity of symptoms and/or functional impairment increases compared to mild PND. Finally, severe PND is defined as having a substantially greater number of symptoms than required to make a diagnosis, whilst being seriously distressing, unmanageable, and markedly interfering with social and occupational functioning (American Psychiatric Association (APA) 2022).

1.3.1.4 Disease course

Disease course in PND has not been well described in the literature and more studies are needed to better understand the disease. However, it is becoming widely recognised that PND, if not addressed, can have long-term effects on mental health and general wellbeing of affected patients, with the literature suggesting that PND patients can suffer from symptoms of depression over prolonged periods of time. A global Systematic Literature Review (SLR) assessed 23 studies published from January 1985 to August 2012 including both community samples and clinical samples. Studies in the community found that 17% to 62% of patients with PND continued to meet the diagnostic criteria for depression at 3 years postnatal. In the clinical setting, these numbers were even higher, with 39% to 58% of patients still reporting depression at 3 to 3.5 years postpartum (Vliegen 2014).

A more recent study by Netsi et al., leveraged data from the Avon longitudinal study of parents and children (ALSPAC) study to examine the association between differing levels of persistence and severity of PND on long-term outcomes in England (Netsi 2018). This is one of the few studies providing UK data on the disease course for PND. The study sample comprised of 9,848 women residing in the former Avon Health Authority in south-west England, with an estimated delivery date between 1 April 1991 and 31 December 1992. The results show that women with PND symptoms in the first year postnatal, especially those with persistent symptoms, are at increased risk of prolonged depression until at least 11 years after childbirth. The study assessed depression at several timepoints up to 11 years after childbirth to establish the natural course of PND. The study defined three levels of PND severity using the following Edinburgh Postnatal Depression Scale (EPDS) scores: 13 to 14 points indicates moderate depression; 15 to 16 points, marked depression; and ≥17 points, severe depression. PND symptoms were considered to be persistent when an individual scored above the defined EPDS thresholds at both the 2- and 8-month postnatal assessment. Across study sample, after the postnatal year there was little change in the mean EPDS scores over time. Compared with patients with EPDS score <13 in the postnatal year, patients with EPDS score ≥13 in the postnatal year had consistently higher EPDS scores for up to 11 years after childbirth that progressed in a stepwise function (Netsi 2018). Therefore, the evidence from ALSPAC study suggests that patients with sub-optimally managed PND (regardless of severity), who have not reached symptom resolution in the short-term, do not improve over time and EPDS scores remain consistently higher compared unaffected individuals.

1.3.2 Epidemiology

PND symptoms are common among postnatal women. Prevalence of PND symptoms ranges from 11-12% in Europe (Dekel 2019, Della Corte 2022), 11-20% in Latin America (Melo 2012, Lara 2015), 15% in Asia (Matsumura 2019, Shi 2021), 13-18% in North America (Bauman 2020, Gheorghe 2021), to 13-35% in the Middle East and North Africa (Ahmed 2012, Alasoom 2014, Dikmen-Yildiz 2017, Roumieh 2019, Ahmed 2020, Alhusaini 2022, Alonazi 2022, Hanach 2023).

In the UK, approximately 15-20% of patients experience symptoms of depression and anxiety in the first year after giving birth (National Institute for Health and Care Excellence 2023a). A study by Petersen et al. (2018) assessed electronic health record data from The Health Improvement Network, a large primary care database in the UK. Diagnoses and symptoms were recorded using Read codes, which can be mapped to International Classification of Diseases-10 (ICD-10). A total of 206,517 patients who have given live birth between 2000–2013 were included. Based on the collected data, 23,623 (11%) had at least one record of depressive diagnosis or symptoms in the year after delivery. Of all individuals included in the study , only 9,005 patients received formal PND diagnosis (prevalence of 4%), 8,815 with “depression” diagnosis (8,815) or a record of “depression symptoms” (11,318) (Petersen 2018).

A more recent publication by Harrison et al. (2023) analysed data from three population-based, maternity surveys carried out in England in 2014 (n=4,571), 2018 (n=4,509), and 2020 (n=4,611). In this study, the EPDS screening tool scores of assessed patients were used to estimate prevalence of PND symptoms. Based on the weighted prevalence estimates for PND symptoms (EPDS score ≥13) across surveys, prevalence increased from 10.3% in 2014 to 16.0% in 2018 (RR: 1.55; 95% CI: 1.36, 1.77) and to 23.9% in 2020 (RR: 1.49; 95% CI: 1.34, 1.66). The survey data from the year 2020, coincided with the Covid-19 pandemic, which appears to have an important negative impact on postnatal women’s mental health. Nevertheless, it is worth noticing an existing trend of increasing prevalence of PND symptoms (an increase of 5.7% between 2014 and 2018), which was likely to have been further accelerated by the pandemic (Harrison 2023).

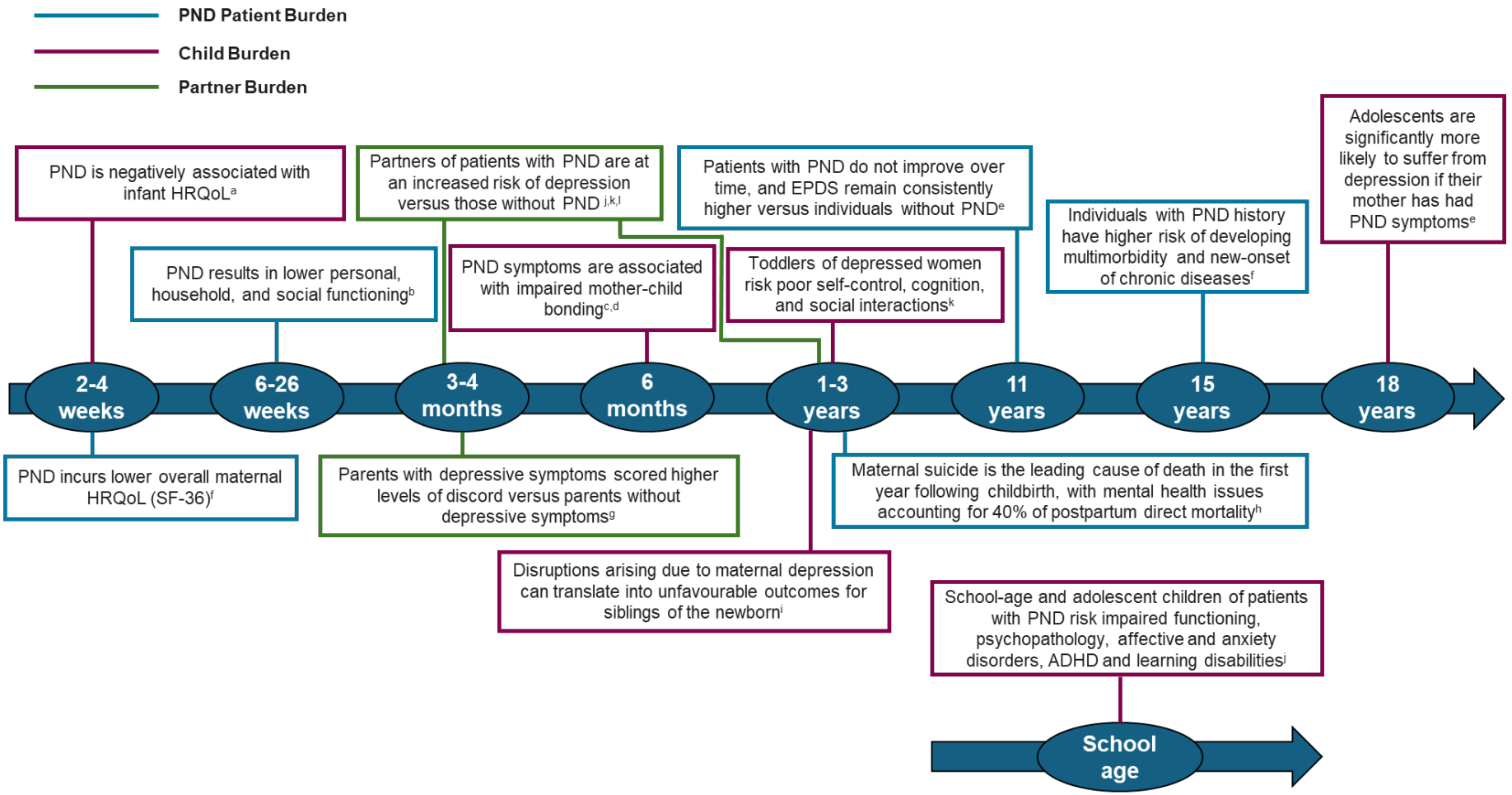
1.3.3 Burden of PND

1.3.3.1 Humanistic burden

PND can be a disabling disease that can result in a substantial burden, and morbid short- and long-term consequences, for patients, children, partners, and their families (Moore Simas 2019). It has been shown that the impact of depressive clinical symptoms extends well beyond abnormalities in mood and neurovegetative symptoms to also include multiple aspects of psychological well-being, as

well as social, role, and physical functioning (Gerbasi 2021). The impact of PND on each population is described in sections below and summarised in the Figure 2.

Figure 2: Humanistic burden of PND: impact on the patients, children and partners across identified studies



Source: [a](Subbiah 2025); [b](Posmontier 2008) ; [c](Moore Simas 2019); [d](Slomian 2019); [e](Netsi 2018); [f](Zhang 2025); [g](Kerstis 2012); [h](Knight 2022); [i](Canadian Paediatric Society 2004); [j](Nishimura 2015); [k](Paulson 2010); [l](Goodman 2004)
Please note that the findings presented in this figure come from several studies with different study populations, follow-up periods and methodologies. As such, caution must be taken when interpreting findings side-by-side. Time horizon is not to scale.
ADHD: Attention Deficit Hyperactivity Disorder; EPDS: Edinburgh Postnatal Depression Scale; HRQoL: Health-Related Quality of Life; PND: Postnatal Depression; SF-36: Short-Form Health Survey - 36 items.

Burden on patients with PND

PND symptoms adversely affect the physical and psychological functioning, well-being, and overall quality of life (QoL) of patients with PND. Compared to unaffected individuals, patients with PND have significantly impaired functioning across various domains including personal ($p<0.001$), social ($p<0.05$), household ($p<0.001$), and overall function ($p<0.001$) (Posmontier 2008). Symptoms of PND differ depending on the severity of illness, with lower overall function predictive of increasing PND severity ($p<0.001$) (Posmontier 2008, Moore Simas 2019). Additionally, PND is associated with lower overall maternal HRQoL, as measured by various instruments including the Short-Form Health Survey-36 items (SF-36), General Health Questionnaire-12 items (GHQ-12), Perceived Maternal Parenting Self-Efficacy (PMP S-E), and Inventory of Functional Status After Childbirth (IFSAC) (Fathi 2018, Moore Simas 2019, Parsa 2019). In the short-term, patients with PND have lower overall mean SF-36 scores compared with unaffected individuals, with statistically significant differences for both physical and mental component subscales at 2 and 4 months postnatal (Moore Simas 2019). Furthermore, symptoms of PND can have long-term consequences on patients' mental health, as there was a significant correlation between PND symptoms at 2, 3, and 6 months postnatal, and maternal health problems at 8-years of follow-up, as assessed using GHQ-12. Additionally, PMP S-E scores (a measure of a parent's belief in their ability to carry out tasks associated with parenting a child) were also significantly associated with PND symptom scores ($p<0.001$) (Moore Simas 2019). Finally, IFSAC scores showed a significant correlation between PND and poorer overall maternal functional status score, social and community engagement, and self-care ($p<0.001$) (Fathi 2018).

More severe PND can also increase the risk of self-harm ideation (Sit 2015, Deligiannidis 2023a). In the 2022 report on "Mothers and Babies: Reducing Risk Through Audits and Confidential Enquiries across the UK", surveillance data was included for patients who died in the first year after pregnancy between 2018 and 2020 in the UK. The report found that deaths from mental-health related causes accounted for nearly 40% of deaths within the first year postnatal, and maternal suicide was found to be the leading cause of direct death during this period (Knight 2022). Additionally, as demonstrated in a Swedish, population-matched cohort study, a total of 86,551 women with PND and 865,510 unaffected women were identified between 2001-2017; during an 18-year follow-up, women with PND had three times higher risk of suicidal behaviour (HR: 3.15; 95% CI: 2.97, 3.35) compared with matched unaffected women. The risk of suicidal behaviour was greatest within 1 year following diagnosis and was approximately seven times higher for women with PND compared with matched unaffected women (HR: 7.20; 95% CI: 6.07, 8.54) (Yu 2024).

PND is also associated with higher risks of chronic diseases and multimorbidity in mid to late life in patients affected by PND. A prospective cohort study of participants from the UK Biobank (a large biomedical database containing health-related data) assessed long-term outcomes of 54,885

participants including 5,106 patients (9.3%) with history of PND. The study demonstrated that PND patients had higher odds of having multimorbidity at baseline (OR: 1.35, 95% CI: 1.27–1.44) and higher risk of developing multimorbidity during the 15-year follow-up (HR: 1.13, 95% CI: 1.08–1.20). PND was also associated with an increased number of new-onset chronic diseases (RR: 1.08, 95% CI=1.05–1.12), which accumulated at a faster annual rate (b=0.009, 95% CI=0.007–0.011), compared to unaffected individuals (Zhang 2025). Other studies also reported an association with PND and long-term risk of developing diabetes mellitus (adjusted OR: 3.0, 95% CI: 1.62-5.74, *p*=0.001) (Atuhaire 2024), cardiovascular disease (HR: 2.42, 95% CI: 2.04-2.86) (Auger 2020), and anxiety and depressive symptoms (Netsi 2018, Slomian 2019, Putnick 2020).

Burden on children

The symptomatic burden of PND can negatively impact infant's HRQoL (Subbiah 2025). A study (n = 1,843) found that PND is negatively associated with infants' HRQoL during the first 12 months of life, with negative impact being observed as early as age of 1-month. Using the Infant and Toddler Quality of Life Questionnaire Short Form-47 (ITQOL SF-47), at 1 month of age, negative associations were observed across all domains of the infant's HRQoL with the strongest impact being observed on domain of infant temperament (standardised regression coefficient: − 0.30; 95%CI: − 0.34; − 0.25) (Subbiah 2025). PND can also negatively impact infant sleep patterns, with the negative impact being especially prominent among children whose mothers have severe and/or chronic PND (Slomian 2019).

As described before, PND can have a negative impact on maternal function, including caring for the child, which can significantly impair mother-infant bonding (Slomian 2019). Women with PND are up to five times more likely to experience poor mother-child bonding, a central psychological process after childbirth that, if interrupted, can lead to negative long-term consequences for a child's development (Noorlander 2008, Slomian 2019). This can result in delayed cognitive development, significantly impaired language development, a higher degree of emotional disorders, lower social engagement, and an increase in behavioural problems among children whose mothers were suffering from PND. Additionally, severe or chronic maternal PND presents a higher risk to children's development compared with milder PND (Slomian 2019).

In the UK setting, the association between PND and long-term child outcomes has been investigated in the long-term observational study leveraging ALSPAC data (Netsi 2018). A total of 9,848 women and 8,287 children from 2016-2017 were included in the sample. The study found that children of patients with persistent PND, particularly when severe, were at greater risk of impaired development and behavioural problems. Compared to women with EPDS score below PND symptom threshold (<13) during the postnatal year, children of women with nonpersistent PND (depression only at 2

months after childbirth) had a statistically significant increase in the risk of behavioural disturbance at 3.5 years, regardless of PND severity. This association was even stronger among children of patients with persistent PND of moderate (OR: 3.04; 95% CI: 2.10, 4.38) or marked severity (OR: 2.84; 95% CI: 1.71, 4.71). Persistent severe PND was associated with worst child outcomes, substantially increasing the risk for behavioural problems at 3.5 years of age (OR: 4.84; 95% CI: 2.94, 7.98), lower mathematics grades at 16 years of age (OR: 2.65; 95% CI: 1.26, 5.57), and higher prevalence of depression at 18 years of age (OR: 7.44; 95% CI: 2.89, 19.11).

Burden on partners and families

PND also puts additional strain on relationships with partners, negatively affecting the partner's own mental health and QoL. Partners of patients experiencing PND face higher rates of stress, anxiety, and depression compared to partners of individuals unaffected by PND (Moore Simas 2019). The rate of paternal depression in the first postnatal year among men whose partners were experiencing PND is higher compared with paternal depression in the general population (24-50% vs 1-26%). Additionally the rate of depression among partners is significantly higher when patients experience severe PND (Goodman 2004). PND has also been associated with decreased marital satisfaction and higher rates of separation and divorce (Kerstis 2012, Letourneau 2012). Additionally, compared to other households, the families of patients experiencing PND are subject to a higher level of disruption due to parenting behaviours and relationship separation (Moore Simas 2019). These disruptions arising due to maternal depression can also translate into unfavourable outcomes for siblings of the newborn. Toddlers and preschoolers of depressed women are at risk for developing poor self-control and difficulties in cognitive functioning and in social interactions with parents and peers. Furthermore, school-age and adolescent children are at risk for impaired adaptive functioning and psychopathology, affective and anxiety disorders as well as attention deficit hyperactivity disorder (ADHD) and learning disabilities (Canadian Paediatric Society 2004).

1.3.3.2 Economic burden

PND, especially among untreated and sub-optimally treated patients, is associated with high economic burden due to high healthcare resource utilisation, as well as decreased productivity (increased absenteeism and presenteeism among patients with PND and their partners) (O'Callaghan 2024). Additionally, the negative long-term impact of PND on patients, children, and families, also leads to substantial indirect economic impact (Bauer 2016b).

In the UK, perinatal mental illnesses (including anxiety and psychosis) is thought to carry a long-term cost to society of an estimated £6.6 billion (2012/2013 prices) for each one-year cohort of births, assuming a prevalence of 10.7% during pregnancy and a prevalence of 7.4% up to 9 months postpartum, using a time horizon of 10 year for estimating the impact (Bauer 2016b). The average

cost to society of one case of PND is approximately £75,000 (2012/13 prices), of which nearly three-quarters of the costs relate to adverse effects on the child (Table 3) (Bauer 2016b). The cost impact of PND on children is estimated to be £52,577 per birth, which can be attributed to factors such as emotional problems, special educational needs, and leaving school without qualifications. The cost impact of PND on women is estimated to be £23,151 with productivity losses accounting for approximately 13% of these costs (£3,028 per case). These estimates are based on the assumptions that 23% of patients with PND do not return to work after birth, only 13% of patients return to full-time employment, and the total productivity loss per year due to depression is 12 weeks (Bauer 2016b).

Table 3: Estimated costs of PND in the UK, per case (in 2012/2013 prices)

	Public sector			Wider society			Total
	Health and social care	Education	Criminal Justice	HRQoL losses*	Productivity losses	Other	
Impact on women	£1,688	-	-	£18,158 £277 (lost life)†	£3,028	-	£23,151
Impact on children	£2,831	£3,744	£1,974	£8,750 £22,157 (lost life)†	£5,661	£7,460	£52,577

Source: (Bauer 2016b)
* Calculated based on health impairments (measured in form of health disutilities) experienced by patients with depression (anxiety, psychosis) during the perinatal period as well as during subsequent years linked to non-remitted mental illness or relapse. Calculated over a 10-year period and assuming a willingness to pay of £25,000 per QALY to monetise the QALY losses
† For a whole life lost through suicide or infanticide, the 'value of a prevented fatality', is estimated at £1,722,000 per case, which is used in UK government policy analysis. The probability of suicide for mothers is 0.049% and risk difference infant death 1.3% vs infants from women without PND.
Abbreviations: HRQoL: Health-Related Quality of Life; PND: Postnatal Depression; QALY: Quality-Adjusted Life Year; UK: United Kingdom.

There is paucity of UK-specific data on the direct medical costs of PND. Outside of UK, there was evidence from analysis of administrative or claims databases in two United States (US)-based studies (Epperson 2020, Moore Simas 2020) and one Australian study (Callander 2021). Households with PND affected patients incur substantial economic burden driven by higher medical and pharmaceutical spending for both the households and the affected patients, and increased healthcare utilisation; e.g. driven by more outpatient visits for PND patients compared to unaffected individuals (Epperson 2020). Healthcare resource use among children of mothers in the US was also significantly higher among children in the PND exposure cohort compared with the non-PND exposure cohort (Moore Simas 2020). While this may not entirely reflect UK clinical practice, it is likely that patients living in the UK, and who are affected by PND also incur additional healthcare utilisation versus unaffected individuals.

1.3.4 Care pathway

The Maternal Mental Health Alliance (MMHA) has been mapping specialist perinatal mental health (PMH) services across the UK since 2015. Their latest report highlighted a significant progress in expanding specialist PMH provision in recent years, which can be attributed primarily to increased government funding and commitments across all four UK nations. England has made the most

progress, with 16% of areas now meeting quality standards and beginning to deliver on NHS Long Term Plan ambitions. However, sustained momentum is needed to address ongoing challenges to ensure all women and families can access high-quality PMH care when needed (Maternal Mental Health Alliance 2023).

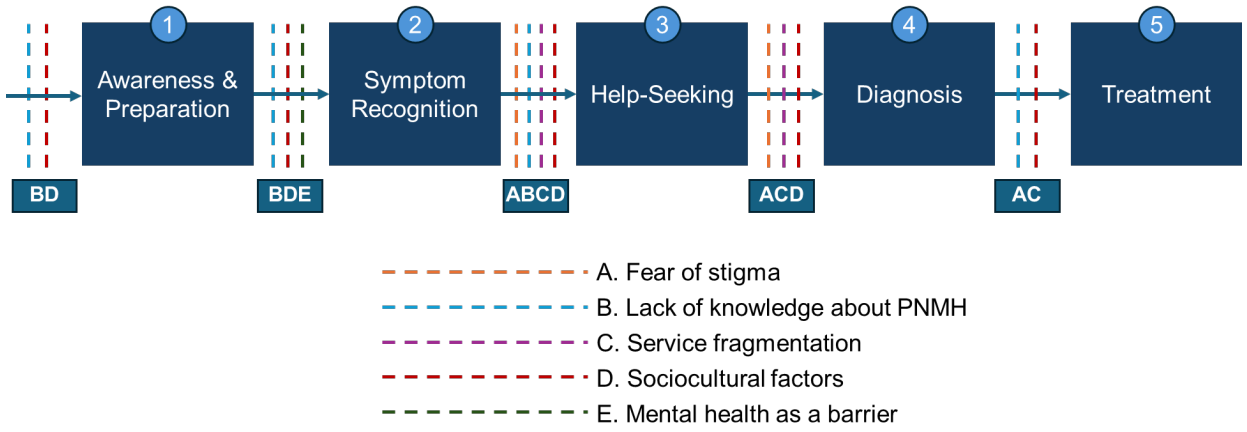
Based on the current clinical journey framework, diagnosis of PND is estimated to be made between 1-week to 6-months from symptom presentation. Most patients are screened by a general practitioner (GP) or allied health professional and are diagnosed by a GP. High-risk patients with pre-existing mental health problems may undergo enhanced screening and receive a diagnosis from a psychiatrist (Biogen 2024c).

Published literature suggest that only between 4% and 15% of women suffering from PND receive a formal diagnosis of PND, suggesting that many patients remain undiagnosed (Jones 2014, Petersen 2018). This is supported by the internal analysis of the Clinical Practice Research Datalink (CPRD) Aurum database conducted by Biogen; based on the data collected between 2017 and 2022 in England, among the 382,639 women who gave birth, only 10,445 (2.73%) had a record of PND diagnosis (Biogen 2025a), suggesting prevalence of PND of less than 3% - far below the 15-20% prevalence rates reported by NICE (National Institute for Health and Care Excellence 2023a) and indicative of significant underdiagnosis problem. Diagnosing PND can be challenging because of normal changes in emotions, responses, mental state, and functioning during the perinatal period. Other conditions such as baby blues (low mood after childbirth, which lasts no longer than 2 weeks), hyperthyroidism or hypothyroidism may contribute to symptoms that are similar to depressive symptoms, but there are also other mental health problems, which need to be differentiated from PND and include postnatal psychosis, bipolar disorder, generalised anxiety disorder, obsessive-compulsive disorder, and post-traumatic stress disorder (National Institute for Health and Care Excellence 2023a, Carlson 2024). The issue of anxiety disorders and depression being under-recognised throughout the pregnancy and postnatal period has been acknowledged in the NICE guidance (CG192) (National Institute for Health and Care Excellence 2020).

There are many reasons for why underdiagnosis of PND remains a significant unmet need and healthcare burden in the UK. Some studies point towards inadequate provision of community mental health services, shortages of health visitors and midwives, and lengthy waiting lists (Bauer 2014); 38% of patients wait over a month to be seen following a referral (Royal College of Obstetricians & Gynaecologists). Additionally, despite demand for perinatal mental health services increasing by 40% between August 2022 - March 2023, there has only been an 8% increase in the number of women accessing these services during the same period (Alliance 2023). There is growing evidence that suggests reasons for difficulties accessing mental health services extend beyond inadequate resources and potentially occur at multiple time points along the care pathway (Sambrook Smith

2019). In particular, the stigma associated with PND exacerbates underdiagnosis and serves as an additional barrier to care. This can include feeling weak, fear of being labelled with a mental health disorder, negative beliefs about antidepressant treatment (ADT), and feeling ashamed about seeking help, especially, since for many patients, experience of PND is the first time they are experiencing a mental health problem (Clement 2015, Bell 2016, Henshaw 2016, Thorsteinsson 2018, Sambrook Smith 2019). Figure 3 summarises some of the barriers within the UK care pathway that contribute to challenges in recognising and addressing perinatal mental health illnesses, leading to delayed referrals and difficulties in accessing appropriate care.

Figure 3: Conceptual model of key barriers in the care pathway to accessing mental health services during the perinatal period



Source: Adapted from Sambrook Smith et al., 2019 (Sambrook Smith 2019)
Abbreviations: PNMH: Perinatal Mental Health

To aid timely and accurate diagnosis of mental health problems in the perinatal period, there are two sets of guidance for care pathways that outline the care provision for expectant and new mothers in the UK: the Perinatal Care Pathways published by NHS England in 2018 (National Health Service 2018) and the Clinical Guideline for Antenatal and Postnatal Mental Health published by NICE in 2014, and most recently updated in 2020 (National Institute for Health and Care Excellence 2020).

1.3.4.1 Assessment and diagnosis

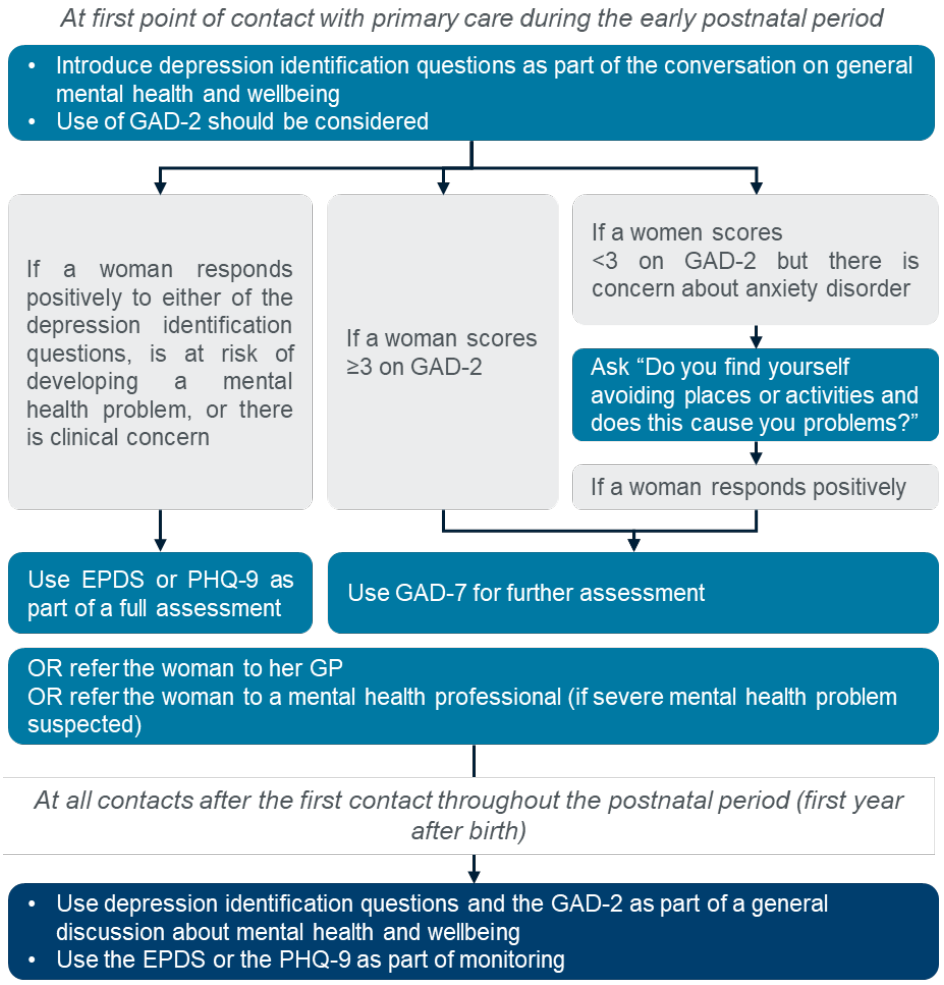
Depending on the circumstances (e.g., a patient is planning pregnancy, is already pregnant or in the postnatal period), patients with suspected or ongoing mental health problems may come into contact with different healthcare professionals in different settings (e.g., mental healthcare visitors in a community setting, GPs in primary care setting or mental health specialists in a secondary/tertiary setting), which are outlined in the NHS-designed Perinatal Care Pathways (National Health Service 2018) (see Figure 21 in Appendix J.1). However, regardless of the individual patient circumstances, it is important that each patient receives the appropriate care, which starts with a timely diagnosis and

delivery of the appropriate treatment. NICE guidelines for the clinical management of antenatal and postnatal mental health (CG192) set out the basis for good quality diagnosis and management of mental health problems in the perinatal period (National Institute for Health and Care Excellence 2020).

According to the NICE guidelines, screening at the first contact with primary and community care and/or during the early postnatal period should be initiated with general discussion on mental health and wellbeing, with use of depression identification questions; use of the General Anxiety Disorder-2 Item (GAD-2) scale should also be considered for anxiety symptoms. Depending on the responses of the patient (i.e., positive responses to depression identification questions and/or score of ≥3 on the GAD-2 scale) and assessment of the healthcare professional regarding the risk of developing mental health problems or presence of a clinical concern, healthcare professionals are advised to conduct a full assessment using additional screening tools such as Patient Health Questionnaire (PHQ-9), General Anxiety Disorder-7 Items (GAD-7) or the PND-specific EPDS (EPDS is the most frequently used depression screening tool in clinical practice). A patient can be referred to a GP for diagnosis and treatment, or – in cases, where a severe mental health problem is suspected – the patient can be referred to a mental health professional in a secondary mental health service setting (Figure 4) (National Institute for Health and Care Excellence 2020). In the most severe cases (severe PND or severe anxiety disorders, postpartum psychosis and other psychotic illnesses), patients may require urgent attention due to the patient or the baby being at risk (Royal College of Psychiatrists 2018). In these cases, women will usually be admitted into mother and baby units (MBU). It is estimated that approximately 3-5% of PND patients will require admission to an MBU (Royal College of Psychiatrists 2021).

For PND, NICE recommends following the diagnostic criteria as used for depression in adults (Diagnostic and Statistical Manual of Mental Disorders-5 [DSM-5] and ICD-11) (National Institute for Health and Care Excellence 2023a). In DSM-5, the patient must have ≥5 symptoms that are present during the same 2-week period, with ≥1 of those symptoms being either depressed mood or loss of interest/pleasure (American College of Obstetricians and Gynecologists (ACOG) , American Psychiatric Association (APA)). These symptoms must cause clinically significant distress or impairment in social, occupational, or other important areas of functioning. For more detail, please see Appendix J.2 PND diagnostic criteria.

Figure 4: Recognising mental health problems in pregnancy and the postnatal period and referral – depression and anxiety disorders



Source: (National Institute for Health and Care Excellence 2020)
Abbreviations: EPDS: Edinburgh Postnatal Depression Scale; GAD-2/7: Generalised Anxiety Disorder 2/7-item; GP: General Practitioner; PHQ-9: Patient Health Questionnaire-9.

Use of appropriate tools, onset of PND symptoms and the duration and impact of these symptoms on daily functioning of affected patients can assist clinicians in making a specific, timely and appropriate diagnosis (Carlson 2024). The above-mentioned tools are summarised in Table 4 below. There are several additional tools, which are used primarily in the research/clinical trial setting; instruments such as the HAMD-17, MADRS, and Clinical Global Impression (CGI) scale are frequently used to assess efficacy of PND interventions (Table 4) (Maurer 2012, Siu 2016, Food and Drug Administration (FDA) June 2018).

Table 4: Assessment tools commonly used to assess mental health disorders

Tool	Description	Interpretation	Use in UK clinical practice/research	Additional information	References
EPDS*	• The EPDS is a 10-item questionnaire that screens for depression and anxiety in patients who are pregnant or have recently given birth	• Range: 0−30 • Scoring above a threshold of 12/13 suggests presence of depressive illness	• Use in UK clinical practice	• EPDS is the most used depression screening tool in perinatal care • EPDS is included in the NICE Guidelines as part of a full assessment for a suspected mental health problem in pregnancy or the postnatal period	(Levis 2020, National Institute for Health and Care Excellence 2020)
GAD-2/7*	• The GAD-2 is an ultra-quick version of the seven-item scale that incorporates the first two questions of the GAD-7 • The GAD-7 is a 7-item scale reflecting how often the patient has been bothered by seven different symptoms of anxiety during the last two weeks with scores ranging from 0 to 3 on all the questions	• Range: GAD-2:0-6, GAD-7:0-21 • The GAD-2/7 measures how often a patient has been bothered by two/seven different anxiety symptoms during the last two weeks with scores ranging from 0 to 3 on each question • For the GAD-7 scores of 5, 10, and 15 are taken as cut off points for mild, moderate, and severe anxiety, respectively	• Use in UK clinical practice	• The GAD-2 questionnaire has been validated in multiple studies and shown to retain the excellent psychometric properties of the GAD-7. As such both are used in primary care patients and by general practitioners	(Sapra 2020)
PHQ-9*	• The PHQ-9 is a self-administered screening tool that assesses the severity of depression • It is a nine-item scale that uses a scoring system of 0–3 to reflect how often a symptom has been	• Range: 0−27 • Scores of 5, 10, 15, and 20 represent mild, moderate, moderately severe, and severe depression, respectively	• Use in UK clinical practice	• PHQ-9 is a common tool used to measure depression severity • PHQ-9 is included in the NICE Guidelines as part of a full assessment for a suspected mental health problem in	(Kroenke 2001, National Institute for Health and Care Excellence 2020)

	experienced in the past week			pregnancy or the postnatal period	
HAMD-17	• HAMD-17 is a questionnaire that clinicians use to assess the severity of depression in patients • The scale consists of 17 items that are scored on a scale of 0 to 52	• Range: 0–52 • Higher scores indicate severe depression	• Use in research only	• HAMD-17 is one of the most widely used tools to measure the effectiveness of antidepressant medication in clinical trials	(Hamilton 1960, Gerbasi 2020)
MADRS	• The MADRS is a 10-item questionnaire that assesses the severity of depression in adults	• Range: 0–60 • Higher scores indicate severe depression	• Use in research only	• MADRS is commonly used in clinical and research settings to assess depression symptoms	(Quilty 2013, Hobden 2017)
CGI-I	• 7-point clinician-rated scale used to compare the patient's overall clinical condition to the 1-week period prior to initiation of medication	• 1–very much improved • 2–much improved • 3–minimally improved • 4–no change from baseline • 5–minimally worse • 6–much worse • 7–very much worse	• Use in research only	• The CGI-I/S are one of the most widely used assessment tools in mental health trials	(Busner 2007, Jones 2019)
CGI-S	• The CGI-S is a 7-point clinician-rated scale used to assess the severity of a patient's mental illness	• 1–normal, not at all ill • 2–borderline mentally ill • 3–mildly ill • 4–moderately ill • 5–markedly ill • 6–severely ill • 7–among the most extremely ill			

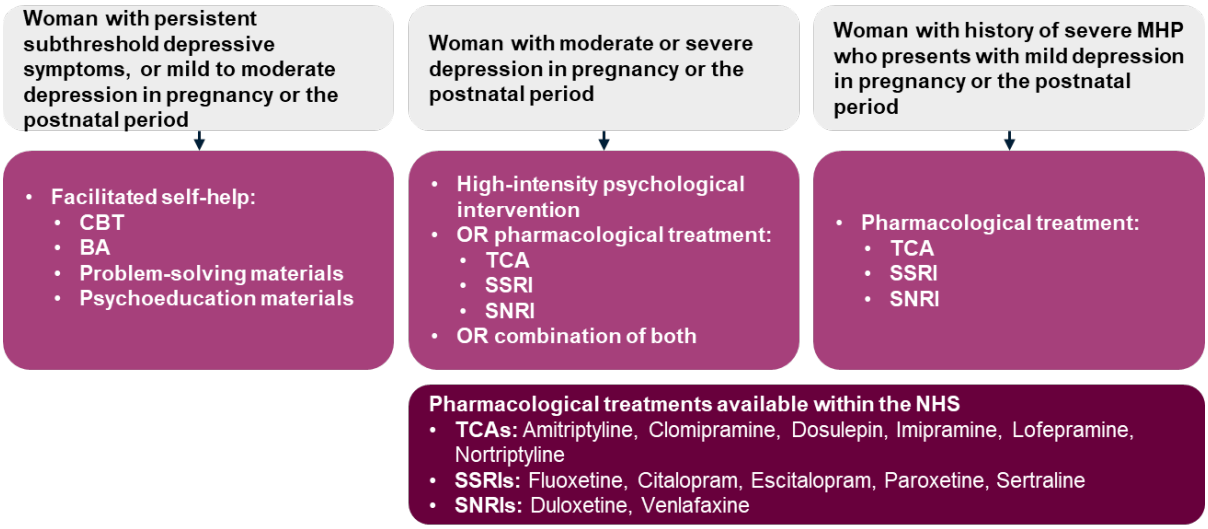
Source: See table
*Recommended for usage in NICE Guideline CG192 for recognising mental health problems in the perinatal period.
Abbreviations: CGI-I: Clinical Global Impression of Improvement scale; CGI-S: Clinical Global Impression of Severity scale; EPDS: Edinburgh Postnatal Depression Scale; GAD-2/7: Generalised Anxiety Disorder 2/7-item; HAMD-17: 17-item Hamilton Rating Scale for Depression; MADRS: Montgomery-Åsberg Depression Rating Scale; NICE: National Institute for Health and Care Excellence; PHQ-9: 9-item Patient Health Questionnaire; UK: United Kingdom.

1.3.4.2 Treatment options across the PND spectrum

Within two weeks of a referral of a patient with a known or suspected mental health problem in the postnatal period, NICE recommends an assessment of potential treatment options, with the aim to provide interventions within one month of the initial assessment (National Institute for Health and Care Excellence 2020).

Currently, in the UK, there are no licensed pharmacological treatments indicated specifically for PND. As per the NICE clinical guidelines, the choice of treatment for PND depends on symptom severity and mental health history of the woman. Treatment options are limited to established clinical management (ECM), which consists of psychological interventions with or without pharmacological treatment with off-label ADTs used in management of MDD (National Institute for Health and Care Excellence 2020). The available treatment options recommended by NICE for management of PND across the disease spectrum are presented in Figure 5 below.

Figure 5: Treatment options across the PND spectrum



Source: (NHS England 2021, National Institute for Health and Care Excellence 2022)
Abbreviations: BA: Behavioural Activation; CBT: Cognitive Behavioural Therapy; MHP: Mental Health Problems; NHS: National Health Service; PND: Postnatal Depression; SNRI: Selective Norepinephrine Reuptake Inhibitor; SSRI: Selective Serotonin Reuptake Inhibitor; TCA: Tricyclic Antidepressant.

1.3.4.3 Proposed clinical pathway of care

Zuranolone is anticipated to be the first approved treatment for PND in the UK.

Zuranolone is expected to fit into the existing clinical pathway as a treatment option to be used alone or as an adjunct to therapeutic alternatives used as part of ECM. As with any novel treatment options, Biogen expects it will initially be mainly prescribed by perinatal psychiatrists within the community perinatal mental health services.

Subject to regulatory approval and following a positive reimbursement decision, zuranolone will be used for treatment of adults with PND, following childbirth.

1.3.5 Unmet needs in PND

1.3.5.1 Lack of Approved Treatment Options Indicated for PND

As described in Section 1.3.4, as many patients never receive formal diagnosis of PND, the majority of those suffering from PND never receive treatment. For those who do receive formal diagnosis, there are currently no oral ADTs with a marketing authorisation for PND in the UK, and there remains a substantial unmet need for pharmacological treatment options indicated for PND (Frieder 2019). Oral ADTs such as selective serotonin reuptake inhibitors (SSRIs), serotonin noradrenaline reuptake inhibitors (SNRIs) and tricyclic antidepressants (TCAs), which are recommended for use in PND by the NHS (NHS England 2021) and NICE guidelines (National Institute for Health and Care Excellence 2022) are being prescribed as off-label treatments. The evidence on efficacy and safety of ADTs in management of PND is limited. A systematic review by Brown et al., analysed data from 11 randomised controlled trials (RCTs) (n=1,016 women) assessing the effectiveness and safety of ADTs in comparison with any other treatment (psychological, psychosocial, pharmacological), placebo, or usual treatment for PND. The meta-analysis showed that that there may be a benefit of SSRIs over placebo in response, remission, and reduction in depressive symptoms, however, the available evidence on the effectiveness and safety of SSRIs for PND was limited. Additionally, the certainty of the evidence for SSRIs was very low due to the inclusion of only 11 studies, and there was no evidence of a difference in acceptability between SSRIs and placebo across four studies (Brown 2021).

1.3.5.2 Efficacy and safety limitations of off-label ADTs

One of the key limitations of the off-label use of treatments in PND, is the slow-acting nature of ADTs, which may take several months for treatment to become effective; e.g., an initial treatment with SSRIs requires on average at least 4 weeks to attain response, and 6 weeks to attain remission (Leuchter 2009), but in several cases can take up to 12 weeks or longer (Trivedi 2006). Additionally, it has been reported that up to 40% of patients do not respond to sequential trials of standard ADTs (Thase 2023). Considering the urgency for symptom relief in patients with PND, the lack of consistent and timely treatment response to oral ADTs emphasises a substantial unmet need in this patient population.

ADTs currently used in clinical practice may also cause a range of side effects including nausea, gastrointestinal issues, insomnia, anxiety, and other side effects (Table 5). This is further exacerbated by the need to take ADTs long-term, i.e., at least 6 months.

Table 5: Common side effects of ADTs by ADT class

Class of ADTs	Side Effects	
SSRIs	• Nausea • Headaches • Gastrointestinal issues • Insomnia • Fatigue • Weight gain • Anxiety • Dizziness	• Dry mouth • Sexual side effects • Increased risk of falls and fractures • Rare but potentially fatal arrythmia • Coagulopathy • Increased risk of bleeding (when used with pain reliefs or blood thinners) • Emotional blunting/indifference
SNRIs	• Nausea and vomiting • Dizziness • Sweating • Fatigue	• Insomnia • Headache • Sexual dysfunction • Emotional blunting/indifference
TCAs	• Dry mouth • Blurry vision • Constipation • Urinary retention	• Weight gain • Sedation • In case of overdose can lead to cardiac arrest

Source: (NHS England 2021, National Institute for Health and Care Excellence 2022).
SSRI: (Richards 2007, 2016, Cosci 2019, Salahudeen 2020, Bodkin 2021, Suchting 2021).
SNRI: (Sansone 2010, Grady 2012, Cosci 2019, Ma 2021).
TCA: (Montano 2023) (Santarsieri 2015).
Abbreviations: ADT: Antidepressant Treatment; SNRI: Selective-Norepinephrine Reuptake Inhibitor; SSRI: Selective Serotonin Reuptake Inhibitor; TCA: Tricyclic Antidepressant.

Due to the lack of rapid response and poor tolerability profile of ADTs, patients frequently do not adhere to treatment (low or non-adherence of ADTs ranging from 46% to 83%) or stop treatment or avoid it completely (Niarchou 2024). In a recent retrospective analysis of Texas Medicaid claims (January 2018 to June 2022), ADT adherence and persistence rates among PND patients were found to be even lower. Among the 15,667 included patients (women aged 12-55 who received an ADT within 90 days after delivery), the adherence rate (proportion of days covered ≥80%) was only 15.9%, and after 270 days of follow-up the mean (SD) adherence was only 43.9 (29.5) days (Pennington 2025).

Low adherence rates are of particular concern as adherence has been proven to be a prerequisite to achieving beneficial treatment outcomes, and failure to comply with medication regimes has been associated with negative consequences, not only on an individual level by jeopardising patients' health but also on the wider community level, amplifying the healthcare costs caused by inadequate disease management (Niarchou 2024). Patients on ADT treatment are at risk of suffering from withdrawal effects, which are common, affecting 27% to 86% of ADT-treated patients (Davies 2019). Symptoms include (but are not limited to) tremors,

sweating, tachycardia, neuralgia, sleep disturbances, worsening anxiety, and mood instability (Carvalho 2016, Cosci 2019), which can last for several months (Taylor 2019).

Zuranolone can address some of the limitations associated with current ADTs. It is a well-tolerated, 14-day treatment with a rapid onset of action (Section 2.6). Short treatment duration could also translate into benefits in form of reduced adverse events experienced by the patient (due to much shorter treatment exposure compared to ADTs) as well as potentially increased treatment adherence (by reducing the need for continued long-term treatment often seen with ADTs) (Kaufman 2022). Due to the rapid symptom resolution (significant improvements observed as early as Day 3 after treatment initiation, see Table 14), treatment with zuranolone is likely to prevent the escalation of symptoms and mitigate the potential long-term consequences for patients and infants compared to established clinical management (ECM) (Netsi 2018, Morgan 2021, Terrone 2023). A new treatment indicated for PND may also help raise awareness of the condition and decrease stigma that may lead to improved detection of the condition.

1.4 Equality considerations

Women's health has historically been an under-resourced and underserved area of healthcare. Concerns have been identified that services for conditions, which only or primarily affect women, are perceived to be of lower priority (Department of Health & Social Care 2022). Although significant progress has been made in expanding specialist PMH provision in recent years, there are still inequalities in the access to high-quality PMH care (Maternal Mental Health Alliance 2023). In the UK, ethnic minority groups have a higher burden of common mental health disorders and women from ethnic minority backgrounds are at greater risk of developing PND when compared to the majority white population (Du Preez 2016, Watson 2019); in particular, Black women are 13% more likely to develop PND than any other race (Impact on Urban Health). Additionally, certain communities, particularly those facing multiple disadvantages and systemic inequalities, encounter additional obstacles in obtaining quality care (Maternal Mental Health Alliance 2023). For example, a 2017 analysis of national commissioning data repository (NCDR) data in the UK found that during the perinatal period, patients from ethnic minorities had significantly lower access to community mental health services and experienced higher percentages of involuntary admissions compared to White British patients (Jankovic 2020); Black patients are the group least likely to initiate treatment and receive follow-up for postnatal mental illness (NHS Devon Partnership 2021). Additionally, younger women may experience heightened stigma and fear of judgment, which can lead to reluctance in seeking help making them particularly vulnerable to PND (Schmidt 2006).

Biogen would like to highlight that Perinatal Care Pathways are complex (see Appendix J.1) and diagnosing patients with PND can be challenging. The mapping of specialist PMH services across England demonstrates that access to these services remains a “postcode lottery”, with many women struggling to access necessary care. Further, only 16% of Specialist PMH Community Teams in England have met perinatal quality network (PQN) Standards Type 1 and have begun delivering on all four main ambitions of the NHS Long Term Plan (Maternal Mental Health Alliance 2023).

The health system needs to continually address these ongoing inequalities to ensure zuranolone is accessible to all patients who may benefit from treatment.

2 Clinical effectiveness

Key points:

- SKYLARK was a randomised, double-blind, placebo-controlled study assessing clinical efficacy and safety of zuranolone (50mg) in PND; the study included a screening period of up to 28 days, a 14-day treatment course, and follow-up through Day 45 (Section 2.3.1)
- SKYLARK met the primary endpoint (HAMD-17 score), demonstrating that compared with placebo, treatment with zuranolone resulted in statistically significant improvements in depressive symptoms at Day 15, with statistically significant improvements observed as early as Day 3, demonstrating the rapid onset of action of zuranolone, and maintained through to Day 45, supporting the durability of the clinical benefits of zuranolone treatment (Section 2.6.1)
- The improvements observed in the investigator-reported HAMD-17 were mirrored by improvements in patients' self-reported assessment of their depressive symptoms as demonstrated by the EPDS (Section 2.6.1)
- Across secondary endpoints, (HAMD-17 at all timepoints, Clinical Global Impression - Severity [CGI-S], Clinical Global Impression - Improvement [CGI-I] response, HAM-A total score, MADRS total score, EPDS total score and PHQ-9), compared with placebo, treatment with zuranolone resulted in statistically significant improvements in depressive and anxiety symptoms (Section 2.6.1)
- The findings were broadly consistent across pre-specified subgroups, supporting efficacy of zuranolone in PND patients with differing demographic and other baseline characteristics (Section 2.8.1)
- As demonstrated by post-hoc analyses, zuranolone treatment offered clinically meaningful patient-level change improvements and group-level difference improvements in HAMD-17 score (Section 2.6.1), and [redacted] (Section 2.8.2.2)
- Zuranolone was generally well tolerated (Section 2.11.1)
- The SKYLARK study protocol allowed background use of ADTs and psychological therapies, therefore, clinical evidence from SKYLARK represents zuranolone with ECM versus ECM alone as outlined in the NICE decision problem (Section 2.3.2.1)

- Results of the ROBIN study, which assessed a different dose of zuranolone, are in line with SKYLARK results and provide supportive evidence on clinical efficacy and safety or zuranolone (Section 2.6.2)

2.1 Identification and selection of relevant studies

An SLR was undertaken to identify clinical trials relevant to the NICE decision problem. The SLR identified RCTs in adults (≥18 years of age) carrying a child through pregnancy to birth, and with a diagnosis of PND. Trials assessing pharmaceuticals in a list of prespecified treatment classes, and pre-specified non-pharmacological treatments were eligible for inclusion. Eligible trials compared any eligible intervention with another eligible intervention, or with any psychological intervention or placebo. Eligible outcomes included depression rating scores, anxiety rating scores, CGI-I, CGI-S, Barkin Index of Maternal Functioning (BIMF), PHQ, Columbia-Suicide Severity Rating Scale (C-SSRS), SF-36, adverse events (AEs), serious AEs, specific AEs, mortality, and withdrawals due to discontinuation. The literature search was conducted on 20 November 2023 and updated on 03 October 2024 to ensure all evidence from database inception until 03 October 2024 was included. The full search strategy and details of the process and methods used to identify and select the clinical evidence relevant to the technology being appraised are summarised in Appendix B. In total, the literature review identified 21 studies (reported across 65 publications).

Two studies were identified that included zuranolone, these were the SKYLARK and ROBIN studies (Deligiannidis 2021, Deligiannidis 2023b). A secondary publication of the SKYLARK study was also identified, but was not included for data extraction as it did not report novel outcomes relevant for the decision problem (Meltzer-Brody 2023). Additionally, the review identified 19 trials evaluating non-approved pharmaceutical interventions (neurosteroids, SSRIs, tricyclics) or psychological therapies (alone or in combination with pharmaceuticals) (Appleby 1997, Cooper 2003, Misri 2004, Wisner 2006, Yonkers 2008, Sharp 2010, Bloch 2012, Myczkowski 2012, Chibanda 2014, Hantsoo 2014, Milgrom 2015, Wisner 2015, Kanes 2017, Marinus Pharmaceuticals 2017, Marinus Pharmaceuticals 2018, Meltzer-Brody 2018, O'Hara 2019, Jiang 2024). Note, one primary study Meltzer-Brody 2018a and Meltzer-Brody 2018b were reported in the same publication (Meltzer-Brody 2018) (Appendix B.4.1).

The results of SKYLARK and ROBIN evaluating zuranolone are discussed in Section 2.6. A feasibility assessment was conducted to assess whether an indirect treatment comparison (ITC) was feasible (see Section 2.10). Details of the studies assessing other pharmacological interventions are summarised in Appendix B.4.

2.2 List of relevant clinical effectiveness evidence

SKYLARK (n=200) and ROBIN (n=153) studies form the key evidence base for the efficacy, safety and tolerability of zuranolone in PND. In line with the guideline on clinical investigation of medicinal products in the treatment of depression from the European Medicines Agency (European Medicines Agency 2023), the SKYLARK and ROBIN studies did not recruit mild patients as it is difficult to demonstrate an effect in this patient population. Biogen acknowledges that the clinical evidence package for zuranolone provides evidence on efficacy of zuranolone in patients with moderate to severe PND (when classified by the baseline HAMD-17 score [Section 2.6] and MADRS score [Section 2.8.2]). Nevertheless, the unique MoA of zuranolone supports its efficacy in all PND patients, especially considering the high disease burden and lack of therapies licensed specifically for PND (Section 1.3.3 and Section 1.3.5). Therefore, Biogen believes that SKYLARK and ROBIN provide clinical evidence on the population relevant to the decision problem (Section 1.1).

SKYLARK has been included in the economic model and forms the basis of the evidence supporting zuranolone in PND, as the studied dose is aligned with the anticipated marketing authorisation of zuranolone. ROBIN, which tested another capsule formulation of a lower dose of 30 mg (not the intended licenced dose), provides supportive evidence on the clinical efficacy and safety or zuranolone. A summary of the clinical evidence for zuranolone is presented below in Table 6.

Table 6: Clinical effectiveness evidence

Study	SKYLARK (NCT04442503)	ROBIN (NCT02978326)
Study design	• Randomised, double-blinded, placebo-controlled study	• Randomised, double-blinded, placebo-controlled study
Population	• Women aged 18 to 45 years who have had MDE that began during the third trimester of pregnancy or ≤4 weeks postnatal • Baseline HAMD-17 total score ≥26 • Concomitant use of ADT was permitted if patients were on a stable dose for ≥30 days prior to first study treatment dose • Participants receiving psychotherapy on a regular schedule for at least 30 days prior to Day 1 are permitted if the participant intends to continue the regular schedule through the follow-up period (Day 45)	• Women aged 18 to 45 years who were ≤6 months postnatal • had an MDE without psychosis (per DSM-5) that began no earlier than the third trimester or ≤4 weeks postnatal • Baseline HAMD-17 total score ≥26 • Concomitant use of ADT was permitted if patients were on a stable dose for ≥30 days prior to first study treatment dose
Intervention(s)	• Zuranolone 50 mg (two 25 mg capsules) orally, once daily in the evenings with fat-containing food for 14 days	• Zuranolone 30 mg (another capsule formulation with a higher relative bioavailability) orally, once daily in the evening with fat-containing food for 14 days

Study	SKYLARK (NCT04442503)	ROBIN (NCT02978326)
Comparator(s)	• Placebo	• Placebo
Indicate if study supports application for marketing authorisation	• Yes	• Yes
Indicate if study used in the economic model	• Yes	• Yes
Rationale if study not used in model	N/A	N/A
Reported outcomes specified in the decision problem (bold text indicate outcomes incorporated into the economic model)	• Depressive symptoms including relapse and remission rates (e.g., HAMD-17 remission) • Severity of depression (e.g., **EPDS**, MADRS, PHQ-9, CGI-S) • Anxiety (e.g., **EPDS**, HAM-A) • Sleep quality (e.g., MADRS – Reduced Sleep, HAMD-17 – Insomnia) • Adverse effects of treatment including adverse effects of treatment discontinuation (e.g., **adverse events**)	• Depressive symptoms including relapse and remission rates (e.g., HAMD-17 remission) • Severity of depression (e.g., EPDS, MADRS, PHQ-9, CGI-S) • Anxiety (e.g., EPDS, HAM-A) • Sleep quality (e.g., MADRS – Reduced Sleep, HAMD-17 – Insomnia) • Adverse effects of treatment including adverse effects of treatment discontinuation (e.g., adverse events) • Health-related quality of life (e.g., SF-36)
All other reported outcomes	N/A	N/A

Source: (Clinical Trials GOV 2023b, Clinical Trials GOV 2023a)
Abbreviations: ADT: Antidepressant Therapy; CGI-S: Clinical Global Impression of Severity Scale; DSM-5: Diagnostic and Statistical Manual of Mental Disorders, 5th Edition; EPDS: Edinburgh Postnatal Depression Scale; HAM-A: Hamilton Rating Scale for Anxiety; HAMD-17: 17-item Hamilton Rating Scale for Depression; MADRS: Montgomery-Åsberg Depression Rating Scale; MDE: Major Depressive Episode; N/A: Not/Applicable; PHQ-9: 9-item Patient Health Questionnaire; SF-36: Short-Form 36 Items.

2.3 Summary of methodology of the relevant clinical effectiveness evidence

SKYLARK and ROBIN were randomised, double-blinded studies, which compared the efficacy, safety and tolerability of zuranolone versus placebo in treatment of PND (Table 7). The clinical evidence from SKLYARK used in the economic model represents zuranolone with ECM versus ECM alone, as the trial protocol allowed background use of a stable dose of ADTs and psychological therapies, supporting the proposed positioning of zuranolone as treatment of PND, which can be used alone or as an adjunct to ECM. This is aligned with the usual care received by patients with PND in the UK.

Table 7: Methodology in clinical trials of zuranolone

Trial	SKYLARK	ROBIN
Trial design	Randomised, double-blinded, placebo-controlled, study	Randomised, double-blinded, placebo-controlled study
Key eligibility criteria for participants	*Inclusion* • Women aged 18 to 45 years • Had an MDE that began during the third trimester of pregnancy or ≤4 weeks postnatal • ≤12 months postnatal • Baseline HAMD-17 total score ≥26 • Ceased lactation or agreed to not provide breastmilk to an infant from Day 1 until 7 days after last dose *Exclusion* • Active psychosis or history of bipolar disorder, psychotic disorders, nonfebrile seizures, schizophrenia and/or schizoaffective disorder • Attempted suicide or risk of suicide associated with the current episode of PND • Recent history or active clinically significant manifestations of metabolic, hepatic, renal, haematological, pulmonary, cardiovascular, gastrointestinal, musculoskeletal, dermatological, urogenital, or neurological disorders; eyes, ears, nose, and throat disorders • BMI ≤18 or ≥45 kg/m^2 • Clinically significant abnormal 12-lead electrocardiogram at Screening or Day 1 • Index pregnancy resulting in miscarriage, still birth, or neonatal/infant death, or terminated parental rights • History of mild, moderate, or severe substance use disorder (including benzodiazepines) diagnosed using DSM-5 criteria in the 12 months prior to Screening • History of sleep apnoea, gastric bypass surgery, gastric sleeve or lap band, or any procedures that interfere with gastrointestinal transit	*Inclusion* • Women aged 18 to 45 years who were ≤6 months postnatal • Had an MDE without psychosis (per DSM-5) that began during the third trimester of pregnancy or ≤4 weeks postnatal • Baseline HAMD-17 total score ≥26 • Ceased lactation or agreed to not provide breastmilk to an infant from Day 1 until 7 days after last dose *Exclusion* • Recent history or active clinically significant manifestations of metabolic, hepatic, renal, haematological, pulmonary, cardiovascular, gastrointestinal, musculoskeletal, dermatological, urogenital, or neurological disorders; eyes, ears, nose, and throat disorders • Known allergy to zuranolone capsule or its excipients • Active psychosis per investigator assessment • Attempted suicide with the current episode of PND • Medical history of seizures • Medical history of bipolar disorder, schizophrenia and/or schizoaffective disorder • History of active alcoholism or drug addiction (including benzodiazepines) in the 12 months prior to screening • For disallowed medications, see 'Trial drugs- disallowed concomitant medications interventions' below

	• Had vagus nerve stimulation, electroconvulsive therapy, or had taken ketamine with the current PND episode • For disallowed medications, see 'Trial drugs – permitted and disallowed concomitant medications/ interventions' below	
Settings and locations where the data were collected	78 sites in the United States, 9 sites in Spain, and 5 sites in the United Kingdom	33 centres in the United States
Trial drugs - Interventions and comparators	Patients were randomised 1:1 to receive zuranolone 50 mg or matching placebo, to be self-administered once daily in the evenings with fat-containing food for 14 days	Patients were randomised 1:1 to receive zuranolone 30 mg (another capsule formulation with a higher relative bioavailability) or matching placebo once daily in the evenings with food for 14 days
Trial drugs - permitted and disallowed concomitant medication/ interventions	Permitted treatments: • Concomitant use of ADT was permitted if patients were on a stable dose for ≥30 days prior to first study treatment dose (Day 1) • Participants receiving psychotherapy on a regular schedule for at least 30 days prior to Day 1 are permitted if the participant intends to continue the regular schedule through the follow-up period (Day 45). Disallowed medication: • Initiation of new psychotropic medications, including antidepressant or antianxiety medications, and any new pharmacotherapy regimen that potentially could have had an impact on efficacy and/or safety endpoints, within 30 days prior to Day 1 through completion of the Day 45 Visit • Use of any benzodiazepines, barbiturates, GABAA modulators (e.g., eszopiclone, zopiclone, zaleplon, and zolpidem) within 28 days prior to Day 1 through completion of the Day 45 Visit (within 30 days prior to Day 1 through completion of the Day 45 Visit for benzodiazepine or GABA modulators with a half-life ≥48 hours [e.g., diazepam]) • First-generation (typical) antipsychotics (e.g., haloperidol, perphenazine) and second generation (atypical) antipsychotics (e.g., aripiprazole, quetiapine) within 14 days prior to Day 1 through completion of the Day 45 Visit	Permitted medication: • Patients who were taking concomitant psychotropic medication were required to be on a stable dose of medication for ≥30 days prior to first study treatment dose (Day 1) Disallowed medication: • Initiation of new psychotropic medications used to treat depressive symptoms, such as antidepressants, atypical antipsychotics, etc, was prohibited within 30 days prior to Day 1 and until completion of the Day 15 assessments. Atypical antipsychotics for treatment of psychotic symptoms were prohibited • Benzodiazepines were to be avoided as much as possible. Eligible subjects taking a stable dose of benzodiazepine at least 30 days prior to Day 1 were allowed on a case-by-case basis following discussion between the investigator and Sage. Subjects may have been permitted to continue to take their current dose of the benzodiazepine (to prevent acute withdrawal), but no new benzodiazepine use was permitted during the study • The use of hypnotics for sleep/insomnia such as Ambien® and trazodone was to be avoided; use of hypnotics was allowed on a case-by-case basis following discussion between the investigator and Sage • Anticonvulsants were prohibited

	• Use of any non-GABA anti-insomnia medications (e.g., melatonin, Benadryl [antihistamines], trazodone) within 14 days prior to Day 1 through completion of the Day 45 Visit • Exposure to another investigational medication or device from 30 days prior to Screening through completion of the Day 45 Visit • Any known strong inhibitors of CYP3A4 within 14 days or 5 half-lives (whichever was longer) within 14 days prior to receiving the first dose of IP through the 14-day Treatment Period • Use of any strong CYP3A inducer, such as rifampin, carbamazepine, enzalutamide, mitotane, phenytoin, or St John's Wort, within 14 days or 5 half-lives (whichever was longer) prior to receiving the first dose of IP through the 14-day Treatment Period • Use of any chronic or as-needed psychostimulants (e.g., methylphenidate, amphetamine) or opioids from within 28 days prior to Day 1 through completion of the Day 45 Visit	• Any known strong inhibitors of CYP3A4 were prohibited within 14 days or 5 half-lives (whichever is longer) prior to receiving the first dose of study drug and throughout the study • Any CYP inducers, such as rifampin, carbamazepine, ritonavir, enzalutamide, efavirenz, nevirapine, phenytoin, phenobarbital or St John's Wort, were prohibited within 14 days or 5 half-lives (whichever is longer) prior to the first dose of study drug and throughout the study
Primary outcomes (including scoring methods and timings of assessments)	• Change from baseline in HAMD-17 total score at Day 15	• Change from baseline in HAMD-17 total score at Day 15
Other outcomes used in the economic model/specified in the scope	• Depressive symptoms including relapse and remission rates (e.g., HAMD-17 remission) • Severity of depression (e.g., EPDS, MADRS, PHQ-9, CGI-S) • Anxiety (e.g., EPDS, HAM-A) • Sleep quality (e.g., MADRS – Reduced Sleep) • Adverse effects of treatment including adverse effects of treatment discontinuation (e.g., adverse events)	• Depressive symptoms including relapse and remission rates (e.g., HAMD-17 remission) • Severity of depression (e.g., EPDS, MADRS, PHQ-9, CGI-S) • Anxiety (e.g., EPDS, HAM-A) • Sleep quality (e.g., MADRS – Reduced Sleep) • Adverse effects of treatment including adverse effects of treatment discontinuation (e.g., adverse events) • Quality of life (SF-36)

Pre-planned subgroups	• Demographic and other baseline characteristics:
	  • Race (White, Black/African American; Other)
	  • Age (18 to 24; 25 to 45)
	  • Country of Origin (United States, Rest of World)
	  • Antidepressant Use at Treatment Start (Yes; No)
		• Demographic and other baseline characteristics:
		  • Race (White, Black/African American; Other)
		  • Age (18 to 24; 25 to 45)
		  • Country of Origin (United States, Rest of World)
		  • Antidepressant Use at Treatment Start (Yes; No)

	• Baseline BMI category (≤18.4, 18.5 to 24.9, 25 to 29.9, ≥ 30 kg/m^2) • Onset of PND (Third Trimester, within 4 Weeks of Delivery) • Family History of PND (Yes; No) • Baseline HAMD-17 Total Score Category • Depression with Elevated Anxiety by Baseline HAM-A Total Score	• Baseline BMI category (≤18.4, 18.5 to 24.9, 25 to 29.9, ≥ 30 kg/m2) • Onset of PND (Third Trimester, within 4 Weeks of Delivery) • Family History of PND (Yes; No)
Post-hoc subgroups	• Severity of PND	N/A

Source: (Deligiannidis 2021, Deligiannidis 2023a)
For a brief description of the scales used to assess the outcomes of patients in the SKYLARK and the ROBIN studies, please see Table 4
For a full list of inclusion and exclusion criteria for SKYLARK and ROBIN, please see Appendix B.6.1 and B.6.2, respectively.
Abbreviations: ADT: Antidepressant Therapy; BMI: Body Mass Index; CGI-S: Clinical Global Impression of Severity Scale; DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition; CYP: Cytochrome P45; EPDS: Edinburgh Postnatal Depression Scale; GABAA: γ-aminobutyric Acid Type-A; HAM-A: Hamilton Anxiety Rating Scale; HAMD-17: 17-item Hamilton Rating Scale for Depression; IP: Investigational Product; MADRS: Montgomery–Åsberg Depression Rating Scale; MDE: Major Depressive Episode; N/A: Not Applicable; PHQ-9: Patient Health Questionnaire-9; PND: Postnatal Depression; SF-36: 36-item Short Form Health Survey.

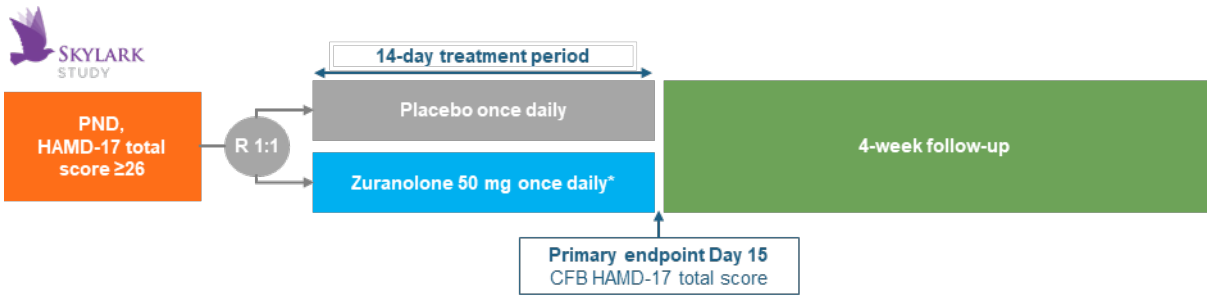
2.3.1 Trial design

The methodologies of each study are described above in Table 7.

2.3.1.1 SKYLARK

SKYLARK was a randomised, double-blinded, placebo-controlled study to evaluate the efficacy and safety of zuranolone 50 mg compared to placebo in adult women with PND (HAMD-17 total score ≥26), consisting of a screening period of up to 28 days, a 14-day double-blinded treatment period, and a follow-up period through Day 45 (Figure 6). SKYLARK included adult women who had a major depressive episode (MDE) that began during the third trimester of pregnancy or less than 4 weeks postnatal. Baseline HAMD-17 total score was ≥26 and concomitant use of ADT was permitted if patients were on a stable dose for ≥30 days (Deligiannidis 2023a). A summary of key eligibility criteria in presented in Table 7. For a full list of inclusion and exclusion criteria for SKYLARK, please see Appendix B.6.1.

Figure 6: SKYLARK study design

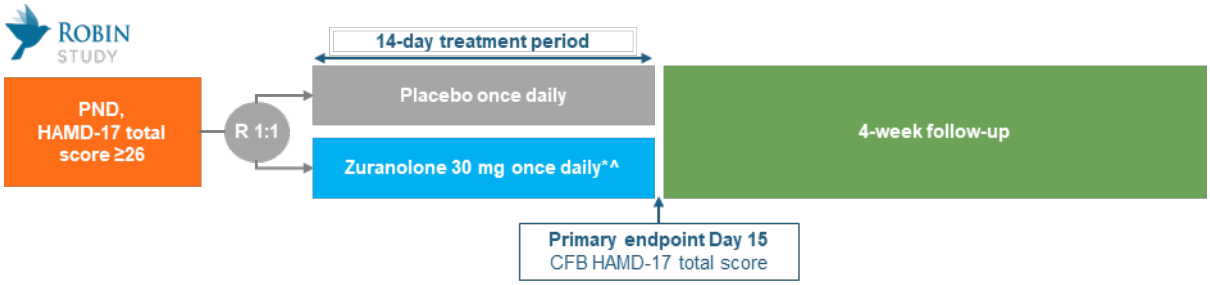


*Zuranolone 50 mg and placebo administered in the evening with fat-containing food; The dose could be reduced to 40 mg as needed based on tolerability
Abbreviations: CFB: Change from Baseline; HAMD-17: 17-item Hamilton Rating Scale for Depression; PND: Postnatal Depression.

2.3.1.2 ROBIN

ROBIN was a randomised, double-blinded, placebo-controlled study to evaluate the efficacy and safety of zuranolone 30 mg (another capsule formulation with a higher relative bioavailability) compared to placebo in the treatment of adult women with PND (HAMD-17 total score ≥26). ROBIN study consisted of a 28 day screening period, a 14-day treatment period, and a follow-up period through Day 45 (Figure 7). ROBIN included adult women who had MDE without psychosis (per DSM-5) that began no earlier than the third trimester of pregnancy or less than 4 weeks postnatal. Baseline HAMD-17 total score was ≥26 and concomitant use of ADT was permitted if patients were on a stable dose for ≥30 days (Deligiannidis 2021). A summary of key eligibility criteria in presented in Table 7. For a full list of inclusion and exclusion criteria for ROBIN, please see B.6.2.

Figure 7: ROBIN study design



*Down-titration to zuranolone 20 mg allowed if 30 mg not tolerated. Zuranolone 30 mg and placebo were administered in the evening with food; ^Patients in the ROBIN study received another zuranolone capsule formulation with a higher relative bioavailability.
Abbreviations: CFB: Change from Baseline; HAMD-17: 17-item Hamilton Rating Scale for Depression; PND: Postnatal Depression.

2.3.2 Trial drugs and concomitant medications

2.3.2.1 SKYLARK

Patients were randomised 1:1 to receive zuranolone or matching placebo. Patients self-administered zuranolone 50 mg (two 25 mg capsules) or matching placebo once daily in the evenings with fat-containing food for 14 days; concomitant use of ADT was permitted if patients were on a stable dose for ≥30 days prior to first study treatment dose, use of non-pharmacological interventions (i.e., psychotherapy) was also permitted, which is aligned with ECM. Patients unable to tolerate zuranolone 50 mg were permitted to decrease their dosage to 40 mg (two 20 mg capsules) for the remainder of the treatment course. A total of 200 patients were randomised, 196 received study intervention, and 170 completed the study.

2.3.2.2 ROBIN

Patients were randomised 1:1 to receive zuranolone or placebo. Patients self-administered study drug (zuranolone 30 mg or placebo); medication was administered orally once daily in the evening with food for 14 days. The dose of zuranolone could be reduced to 20 mg based upon patient tolerability. A total of 153 patients were randomised, 151 received study intervention, and 142 completed the study.

2.3.3 Trial outcomes

The pre-specified primary, secondary and exploratory outcomes for SKYLARK and ROBIN studies are summarised in Table 8.

Table 8: Outcomes in zuranolone clinical trials

Outcomes	SKYLARK	ROBIN
Primary outcome	• CFB in HAMD-17 total score at Day 15	• CFB in HAMD-17 total score at Day 15
Secondary outcomes	Key secondary efficacy endpoints: • CFB in HAMD-17 total score at Days 3, 28, and 45 • CFB in CGI-S score at Day 15 Other secondary efficacy endpoints: • CFB in HAMD-17 subscales and individual item scores at Day 15 and other time points • HAMD-17 response and remission at Days 15 and 45 • CGI-I response at Day 15 • CFB in HAM-A total score at Day 15 • CFB in MADRS total score at Day 15 • CFB in EPDS total score at Day 15 • CFB in PHQ-9 total score Safety evaluation: • Safety and tolerability • TEAEs • Vital signs • Clinical laboratory evaluations • ECG parameters • Suicidal ideation and behaviours (as assessed by the C-SSRS) • Withdrawal symptoms (as assessed by the PWC-20)	Secondary efficacy endpoints: • CFB in HAMD-17 total score at Days 3, 8, 21, and 45 • HAMD-17 response and remission • CFB in HAMD-17 subscales and individual item scores at Day 15 and other time points • CFB in MADRS total score at Day 15 and other time points • CFB in HAM-A total score at Day 15 and other time points • CGI-I response • MADRS response and remission Other efficacy endpoints: • HRU • CFB in EPDS total score • CFB in PHQ-9 total score • CFB in BIMF total and subscale scores • CFB in SF-36 Safety evaluation: • Safety and tolerability • AEs • Vital signs • Clinical laboratory evaluations • ECG parameters • Suicidal ideation and behaviours (as assessed by the C-SSRS)

Source: (Deligiannidis 2021, Deligiannidis 2023a)
For a brief description of the scales used to assess the outcomes of patients in the SKYLARK and the ROBIN studies, please see Table 4
Abbreviations: AE: Adverse Event; BIMF: Barkin Index of Maternal Functioning; CFB: Change from Baseline; CGI-I: Clinical Global Impression – Improvement; CGI-S: Clinical Global Impression of Severity Scale; C-SSRS: Colombia Suicide Severity Rating Scale; ECG: Echocardiogram; EPDS: Edinburgh Postnatal Depression Scale; HAM-A: Hamilton Anxiety Rating Scale; HAMD-17, 17-item Hamilton Rating Scale for Depression; HRU: Healthcare Resource Utilisation; MADRS: Montgomery–Åsberg Depression Rating Scale; PHQ-9: Patient Health Questionnaire 9; PWC-20: Physician Withdrawal Checklist; SF-36: 36-item Short Form Health Survey; TEAE: Treatment-Emergent Adverse Event.

2.3.4 Patient characteristics

2.3.4.1 SKYLARK

Demographic and baseline characteristics of patients enrolled into SKYLARK were generally well balanced between treatment groups (Deligiannidis 2023a). Any differences in patient characteristics were small, and therefore, unlikely to influence study results (Food and Drug

Administration (FDA) 2022). The mean age was 30.5 years (SD 5.93), and majority ([redacted]) of patients were in the 25 to 45-year age range (Sage Therapeutics 2022a). A concomitant use of ADTs at a stable dose was permitted; 15% of patients in both the zuranolone and placebo groups were taking another ADT at baseline. The majority of patients in the study (approximately 85% in both arms) received zuranolone as monotherapy; this is expected in the clinical population with PND, which is often untreated at the time of symptom onset, as for many patients, PND is the first encounter with mental health problems (Biogen 2024d). Baseline disease severity was similar between the two treatment groups with a mean baseline HAMD-17 score of approximately 29 in each group. PND onset was in the third trimester in 33% of subjects and within 4 weeks of delivery in 67% of patients, with similar rates between both treatment groups (Food and Drug Administration (FDA) 2022, Deligiannidis 2023a). Table 9 presents a summary of the baseline patient demographics in SKYLARK.

Table 9: SKYLARK: Patient Demographics and Baseline Characteristics (Safety/Full Analysis Set)

Characteristics	Zuranolone 50 mg (n=98)	Placebo (n=98)
Age, mean (SD), years	30.0 (5.9)	31.0 (6.0)
Age group, years, n (%)		
18-24	[redacted]	[redacted]
25-45	[redacted]	[redacted]
Race, n (%)		
White	68 (69.4)	69 (70.4)
Black/African American	25 (25.5)	18 (18.4)
Other[a]	5 (5.1)	11 (11.2)
Hispanic/Latina ethnicity, n (%)	33 (33.7)	42 (42.9)
BMI, mean (SD), kg/m2	30.9 (6.3)	29.6 (6.3)
HAMD-17 total score, mean (SD)	28.6 (2.5)	28.8 (2.3)
CGI-S score, mean (SD)	5.0 (0.7)	4.9 (0.6)
Country, n (%)		
USA	95 (96.9)	96 (98.0)
Rest of the world	3 (3.1)	2 (2.0)
Baseline ADT use, n (%)	15 (15.3)	15 (15.3)
History of PND, n (%)		
First episode	81 (82.7)	87 (88.8)
Recurrent episode	17 (17.3)	11 (11.2)
Onset of PND, n (%)		
Third trimester	34 (34.7)	31 (31.6)
≤4 weeks postnatal	64 (65.3)	67 (68.4)
Characteristics	**Full Analysis Set**	
	Zuranolone 50 mg (N=98)	**Placebo (N=98)**
History of MDD[b], n (%)	[redacted]	[redacted]

Source: (Sage Therapeutics 2022a, Deligiannidis 2023a, Biogen 2025c)
The safety set was composed of patients who received ≥1 dose of the assigned blinded treatment.
[a]Other included Asian, American Indian, Alaska Native, Native Hawaiian/Pacific Islander, multiple, other race, and/or not reported.

[b] [redacted]

Abbreviations: ADT: Antidepressant Therapy; BMI: Body Mass Index; CGI-S: Clinical Global Impressions – Severity; HAMD-17: 17-item Hamilton Rating Scale for Depression; PND: Postnatal Depression; SD: Standard Deviation; USA: United States of America.

The patient population in SKYLARK included patients with severe PND when classified by the baseline HAMD-17 score only (in line with the inclusion criteria of baseline HAMD-17 total score ≥26); however, when classified by the MADRS score, the patient population in SKYLARK included a mix of patients with moderate and severe disease (for more detail, please see Section 2.8.2.2). While Biogen acknowledges that SKYLARK does not assess efficacy of zuranolone in patients with mild PND, who are part of the target population for this submission, the unique MoA of zuranolone supports its efficacy in all PND patients, irrespective of disease severity.

The UK-specific information on the demographics and other baseline characteristics of PND patients is limited, but the SKYLARK population is generally expected to be similar to the population of PND patients in the UK, with majority of patients to be aged 25-45 years (Harrison 2023, Biogen 2024d), majority of patients (>80%) experiencing their first PND episode, and low proportion of patients with ADT use (Biogen 2024d). Finally, it is not anticipated that the pathophysiology of PND would differ by region (Stewart 2016).

2.3.4.2 ROBIN

The baseline patient demographics in ROBIN are summarised below in Table 10.

Table 10: ROBIN: Patient Demographics and Baseline Characteristics (Efficacy Set)

Characteristics	Zuranolone 30 mg (n=76)	Placebo (n=74)
Age, mean (SD), years	29.3 (5.4)	27.4 (5.3)
Race, n (%)		
White	44 (58)	40 (54)
African American	31 (41)	31 (42)
Other[a]	1 (1)	3 (4)
Hispanic/Latino ethnicity, n (%)	16 (21)	18 (24)
Weight, mean (SD), kg	85.1 (19)	80.2 (24)
HAMD-17 total score, mean (SD)	28.4 (2)	28.8 (2)
Baseline ADT use, n (%)	16 (21)	13 (18)
Family history of PND, n (%)	10 (13)	10 (14)
Onset of PND, n (%)		
Third trimester	32 (42)	31 (42)
≤4 weeks after delivery	44 (58)	43 (58)
History of MDD, n (%)	[redacted]	[redacted]

Source: (Deligiannidis 2021, Biogen 2025c)
[a]Other included Asian, Native Hawaiian/Pacific Islander, and more than 1 race.
Abbreviations: ADT: Antidepressant Therapy; HAMD-17: 17-item Hamilton Rating Scale for Depression; MDD: Major Depressive Disorder; PND: Postnatal Depression; SD: Standard Deviation.

2.4 Statistical analysis and definition of study groups in the relevant clinical effectiveness evidence

2.4.1 SKYLARK

The primary objective of the SKYLARK study was to evaluate the efficacy and safety of zuranolone 50 mg versus placebo in adults with PND (Sage Therapeutics 2022a, Deligiannidis 2023b). To achieve this, the efficacy analysis was conducted in the Full Analysis Set. A sample size of [redacted] evaluable patients per treatment group provided 90% power to detect a placebo-adjusted treatment difference of 4.0 for the primary endpoint (Change from baseline in HAMD-17 total score at Day 15), and a Standard Deviation (SD) of 8 points using a 2-sided t-test at an alpha level of 0.05. Assuming a 10% dropout rate and 1:1 randomisation, approximately [redacted] randomised patients were required to obtain [redacted] evaluable patients per treatment group. The Safety Set included all patients who received at least one dose of study intervention. The Full Analysis Set was defined as all randomised patients who received at least one dose of study intervention with both a valid baseline and at least one post-baseline score on the HAMD-17, HAM-A, MADRS, CGI-S, EPDS, and PHQ-9, or at least one post-baseline CGI-I score. The primary endpoint was analysed utilising a mixed effects model for repeated measures, which included treatment, baseline HAMD-17 score, baseline ADT use, assessment timepoint, and timepoint–by-treatment as explanatory variables (all explanatory variables were treated as fixed effects). Multiplicity adjustment for hypothesis testing of the key secondary endpoints was conducted using a fixed sequence strategy. If the primary endpoint was statistically significant at a two-sided 0.05 level, hypothesis testing for the key secondary endpoints was performed at a 5% level of significance. If an endpoint was not significant at this level, the next endpoint in the sequence was to be interpreted with a nominal p value. Other secondary endpoints were not adjusted for multiplicity and are to be interpreted with nominal p values, and all comparisons are considered descriptive. Effect sizes were reported by Cohen's d values, which were calculated for each study visit as the observed mean difference between measures in the zuranolone and placebo groups divided by the pooled standard deviation. Statistical analyses were performed using statistical analysis system (SAS), version 9.4 (SAS 2024).

Definitions of study groups/analysis populations included in the submission are summarised in Table 11.

Table 11: SKYLARK: Definition and number of participants by analysis set

Analysis Set	Definition	Number of participants	
		Zuranolone 50 mg	Placebo
Randomised Set	All participants who were randomised	99	101
Safety Set	All participants who were administered study intervention	98	98
Full Analysis Set	All randomised participants who were administered IP with a valid baseline total score and at least 1 post-baseline total score in at least 1 of HAMD-17, HAM-A, MADRS, CGI-S, EPDS, or PHQ-9 or at least 1 postbaseline value of CGI-I	98	97

Source: (Deligiannidis 2023b)
Abbreviations: CGI-I: Clinical Global Impression – Improvement; CGI-S: Clinical Global Impression – Severity; EPDS: Edinburgh Postnatal Depression Scale; HAM-A: Hamilton Anxiety Rating Scale; HAMD-17: 17-item Hamilton Rating Scale for Depression; IP: Investigational Product; MADRS: Montgomery-Åsberg Depression Rating Scale; PHQ-9: 9-item Patient Health Questionnaire.

2.4.2 ROBIN

The primary objective of the ROBIN study was to evaluate the efficacy and safety of zuranolone 30 mg (capsule formulation with a higher relative bioavailability) compared to placebo in the treatment of adult women with PND (Sage Therapeutics 2019, Deligiannidis 2021). To achieve this, the efficacy analysis was conducted using the Efficacy Set, with data collected between December 2018 and March 2019. A sample size of approximately [redacted] evaluable patients per group provided 90% power to detect a treatment difference of 4.0 between the zuranolone and the placebo groups for the primary endpoint (Change from baseline in HAMD-17 total score at Day 15), and a common SD of 7 points using a 2-sided *t*-test at an alpha level of 0.05. Assuming a 10% dropout rate and 1:1 randomisation ratio, approximately [redacted] randomised patients per treatment group were required to obtain [redacted] evaluable patients. The safety set included all patients who received at least one dose of the double-blind study drug. The Efficacy Set, based on intention-to-treat, included all patients in the safety set who had a valid baseline and at least one post-baseline efficacy assessment. A mixed-effects model for repeated measures including treatment, baseline HAMD-17 total score, baseline antidepressant use, assessment time point, and time point by treatment as explanatory variables was used for analysis of the primary endpoint. An unstructured covariance structure was used to model the within-patient errors. Secondary (HAMD-17, MADRS, HAM-A scores change from baseline), and other endpoints (BIMF score) were assessed using the same mixed-effects model for repeated measures as the primary endpoint. Model based point estimates (i.e., least-squares means), 95% CIs, and p values were reported. Treatment comparisons were made between placebo and treatment groups. Categorical secondary endpoints (HAMD-17 response and remission and CGI-I response) were analysed by a generalised estimating equations approach. ORs, 95% CIs, and p values

were reported. P values and 95% CIs for secondary endpoints, which are highly associated with the primary endpoint, were not adjusted for multiplicity, and therefore reported as point estimates. Statistical analyses used SAS statistical software version 9.4 (SAS Institute). No imputation was used to estimate missing data.

Definitions of study groups/analysis populations included in the submission are summarised in Table 12.

Table 12: ROBIN: Definition and number of participants by analysis set

Analysis Set	Definition	Number of participants	
		Zuranolone 50 mg	Placebo
All Randomised Set	All participants who were randomised	77	76
Safety Set	All participants who were administered study intervention	78	73
Efficacy Set*	All participants in the All Randomised Set who completed at least 1 day of zuranolone and had a valid baseline and at least 1 post-baseline efficacy assessment	76	74

Source: (Sage Therapeutics 2019, Deligiannidis 2021)
* Note that the definition of the Efficacy Set in the ROBIN study corresponds to the definition of the Full Analysis Set in the SKYLARK study.

2.5 Critical appraisal of the relevant clinical effectiveness evidence

The Downs and Black Checklist for the assessment of the methodological quality of both the randomised and non-randomised healthcare intervention studies was used to evaluate the quality of SKYLARK and ROBIN. The trial design and methodology of both SKYLARK and ROBIN scored 21/27 (Table 13). For the full assessment, please see Appendix B.5.

Table 13: Black and Downs quality checklist for relevant clinical effectiveness evidence

Study reference	SKYLARK	ROBIN
Reporting	9/11	9/11
External validity	1/3	1/3
Bias	7/7	7/7
Confounding	4/6	4/6
Total	**21/27**	**21/27**

While overall both studies were considered to be at a low risk of bias, the potential risk of functional unblinding was raised by the FDA. This is because in SKYLARK, patients receiving zuranolone were more likely to discontinue treatment due to AEs (4%) compared to patients receiving placebo (2%). Additionally, there was an imbalance of patients treated with zuranolone (14%) who had dose reductions (most often secondary to CNS depressant effect-

related AEs) compared to patients receiving placebo (1%). In ROBIN, one zuranolone-treated patient discontinued treatment because of an AE compared to no patients in placebo group, and the imbalance of patients requiring dose reductions due to AEs between the two groups (4% in zuranolone group vs 0% in placebo group; for more detail on patient disposition, please see Appendix B.4.2). While this could have compromised study integrity and interpretability of findings, the FDA concluded that the robust efficacy findings for zuranolone in PND overcome confounding from this potential bias (Food and Drug Administration (FDA) 2022).

2.6 Clinical effectiveness results of the relevant studies

Evidence supporting the clinical effectiveness of zuranolone in PND is provided primarily by the SKYLARK study with additional supportive evidence provided by the ROBIN study. Safety data from both studies is used to support the tolerability of zuranolone. Clinical evidence from SKYLARK represents zuranolone with ECM versus ECM alone, as the trial protocol allowed background use of stable dose of ADTs and psychological therapies, which is aligned with the usual care received by patients with PND in the UK.

2.6.1 SKYLARK

The CONSORT flow diagram for SKYLARK is presented in Appendix B.4.2.

2.6.1.1 Primary endpoint

Change from baseline in HAMD-17 total score at Day 15

In SKYLARK study, the primary endpoint of change from baseline in HAMD-17 score at Day 15 was met. The least-square means (LSM) change from baseline was −15.6 for the zuranolone 50 mg group compared with −11.6 for the placebo. A statistically significant improvement in depressive symptoms at Day 15 was observed in the zuranolone 50 mg group with an LSM treatment difference of −4.0 (95% CI: −6.3, −1.7; p=0.001) (Table 14, Figure 8) (Deligiannidis 2023b).

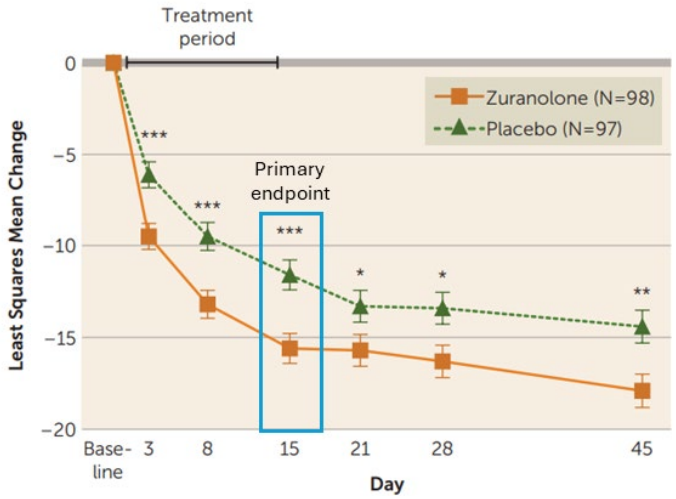
A statistically significant improvement in change from baseline in HAMD-17 total score was observed in the zuranolone 50 mg group compared with the placebo group at additional timepoints (key secondary endpoint). Changes were observed as early as Day 3 (LSM difference: −3.4 [95% CI: −5.4, −1.4]; p=0.001), and were maintained through to Day 45 (LSM difference: −3.5 [95% CI: −6.0, −1.0]; p=0.007) (Table 14, Figure 8) (Deligiannidis 2023b).

Table 14: SKYLARK: LSM change from baseline in HAMD-17 score (Full Analysis Set)

Study visit	Zuranolone 50 mg N=98		Placebo N=97		LSM difference			p-value
	Mean	SE	Mean	SE	Mean	95% CI lower	95% CI upper	
Baseline[a]	28.6	2.49	28.8	2.34	-	-	-	-
Day 3[b]	-9.5	0.70	-6.1	0.71	-3.4	-5.4	-1.4	0.001
Day 8	-13.2	0.76	-9.5	0.76	-3.7	-5.8	-1.6	0.001
Day 15[c]	**-15.6**	**0.82**	**-11.6**	**0.82**	**-4.0**	**-6.3**	**-1.7**	**0.001**
Day 21	-15.7	0.87	-13.3	0.87	-2.5	-4.9	-0.1	0.045
Day 28[b]	-16.3	0.88	-13.4	0.88	-2.9	-5.4	-0.5	0.020
Day 45[b]	-17.9	0.90	-14.4	0.90	-3.5	-6.0	-1.0	0.007

Source: (Deligiannidis 2023b)
Results are presented as LSM and SE CFB unless otherwise noted.
[a] Baseline values are reported as mean ± SD.
[b] Key secondary endpoint.
[c] Primary endpoint.
Primary and key secondary endpoints were adjusted for multiplicity. All other secondary endpoints were not adjusted for multiplicity and are to be interpreted with nominal p-values.
Abbreviations: CFB: Change form Baseline; CI: Confidence Interval; HAMD-17: 17-item Hamilton Depression Rating Scale; LSM: Least Squares Mean; SE: Standard Error.

Figure 8: SKYLARK: LSM change from baseline in HAMD-17 score (Full Analysis Set)[a]



Source: (Deligiannidis 2023b)
[a] The primary endpoint was change from baseline in score on the HAMD-17 at day 15, and the key secondary endpoints included change from baseline in HAMD-17 score at days 3, 28, and 45. Multiplicity was accounted for when analysing primary and key secondary endpoints. All other secondary endpoints were not adjusted for multiplicity and are to be interpreted with nominal p values. Error bars indicate standard error. *p<0.05. **p<0.01. ***p<0.001.
Abbreviations: HAMD-17: 17-item Hamilton Depression Rating Scale; LSM: Least Squares Mean.

2.6.1.2 Key secondary endpoints

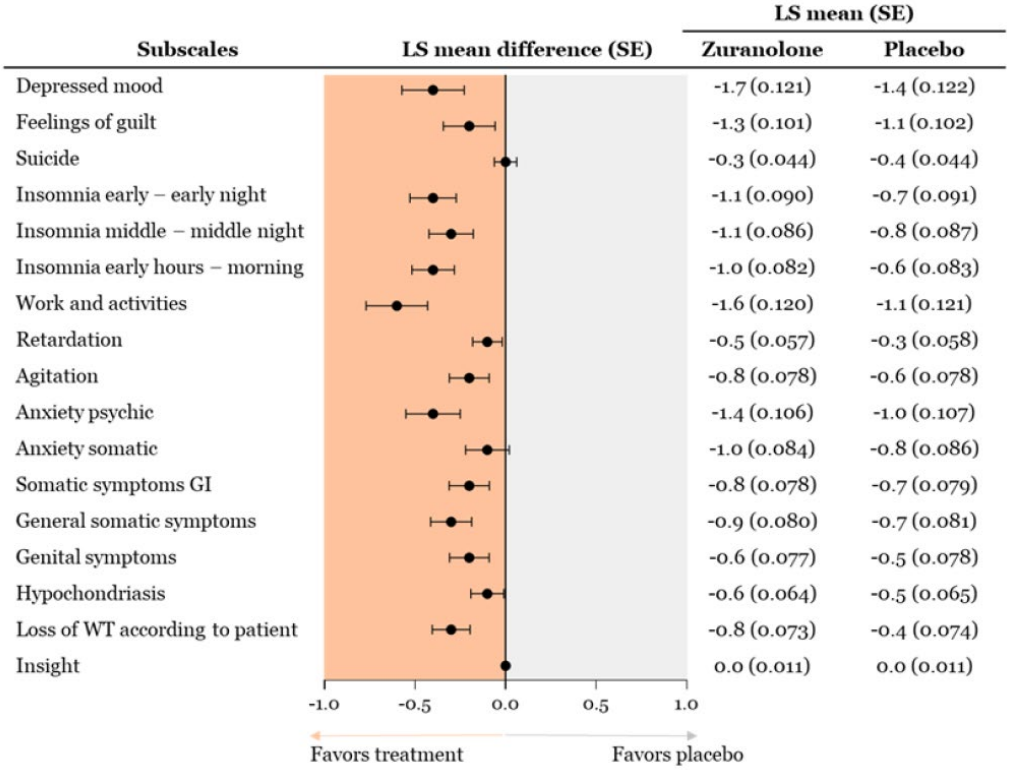
Change from baseline in HAMD-17 total score

The effects of zuranolone versus placebo on the HAMD-17 total score were rapid (by day 3), and were sustained at all measured timepoints through day 45 (Section 2.6.1.1, Table 14, Figure 8).

Change from baseline in HAMD-17 subscales and individual item scores at Day 15

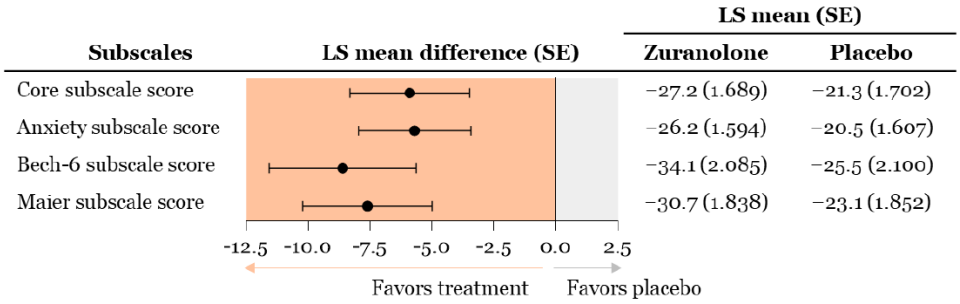
Consistent with the overall improvements in HAMD-17 total scores at Day 15, treatment with zuranolone 50 mg was favoured versus placebo across 16 out of 17 individual HAMD-17 item scores (change from baseline at Day 15, Figure 9). Additionally, at Day 15, compared with placebo-treated patients, patients treated with zuranolone 50 mg demonstrated significantly greater improvement in each HAMD subscale, indicating that the greater decrease in HAMD-17 total score observed with zuranolone 50 mg is due to a decrease across various symptoms associated with depression (change from baseline at Day 15, Figure 10) (Deligiannidis 2023b).

Figure 9: Change from baseline in HAMD-17 individual items at Day 15 (Full Analysis Set)



Source: (Deligiannidis 2023b)
Abbreviations: CI: Confidence Interval; GI: Gastrointestinal; HAMD-17: 17-item Hamilton Depression Rating Scale; LS: Least Squares; SE: Standard Error; WT: Weight.

Figure 10: Change from baseline in HAMD-17 subscales at Day 15 (Full Analysis Set)



Source: (Deligiannidis 2023b)
Abbreviations: HAMD-17: 17-item Hamilton Depression Rating Scale; LS: Least Squares; SE: Standard Error.

Change from baseline in CGI-S score at Day 15

Treatment with zuranolone 50 mg resulted in a statistically significant greater improvement in change from baseline in CGI-S score compared with the placebo group at Day 15 (LSM difference: −0.6 [95% CI: −0.9, −0.2]; p=0.005) (Table 15) (Deligiannidis 2023b).

Table 15: SKYLARK: LSM change from baseline in CGI-S (Full Analysis Set)

Study visit	Zuranolone 50 mg N=98		Placebo N=97		LSM difference			p-value
	Mean	SE	Mean	SE	Mean	95% CI lower	95% CI upper	
Baseline[a]	5.0	0.66	4.9	0.58	-	-	-	-
Day 3	-1.0	0.10	-0.7	0.10	-0.2	-0.5	0	0.093
Day 8	-1.7	0.12	-1.1	0.12	-0.6	-0.9	-0.3	0.0004
Day 15[b]	**-2.2**	**0.14**	**-1.6**	**0.14**	**-0.6**	**-0.9**	**-0.2**	**0.005**
Day 21	-2.3	0.14	-1.9	0.14	-0.5	-0.9	-0.1	0.022
Day 28	-2.3	0.14	-1.9	0.14	-0.4	-0.8	0	0.044
Day 45	-2.6	0.14	-2.1	0.14	-0.5	-0.9	-0.1	0.012

Source: (Deligiannidis 2023b)
[a] Baseline values are reported as mean ± SD.
[b] Key secondary endpoint.
Multiplicity was accounted for when analysing primary and key secondary endpoints. All other secondary endpoints were not adjusted for multiplicity and are to be interpreted with nominal p values.
Abbreviations: CI: Confidence Interval; CGI-S: Clinician Global Impressions – Severity; LSM: Least Squares Mean; SE: Standard Error.

2.6.1.3 Other secondary endpoints

All subsequent secondary efficacy endpoints were not adjusted for multiple comparisons; therefore, they were interpreted with nominal p-values (Deligiannidis 2023b). Across all these endpoints, numerically greater improvements were observed in patients treated with zuranolone 50 mg compared with patients treated with placebo.

HAMD-17 Response Rate and Remission Rate

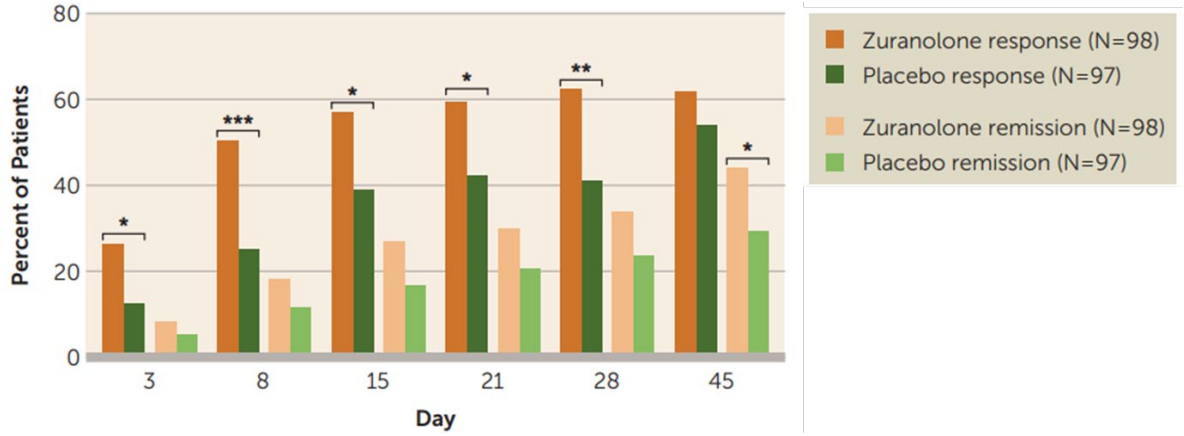
Assessment of the proportion of patients who achieved a HAMD-17 response demonstrated that a numerically greater percentage of patients treated with zuranolone 50 mg achieved a response compared with patients treated with placebo through Day 45, with nominal significance observed at Day 3 (26.5% vs. 12.5%; nominal P=0.012) and Day 15 (57.0% vs. 38.9%; nominal P=0.021) (Table 16, Figure 11) (Deligiannidis 2023b). Additionally, a numerically greater percentage of patients treated with zuranolone 50 mg achieved HAMD-17 remission compared with patients treated with placebo from Day 15 (26.9% vs. 16.7%; nominal P=0.111) through Day 45 (44.0% vs. 29.4%; nominal P=0.023) (Table 16) (Deligiannidis 2023b).

Table 16: SKYLARK: HAMD-17 response and remission over time (Full Analysis Set)

Study visit	Zuranolone 50 mg N=98			Placebo N=97			Odds ratio			Nominal p-value
	n	N	%	n	N	%	Mean	95% CI Lower	95% CI Upper	
HAMD-17 response [a]										
Day 3	26	98	26.5	12	96	12.5	2.64	1.24	5.61	0.012
Day 8	47	93	50.5	24	95	25.3	2.99	1.61	5.54	0.001
Day 15	**53**	**93**	**57.0**	**35**	**90**	**38.9**	**2.02**	**1.11**	**3.67**	**0.021**
Day 21	50	84	59.5	35	83	42.2	2.15	1.17	3.96	0.014
Day 28	48	77	62.3	35	85	41.2	2.39	1.29	4.43	0.006
Day 45	52	84	61.9	46	85	54.1	1.53	0.84	2.81	0.166
HAMD-17 remission [b]										
Day 3	8	98	8.2	5	96	5.2	1.64	0.51	5.25	0.406
Day 8	17	93	18.3	11	95	11.6	1.68	0.74	3.81	0.218
Day 15	**25**	**93**	**26.9**	**15**	**90**	**16.7**	**1.78**	**0.88**	**3.62**	**0.111**
Day 21	25	84	29.8	17	83	20.5	1.55	0.78	3.06	0.210
Day 28	26	77	33.8	20	85	23.5	1.50	0.76	2.99	0.247
Day 45	37	84	44.0	25	85	29.4	2.08	1.11	3.92	0.023

Source: (Deligiannidis 2023b)
[a] HAMD-17 response was defined as a ≥50% reduction from baseline in HAMD-17 score. HAMD-17 response data were not adjusted for multiplicity and are to be interpreted with nominal p-values.
[b] HAMD-17 remission was defined as a HAMD-17 score ≤7. HAMD-17 remission data were not adjusted for multiplicity and are to be interpreted with nominal p-values.
Abbreviations: CI: Confidence Interval; HAMD-17: 17-item Hamilton Depression Rating Scale.

Figure 11: SKYLARK: HAMD-17 response and remission (Full Analysis Set)[a]



Source: (Deligiannidis 2023b)
[a] Response was defined as a reduction of 50% from baseline in score on the HAMD-17, and remission was defined as a HAMD-17 score <7. Secondary endpoints were not adjusted for multiplicity and are to be interpreted with nominal p values. *p<0.05. **p<0.01. ***p<0.001.
Abbreviations: HAMD-17: 17-item Hamilton Depression Rating Scale.

CGI-I Response Rate

Assessment of the proportion of patients who achieved a CGI-I response by Day 15 demonstrated that a numerically greater percentage of patients treated with zuranolone 50 mg achieved a response compared with patients treated with placebo (66.7% vs. 46.7%; nominal P=0.0089) (Table 17) (Sage Therapeutics 2022a, Deligiannidis 2023b).

Table 17: SKYLARK: CGI-I Response by Study Visit (Full Analysis Set)

Study visit	Zuranolone 50 mg N=98		Placebo N=97		Treatment difference (zuranolone vs placebo)		
	n	Responders N (%)	n[a]	Responders[b] N (%)	OR[c]	95% CI for OR	Nominal p-value[d]
Day 3	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Day 8	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Day 15	[redacted]	**62 (66.7)**	**90**	**42 (46.7)**	**2.23**	**1.223, 4.072**	**0.0089**
Day 21	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Day 28	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Day 45	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Source: (Sage Therapeutics 2022a, Deligiannidis 2023b)
[a] Number of participants with non-missing CGI-I score at the visit as well as the baseline.
[b] CGI-I response is defined as having a CGI-I score of "very much improved" or "much improved". The denominator of percentage is the number of participants with non-missing CGI-score at the visit.
[c] OR is the estimate of the odds of having CGI-I response for participants treated with SAGE-217 relative to that for participants treated with placebo.
[d] Reported p-values are nominal.
Abbreviations: CI: Confidence Interval; CGI-I: Clinical Global Impression – Improvement; OR: Odds Ratio.

HAM-A Total Score

Numerically greater improvements in anxiety symptoms, as measured by HAM-A, were observed for the zuranolone 50 mg group compared with the placebo group as early as Day

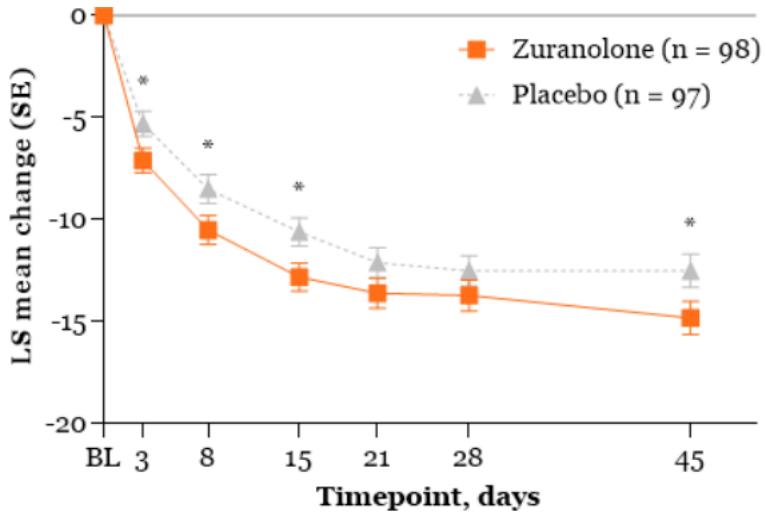
3 (nominal P=0.037), which was sustained through Day 15 (nominal P=0.024) and Day 45 (nominal P=0.050) (Table 18, Figure 12) (Deligiannidis 2023b).

Table 18: SKYLARK: LSM change from baseline in HAM-A total score (Full Analysis Set)

Study visit	Zuranolone 50 mg N=98		Placebo N=97		LSM difference			Nominal p-value
	Mean	SE	Mean	SE	Mean	95% CI lower	95% CI upper	
Baseline[a]	24.4	6.01	24.7	5.96	-	-	-	-
Day 3	-7.1	0.61	-5.3	0.61	-1.8	-3.5	-0.1	0.037
Day 8	-10.5	0.70	-8.5	0.70	-2.0	-4.0	0	0.045
Day 15	**-12.8**	**0.69**	**-10.6**	**0.70**	**-2.2**	**-4.2**	**-0.3**	**0.024**
Day 21	-13.6	0.74	-12.1	0.74	-1.4	-3.5	0.6	0.171
Day 28	-13.7	0.76	-12.5	0.75	-1.2	-3.3	0.9	0.248
Day 45	-14.8	0.81	-12.5	0.81	-2.3	-4.5	0	0.050

Source: (Deligiannidis 2023b)
Results are presented as LSM and SE CFB unless otherwise noted.
[a]Baseline values are reported as mean ± SD. Primary and key secondary endpoints were adjusted for multiplicity. All other secondary endpoints were not adjusted for multiplicity and are to be interpreted with nominal p-values.
Abbreviations: CFB: Change from Baseline; CI: Confidence Interval; HAM-A: Hamilton Anxiety Rating Scale; LSM: Least Squares Mean; SE: Standard Error.

Figure 12: SKYLARK: LSM change from baseline in HAM-A total score (Full Analysis Set)



Source: (Deligiannidis 2023b)
p-values designated as: *<0.05. Secondary endpoints were not adjusted for multiplicity and are to be interpreted with nominal p-values.
Abbreviations: BL: Baseline; HAM-A: Hamilton Rating Scale for Anxiety Total Score; LSM: Least Squares Mean; SE: Standard Error.

MADRS Total Score

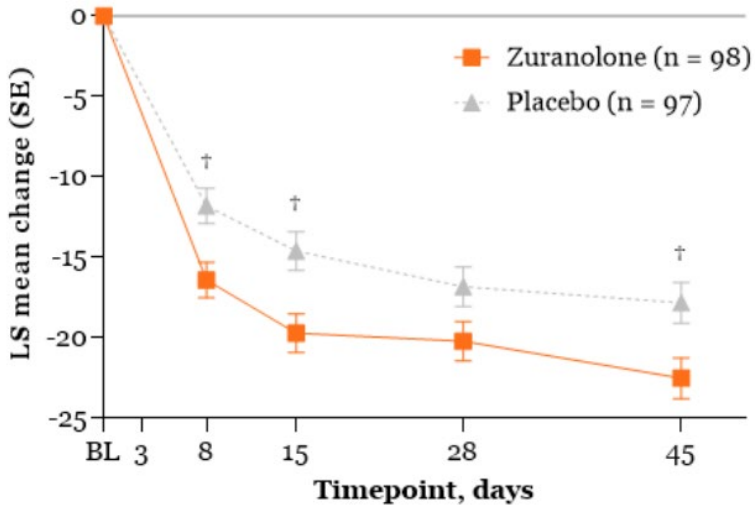
Analysis using MADRS also demonstrated numerically greater improvements in depressive symptoms for the zuranolone 50 mg group compared with the placebo group at Day 15 (LSM difference [95% CI]: −5.1 [−8.4, −1.7]; nominal P=0.003) (Table 19, Figure 13) (Deligiannidis 2023b).

Table 19: SKYLARK: LSM change from baseline in MADRS total score (Full Analysis Set)

Study visit	Zuranolone 50 mg N=98		Placebo N=97		LSM difference			Nominal p-value
	Mean	SE	Mean	SE	Mean	95% CI lower	95% CI upper	
Baseline[a]	35.5	5.37	35.0	4.81	-	-	-	-
Day 3	-16.4	1.11	-11.8	1.10	-4.6	-7.7	-1.5	0.004
Day 15	**-19.7**	**1.20**	**-14.6**	**1.21**	**-5.1**	**-8.4**	**-1.7**	**0.003**
Day 28	-20.2	1.23	-16.8	1.22	-3.4	-6.8	0	0.051
Day 45	-22.5	1.27	-17.8	1.28	-4.7	-8.3	-1.1	0.010

Source: (Deligiannidis 2023b)
Results are presented as LSM and SE CFB unless otherwise noted.
[a]Baseline values are reported as mean ± SD.
Primary and key secondary endpoints were adjusted for multiplicity. All other secondary endpoints were not adjusted for multiplicity and are to be interpreted with nominal p-values.
Abbreviations: CFB: Change form Baseline; CI: Confidence Interval; LSM: Least Squares Mean; MADRS: Montgomery-Åsberg Depression Rating Scale; SE: Standard Error.

Figure 13: SKYLARK: LSM change from baseline in MADRS total score (Full Analysis Set)



Source: (Deligiannidis 2023b)
p-values designated as: †<0.01. Secondary endpoints were not adjusted for multiplicity and are to be interpreted with nominal p-values.
Abbreviations: BL: Baseline; LSM: Least Squares Mean; MADRS: Montgomery-Åsberg Depression Rating Scale; SE: Standard Error.

Across the individual items of the MADRS, treatment with zuranolone 50 mg [redacted] (Table 20). The greatest improvement was observed on MADRS [redacted] item. For more detail on [redacted], see Appendix B.7.1.1.

Table 20: SKYLARK: LSM change from baseline in MADRS individual items at Day 15 (Full Analysis Set)

Item	Zuranolone 50 mg N=98		Placebo N=97		LSM difference			Nominal p-value
	Mean	SE	Mean	SE	Mean	95% CI lower	95% CI upper	
Reported sadness	[redacted]							
Apparent sadness	[redacted]							
Inner tension	[redacted]							
Reduced sleep	[redacted]							
Reduced appetite	[redacted]							
Concentration difficulties	[redacted]							
Lassitude	[redacted]							
Inability to feel	[redacted]							
Pessimistic thoughts	[redacted]							
Suicidal thoughts	[redacted]							

Source: (Sage Therapeutics 2022b)
Results are presented as LSM and SE CFB unless otherwise noted.
[a]Baseline values are reported as mean ± SD.
Primary and key secondary endpoints were adjusted for multiplicity. All other secondary endpoints were not adjusted for multiplicity and are to be interpreted with nominal p-values.
Abbreviations: CI: Confidence Interval; CFB: Change From Baseline; LSM: Least Squares Mean; MADRS: Montgomery-Åsberg Depression Rating Scale; SE: Standard Error.

Change From Baseline in EPDS Total Score

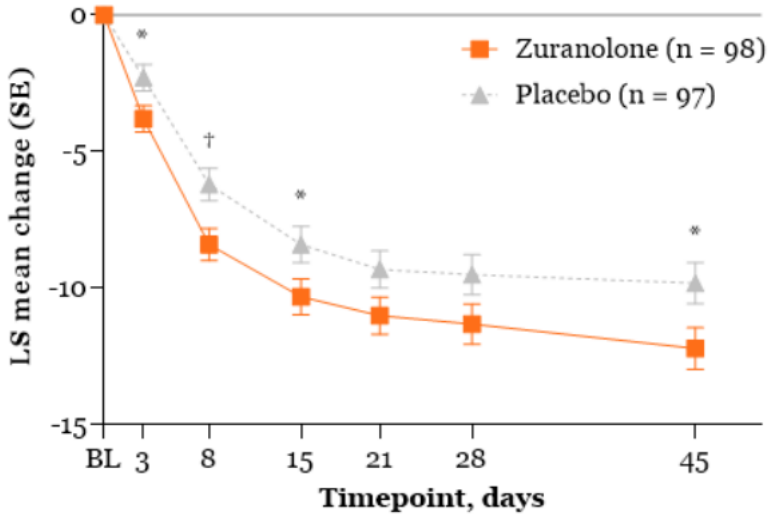
A numerically greater percentage of zuranolone-treated patients experienced greater improvement in EPDS scores at all time points compared with placebo-treated patients, with nominal significance achieved as early as at Day 3 (LSM difference [95% CI]: −1.5 [−2.9, −0.1]; nominal P=0.03), through to Day 8 (LSM difference [95% CI]: −2.2 [−3.8, −0.5]; nominal P=0.01), Day 15 (LSM difference [95% CI]: −2.0 [−3.8, −0.1]; nominal P=0.04), and Day 45 (LSM difference [95% CI]: −2.4 [−4.5, −0.3]; nominal P=0.03) (Table 21, Figure 14) (Sage Therapeutics 2022a, Deligiannidis 2023b).

Table 21: SKYLARK: LSM change from baseline in EPDS total score (Full Analysis Set)

Study visit	Zuranolone 50 mg N=98		Placebo N=97		LSM difference			Nominal p-value
	Mean	SE	Mean	SE	Mean	95% CI lower	95% CI upper	
Day 3	-3.8	0.49	-2.3	0.49	-1.5	-2.9	-0.1	0.032
Day 8	-8.4	0.60	-6.2	0.59	-2.2	-3.8	-0.5	0.010
Day 15	**-10.3**	**0.66**	**-8.4**	**0.66**	**-2.0**	**-3.8**	**-0.1**	**0.038**
Day 21	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Day 28	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Day 45	-12.2	0.76	-9.8	0.76	-2.4	-4.5	-0.3	0.028

Source: (Sage Therapeutics 2022a, Deligiannidis 2023b)
Abbreviations: CI: Confidence Interval; EPDS: Edinburgh Postnatal Depression Scale; LSM: Least Squares Mean; SE: Standard Error.

Figure 14: SKYLARK: LSM change from baseline in EPDS total score (Full Analysis Set)



Source: (Deligiannidis 2023b)
p-values designated as: *<0.05, †<0.01. Secondary endpoints were not adjusted for multiplicity and are to be interpreted with nominal p-values.
Abbreviations: BL: Baseline; EPDS: Edinburgh Postnatal Depression Scale; LSM: Least Squares Mean; SE: Standard Error.

Change From Baseline in PHQ-9

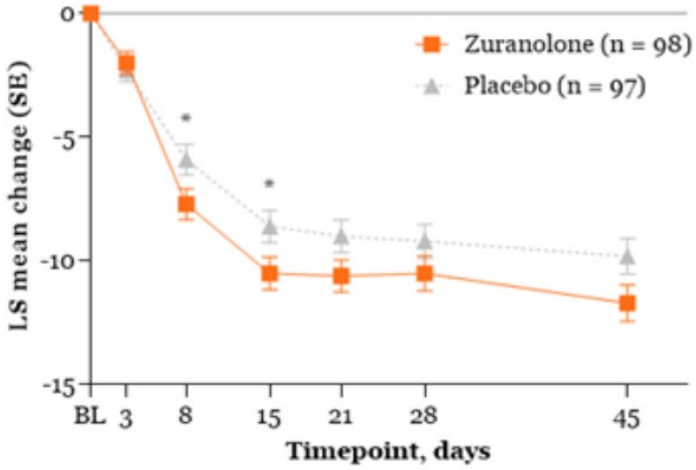
A numerically greater percentage of zuranolone-treated patients experienced improvements in PHQ-9 at all time points except Day 3, with nominal significant improvements observed at Day 8 (nominal P<0.05) and Day 15 (nominal P<0.05) (Table 22, Figure 15) (Sage Therapeutics 2022a, Deligiannidis 2023b).

Table 22: SKYLARK: LSM change from baseline in PHQ-9 total score (Full Analysis Set)

Study visit	Zuranolone 50 mg N=98		Placebo N=97		LSM difference			Nominal p-value
	Mean	SE	Mean	SE	Mean	95% CI lower	95% CI upper	
Day 3	[redacted]							
Day 8	[redacted]							
Day 15	[redacted]							
Day 21	[redacted]							
Day 28	[redacted]							
Day 45	[redacted]							

Source: (Sage Therapeutics 2022a)
Abbreviations: CI: Confidence Interval; LSM: Least Squares Mean; PHQ-9: 9-item Patient Health Questionnaire Total Score: Standard Error.

Figure 15: SKYLARK: LSM change from baseline in PHQ-9 total score (Full Analysis Set)



Source: (Deligiannidis 2023b)
p-values designated as: *<0.05. Secondary endpoints were not adjusted for multiplicity and are to be interpreted with nominal p-values.
Abbreviations: BL: Baseline; LSM: Least Squares Mean; PHQ-9: 9-item Patient Health Questionnaire Total Score; SE: Standard Error.

2.6.1.4 Post-hoc analyses

Determining clinically meaningful improvement in depressive symptoms among zuranolone-treated patients

Post-hoc analyses were conducted to assess the clinical meaningfulness of the improvements offered by zuranolone treatment. Based on pooled data from SKYLARK and ROBIN, anchor-based methods were used to the determine patient-level meaningful change threshold (MCT) (Biogen 2024b). The MCT represents the change in HAMD-17 that is considered meaningful at the patient-level. Anchor-based methods estimate the MCT by using an “anchor”, i.e. a variable that acts as a reference to interpret the size of the effect. [redacted] was chosen as the

anchor [redacted] and change in HAMD-17. MCTs were, therefore, estimated as the change in HAMD-17 relative to a [redacted], which was estimated as a [redacted]-point improvement in HAMD-17 from baseline. The minimally important difference (MID) for between group differences was estimated as ½ standard deviation (½ SD=[redacted]) and 1 standard error of measurement (SEM =[redacted]), with a mean of these two measures ([redacted]) used as the base case (see Appendix B.7.1.2.1 for more detail on methodology).

The estimated MCT and MID were then applied to SKYLARK data to identify differences between treatment arms.

Patient-level change in SKYLARK: HAMD-17 MCT

After applying HAMD-17 PND MCT [redacted] to SKYLARK data, at the individual patient level, [redacted] (Figure 16) (Biogen 2024b). These results are aligned with the response rates presented in Table 16.

Additionally, [redacted] (Figure 17).

Figure 16: Patient-level change in SKYLARK: Cumulative proportion of individual patients achieving HAMD-17 MCT

Source: (Biogen 2024b)
Abbreviations: CFB: Change from Baseline; HAMD-17: 17-item Hamilton Depression Rating Scale; MCT: Meaningful Change Threshold.

Figure 17: Patient-level change in SKYLARK: Proportion of individual patients achieving HAMD-17 MCT

Source: (Biogen 2024b)
*Statistical significance (p≤0.05) between placebo and zuranolone was determined using Fisher's exact test
Abbreviations: CFB: Change from Baseline; HAMD-17: 17-item Hamilton Depression Rating Scale; MCT: Meaningful Change Threshold.

Group-level difference in SKYLARK: HAMD-17 MID

After applying HAMD-17 MID estimates to SKYLARK data (estimate of), [redacted] (Figure 18) (Biogen 2024b).

Figure 18: Group-level difference in SKYLARK: Comparison of between-group LSM difference with HAMD-17 MID

Source: (Biogen 2024b)
Abbreviations: CFB: Change from Baseline; HAMD-17: 17-item Hamilton Depression Rating Scale; LSM: Least Squares Mean; MID: Minimal Important Difference; SD: Standard Deviation; SEM: Standard Error of Measurement.

Impact of time to zuranolone treatment initiation on efficacy

An exploratory post-hoc analysis was conducted to determine whether time to zuranolone treatment initiation (< 6 months or ≥ 6 months since PND onset) had an impact on efficacy. Overall, the results of this subgroup analysis demonstrated that treatment with zuranolone 50 mg provided similar efficacy results in patients treated < 6 months or ≥ 6 months since PND onset (Figure 19; see Appendix B.7.1.2.2 for additional details) (Peeper 2023).

Figure 19: CFB in HAMD-17 total score over time

Source: (Peeper 2023)
Abbreviations: CFB: Change from Baseline; HAMD-17: 17-item Hamilton Depression Rating Scale; LS: Least Squares; SE: Standard Error.

2.6.2 ROBIN

The CONSORT flow diagram for ROBIN is presented in Appendix B.4.2.

2.6.2.1 Primary endpoint

Change from baseline in HAMD-17 total score at Day 15

In ROBIN, the primary endpoint of change from baseline in HAMD-17 score at Day 15 was met. The LSM change from baseline was −17.8 for the zuranolone 30 mg group compared with −13.6 for the placebo group. A statistically significant improvement in depressive symptoms was observed in the zuranolone 30 mg with an LSM treatment difference of −4.2 (95% CI: −6.9, −1.5; p=0.003) (Table 23, Figure 20) (Deligiannidis 2021).

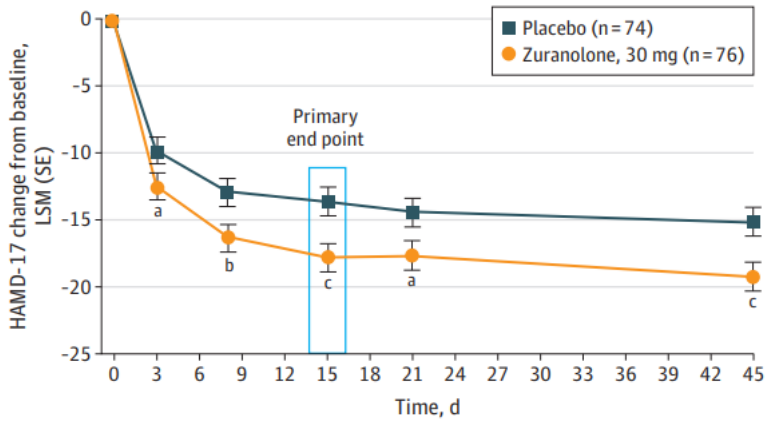
Numerically greater improvements in change from baseline in HAMD-17 total score was observed in the zuranolone 30 mg group compared with the placebo group at additional timepoints (key secondary outcome). Changes were observed as early as Day 3 (LSM difference [95% CI]: −2.7 [−5.1, −0.3]; nominal P=0.025), and were maintained through to Day 45 (LSM difference [95% CI]: −4.1 [−6.7, −1.4]; nominal P=0.003) (Table 23) (Deligiannidis 2021).

Table 23: ROBIN: LSM change from baseline in HAMD-17 score (Efficacy Set)

Study visit	Zuranolone 30 mg N=76		Placebo N=74		LSM difference			p-value
	Mean	SE	Mean	SE	Mean	95% CI lower	95% CI upper	
Day 3	-12.5	0.93	-9.8	0.95	-2.7	-5.1	-0.3	0.025
Day 8	-16.3	1.00	-12.9	1.03	-3.4	-6.0	-0.8	0.011
Day 15[a]	**-17.8**	**1.04**	**-13.6**	**1.07**	**-4.2**	**-6.9**	**-1.5**	**0.003**
Day 21	-17.6	1.09	-14.4	1.11	-3.1	-6.0	-0.3	0.032
Day 45	-19.2	1.02	-15.1	1.06	-4.1	-6.7	-1.4	0.003

Source: (Deligiannidis 2021)
Results are presented as LSM and SE CFB unless otherwise noted.
Primary and key secondary endpoints were adjusted for multiplicity. All other secondary endpoints were not adjusted for multiplicity and are to be interpreted with nominal p-values.
[a] Primary endpoint.
Abbreviations: CFB: Change form Baseline; CI: Confidence Interval; HAMD-17: 17-item Hamilton Depression Rating Scale; LSM: Least Squares Mean; SE: Standard Error.

Figure 20: ROBIN: LSM change from baseline in HAMD-17 score (Efficacy Set)



Source: (Deligiannidis 2021)
Treatment with zuranolone, 30 mg, achieved the primary end point of a significant change from baseline HAMD-17 total score at day 15 compared with placebo using mixed-effects model for repeated measures. HAMD-17 total score at time points other than day 15 were secondary end points, not adjusted for multiplicity and therefore reported as point estimates, which also showed sustained improvements for the zuranolone group compared with the placebo group. [a] P=0.03. [b] P=0.01. [c] P=0.003.
Abbreviations: CI: Confidence Interval; D: Day; HAMD-17: 17-item Hamilton Depression Rating Scale; LSM: Least Squares Mean. SE: Standard Error.

2.6.2.2 Secondary endpoints

Data for the secondary outcome change from baseline in HAMD-17 total score is presented in Section 2.6.2.1. Data for all other secondary endpoints of the ROBIN study (HAMD-17 response rate and remission rate, change from baseline in HAMD-17 subscales and individual item scores at Day 15, MADRS total score, CGI-I response rate, HAM-A total score) is presented in the Appendix B.7.2.1.

2.6.2.3 Other efficacy endpoints

Data for change from baseline in EPDS total score and SF-36 are presented below. Data for change from baseline in BIMF total score is presented in the Appendix B.7.2.2.

Change From Baseline in EPDS Total Score

Compared with placebo-treated patients, patients treated with zuranolone 30 mg demonstrated [redacted] (LSM difference [redacted] (Table 24) (Sage Therapeutics 2019).

Table 24: ROBIN: LSM change from baseline in EPDS total score (Efficacy Set)[a]

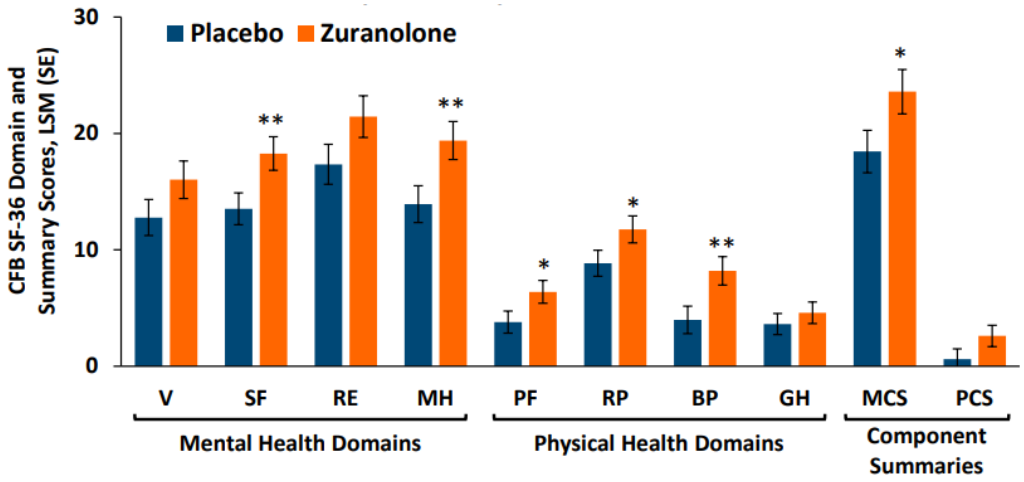
Study visit	Zuranolone 30 mg N=76		Placebo N=74		LSM difference			Nominal p-value
	Mean	SE	Mean	SE	Mean	95% CI lower	95% CI upper	
Day 3	[redacted]							
Day 8	[redacted]							
Day 15	[redacted]							
Day 21	[redacted]							
Day 45	[redacted]							

Source: (Sage Therapeutics 2019)
[a]Estimates for LSM, 95% CI for LSM, mean difference from placebo, 95% CI for mean difference from placebo, and p-value for mean difference from placebo are from a mixed model for repeated measures with treatment, baseline EPDS total score, baseline antidepressant use, assessment time point, and time point-by-treatment interaction as fixed effects. An unstructured covariance matrix was used to model the within subject correlation.
Abbreviations: CI: Confidence Interval; EPDS: Edinburgh Postnatal Depression Scale; LSM: Least Squares Mean; SE: Standard Error.

SF-36

By Day 45, patients receiving zuranolone achieved improvements compared with placebo across five domains of the SF-36, including Social Functioning, Mental Health, Physical Functioning, Role Physical, and Bodily Pain and in the Mental Component Summary score; CFB in other SF-36 domains was numerically greater with zuranolone versus placebo (Figure 21) (Deligiannidis 2023a).

Figure 21: ROBIN: SF-36 domain and summary scores at Day 45



Source: (Deligiannidis 2023a)
Abbreviations: BP: Bodily Pain; CFB: Change from Baseline; GH: General Health; LSM: Least Squares Mean; MCS: Mental component Score; MH: Mental Health; PCS: Physical Component Score; PF: Physical Functioning; RE: Role Emotional; RP: Role Physical; SE: Standard Error; SF: Social Functioning; SF-36: 36-item Short Form Health Survey; V: Vitality.

2.6.2.4 Post-hoc analyses

Post-hoc analyses of sleep quality among participants in the ROBIN study are presented in the Appendix B.7.2.3.

2.7 Subsequent treatments used in the relevant studies

No subsequent treatments were used in the SKYLARK and the ROBIN studies.

2.8 Subgroup analysis

2.8.1 Pre-specified subgroup analysis

Efficacy of zuranolone was assessed in a number of pre-specified subgroups in SKYLARK and ROBIN, which were based on demographic and other baseline characteristics including race, age, country of origin, ADT use at baseline, baseline BIM, onset of PND, family history of PND in both ROBIN and SKYLARK. Additionally, in SKYLARK, the subgroups included baseline HAMD-17 total score and depression with elevated anxiety (assessed by baseline HAM-A total score (for more detail, see Table 7) (Deligiannidis 2021, Sage Therapeutics 2022a, Deligiannidis 2023b).

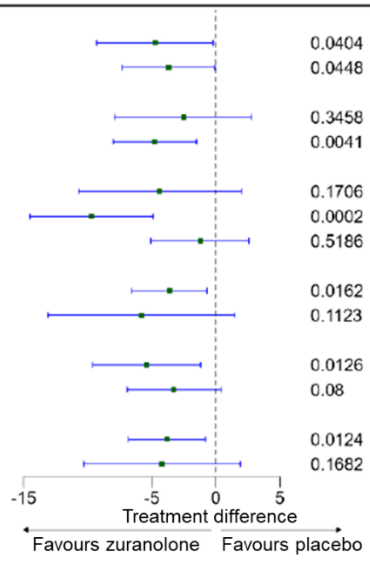
In both SKYLARK and ROBIN, the results of subgroup analyses were similar to the overall HAMD-17 total score analyses. In SKYLARK, the LS mean difference favoured zuranolone in all subgroups, except for the ADT use at baseline subgroup (Figure 22). However, a majority of participants did not use ADTs at baseline (placebo: 84.7%; zuranolone: 84.7%, Table 9). In ROBIN, zuranolone was favoured in all subgroups including ADT use at baseline, with results being statistically significant in 7 out of 14 subgroups (Figure 23) (Sage Therapeutics 2019, Food and Drug Administration (FDA) 2022, Sage Therapeutics 2022a). It is important to note that several subgroups included small number of patients, and neither study was powered to detect efficacy differences in these subgroups.

Figure 22: SKYLARK: Forest plot of LS mean difference in change from baseline in HAMD-17 total score by subgroups at Day 15 (Full Analysis Set)

Source: (Sage Therapeutics 2022a)
Abbreviations: BMI: Body Mass Index; HAM A: Hamilton Rating Scale for Anxiety; HAMD-17: 17-item Hamilton Rating Scale for Depression; LCL: Lower Control Limit; LS: Least Squares; NE: Not Estimable; PPD: Postpartum Depression; SE: Standard Error; UCL: Upper Control Limit

Figure 23: ROBIN: Forest plot of LS mean difference in change from baseline in HAMD-17 total score by subgroups at Day 15 (Efficacy Set)

Variable	Count	Placebo LS mean (SE) CFB	Zuranolone LS mean (SE) CFB	Treatment difference (95% CI)	P-value
Race					
Black or African American	62	-13.2 (1.59)	-18 (1.59)	-4.7 (-9.3, -0.2)	0.0404
White	81	-13.7 (1.3)	-17.4 (1.25)	-3.7 (-7.3, -0.1)	0.0448
Age					
18-24 years	42	-14 (1.68)	-16.5 (2.04)	-2.5 (-7.9, 2.8)	0.3458
25-45 years	105	-13.3 (1.2)	-18.1 (1.1)	-4.8 (-8, -1.5)	0.0041
Baseline BMI (kg/m^2)					
18.5-24.9	31	-12.3 (2)	-16.7 (2.36)	-4.4 (-10.7, 2)	0.1706
25-29.9	40	-10.5 (1.69)	-20.3 (1.67)	-9.7 (-14.5, -4.9)	0.0002
>=30	74	-15.5 (1.43)	-16.8 (1.29)	-1.2 (-5.1, 2.6)	0.5186
Baseline Anti-depressant Use					
No	119	-13.5 (1.05)	-17.2 (1.06)	-3.6 (-6.6, -0.7)	0.0162
Yes	28	-13.9 (2.61)	-19.7 (2.39)	-5.8 (-13.1, 1.5)	0.1123
PND Onset					
3rd Trimester	63	-13.2 (1.49)	-18.5 (1.47)	-5.4 (-9.6, -1.2)	0.0126
Within 4 Weeks of Delivery	84	-13.9 (1.3)	-17.1 (1.29)	-3.3 (-6.9, 0.4)	0.08
Family History of PND					
No	127	-13.6 (1.08)	-17.5 (1.06)	-3.8 (-6.8, -0.8)	0.0124
Yes	20	-14.2 (2.04)	-18.4 (2.04)	-4.2 (-10.3, 1.9)	0.1682



Source: (Food and Drug Administration (FDA) 2022)
Estimates for LS mean difference from placebo, 95% CI and p value for mean difference from placebo are from a mixed model for repeated measures with treatment, baseline results, assessment timepoint, and timepoint-by-treatment interaction as fixed effects.
Abbreviations: BMI: Body Mass Index; CI: Confidence Interval; CFB: Change from Baseline; HAMD-17: 17-item Hamilton Rating Scale for Depression; LS: Least Squares; PND: Postnatal Depression; SE: Standard Error.

2.8.2 Post-hoc subgroup analysis

2.8.2.1 Prior MDD

SKYLARK was not designed or powered to demonstrate a treatment difference in a subgroup of patients with MDD and MDD was not a standalone variable for collection in the study. The analysis of history of MDD was conducted as part of the post-hoc analyses and was extracted from across 'Other' fields in data collection sheets [redacted]. In SKYLARK, [redacted] of participants were identified as having prior MDD history. LSM change from baseline of HAMD-17 total score [redacted] (Figure 24). [redacted]

Figure 24: SKYLARK: Forest Plot of LSM (95% CI) Treatment Difference in Change from Baseline in HAMD-17 Total Score by History of MDD Subgroups at Day 15 and Day 45 (Full Analysis Set)

[redacted]

Results are from a mixed model for repeated measures (MMRM) with treatment, baseline HAMD-17 total score, antidepressant use at baseline, assessment time point, and time point-by-treatment interaction as fixed effects.
[1] Post hoc subgroup analysis.
Source: (Biogen 2025c)
Abbreviations: CI: Confidence interval; HAMD-17: 17-item Hamilton Rating Scale for Depression; LSM: Least-squares mean; MDD: Major Depressive Disorder.

In ROBIN, [redacted] had history of MDD. [redacted] LSM change from baseline of HAMD-17 total score [redacted] (Figure 25).

Figure 25: ROBIN: Forest Plot of LSM (95% CI) Treatment Difference in Change from Baseline in HAMD-17 Total Score by History of MDD Subgroups at Day 15 and Day 45 - Efficacy Set (Part B)

For subgroup analyses, results are from a mixed model for repeated measures (MMRM) with treatment, baseline result, assessment time point, and time point-by-treatment interaction as fixed effects. For Overall analysis only, an additional fixed effect term for antidepressant use at baseline was included.
[1] Post hoc subgroup analysis.
Source: (Biogen 2025c)
Abbreviations: CI: Confidence interval; HAMD-17: 17-item Hamilton Rating Scale for Depression; LSM: Least-squares mean; MDD: Major Depressive Disorder.

The heterogeneity between subgroup results observed within SKYLARK and compared to ROBIN, is likely due to the limitations of the sample size and is likely further impacted by the way the data were collected in SKYLARK (i.e., MDD was not a standalone variable for collection).

2.8.2.2 By PND severity

While the baseline HAMD-17 total score was utilised for assessing eligibility for study inclusion, the MADRS was an additional clinician-rated scale used in both SKYLARK and ROBIN studies, which can be used to further assess both baseline depression severity and the effect of zuranolone by baseline severity. The MADRS score provides an additional means of categorising depression severity that focuses more on psychological symptoms, with less emphasis on somatic or anxiety-related symptoms found in the HAMD-17.

Literature-established thresholds for the MADRS were used to identify moderate (total score ≤34) and severe (total score ≥35) baseline depression severity subgroups in SKYLARK (moderate PND: n= ; severe PND: n=) and in ROBIN (moderate PND: n= ; severe PND: n=) (Snaith 1986, Müller 2000).

[redacted] (see Figure 26 and Figure 27)

Figure 26: SKYLARK: Forest plot of LSM (95% CI) treatment difference in change from baseline in HAMD-17 total score by MADRS severity subgroups at Day 15 and Day 45 (Full Analysis Set)

Source: (Biogen 2025d)
[1] Post-hoc subgroup analysis
Results are from a MMRM with treatment, baseline HAMD-17 total score, antidepressant use at baseline, assessment time point, and time point-by-treatment interaction as fixed effects.
Abbreviations: CI: Confidence Interval; HAMD-17: 17-item Hamilton Depression Rating Scale; LS: Least Squares Mean; MADRS: Montgomery-Åsberg Depression Rating Scale; MMRM: Mixed Model for Repeated Measures.

Figure 27: ROBIN: Forest Plot of LSM (95% CI) Treatment Difference in Change from Baseline in HAMD-17 Total Score by MADRS Severity Subgroups at Day 15 and Day 45 (Efficacy Set)

Source: (Biogen 2025d)
[1] Post hoc subgroup analysis
For subgroup analyses, results are from a MMRM with treatment, baseline result, assessment time point, and time point-by-treatment interaction as fixed effects. For Overall analysis only, an additional fixed effect term for antidepressant use at baseline was included.
Abbreviations: CI: Confidence Interval; HAMD-17: 17-item Hamilton Depression Rating Scale; LS: Least Squares Mean; MADRS: Montgomery-Åsberg Depression Rating Scale; MMRM: Mixed Model for Repeated Measures.

2.9 Meta-analysis

Two meta-analyses aiming to determine the efficacy of zuranolone for treating PND and MDD are available in the literature and provide supportive evidence on zuranolone as a new therapeutic option offering improvements in depressive symptoms (Raja 2024, Winslow 2024). However, neither of the published meta-analyses is relevant for the NICE decision problem, as both analyses consider populations outside of the target patient population (mix of MDD and PND patients) and assess different doses of zuranolone (mix of 20 mg and 30 mg for MDD and mix of 30 mg and 50 mg for PND). Meta-analysis requires two or more studies that contain the intervention of interest (zuranolone) at the indicated dose (50 mg) and patient population (adults with PND). Therefore, a relevant meta-analysis was not possible as SKYLARK is the only study meeting these criteria; ROBIN assessed zuranolone with a dose and formulation (30 mg capsules), which is not relevant for the decision problem (Section 1.1).

2.10 Indirect and mixed treatment comparisons

The SKYLARK study assessed the efficacy of zuranolone plus ECM versus ECM alone, with ECM consisting of ADTs and/or psychological therapies (e.g., CBT, facilitated self-help). As ECM can consist of a single treatment with an ADT, development of an ITC comparing the efficacy and safety of zuranolone versus ADTs for the treatment of PND was explored as supportive evidence.

2.10.1 Published ITCs of zuranolone versus ECM

A recent ITC was published by Meltzer-Brody et al. (2024), which investigated the relative clinical efficacy of zuranolone in PND versus SSRIs and combination therapies (SSRIs + psychological therapies) used for PND in the US (Meltzer-Brody 2024). The ITC included RCTs for zuranolone and SSRIs that were identified through a systematic review (last update in 2023) and were then used to construct evidence networks, linking via common comparator arms (i.e., placebo arms of the RCTs). However, due to heterogeneity in placebo responses among the identified RCTs, the authors concluded that the placebo arm from the SKYLARK study was not suitable as a common comparator within an ITC. To overcome this, the authors took an unanchored approach and conducted a matching-adjusted indirect comparison (MAIC). For the MAIC, as the placebo arm of the SKYLARK study was omitted, the zuranolone treatment arm was matched to the placebo arm of an SSRI study, to provide an assessment of the relative effect of zuranolone versus placebo. The population overlap through their unanchored MAIC analysis was limited, which indicates presence of substantial heterogeneity across studies. As a result, within the MAIC, the sample size for SKYLARK was reduced by 91.8% for the HAMD-17 analysis and by 60.4% for the EPDS analysis. No MAIC could be

performed for other clinical outcomes. The ITC reported larger changes from baseline in EPDS and HAMD.17 scores for zuranolone compared to SSRIs, indicating that zuranolone treatment offers greater improvements in depression symptoms among PND patients. However, authors of the study acknowledged many limitations with their analysis, including the paucity of evidence for ECM from RCTs, high heterogeneity due to differences in time of study, study design, timepoint of assessment, inclusion and exclusion criteria, and patient characteristics (e.g., baseline HAMD-17 scores) (Meltzer-Brody 2024).

2.10.2 ITC feasibility assessment

To assess the feasibility of conducting an ITC to support this submission, an extensive evaluation of the evidence base supporting the efficacy of SSRI for treatment of PND has been conducted (for details of the identified studies, please see Appendix B.4) and compared with the evidence base supporting zuranolone (i.e., SKYLARK and ROBIN). The following key issues have been identified:

1. **Older studies with small sample sizes.** Small sample sizes in the SSRIs studies would lead to lack of precision in outcomes reported in these studies (e.g., Appleby 1997 (Appleby 1997) and Bloch 2012 (Bloch 2012) had approximately 20 patients per arm vs >90 and >70 in SKYLARK and ROBIN, respectively). Some of the older studies may not align with current clinical practices or the rigorous standards of recent trials like zuranolone studies, diminishing their reliability in comparison. Specifically, how patients were 'treated' with standard of care over the years has changed and different pharmacological and non-pharmacological therapies will have been available to the patients in the studies and is likely to have impacted the inclusion/exclusion criteria (and therefore baseline characteristics, see below) of patients across the trials. Additionally, standards in reporting the findings of the studies have significantly improved over time.

2. **Substantial clinical heterogeneity across the studies, particularly in terms of baseline patient characteristics and study design.** Unlike SKYLARK and ROBIN, most identified studies were conducted in only one or two centres. Although this is to be expected with investigator-initiated studies using treatments off-label, it does impact the generalisability of the findings of these trials. The maximum length of time postpartum during which participants could be recruited varied across studies, ranging from a maximum of 2 months in Bloch et al. (2012) (Bloch 2012) to 12 months in SKYLARK. The specific versions of the questionnaires used to measure HAMD differed across studies, which would affect comparability (due to different

number of questions in each versions, different maximum possible total scores), resulting in difficulty making like for like comparisons of absolute changes from baseline. Ultimately, taken together, these differences would complicate the assumptions necessary for a valid ITC, such as homogeneity, similarity, and consistency, leading to questionable reliability and generalisability of the results.

3. **Difference in baseline severity, assessed by HAMD-17, of the patients included/excluded from the trials.** By design, ROBIN and SKYLARK recruited patients with more severe depression, with reported higher baseline HAMD-17 scores (28.4–28.8) and EPDS scores (19.9–21.1) compared with the identified SSRI studies (HAMD: 13.8–24.7; EPDS: 16.05–18.4). These disparities are critical as they can act as treatment effect modifiers, influencing outcomes and complicating comparability across studies.

4. **Availability of data from early timepoints.** ROBIN and SKYLARK were designed to regularly assess clinical effectiveness in the first 2 weeks (when treatment was given) and through to Day 45. The older studies did not regularly collect data in the first month or two (as efficacy responses were expected to stabilise later in the trial due to delayed onset of action characteristic of SSRIs). Most other studies only reported outcomes at the time of their primary endpoints (~2 or 3 months), when maximum clinical effect is reached (for the SSRIs). To make comparisons of zuranolone versus SSRIs at the early timepoints, imputation of outcomes for the SSRI studies would be required, as data were not collected for SSRIs at as many early timepoints as SKYLARK/ROBIN. Similarly, making comparisons at the study end ('last observations') would require a last observation carried forward approach for the relative treatment differences within study, which likely would not be a reliable assumption. Data imputations always lead to more uncertainty in the results and compromises the generalisability of the findings

5. **Lack of a truly representative 'placebo' common comparator in the evidence base.** An ITC requires a network of trials with common comparators across the studies, such that the relative treatment effects for (in this case) the active treatments are all generally being compared to the same treatment (placebo). The requirement of being the common node in a network is that a patient receiving placebo in one trial would experience similar outcomes had they been included in a placebo arm of one of the other trials. Some natural trial-level variation is expected, but the assumption is that such trial differences cancel out within trial when relative treatment effects are used in ITCs. An assumption that all placebos are similar

across the trials of SSRIs and zuranolone would be questionable. This is illustrated by the differences in placebo outcomes observed across the trials; in SKYLARK and ROBIN the placebo group performs quite a lot better than the other trials in the network. There could be many reasons for this, including (among others): i) changes in the standard of care over time and the other trials beyond zuranolone are substantially older studies; ii) variable use of, and potentially reporting of, background ADTs and non-pharmacological therapies across the studies. Naturally, including studies from different eras (e.g., 1997 for Appleby and 2022 for SKYLARK) is a concern for generalisability and appropriateness of comparison. Where possible, comparison of the absolute outcomes in the placebo arms shows large differences across the studies in the ITCs. It is not possible to specify exactly why these differences occur, but it is likely to be a combination of some/all the issues raised in the points above (including study design, patients included, standard of care, collection/reporting of outcomes, imputations). It is likely that at least some of these challenges/limitations are treatment effect modifiers which cannot be controlled for the ITCs.

2.10.3 Conclusions

In summary, given the severe limitations of the evidence base identified during an ITC feasibility analysis conducted for this submission, and highlighted by Meltzer-Brody et al. (2024) (Meltzer-Brody 2024), as well as lack of any additional relevant RCTs (as demonstrated in the clinical SLR, see Section 2.1), it is not feasible to conduct an ITC, which would allow for robust conclusions and would be considered as sufficient for decision-making. Finally, Biogen believes that the in-trial comparison from the SKYLARK study is relevant and sufficient for the purpose of the NICE decision problem (Section 1.1), as background use of psychological therapies and stable dose ADTs was permitted in both study groups, providing evidence on efficacy of zuranolone plus ECM versus ECM alone.

2.11 Adverse reactions

2.11.1 Overview of AEs in SKYLARK

Zuranolone was generally well tolerated (Deligiannidis 2023b). The majority of treatment-emergent adverse events (TEAEs) were mild (zuranolone 50 mg: 33.7%; placebo: 39.8%) to moderate (zuranolone 50 mg: 29.6%; placebo: 12.2%) in severity (Table 25). On-treatment TEAEs were reported by 60.2% (n=59) and 41.8% (n=41) of patients receiving zuranolone 50 mg and placebo, respectively. Post-treatment TEAEs were reported by 21.4% (n=21) and 27.6% (n=27) of patients receiving zuranolone 50 mg and placebo, respectively. TEAEs

leading to dose reduction were reported for 16 (16.3%) patients in the zuranolone 50 mg group and 1 (1.0%) patient in the placebo group. 15 of these patients (zuranolone, n=14; placebo, n=1) completed the study. TEAEs leading to dose reduction in more than one patient receiving zuranolone 50 mg included somnolence (7.1%, n=7), dizziness (6.1%, n=6), and sedation (3.1%, n=3). The TEAE leading to dose reduction in the placebo group was anxiety. Of the patients who experienced TEAEs leading to discontinuation of the study drug, 4.1% (n=4) were in the zuranolone 50 mg group and 2.0% (n=2) were in the placebo group.

The most frequently reported TEAEs (≥5.0%) in the zuranolone 50 mg group were somnolence (26.5%), dizziness (13.3%), sedation (11.2%), headache (9.2%), diarrhoea (6.1%), nausea (5.1%), urinary tract infection (5.1%), and COVID-19 (5.1%). The most frequently reported TEAEs (≥5.0%) in the placebo group were headache (13.3%), dizziness (10.2%), nausea (6.1%), and somnolence (5.1%) (Table 25). Two serious adverse events (SAEs) were reported, both of them in patients in the zuranolone 50 mg group, one during the treatment course and one during the post-treatment period, and both were considered unrelated to study drug.

Table 25: SKYLARK: Summary of TEAEs (Safety Set)

Category, n (%)	Zuranolone 50 mg (n=98)	Placebo (n=98)
TEAEs	65 (66.3)	52 (53.1)
On-treatment TEAE	59 (60.2)	41 (41.8)
Post-treatment TEAE	21 (21.4)	27 (27.6)
Maximum severity of TEAE		
Mild	33 (33.7)	39 (39.8)
Moderate	29 (29.6)	12 (12.2)
Severe	3 (3.1)	1 (1.0)
TEAE leading to dose reduction[a]	16 (16.3)	1 (1.0)
TEAE leading to treatment discontinuation[b]	4 (4.1)	2 (2.0)
TEAE leading to withdrawal from the study[c]	1 (1.0)	1 (1.0)
SAE[d]	2 (2.0)	0
Death	0	0
TEAEs with ≥5% in any group, through Day 45		
Somnolence	26 (26.5)	5 (5.1)
Dizziness	13 (13.3)	10 (10.2)
Sedation	11 (11.2)	1 (1.0)
Headache	9 (9.2)	13 (13.3)
Diarrhoea	6 (6.1)	2 (2.0)
Nausea	5 (5.1)	6 (6.1)
Urinary tract infection	5 (5.1)	4 (4.1)
COVID-19	5 (5.1)	0

Source: (Deligiannidis 2023b)
[a] The most common adverse events leading to dosage reduction in the zuranolone group included somnolence, dizziness, and sedation.
[b] Discontinuation of blinded treatment occurred during the 14-day treatment course. Patients who discontinued treatment early could complete the remaining study visits through Day 45, unless the patient withdrew consent.
[c] Withdrawal from the study could occur at any time from day 1 through day 45 for any reason. AEs leading to withdrawal from the study also resulted in treatment discontinuation.
[d] One patient (age 36 years) experienced three SAEs on treatment: severe upper abdominal pain on Day 8, which resulted in hospitalisation and resolved on Day 38, and moderate hypertension and peripheral oedema on Day 14, which resolved on Day 15; the patient remained on treatment throughout and the dosage was not reduced; all three SAEs were reported as not related to zuranolone by the investigator. A different patient (also age 36 years) experienced an SAE of perinatal depression secondary to methamphetamine use on day 22 during the posttreatment follow-up; the event was assessed as not related to study drug by the investigator and resolved on day 25.
Abbreviations: COVID-19: Coronavirus-19; SAE: Serious Adverse Event; TEAE: Treatment-Emergent Adverse Event.

[redacted] patients in the zuranolone 50 mg group and [redacted] patients in the placebo group reported ≥1 treatment-related TEAE as assessed by the investigator (Sage Therapeutics 2022a). Of the treatment-related TEAEs reported by ≥5% of patients in either group, somnolence, dizziness, sedation, and diarrhoea were reported by more participants in

the zuranolone 50 mg group compared with the placebo group; headache was reported by more participants in the placebo group compared with the zuranolone 50 mg group.

No loss of consciousness and no clinically significant changes in vital signs, echocardiograms (ECGs), or clinical laboratory parameters were reported (Deligiannidis 2023b). No signals of increased suicidal ideation or behaviour (as assessed by the C-SSRS) were observed. No evidence of withdrawal symptoms (as assessed by the Physician Withdrawal Checklist [PWC-20] and TEAEs) was observed after the discontinuation of zuranolone 50 mg.

2.11.2 Overview of AEs in ROBIN

Zuranolone was generally well tolerated (Sage Therapeutics 2019, Deligiannidis 2021). Overall, [redacted] of patients reported ≥1 TEAE: 60% in the zuranolone 30 mg group and 52% in the placebo group (Table 26). In both treatment groups, most TEAEs were mild (zuranolone 30 mg: [redacted]; placebo: [redacted]) to moderate (zuranolone 30 mg: [redacted]; placebo: [redacted]) in severity. Three patients in the zuranolone 30 mg group experienced severe TEAEs (sedation, n=1; confusional state, n=1; migraine, n=1), and 3 patients in the placebo group experienced severe TEAEs (back pain/muscle spasms, headache/oropharyngeal pain, menorrhagia, n=1 for each).

The most frequently reported TEAEs (≥5%) in the zuranolone 30 mg group were somnolence (15.4%), headache (9%), dizziness (7.7%), upper respiratory tract infection (7.7%), diarrhoea (6.4%), and sedation (5.1%). The most frequently reported TEAEs (≥5%) in the placebo group were somnolence (11%), headache (12.3%), nausea (8.2%), dizziness (5.5%), vomiting (5.5%), abnormal dreams (5.5%), and hyperhidrosis (5.5%) (Table 26). Two SAEs were reported. One patient in the zuranolone 30 mg group experienced a confusional state starting on Day 3 which resulted in hospitalisation, involving inability to remember the exact sequence of the day's events along with sedation. The SAE resolved within 7 hours and the patient completed the treatment period on a reduced dose of 20 mg without any further symptoms during the study. One patient in the placebo group had an SAE of pancreatitis (cholelithiasis pancreatitis) which occurred during the follow-up period. Following a cholecystectomy, the SAE was considered resolved. In the zuranolone 30 mg group, one patient discontinued treatment due to severe intermittent sedation.

Table 26: ROBIN: Summary of TEAEs (Safety Set)

Category, n (%)	Zuranolone 30 mg (n=78)	Placebo (n=73)
≥1 TEAE	47 (60.3)	38 (52.1)
Maximum severity of TEAE		
Mild	[redacted]	[redacted]
Moderate	[redacted]	[redacted]
Severe	3 (3.8)	3 (4.1)
TEAE leading to treatment discontinuation	1 (1.3)	0
TEAE leading to death	0	0
≥1 SAE	1 (1.3)	1 (1.4)
TEAEs with ≥5% in any group		
Somnolence	12 (15.4)	8 (11.0)
Headache	7 (9.0)	9 (12.3)
Dizziness	6 (7.7)	4 (5.5)
Upper respiratory tract infection	6 (7.7)	1 (1.4)
Diarrhoea	5 (6.4)	2 (2.7)
Sedation	4 (5.1)	0
Nausea	3 (3.8)	6 (8.2)
Vomiting	1 (1.3)	4 (5.5)
Abnormal dreams	0	4 (5.5)
Hyperhidrosis	0	4 (5.5)

Source: (Sage Therapeutics 2019, Deligiannidis 2021)
Abbreviations: SAE: Serious Adverse Event; TEAE: Treatment-Emergent Adverse Event.

No clinically significant changes in vital signs, ECG, or clinical laboratory parameters were reported (Sage Therapeutics 2019, Deligiannidis 2021). No signals of increased suicidal ideation or behaviour (as assessed by the C-SSRS) were observed.

2.11.3 Additional safety information

Other warnings and precautions for use of zuranolone may [redacted] (Biogen 2025e).

2.11.3.1 Impaired ability to drive or engage in other potentially hazardous activities

[redacted]

[redacted] (Biogen 2025e).

[redacted] (Biogen 2025e).

2.11.3.2 CNS depressant effect

[redacted] (Biogen 2025e).

2.11.3.3 Pregnancy

[redacted] (Biogen 2025e).

[redacted] (Biogen 2025e).

2.11.3.4 FDA Schedule IV substance designation

The US Drug Enforcement Administration designated zuranolone as a Schedule IV substance due to human abuse potential. This designation was reflected on the FDA label (Food and Drug Administration (FDA) 2022).

In a human abuse potential study, single oral doses of 30 mg, 60 mg, and 90 mg of zuranolone were compared to single oral doses of alprazolam (1.5 mg and 3 mg) and placebo in healthy, nondependent individuals with a history of recreational CNS depressant use. The study demonstrated that zuranolone has dose-dependent abuse potential comparable to alprazolam, a Schedule IV benzodiazepine, and greater abuse potential than placebo on

positive subjective measures of “drug liking”, “overall drug liking”, “take drug again”, “high”, and “good drug effects” (Food and Drug Administration (FDA) 2022).

In the UK, Schedule IV designation has not yet been confirmed, and Biogen are in ongoing conversations with the Home Office on the matter.

2.12 Ongoing studies

There are currently no ongoing studies investigating clinical efficacy and safety of zuranolone in PND.

2.13 Additional evidence

Biogen is committed to further expanding the evidence base supporting use of zuranolone for treatment of adults with PND. As part of this commitment, there are several ongoing studies focusing on different outcomes of zuranolone treatment.

Biogen conducted a retrospective analysis using the [redacted]. Women [redacted] were included in the analysis. [redacted]. The analysis evaluated the proportion of zuranolone users who [redacted]

Among the [redacted] individuals [redacted] The findings were consistent when the sample was limited to [redacted] participants with [redacted] (Biogen 2025b). [redacted] (Sage Therapeutics 2022a, Deligiannidis 2023b)

2.14 Innovation

The innovative nature of zuranolone has been recognised by international and national regulators. In February 2023, zuranolone was given a priority review by the US Food and Drug Administration (FDA), acknowledging the potential of zuranolone to procure a significant improvement in the safety or effectiveness for treatment of PND (FDA 2018, Biogen 2023). This is further supported by receiving an Innovation Passport granted by the MHRA as part of the innovative licensing access pathway (ILAP), demonstrating zuranolone to be an innovative product that addresses area of high unmet need and offering potential benefits to patients in this vulnerable patient population (GOV.UK 2021).

PND is a disease area with a long-standing unmet need, with no pharmacological treatments available that are specifically indicated for the treatment of PND in the UK, and the existing pharmacological treatment landscape limited to off-label ADTs with a multitude of limitations including limited evidence to support safety and efficacy in PND patients, a lack or slow responses to treatment, and poor adherence (Section 1.3.5)**.**

The key innovations for patients relate to the drug's unique posology and novel mechanism of action:

- **Unique mechanism of action:** zuranolone is the first-in-class molecule in the UK that is clearly distinguished from currently available ADTs in its design and mechanism of action. Zuranolone is a synthetic neuroactive steroid closely resembling allopregnanolone – an endogenous neuroactive steroid produced in the brain – which is able to enhance GABAergic neurotransmission at both synaptic and extra-synaptic $GABA_A$ receptors. This broad activity and the dual modulation of phasic and tonic inhibition make zuranolone uniquely positioned to addresses the excitation-inhibition imbalance, which is postulated to be one of the core components of the pathophysiology of PND (Deligiannidis 2023b).
- **Rapid onset of action:** a key limitation with current ADTs is their lack of rapid symptom control, with ADT treatment frequently taking weeks or even up to several months to become effective. The novel mechanism of action of zuranolone translates into rapid onset of clinical effect, with significant improvements in depressive symptoms being observed as early as Day 3 of the 14-day course of oral treatment (Kaufman 2022). The rapid-acting nature of zuranolone addresses the current unmet need, allowing for quick resolution of symptoms. This is particularly important as it contributes to improving patients' functioning, QoL, and allows for development of the bond with the child (Posmontier 2008, Moore Simas 2019, Slomian 2019). Unresolved symptoms can have long-lasting effects on the patient and child (Netsi 2018). Early

intervention and timely symptom resolution may also help to prevent deterioration of the patient's state and admission to MBUs, which are associated with substantial economic burden.

- **Short-course of treatment:** current ADTs require long-term use to exert the therapeutic effect and are associated with multiple safety concerns including sexual difficulties, weight gain, emotional numbness, dry mouth, agitation, drowsiness, and withdrawal effects (Cartwright 2016, Braund 2021). Because of the lack of rapid response and the side effects associated with current treatment options, the discontinuation rates are high (>50%), leading to many patients not benefiting from the treatments (De Crescenzo 2014). Specifically, among PND patients the ADT adherence rate is estimated to be only 15.9% (Pennington 2025). Zuranolone in uniquely positioned to address these unmet needs as it is a short, 14-day treatment course with a favourable tolerability profile, which may encourage more PND patients to initiate and persist with the treatment; as demonstrated in the zuranolone clinical development programme, 91% of patients persisted with zuranolone treatment (Deligiannidis 2021, Deligiannidis 2023b).
- **Potential to improve healthcare equality:** inequalities remain in access to high-quality PMH care in the UK. Those from ethnic minorities, younger women, and those in rural communities may be more likely to have restricted access to PMH services (Schmidt 2006, Jankovic 2020, Maternal Mental Health Alliance 2023). The convenience of short-course oral treatment with zuranolone, could potentially translate into increased health equality by offering a treatment option to patient groups impacted by socio-economic inequalities or who live in remote areas with limited access to full services offered by NHS England and for whom the requirements for frequent visits to specialist care facilities may be restrictive.

Finally, introduction of zuranolone for the treatment of PND would align with the UK policy. Mental health was specifically named as a substantial healthcare challenge in the NHS long term plan published in 2019 (National Health Service 2019), the first-ever government-led Women's Health Strategy for England programme initiated in 2022 (Department of Health & Social Care 2022), and 3-year delivery plan for maternity and neonatal services in 2023 (National Health Service England 2023). The NHS long term plan committed to improving access to and the quality of perinatal mental health care, including providing access to specialist community care for at least 66,000 patients with moderate to severe perinatal mental health needs. The plan committed to increased access to evidence-based specialist care for patients with moderate to severe perinatal mental health difficulties, expanded access to evidence-based psychological therapies within specialist perinatal mental health services, and

increased access to psychological support and therapy through maternity outreach clinics (National Health Service 2019). Additionally, the Women's Health Strategy made several key commitments: ensure patients are treated as equal partners in the perinatal period; £100 million investment for perinatal mental health and parent-infant relationship support; increase provision of specialist community perinatal mental health services; increase access to NHS psychological therapies; build greater knowledge of perinatal mental health among healthcare professionals through training (Department of Health & Social Care 2022).

2.15 Interpretation of clinical effectiveness and safety evidence

2.15.1 Statement of the principal findings from the clinical evidence highlighting the clinical benefits and harms of the technology

The clinical programme shows that zuranolone 50 mg provides rapid symptom resolution and clinically relevant benefits for patients suffering from PND, whilst being well tolerated, and offering a convenient short-course oral administration that may lead to better persistency to complete the treatment and less time with side effects.

In the SKYLARK study, compared to placebo, patients treated with zuranolone 50 mg achieved statistically significant improvement in depressive symptoms at Day 15, as measured by the LS mean change from baseline in HAMD-17 total score (p=0.001). Significant improvements in depressive symptoms were achieved as early as Day 3 (p=0.001), highlighting the rapid onset of action, and were sustained through Day 45 (p=0.007), demonstrating durability of treatment effect. A similar pattern was observed across the secondary outcomes of the SKYLARK study, demonstrating that treatment with zuranolone leads to improvements in symptoms of depression and anxiety, as assessed by EPDS, MADRS, PHQ-9, CGI-S, HAM-A (Deligiannidis 2021, Sage Therapeutics 2022a, Deligiannidis 2023b). The SKYLARK study protocol allowed background use of ADTs and psychological therapies, therefore, clinical evidence from SKYLARK represents zuranolone with ECM versus ECM alone as outlined in the NICE decision problem. The results of the ROBIN study are in line with the principal findings from SKYLARK and provide supportive evidence on efficacy of zuranolone (Sage Therapeutics 2019, Deligiannidis 2021).

In the clinical programme, zuranolone was well tolerated as demonstrated by low discontinuation rates (Deligiannidis 2021, Deligiannidis 2023b). Additionally, most TEAEs were mild to moderate in severity and included (incidence ≥5% and greater than placebo) somnolence, dizziness, diarrhoea, fatigue, nasopharyngitis, and urinary tract infection. No

increase in suicidal ideation or behaviour (as measured by the C-SSRS) was reported in either study.

2.15.2 Strengths and limitations of the clinical evidence base for the technology

The clinical programme for zuranolone (SKYLARK and ROBIN) constitutes the largest clinical studies conducted in patients with PND. Both studies are considered to be well balanced between treatment groups in terms of demographic characteristics, and the groups were ethnically and racially diverse. Additionally, the studies were considered to have an overall low risk of bias (Table 13). While the potential risk of functional unblinding was raised by the FDA (due to imbalance in the number of patients requiring dose reductions due the AEs between the two study arms), the agency concluded that the robust efficacy findings from SKYLARK and ROBIN overcome confounding from this potential bias (Section 2.5) (Food and Drug Administration (FDA) 2022), and provide some of the most substantial results to date for improving depressive symptoms in women with PND.

The evidence base for zuranolone presented in this submission is relevant to the decision problem (Section 1.1) as it provides information on clinically relevant outcomes such as resolution of symptoms of depression (as demonstrated by HAMD-17, MADRS, CGI-S – tools commonly used in clinical research, as well as EPDS, PHQ-9 – tools used in the clinical practice in the UK and recommended by NICE (National Institute for Health and Care Excellence 2020)), anxiety (as demonstrated by EPDS, HAM-A), as well as other benefits that may be relevant for patients suffering from PND, such as sleep quality (as demonstrated by MADRS – Reduced Sleep, HAMD-17 – Insomnia).

Finally, the clinical benefits of zuranolone demonstrated in the clinical trial programme are relevant for both clinicians and patients; the improvements demonstrated in the primary endpoint assessed by investigator-reported HAMD-17 were mirrored by patients' self-reported assessment of their depressive symptoms by the EPDS.

There are some limitations to the clinical evidence base provided by the SKYLARK study. Most ADT trials are associated with some general limitations, including a short follow-up duration that rarely exceeds 8-weeks (Ioannidis 2008). Additionally, as with any clinical trials where certain inclusion/exclusion criteria are applied (e.g., inclusion criteria based on HAMD-17 total score, exclusion of patients with a history of bipolar disorder and/or psychotic disorders), the SKYLARK study population might not represent the broad PND patient population in the clinical setting (Deligiannidis 2023b). Recruited sample in SKYLARK was primarily from the US, although we do not anticipate the pathophysiology of PND would differ

by region (Stewart 2016, Deligiannidis 2023b). The safety of breastfeeding while being treated with zuranolone remains unknown due to study inclusion criteria (Deligiannidis 2023b).

Biogen is aware of the limitations of the data package supporting the broad population (adults with PND) given there is no data demonstrating the efficacy of zuranolone in patients with mild disease. However, Biogen expects that the unique MoA of zuranolone supports its efficacy in all PND patients, irrespective of disease severity. Additionally, zuranolone is expected to initially be prescribed in a specialist setting, mainly by perinatal psychiatrists within the community perinatal mental health services, and therefore, will most likely be used for the treatment of patients with moderate to severe PND.

Finally, although the development of an ITC was explored as supportive evidence, Biogen concluded that it was not feasible to conduct an ITC for two main reasons: 1) the published ITC by Meltzer-Brody et al. (2024) (Meltzer-Brody 2024) highlighted severe limitations of the evidence base; 2) there was a lack of any additional relevant RCTs (as demonstrated in the clinical SLR, see Section 2.1). However, as background use of psychological therapies and stable dose ADTs was permitted in both study groups in the SKYLARK study, the study provides evidence on efficacy of zuranolone plus ECM versus ECM alone. Therefore, Biogen believes that the in-trial comparison from SKYLARK is relevant and sufficient for the purpose of the NICE decision problem (Section 1.1).

3 Cost-effectiveness

3.1 Published cost-effectiveness studies

An SLR was undertaken to identify economic evaluations (cost-effectiveness analyses [CEAs]) relevant to the NICE decision problem. The SLR included critical appraisal of identified CEAs, and assessed the economic model structures, data used to inform cost and resource use, and utility values associated with assessment of treatments used for management of PND. The SLR was conducted iteratively, with the original searches conducted between 23 November and 04 December 2023, and an update conducted between September and October 2024. The searches were restricted to studies from 2013 to date.

The approaches used to identify studies in the review, including methodology, and a full description and quality assessment of studies considered relevant to decision-making in England are provided in Appendix E.2.

The review identified three economic evaluations of interventions for PND, with two of those evaluations being cost-utility analyses for zuranolone in the US (O'Callaghan 2024) and brexanolone in the US (Eldar-Lissai 2020) and the other being a simplistic cost-benefit analysis of non-pharmaceutical interventions in the UK (Bauer 2016a). No relevant NICE technology appraisals for interventions for the treatment of PND were identified. The NICE clinical guideline on PND (CG192) was not considered relevant as it does not include an economic model (National Institute for Health and Care Excellence 2020).

An analysis of the cost-effectiveness of zuranolone versus SSRIs for the treatment of PND in the US was evaluated by O'Callaghan et al., (2024). The model compared zuranolone with SSRIs over an 11-year time horizon to align with the 11-year follow-up available in Netsi et al., (2018) (Netsi 2018). The model used in the analysis was a Markov model with five health states representing depressive symptom ranges based on EPDS scores, and categorised into PND grades of “none”, “minimal”, “mild”, “moderate” and “severe”. Short-term relative effectiveness was based on EPDS evidence from SKYLARK and other trials (Meltzer-Brody 2024). Short-term EPDS effectiveness was linked to long-term patient outcomes using published evidence of EPDS trajectories over time from the ALSPAC cohort (O'Callaghan 2024). The O'Callaghan et al., (2024) paper uses results from ALSPAC data up to 11 years as reported in the Netsi study (Netsi 2018) to categorise patients at the end of the short-term period by using the patient’s EPDS score at day 45 to place them into five EPDS categories. This is a key limitation of the O'Callaghan et al., (2024) paper, as it results in an instantaneous and potentially implausible large shift towards higher EPDS severity levels, particularly for

patients who are near the EPDS limits of each category. In order to link the short-term and long-term outcomes, the proportion of patients in each severity category was calculated using the last observed EPDS score from the ITC and was used to calculate the long-term outcomes to 11-years. Additionally, by using a 5-state Markov model, the O'Callaghan study was unable to granularly capture differences in EPDS levels of this heterogenous patient population as patients were grouped into different severity categories rather than explicitly estimating the EPDS scores in each model cycle.

Eldar-Lissai et al., (2020), evaluated the cost-effectiveness of brexanolone injection versus SSRIs for the treatment of moderate to severe PND in the US over an 11-year time horizon. Similar to the O'Callaghan et al., 2024 (O'Callaghan 2024), the model was a health state transition model using EPDS-based PND severity states were defined as “none”, “minimal”, “mild”, “moderate”, and “severe” (Eldar-Lissai 2020). Clinical and economic outcomes were based on the EPDS scores, with efficacy informed by ITC and real-world evidence from an 18-year longitudinal study. As the long-term data and model structure aligns with O'Callaghan et al., (2024), the same limitations apply; namely the implausible shift to higher EPDS scores at the end of the short-term model, and a lack of granularity when estimating the EPDS scores in each cycle.

Bauer et al., (2016) (Bauer 2016a) report a UK cost-benefit analysis assessing the case for best practice for perinatal mental health care. It included a SLR and explored the costs and benefits of early non-pharmaceutical interventions to prevent or reduce the symptoms of PND in mothers in the UK (Bauer 2016a). As Bauer et al., (2016) only presented a cost-benefit analysis of non-pharmacological interventions, this is not appropriate for this analysis as cost-benefit analysis is not aligned with the NICE reference case.

Table 27 presents a summary list of the published cost-effectiveness studies and Appendix E provides more detail on the three included economic evaluations.

Table 27: Summary list of published cost-effectiveness studies

Study (year)	Summary of model	Patient population (average age in years)	QALYs (intervention, comparator)	Costs (currency) (intervention, comparator)	ICER (per QALY gained)
O'Callaghan (2024) (O'Callaghan 2024) US societal perspective	Cost-utility analysis A 6-state Markov model with five 'alive' states defined by EPDS scores and a dead state. 11-year time horizon.	PND (mean age: 28)	Mother, child and partner: • Zuranolone: 21.981 • SSRIs: 21.898	Total: • Zuranolone: $200,420 • SSRIs: $192,621	$94,741 per QALY.
Eldar-Lissai (2020) (Eldar-Lissai 2020) US payer perspective	Cost-utility analysis A 6-state Markov model with five 'alive' states defined by EPDS scores and a dead state. 11-year time horizon	PND (mean age: 28)	Mother and Child: • Brexanolone: 14.37 • SSRIs: 14.085	Total: • Brexanolone: $110,011 • SSRIs: $79,479	$106,662 per QALY.
Bauer (2016) (PSSRU) (Bauer 2016a) UK payer, societal	Cost-benefit analysis A simple decision tree with health and social care, government and societal benefits applied to efficacy of intervention.	PND (mean age: NR)	N/A	N/A	Each of the interventions included in the economic analysis led to positive net benefits from a societal perspective.

Source: Cited within table.
Abbreviations: ICER: Incremental Cost-effectiveness Ratio; EPDS: Edinburgh Postnatal Depression Scale; N/A: Not Applicable; NR: Not Reported; PND: Postnatal Depression; PSSRU: Personal Social Services Research Unit; QALYs: Quality-adjusted Life Years; SSRI: Selective Serotonin Reuptake Inhibitor; UK: United Kingdom; US: United States.

3.2 Economic analysis

As described above, none of the three identified economic evaluations of PND interventions, identified as part of the SLR, were aligned with the NICE reference case. Of the three studies, only one included zuranolone but incorporated inputs relevant only in the US context (O'Callaghan 2024). The only study with the UK perspective, presented a simplistic cost-benefit analysis of non-pharmacological interventions (zuranolone was not included) (Bauer 2016a), which is not in alignment with the NICE reference case either. Additionally, the model

structures employed in O'Callaghan et al., (2024) and Bauer et al., (2016) were unable to granularly capture the change in EPDS over time, and the use of Netsi data (Netsi 2018) leads to implausible changes in the EPDS trajectories.

Therefore, a *de novo* model-based CEA was developed to evaluate the clinical and cost-effectiveness of zuranolone with ECM versus ECM alone for treatment of PND.

3.2.1 Patient population

The patient population considered in the economic evaluation is adults with PND. This reflects the anticipated licence for zuranolone and aligns with the final NICE scope for this appraisal (see Section 1.1).

3.2.2 Model structure

The CEA conducted for this appraisal comprises a patient-level simulation tracking individual simulated patients' PND (based on EPDS scores), cost, and health outcomes. The EPDS was chosen as the preferred endpoint to assess PND symptoms because it is a commonly used to measure PND in clinical practice, and was the tool used to collect patient outcomes in ALSPAC, which is the best available source of data on the long-term impact of PND on patients. ALSPAC is a transgenerational prospective observational study investigating influences on health of mothers as well as health and development outcomes of children across the life course (Boyd 2013, Fraser 2013). It considers multiple genetic, epigenetic, biological, psychological, social and other environmental exposures in relation to a similarly diverse range of health, social and developmental outcomes. The study included pregnant women residing in Avon, UK with expected dates of delivery between 1st April 1991 and 31st December 1992. A total of 14,833 unique women were enrolled in ALSPAC as of September 2021. The model applies a 29-year time horizon to reflect the duration of data available for the ECM arm in the ALSPAC database.

A patient-level model was deemed more appropriate for this appraisal. However, alternative model structures utilised by other studies were considered at model conceptualisation stage. While the economic evaluation SLR identified two cost-utility analyses for PND which implemented a Markov model (Eldar-Lissai 2020, O'Callaghan 2024), a key limitation of both models was the potentially implausible assumptions linking the short-term trial data with longer-term EPDS trajectories. Both studies used Netsi et al., (2018) (Netsi 2018) to categorise patients at the end of the short-term trial period; however, this results in an instantaneous and implausible shift in health state occupation, when the trial data informing the short-term model reaches its end. While this limitation could be overcome by tracking

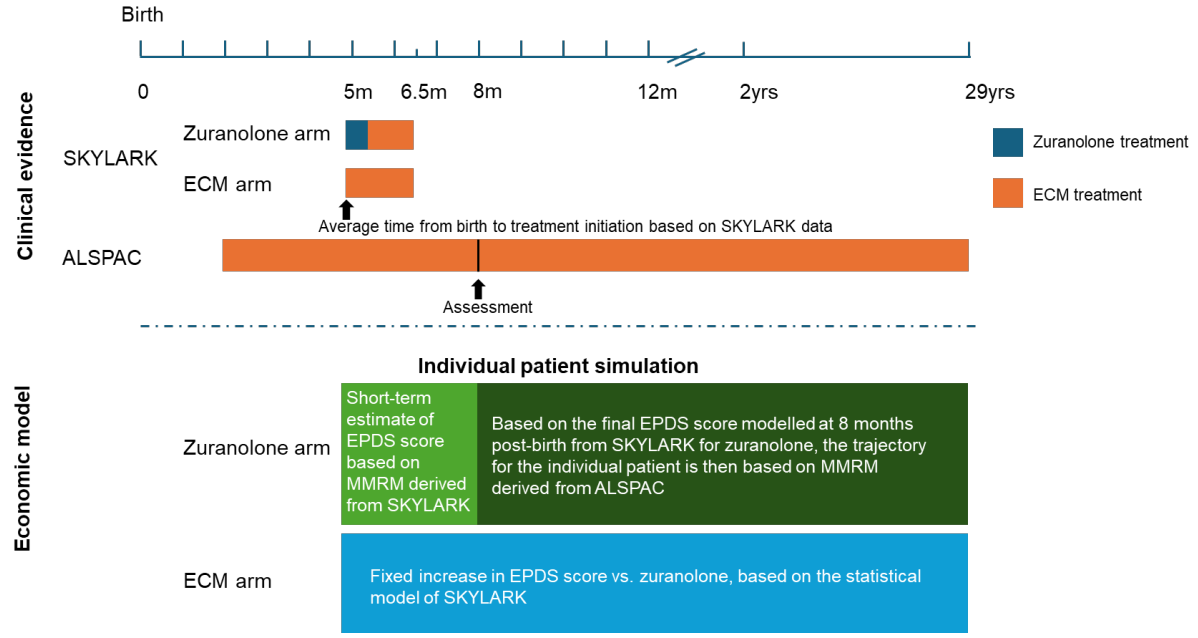
patient history and preparing more granular EPDS categories, this approach would be burdensome to implement within a cohort-level model.

Another challenge with the cohort approach is the potential for non-linearity in outcomes with respect to heterogenous population characteristics. This concern is in line with the NICE decision support unit (DSU) technical support document (TSD) 15 on patient-level models, which cites non-linearity as an important rationale for choosing between a cohort based and patient-level model (Decision Support Unit. 2022). Data from SKYLARK suggests that the patient characteristics of age and disease severity at trial baseline are predictive of the EPDS scores over time. Both are included in the mixed model for repeated measures (MMRM) that was developed to predict EPDS after zuranolone or ECM treatment initiation (see Section 3.3.2). Age is accounted for as a categorical variable with a threshold at ≤24 and >24 years. The decision to categorise age using a binary split at 24 years was based on the pre-specified analysis plan for the SKYLARK study, which divided participants into two subgroups: 18-24 years and 24-45 years. This classification was predefined in the clinical study report (CSR) and was therefore, maintained in the modelling approach for consistency.

Although age was initially tested as a continuous variable, its predictive performance was found to be inferior in comparison to the categorical approach. The binary split at 24 years was ultimately retained as it aligned with the study design and provided better model performance in terms of predicting EPDS outcomes. Using a patient-level model accounts for this and has the additional advantage that patient populations that are different from the study population can be more easily handled than in a cohort approach.

Each simulation produces results for both the zuranolone and ECM arm. The model leverages data from the SKYLARK study in the short-term phase of the model (up to month 3), with data from the ALSPAC database used to inform the long-term EPDS projection for zuranolone. The key strength of the ALSPAC analysis is that it allows for more plausible extrapolations between the short- and long-term EPDS extrapolations (further details in Section 3.3.4) compared to using the Netsi data (as in O'Callaghan et al., (2024)), which requires distributing patients into separate EPDS categories depending on the last observed EPDS value in the short-term model (Section 3.1). For the ECM arm of the model, a fixed increase in EPDS is applied onto the zuranolone EPDS estimations for the duration of the model (Figure 28).

Figure 28: Schematic of data used in the economic model



Abbreviations: ALSPAC: Avon Longitudinal Study of Parents and Children; ECM: Established Clinical Management; EPDS: Edinburgh Postnatal Depression Scale; m: months; MMRM: Mixed Model for Repeated Measures; yrs: years.

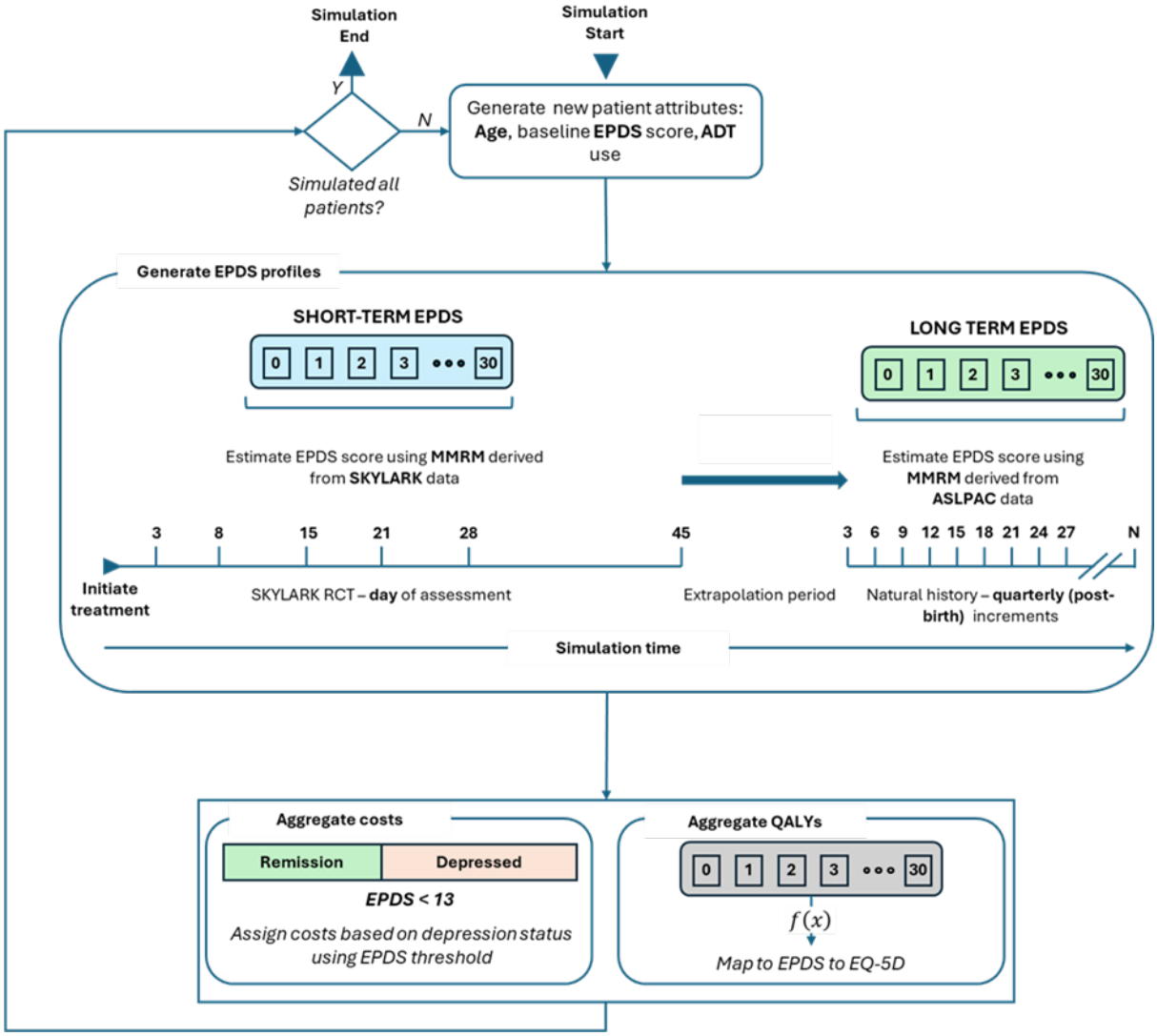
The model base-case assumes a time horizon of 29 years which aligns with the duration of data on ECM available from the ALSPAC analysis. Model time was incremented in intervals that were variable up to three months from model baseline (Table 28), to allow alignment with reporting time points in SKYLARK, thereafter, the cycle length was three months.

Table 28: Model cycle length in the first 90 days

Cycle	Start period (months)	End period (months)	Cycle length (months)
0	0	0.10	0.10
1	0.10	0.26	0.16
2	0.26	0.49	0.23
3	0.49	0.69	0.20
4	0.69	0.92	0.23
5	0.92	1.48	0.56
6	1.48	3.00	1.52

Half-cycle correction was not included in the model as the ASLPAC data showed that there were minimal changes to EPDS scores over time, and therefore, the gradient of change is not large enough to require half-cycle correction. In line with the NICE reference case, costs and health outcomes were discounted annually at 3.5%, and cost-effectiveness was assessed in terms of the cost per QALY gained (National Institute for Health and Care Excellence. 2022). The detailed model structure for the zuranolone arm is presented in Figure 29.

Figure 29: Detailed model overview for zuranolone arm



Abbreviations: ADT: Antidepressant Treatment; ALSPAC: Avon Longitudinal Study of Parents and Children; EPDS: Edinburgh Postnatal Depression Scale; EQ-5D: EuroQoL 5-Dimension; MMRM: Mixed Model for Repeated Measures; RCT: Randomised Controlled Trial.

The model uses an iterative approach within Excel, managed by a visual basic for applications (VBA) loop, to capture individual patient trajectories (Figure 29). This methodology integrates patient-specific factors and treatment effects. Each patient is simulated individually, with their journey through the model guided by baseline characteristics and subsequent predictions (see Section 3.3.1 to Section 3.3.4).

The model starts at randomisation of SKYLARK, i.e., on average 5 months post-birth. Short-term EPDS scores are modelled at specific intervals (days 3, 8, 15, 21, 28, and 45) based on data from the SKYLARK study. Beyond day 45, individual EPDS scores for both treatment arms are extrapolated to month 3 post treatment initiation using the short-term EPDS model (Section 3.3.2), to align with the timing of the 8-month post-birth assessment in the ALSPAC study (Biogen 2024a).

For each patient that is simulated in the zuranolone arm, from the 8-month post-birth point, EPDS scores are projected using a separate long-term prediction model derived from the ALSPAC database, updated quarterly. A fixed increase in EPDS score is applied onto the zuranolone EPDS score to generate the long-term EPDS estimate for patients treated with ECM. The effect of baseline ADT use is incorporated via its inclusion as a binary covariate in the short-term model, allowing for the simulation to apply any prognostic effect that is evident from the SKYLARK data (Section 3.3.2).

In the model, patients are classified as having PND or in being in remission depending on whether their EPDS score has met the remission threshold. In the base case, the remission threshold is <13 EPDS e.g., if a patient has an EPDS of less than 13, they are in remission. While thresholds of <13 and <10 are commonly used in the literature and clinical practice, the threshold of <13 was chosen following discussions with clinicians. A scenario analysis has been conducted using a remission threshold of <10.

Costs are applied throughout the simulation for both zuranolone and ECM patients, varying based on whether the patient meets the EPDS threshold for depressive symptoms (≥13), with detailed cost structures provided in Section 3.5.3. EPDS scores are mapped to SF-6D utility values from the ROBIN study (Section 3.4.2), and quality-adjusted life years (QALYs) are calculated over the simulation period. Mortality is modelled using general population lifetables in the base case. A scenario analysis has been conducted that incorporates excess mortality for depressed patients via standardised mortality ratios.

Table 29: Features of the economic analysis

Factor	Current evaluation, chosen values	Current evaluation, justification
Time horizon	29 years	The ALSPAC database includes data on the EPDS of patients treated with ECM for 29 years
Short-term effectiveness	Post-hoc SKYLARK analysis of EPDS	The SKYLARK study provides the most robust estimates for the short-term EPDS scores for zuranolone with ECM and ECM alone. The ALSPAC database is the best source of long-term projections of patients with PND and uses EPDS as the measurement of PND. Therefore, the use of EPDS from SKYLARK allows for the short-term model to be linked with the long-term ALSPAC analysis.
Long-term EPDS estimation	ALSPAC University of Bristol analysis	The ALSPAC database is a real-world UK data source that allows for an estimate of the EPDS trajectory over a 29-year time horizon based on the EPDS score and age.
Long-term comparative effectiveness	Difference in EPDS score at month 8 months post-birth carried forward	The SKYLARK study provides the most robust estimates for comparative effectiveness of zuranolone with ECM and ECM alone. As RWE indicates that the EPDS score trajectory is dependent on the EPDS score at 8 months, the comparative effectiveness is maintained over the model time horizon
Source of patient utilities	ROBIN SF-6D	The population is well matched to the population of SKYLARK
Source of mother and infant HRU resource use	Petrou 2002	Petrou 2002 provides the most recent estimate of granular HRU for patients diagnosed with PND in the UK.
Source of mother and infant HRU costs	NHS reference costs 2023/2024 and PSSRU unit cost report 2023	NHS specific costs.

Source: Cited within table.
Abbreviations: ALSPAC: Avon Longitudinal Study of Parents and Children; ECM: Established Clinical Management; EPDS: Edinburgh Postnatal Depression Scale; HRU: Healthcare Resource Utilisation; NHS: National Health Service; PND: Postnatal Depression; PSSRU: Personal Social Services Research Unit; RWE: Real-World Evidence; SF-6D: Short Form 6-Dimension; UK: United Kingdom.

3.2.3 Intervention technology and comparators

3.2.3.1 Intervention

The intervention evaluated in this analysis is zuranolone 50 mg (two 25 mg capsules), taken once daily for 14 days (Biogen 2024g). In SKYLARK study and in practice, dose reductions were and are permitted. In the event when dose reduction is needed, an additional pack of zuranolone is required to reduce the dose from 50 mg (two 25 mg tablets) to 40 mg (two 20 mg tablets). Therefore, the model applies an additional cost for a percentage of patients who require a dose reduction from 50 mg to 40 mg (Section 3.5.2).

Concomitant use of ADTs and physiological therapies was permitted in SKYLARK study (Deligiannidis 2023a), reflecting the expected clinical practice and the place of zuranolone

within it. The analysis captures expected costs of concomitant ADTs in line with SKYLARK data, but other healthcare resource use assumptions are based on external sources (see Section 3.5.3).

3.2.3.2 Comparators

The primary comparator evaluated in this analysis is ECM. The effects of ECM are assumed to be captured by evidence from the control arm of SKYLARK. There is no treatment acquisition cost assumed for ECM beyond that of concomitant ADTs (see Section 3.5.2).

3.3 Clinical parameters and variables

3.3.1 Patient characteristics

At model initialisation, patient age and EPDS scores are independently sampled from normal distributions, with a mean age of 30.50 years and a mean EPDS score of 21.1, consistent with the patient population described in SKYLARK (Table 30) (Deligiannidis 2023a). Standard deviations define the variability in these distributions, while SEs are used to vary the means in the probabilistic sensitivity analysis (PSA). Baseline ADT use is determined via Bernoulli sampling, with SEs similarly used to vary the mean in the PSA. ADT use is assumed independent of patient age and EPDS, validated by exploratory analysis suggesting the demographics of the ADT and non-ADT groups are sufficiently similar (see Appendix K.2). These patient characteristics are used to inform the trial based predictive equations for EPDS and subsequently, for patient age, the utility estimations.

Table 30: Patient baseline characteristics, SKYLARK (Full Analysis Set)

Variable at baseline	Mean	SD	SE
Age in years	30.50	5.93	0.42
EPDS score*	21.10	3.68	0.26
ADT at baseline (%)	15.30	--	0.03

Source: (Sage Therapeutics 2022a, Deligiannidis 2023a)
*: zuranolone arm only
Abbreviations: ADT: Antidepressant Therapy; EPDS: Edinburgh Postnatal Depression Scale; SD: Standard Deviation; SE: Standard Error.

3.3.2 Short-term EPDS simulation

To model short-term EPDS in the economic model, statistical analysis of the SKYLARK data was performed, with a summary of the methodology presented in sections below. For details on the EPDS data collection, see Section 2.3 and Section 2.4.

3.3.2.1 Patient selection for analysis

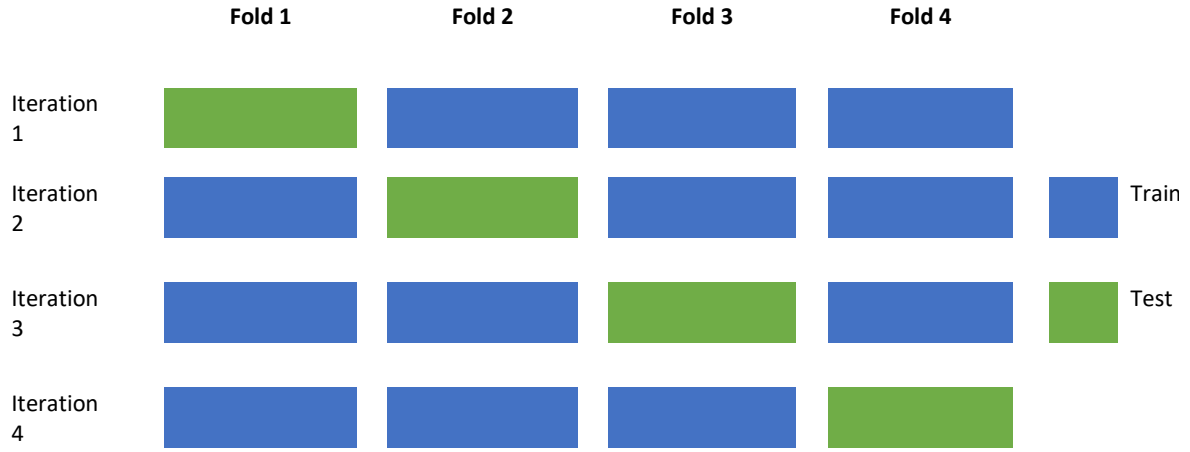
The analyses were based on patients in the Full Analysis Set of SKYLARK. This included patients with a baseline EPDS score and at least one post-baseline score, ensuring that only patients with sufficient data were included (Table 11).

3.3.2.2 Predictive modelling for the economic model

To support the economic model, a reliable predictive model of change in EPDS was required to estimate and extrapolate short-term EPDS values. To assess predictive accuracy of the models, a common k-fold cross-validation approach was used to evaluate the model performance. For this approach, the dataset was split into four folds, i.e., subsets, preserving, as closely as possible, the randomisation process by balancing zuranolone with ECM and ECM patients within each subset. The use of four subsets was determined based on the SKYLARK sample size and the expected number of covariates in the regression models. Given the relatively small patient population in SKYLARK, separate confirmation sets were not used.

In each iteration of the cross-validation process, one subset was omitted and reserved for testing, and the models were trained on, i.e., fitted to, the remaining subsets. This process was repeated four times, ensuring that a different subset was omitted on each iteration, so that all subsets were used for testing once (Figure 30). At each iteration, the performance of the predictive models was assessed by comparing the average predictions of the fitted models to the observed data from the omitted subset at day 3, 8, 15, 21 and 45 using root mean square error (RMSE), mean absolute error (MAE), and mean error (ME). This iterative approach ensured a robust evaluation of predictive accuracy of the short-term EPDS model.

Figure 30: 4-fold cross validation strategy



EPDS data in the SKYLARK study were measured repeatedly over time, following the pre-specified schedule. To account for the clustered nature of the data (where individual patients contributed multiple EPDS observations), linear MMRMs with random intercepts were employed. MMRMs are particularly well-suited for analysing time-varying covariates, such as time-on-treatment, which is relevant in SKYLARK, as patients received treatment for 15 days followed by a monitoring period without active intervention (up to Day 45).

In these models, the EPDS score served as the dependent variable. A random intercept was included to account for the correlation of responses within patients, with visit number used as the repeated measure. An unstructured covariance matrix was applied to model the variance of the random effect component, ensuring flexibility in capturing the variability of responses over time.

Covariates were included in the regression to control for relevant patient characteristics. These variables were selected based on analyses from the clinical study report and the requirements of the economic model (Table 31). To refine the analysis, univariate models were initially tested using the variables identified in the literature.

Table 31: Summary of variables included in the regression model development

Variable	Characteristics
Age of patients	Categorical (≤24 vs >24)
Baseline ADT use	Categorical (Yes vs no)
Baseline HAM-A	Numerical (0-56)
Time	Numerical (months)
Treatment arm	Categorical (zuranolone vs ECM)
Indicator variable for on treatment vs off-treatment period*	Categorical (days 1-14 vs days 15-45)

* This applies only for the zuranolone arm
Abbreviations: ADT: Antidepressant Treatment; ECM: Established Clinical Management; HAM-A: Hamilton Rating Scale for Anxiety.

For all univariate models, baseline EPDS score was included as an independent covariate to control for differences in baseline EPDS scores. This variable was rebased to represent a measure of patient frailty relative to the rest of the population whereby a value of 0 represented average baseline EPDS for the population.

Based on the findings from the univariate models, multivariate models were then developed using the EPDS score as the dependent variable. The Collett framework was applied to guide variable selection, following these steps (Collett 2023):

1. **Univariate Analysis:** Variables from the univariate models were evaluated using a pre-specified p-value threshold of 0.10 to identify those with potential associations with the EPDS score.

2. **Backward Selection:** Variables identified in step 1 were included in a backward selection process, with a stricter p-value threshold of 0.05 used to refine the model.

3. **Forward Selection:** Using the model from step 2, a forward selection process was conducted with a p-value threshold of 0.05. This step included variables not initially found to be significant in the univariate analyses.

4. **Interaction Effects:** Interaction terms, such as the interaction between time and the on-treatment indicator, were added during a final stepwise variable selection process.

Both time and baseline EPDS score were included in all models, regardless of their statistical significance. Time was retained due to its critical role in meeting the requirements of the economic model, while baseline EPDS score was essential for ensuring interpretability of results. ADT use was also retained in the base case as it allowed the ECM arm in the economic model to better reflect current ECM. A separate model that allowed the covariate for ADT use to be free, i.e. drop out of the model during the covariate selection process was also fitted as statistical analysis suggested that the variable was not significant. The results of this alternative model are presented in Appendix K.2.

Table 32 summarises the possible model specifications, excluding interaction effects, under the different scenarios evaluated.

Table 32: Possible specifications of the models investigated

State	Base case model	Alternative model
Baseline EPDS score	Forced	Forced
Time	Forced	Forced
Baseline ADT use	Forced	Free
Age of patients	Free	Free
Baseline HAM-A	Free	Free
Treatment arm	Free	Free
Indicator for on-treatment period	Free	Free

Abbreviations: ADT: Antidepressant Treatment; EPDS: Edinburgh Postnatal Depression Scale; HAM-A: Hamilton Anxiety Rating Scale.

3.3.2.3 Short-term EPDS simulation descriptive statistics

The change in EPDS total scores over time, stratified by treatment arm from SKYLARK are presented in Section 2.6.1.3 (Figure 14). Briefly, for both ECM- and zuranolone-treated patients, EPDS scores decreased throughout the study period. A consistently lower EPDS was observed for zuranolone-treated patients from Day 8, with the EPDS difference between treatment arms maintained after Day 15 when patients stopped receiving treatment.

Table 33 summarises compliance data for EPDS assessments in SKYLARK Completion rates reflect the proportion of questionnaires completed by patients at each visit among those who were alive, still followed up in the trial, and had provided a baseline EPDS record. Compliance rates indicate the proportion of completed questionnaires that were usable for analysis (e.g., where all EPDS dimensions were fully completed). Overall, the completion and compliance rates were high throughout the study.

Table 33: EPDS compliance data across treatment arms in SKYLARK

Visit	Completion – Zuranolone N (%)	Completion – ECM N (%)
Day 1	[redacted]	[redacted]
Day 3	[redacted]	[redacted]
Day 8	[redacted]	[redacted]
Day 15	[redacted]	[redacted]
Day 21	[redacted]	[redacted]
Day 28	[redacted]	[redacted]
Day 45	[redacted]	[redacted]

Abbreviations: ECM: Established Clinical Management; EPDS: Edinburgh Postnatal Depression Scale.

3.3.2.4 Base case model selection – forced ADT use

Table 34 presents the results of the models obtained in each of the four iterations, In each iteration, a different subset was omitted, and the models were fitted to the remaining subsets using the Collett framework outlined above to guide variable selection. The average predictions of the fitted models were then compared to the observed data from the omitted subset. This process was repeated four times, with each row in the table showing the final model specification achieved in each iteration along with the average RMSE, MAE and ME of the predictions compared to the omitted subset. The base case model specification was then selected as the model that achieved the lowest overall errors, which was the model from the third iteration as it achieved the lowest RMSE and MAE out of the four iterations.

Table 34. Best fitting base case model for each iteration of the 4-fold cross-validation including interaction terms

Iteration	Specification	RMSE	MAE	ME
1	Baseline EPDS + Time + ADT use + Treatment arm + On-treatment indicator + Age group + Time x On-treatment	[redacted]	[redacted]	[redacted]
2	Baseline EPDS + Time + ADT use + On-treatment indicator + Age group + Time x On-treatment	[redacted]	[redacted]	[redacted]
3	**Baseline EPDS + Time + ADT use + Treatment arm + On-treatment indicator + Age group + Time x On-treatment**	[redacted]	[redacted]	[redacted]
4	Baseline EPDS + Time + ADT use + Treatment arm + On-treatment indicator + Age + Time x On-treatment	[redacted]	[redacted]	[redacted]

Bold model indicates best performing among base case models.
Abbreviations: ADT: Antidepressant Treatment; EPDS: Edinburgh Postnatal Depression Scale; MAE: Mean Absolute Error; ME: Mean Error; RMSE: Root Mean Square Error.

The best-fitting model was then fitted to the full SKYLARK dataset, and its fixed-effect coefficients are presented in Table 35. Notably, the coefficient associated with ADT use is positive, which might initially appear counterintuitive, as patients treated with ADTs were observed to have higher EPDS scores. It is important to emphasise that the study was not designed to assess the causal effect of ADTs use on EPDS, and no causal inference should be drawn from this association. Importantly, this coefficient is not statistically significant at the 0.05 threshold and should not be interpreted as evidence of an association. A model excluding this fixed effect was also fitted as an alternative to the base case and the results are presented in Appendix K.2.

Table 35. Fixed-effect coefficients from the best fitting base case model

Visit	Number of patients	Number of observations	Estimates	SE	p-value
Intercept	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Baseline EPDS	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Time (days)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
ADT use – Yes	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Age group – >24 years	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Treatment arm – zuranolone	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
On-treatment period – Days 1-14	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Time x on-treatment period interaction	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Abbreviations: ADT: Antidepressant Treatment; EPDS: Edinburgh Postnatal Depression Scale; SE: Standard Error.

Figure 31 presents the average of the model predictions of EPDS per patient against the observed EPDS score by treatment arm through time. The model predictions have good face validity against the observed data. Predictions for Day 8 are slightly higher than the corresponding observed scores, but the difference is of the same magnitude for each treatment arm.

Figure 31. Average best-fitting base case model predictions per patient against observed EPDS score by treatment arm through time

The placebo arm corresponds to ECM, and the zuranolone arm corresponds to zuranolone + ECM.
Abbreviations: ECM: Established Clinical Management; EPDS: Edinburgh Postnatal Depression Scale.

3.3.2.5 Alternative model selection – free ADT use

In comparison to the base case model where ADT use was retained (allowing the ECM arm in the economic model to better reflect current ADT usage), an alternative model that allowed the covariate for ADT use to be free, and, therefore, removed from the model due to statistical insignificance, was also estimated. This model estimated similar coefficients for the covariates excluding ADT use, and estimated similar predictions to the base case model with only marginal improvements in RMSE and ME. The goodness-of-fit in terms of Akaike Information Criterion (AIC) was slightly improved with the inclusion of ADT use but the Bayesian

Information Criterion (BIC), which penalises model complexity, was better for the alternative model. The full results of this model are presented in Appendix K.2.2

3.3.2.6 Conclusions

The alternative model specification with fewer constraints on factor selection achieved the lowest BIC and offered marginally better predictive performance for EPDS scores in terms of RMSE and ME. The model including ADT use, however, better aligned with the needs of the economic model, allowing ECM to better reflect current standard of care, and for this reason it was selected for the base case.

Overall, the predictive performance of the models was satisfactory, capturing the EPDS trajectory over time effectively despite the slight overestimation at Day 8. The selected models provided a robust basis for accurately predicting EPDS changes. Whilst more complex models could have been employed to potentially improve the Day 8 prediction (e.g., through the inclusion of non-linear terms), the additional complexity was considered unwarranted given the overall predictive accuracy of the estimated linear MMRM. Simpler models were also ruled out due to the requirements of the patient simulation approach in the economic model and the clustered nature of the data.

3.3.3 Long-term EPDS simulation

Within the economic model, long-term EPDS is predicted using a model based on data from the ALSPAC cohort. ALSPAC data are only available for patients treated with ECM, however, as zuranolone treatment is only for 14 days and the impact in the short-term will predict long-term outcome, the long-term trends are assumed to be the same in both treatment arms. The fixed treatment effect from the short-term EPDS simulation is assumed to also apply in the long-term. In the economic model this is achieved by projecting EPDS for zuranolone with ECM patients, then adding the fixed treatment effect to the projected EPDS scores to estimate the EPDS for ECM (Figure 28 and Table 35).

The long-term model based on ALSPAC data is used to predict EPDS in the economic model from 90 days post treatment initiation, which on average is equal to 8 months post-birth, coinciding with the 8-month data point from ALSPAC. There is one earlier data point in the ALSPAC study (2 months post-birth), which coincides with a time point prior to the average timepoint (5 months post-birth) that patients received treatment in SKYLARK. Individuals from the ALSPAC study were included in the analysis set if they had an EPDS≥13 at the earliest postnatal time point of 2 months as well as non-missing data at 8 months, resulting in a sample size for the analysis of [redacted].

The approach used in the base case involves a *de novo* analysis (commissioned by Biogen, see Appendix K.3.1) of the current ALSPAC dataset, which allows for a granular estimation of the long-term EPDS of patients based on their baseline characteristics for a follow-up of 29 years, allowing for a greater understanding of how EPDS changes with time. It also explores the potential of covariates predictive of long-term EPDS score, and it uses the full follow-up from the ALSPAC cohort (up to year 29), facilitating longer-term extrapolation of EPDS scores over time.

The long-term extrapolations of EPDS were generated using a similar approach to the short-term extrapolations. Predictions were generated by fitting an MMRM to the ALSPAC data including the same baseline covariates as considered for the short-term EPDS models fitted to SKYLARK data (Section 3.3.2). Of note, in the analysis of ALSPAC data, age of the mother was not found to have a significant association with the EPDS score. However, for consistency with the short-term predictors, one model including age and another model excluding age were considered in the analyses. Predictive performance of the models was evaluated by cross-validation, using a split of 80% of the sample used as the training sample and the remaining 20% as the testing sample, which is a common way to split datasets for cross-validation purposes. Predictive performance was determined using the same estimators presented for the short-term analysis, i.e. RMSE, MAE and ME. The model without age was chosen for the base case as it had better face validity on its predictions of EPDS through time versus the observed data and achieved noticeably lower ME and marginally lower RMSE and MAE as presented in Table 36.

Table 36: Summary of the predictive performance of the MMRM on ALSPAC data

Model	RMSE	MAE	ME
Model with age	[redacted]	[redacted]	[redacted]
Model without age	[redacted]	[redacted]	[redacted]

Abbreviations: ALSPAC: Avon Longitudinal Study of Parents and Children; MAE: Mean Absolute Error; ME: Mean Error; MMRM: Mixed Model for Repeated Measures; RMSE: Root Mean Square Error.

The coefficients associated with the *de novo* MMRM model without age are shown in Table 37. This model allows for the prediction of EPDS over the long-term by using the predicted EPDS score at 8 months from the short-term predictions to inform the EPDS score at subsequent time points. Goodness of fit of the predictions against the observed data from the fitting samples are presented in Figure 32. Numerical variations were noted, but the predictions are generally consistent with the observed data when inspected visually, taking into consideration the confidence and credible intervals of the observations and predictions respectively.

Table 37: Long-term depression score (EPDS) coefficients using MMRM

Model	Coefficients	Estimate	SE	p-value
Random effects (SD)	Intercept	[redacted]	[redacted]	[redacted]
Fixed effects (value)	Intercept	[redacted]	[redacted]	[redacted]
	EPDS at 8 months	[redacted]	[redacted]	[redacted]
	Time	[redacted]	[redacted]	[redacted]
	Time^2	[redacted]	[redacted]	[redacted]

Source: (Biogen 2024a)
Abbreviations: EPDS: Edinburgh Postnatal Depression Scale; MMRM: Mixed Model for Repeated Measures; SD: Standard Deviation; SE: Standard Error.

Figure 32: Average of patients' EPDS prediction against average observed EPDS in the testing sample through time

Abbreviations: EPDS: Edinburgh Postnatal Depression Scale.

3.3.4 Efficacy of treatments

At 3 months (8 months postpartum), EPDS score of a simulated zuranolone patient transitions to the long-term model for further projection. This section describes the approach, assumptions and data involved.

Between Day 45, which represents the last data point available from the short-term dataset (SKYLARK), and 8 months, which is the baseline time point in the long-term dataset (ALSPAC), each patient's EPDS score is extrapolated using the short-term EPDS model. The short-term model can be written as

$$EPDS_i(t) = b_i + \beta_0 + \beta_1 EPDS_i(0) + \beta_2 ADT\ use + \beta_3 treatment\ arm + \beta_4 ontreatment + \beta_5 age + \beta_6 t,$$

where b_i is the random intercept for individual i, β_0 is the fixed effect intercept, $EPDS_i(0)$ is the baseline EPDS for individual i and $\beta_1 - \beta_6$ are the coefficients for each covariate estimated by the short-term MMRM. The EPDS for individual i at 8 months can, therefore, be estimated by setting t equivalent to 8 months in the above equation and setting the covariates to reflect the individual's characteristics. The intercept and fixed effects coefficients for the short-term model are presented in Table 35.

Long-term EPDS assumptions on both the zuranolone and ECM arms are informed in the base case by the ALSPAC data and applied for the duration of the model time horizon. Between 3 months (8 months postpartum) and 29 years, the estimated change in EPDS from one time period to the next is predicted for zuranolone with ECM by the long-term EPDS model. The fixed treatment effect estimated in the short-term model (Table 35) is then applied to the zuranolone predictions to generate EPDS values for ECM. The change in EPDS from one period to the next is estimated from the long-term MMRM using the equation

$$EPDS_i\ (t_1) - EPDS_i(t_0) = \beta_2(t_1 - t_0) + \beta_3(t_1^2 - t_0^2),$$

Where t_1 and t_0 are time periods, and β_2 and β_3 are the coefficients for the time and time^2 covariates, respectively, estimated by the long-term MMRM. The coefficients from the long-term MMRM model are presented in Table 37. Applying the change in EPDS from one period to the next gives more plausible extrapolations than using the long-term MMRM to predict the absolute value of EPDS. This is because estimating separate MMRMs for the long-term and short-term assumes the random intercepts are independent between each model, but correlated within each individual, which is not a plausible assumption. When taking the change from one period to the next, however, the random effect cancels out, resulting in more plausible predictions of EPDS.

3.3.5 Adverse events

The most common any grade AEs (occurring in ≥5% of patients and higher in zuranolone group than ECM group) observed during the SKYLARK study among patients treated with zuranolone were somnolence, dizziness, diarrhoea, fatigue, and urinary tract infection (Biogen 2024g). Events are included in both zuranolone and ECM arms of the model. Events in SKYLARK were reported as the sum of on-treatment (days 1-14) and off-treatment (days 15-45) and are reported as probabilities (Table 38) (Biogen 2024g). As SSRIs are included in both arms at the same percentage, AEs occurring with sertraline and fluoxetine have not been included in the economic model.

Table 38: Adverse reactions that occurred in ≥5% of patients with PND treated with 50 mg of zuranolone and greater than in patients treated with ECM

AEs (≥5% of patients)	Zuranolone N = 98		ECM N = 98		Source
	Mean (SE)	N	Mean (SE)	N	
Somnolence[a]	36% (0.05)	35	6% (0.02)	6	Zuranolone prescribing information (Biogen 2024g). Assumes UK label is the same as the FDA.
Dizziness[b]	13% (0.03)	13	9% (0.03)	9	
Diarrhoea	6% (0.02)	6	2% (0.01)	2	
Fatigue[c]	5% (0.02)	5	2% (0.01)	2	
Urinary tract infection	5% (0.02)	5	4% (0.02)	4	

Source: Cited within table.
[a] Somnolence includes sedation and hypersomnia; [b] Dizziness includes vertigo; [c] Fatigue includes asthenia.
Abbreviations: AE: Adverse Event; ECM: Established Clinical Management; FDA: United States Food and Drug Administration; PND: Postnatal Depression; SE: Standard Error; UK: United Kingdom.

3.3.6 Mortality

In the base case, mortality is applied as per the general population from lifetables published by the Office for National Statistics (ONS) (Office for National Statistics. 2024), as there is no direct evidence that treatment with zuranolone impacts mortality.

In a scenario analysis, an adjustment is made for higher mortality rates in patients with depressive disorders. For this, an increased mortality is applied when a patient has an EPDS score of ≥13. The source for the standardised mortality ratio is a publication by Angst et al. (1999) (Angst 1999), a literature review on the mortality and suicide risk in patients with MDD. The authors identified 11 studies that reported standardised mortality ratios for patients with depressive disorders compared to the general population, with the results ranging between 1.5 and 2.0 (Angst 1999). For the scenario analysis, a ratio of 2.0 is applied in the economic model quarterly, dependent on whether a patient has an EPDS score above or below the threshold set for depression/remission (remission: EPDS <13 in the base case).

3.4 Measurement and valuation of health effects

No generic, preference-based measure of HRQoL was collected in SKYLARK. An economic SLR (see Appendix F and Section 3.4.3) and ad-hoc literature searches were performed to identify approaches to relate EPDS to HRQoL. Utility data to inform the model was derived from the ROBIN study of zuranolone 30 mg (another capsule formulation with increased bioavailability) which collected SF-36 responses (Deligiannidis 2021).

3.4.1 Health-related quality-of-life data from clinical trials

SF-36 data from patients in the ROBIN study are used to derive SF-6D utility values (see Section 3.4.2) to populate the economic model. The population is well matched to the population of SKYLARK, which makes it a good dataset to derive utility data. The ROBIN study is described in detail in Section 2.3.1.2.

3.4.2 Mapping

In terms of HRQoL data analysis, SF-36 data from ROBIN was converted to SF-6D utility values and then values from each arm were pooled, the mean and SEs were estimated using the UK (2002) tariff (Brazier 2002). The mean SF-6D utility was derived for five EPDS groupings. A summary of the mean and standard errors utility values derived from pooling both arms from the ROBIN study are provided in Table 39.

Table 39: Summary of utility values (SF-6D from ROBIN)

EPDS score	Mean	SE
EPDS 0	0.935	0.020
EPDS 1 to 6	0.795	0.012
EPDS 7 to 13	0.659	0.008
EPDS 14 to 18	0.573	0.008
EPDS 19 to 30	0.520	0.006

Source: (O'Callaghan 2024).
Abbreviations: EPDS: Edinburgh Postnatal Depression Scale; SE: Standard Error; SF-6D: Short-Form Six-Dimension.

3.4.3 Health-related quality-of-life studies

An SLR was undertaken to identify studies reporting health state utility values (HSUVs) in patients with PND. The SLR identified studies of patients aged ≥18 years carrying a child through pregnancy and diagnosed with PND of any severity. Economic evaluations and health technology assessments (HTAs), utility elicitation studies, trials and other studies collecting and reporting utilities (case reports were excluded), and retrospective studies were included. Eligible outcomes included EQ-5D data (both EQ-5D 3L and EQ-5D 5L), SF-6D, 15D, health utilities index (HUI), and any other preference-based utility data. Studies reporting utilities that have been mapped from other tools (e.g. SF-36) were also eligible. The original literature search was conducted between 27 November 2023 and 4 December 2023, and updated on 03 October 2024, with studies from 2003 to date. The SLR identified five studies, however none of the studies mapped EPDS scores to utility values and therefore could not be leveraged in the economic model. The full search strategy, methodology and summary of the results are summarised in Appendix F.

3.4.4 Adverse reactions

Disutility values were applied in the economic model for AEs sourced from UK catalogues to align with NICE reference case and UK base-case. Sullivan (2011) and Nafees (2017) has been accepted for AE disutility sources for fatigue and diarrhoea, respectively, in a previous NICE submission (TA860) (National Institute for Health and Care Excellence 2023b), while the disutility from UTI has been sourced from data reported in an RTC commissioned by the National Institute for Health and Care Research (Pickard 2012). It has been assumed that dizziness has the same disutility as nausea. In the base case, annual disutility values were applied in each cycle that these events occur (Table 40), i.e., until cycle 3 of the model, corresponding with the zuranolone treatment duration of 14 days.

Table 40: Summary of annual disutility values applied for treatment AEs

Disutility	Mean	SE	Source
Somnolence	0.085	0.009	(O'Callaghan 2024)
Dizziness	0.025	0.005	Assumed to be the same as nausea (Nafees 2017)
Diarrhoea	0.073	0.017	(Sullivan 2011)
Fatigue	0.041	0.008	(Nafees 2017)
Urinary tract infection	0.006	0.002	(Pickard 2012)

Source: Cited within table.
Abbreviations: AE: Adverse Events: SE: Standard Error.

3.4.5 Health-related quality-of-life data used in the cost-effectiveness analysis

A summary of the utility values used for the present CEA is presented in Table 41.

Table 41: Summary of utility values for cost-effectiveness analysis

State	Mean	SE	Reference in submission (section and page number)	Justification	Source
ROBIN EPDS utility					
EPDS 0	0.935	0.020	Section 3.4.2 page 115	The population in the ROBIN trial is well matched to the population of SKYLARK	(O'Callaghan 2024)
EPDS 1 to 6	0.795	0.012			
EPDS 7 to 13	0.659	0.008			
EPDS 14 to 18	0.573	0.008			
EPDS 19 to 30	0.520	0.006			
AE disutility					
Somnolence	0.085	0.009	Section 3.4.4, page 116	Utility decrements sourced from UK catalogues to align with NICE reference case and UK base-case	(O'Callaghan 2024)
Dizziness	0.025	0.005			Assumed to be the same as nausea (Nafees 2017)
Diarrhoea	0.073	0.017			(Sullivan 2011)
Fatigue	0.041	0.008			(Nafees 2017)

Abbreviations: AE: Adverse Event; EPDS: Edinburgh Postnatal Depression Scale; NICE: National Institute for Care and Health Excellence; SE: Standard Error; UK: United Kingdom.

3.5 Cost and healthcare resource use identification, measurement and valuation

3.5.1 Resource identification, measurement and valuation studies

An SLR was undertaken to identify studies reporting costs and healthcare resource use associated with PND. The SLR identified studies of patients aged 18 years and older carrying a child through pregnancy and diagnosed with PND of any severity, published since 2003. Eligible studies included burden/cost of illness studies, studies reporting resource utilisation and costs, observational studies, and economic evaluations and HTAs. Eligible outcomes included monetary costs, specific non-monetary costs (productivity losses for patients/parents/caregivers reported as absenteeism, presenteeism, or days lost from work), and non-monetary healthcare resource use associated with PND.

The original literature search was conducted between 27 November 2023 and 4 December 2023 and updated on 03 October 2024 to ensure that all evidence from 2003 until 03 October 2024 was included. The search did not identify any UK source with disaggregated costs that would be relevant to the model, The full search strategy and details of the process and

methods used to identify and select the cost and healthcare resource studies relevant to the technology being appraised as well as the results are summarised in Appendix G.

3.5.2 Intervention and comparators' costs and resource use

Acquisition costs for zuranolone and SSRIs are shown in Table 42. For zuranolone and ECM, the cost of concomitant SSRIs is included, with 15.3% of the modelled population having concomitant use as per SKYLARK (Table 9). Based on analysis of CPRD data, it is assumed that [redacted] of this will be sertraline and [redacted] of it will be fluoxetine (Biogen 2025a). It is assumed that the treatment duration for both fluoxetine and sertraline is 6 months, and that the mean dose for sertraline is 67.5 mg/day (Bloch 2012) and for fluoxetine 20 mg/day (NHS England. 2022). Treatment duration, dosing, and costs are assumed to be the same as for the ECM only comparator arms in the model as the add-on ECM in the zuranolone arm. All treatments listed below are taken orally and therefore it is assumed that no administration costs apply. In SKYLARK, [redacted] of zuranolone patients started on a 50 mg dose, with the option for dose reduction to 40 mg (further details below). Zuranolone is expected to initially be prescribed by perinatal psychiatrists within the community perinatal mental health services.

Table 42: Treatment acquisition costs

Drug	Defined dose	Tablet dose	Pack size	Cost per pack	Source	Treatment duration[a]	Net cost per treatment course
Zuranolone							
Zuranolone[c]	50mg	28 x 25mg	28	[redacted]	Biogen	14 days	[redacted]
	40mg	28 x 20mg	28	[redacted]	Biogen	14 days	[redacted]
SSRIs							
Sertraline[d]	50mg	28 x 50mg	28	£0.56	eMIT national database (GOV.UK. 2024)	6 months	£4.93
Fluoxetine[d]	20mg	30 x 10mg	30	£0.40		6 months	£2.46

Source: Cited within table.
[a]Treatment duration: 14 days for zuranolone at 50mg/day based on treatment protocol in SKYLARK (Biogen 2024g); 6 months for sertraline at 67.5mg/day (Bloch 2012) and fluoxetine at 20mg/day (NHS England. 2022).
[b]Zuranolone cost is listed as gross cost including a discount of [redacted]%.
[c]The proportion of patients receiving 50 mg and 40 mg zuranolone was assumed at [redacted]%, and [redacted]% respectively; 40 mg dose was used only if a dose reduction was required
[d]The proportion of patients using concomitant antidepressants was assumed at 15.3% in line with SKYLARK; among which, [redacted]% and [redacted]% of patients received sertraline and fluoxetine, respectively.
Abbreviations: eMIT: Electronic Marketing Information Tool; mg: Milligram; SSRIs: Selective Serotonin Reuptake Inhibitors.

As mentioned in Section 3.2.3.1, patients treated with zuranolone 50 mg were permitted dose reductions to 40 mg in SKYLARK (dose reductions are also allowed in clinical practice (Biogen 2025e). An assumption has been made that the additional costs associated with the dose adjustment are applied in cycle one, as this is approximately halfway through the zuranolone treatment course (cycle one lasting from Day 3 to Day 8). In the base case, based on prescription data in the US, it is assumed that [redacted]% of patients require a dose adjustment (Biogen 2024f), and therefore the cost of an additional pack of zuranolone is applied to these patients (Table 43). Additionally, an assumption has been made that one perinatal psychiatrist visit is required to prescribe the additional pack of zuranolone, with the associated cost also applied in the first cycle. Scenario analysis has been conducted with zuranolone dose reduction turned off.

Table 43: Zuranolone dose reduction parameters

Parameter	Mean	SE	Source
Cost of additional zuranolone pack, 40 mg	[redacted]	[redacted]	Biogen
Percentage of patients requiring dose reduction	[redacted]	[redacted]	Biogen (Biogen 2024f)
Number of perinatal psychiatrist visits	1.00	0.20	Assumption
Cost of perinatal psychiatrist visits	£136.40	27.28	NHS costs 2023/24 (NHS England. 2024)

Source: See table
Abbreviations: NHS: National Health Service; SE: Standard Error; USA: United States of America

3.5.3 Health-state unit costs and resource use

3.5.3.1 HRU and costs (up to 18 months post-birth)

Background disease-related resource use in the model encompasses costs from primary, secondary, and community care services. The model applies PND-related costs for the first 18 months postpartum, followed by data from patients with MDD thereafter. For the first 18 months postpartum, healthcare resource utilisation (HRU), along with associated healthcare costs, is included in the model, as detailed in Table 44. These 18-month data were adjusted to account for the time since birth in SKYLARK (average of 5-months post-birth), and then adjusted further to ensure the costs reflected the duration of time per cycle e.g., variable cycle length in the first three months of the model and 3-monthly cycles thereafter. For example, in cycle one it is assumed there are 5 days, therefore the 18-month data, after an initial adjustment to reflect 5-months post-partum was adjusted to further to reflect the costs for 5 days: total HRU cost in 18 months * 12/18 (annualise the 18-month costs) * 0.16 (5-days divided by 365.25 days in a year divided by 12 months in a year).

As stated in Section 3.5.1, the SLR did not identify any UK source relevant to the model. Therefore, ad hoc literature searches were conducted. The ad hoc search identified a study by Petrou et al. (2002) as a key source of granular resource use data. As Petrou et al. (2002) was published in 2002, it was outside of the date range of the SLR (2003 until 2024) and therefore was not captured. However, as it is the best data source available, it has been used in the economic analysis. Petrou et al. (2002) estimated the economic costs in 206 women recruited from antenatal clinic in Reading, UK between May 1997 to April 1999. Data on health (hospital and community) and social care services use for mothers and their infant was collected through face-to-face interviews 18 weeks, 12 and 18 months post-partum. The sample included 70 women with PND and 136 women without PND. The costs for women with PND was significantly higher than for women without PND. The mean cost difference reached statistical significance for community care services but not for other categories of services. The economic costs were also higher for women with extended experiences of PND. The data is categorised HRU into PND and no-PND groups, determined using Structured Clinical Interview for DSM-III-R diagnoses. In the model, decision rules were required to categorise patients as PND or not, based on EPDS thresholds. In the economic analysis for this submission, PND is defined as an EPDS score of 13 or more (Section 3.2.2). As Petrou et al. (2002) did not define PND according to EPDS threshold, an assumption is made that the definition of PND in the study and model is aligned.

The resource use data from Petrou et al. (2002) served as the primary evidence for the base case model. However, due to the age of the Petrou study and significant changes in maternity services since 2002, additional feedback was gathered through one-to-one interviews with UK clinicians specialising in PND (Biogen 2024e). This feedback emphasised that the resource use in current practice has evolved significantly and is different than the reported resource use from Petrou et al. (2002). The age of the Petrou data posed a limitation, as noted in clinician feedback. However, no more recent data with equivalent granularity exists, necessitating its continued use in the model. In a scenario analysis, alternative HRU inputs were developed using specific feedback from two UK clinicians on their estimates of resource use based on clinical experience which is applied to both patients treated with zuranolone and ECM (Table 45).

Unit costs were updated using more recent data from the NHS Cost Collection and the PSSRU. All costs are presented in source-year prices and inflated to 2024 values using the ONS CPI – Health inflation index (Office for National Statistics. 2025).

Table 44: Healthcare resource use and costs, 18 months post-birth, base case

Resource	PND N=78		No PND N=134		Source	Unit cost (£)	Source
	Mean	SE	Mean	SE			
Community care visits							
Midwifery	6.96	1.10	7.23	0.43	(Petrou 2002)	158.39	NHS costs 2023/24 (NHS England. 2024)
GP contacts	7.24	0.58	4.87	0.27		49.00	PSSRU 2023 (Personal Social Services Research Unit. 2023)
Practice nurse	1.14	0.28	0.99	0.16		13.25	
Practice counsellor	0.76	0.21	0.29	0.16		20.50	
Health visitor	6.32	0.32	6.63	0.22		158.39	NHS costs 2023/24 (NHS England. 2024)
Home help	0.03	0.02	0.07	0.07		27.00	PSSRU 2023 (Personal Social Services Research Unit. 2023)
Social worker	2.41	1.21	0.65	0.31		53.00	
Physiotherapist	2.08	0.65	1.23	0.26		20.50	
Community psychiatric nurse	3.01	1.06	1.27	0.48		16.00	
Community psychologist	0.97	0.51	0.35	0.17		31.50	
Other community mental health	0.44	0.22	0.21	0.10		31.50	
Other community care	2.85	0.94	1.98	0.58		27.00	
Total costs	**£2,873**		**£2,631**		-	-	-
Day care service visits							
Hospital admissions	0.24	0.08	0.24	0.08	(Petrou 2002)	342.26	NHS costs 2023/24 (NHS England. 2024)
Community day care	0.03	0.03	0.16	0.13		46.00	PSSRU 2023 (Personal Social Services Research Unit. 2023)
Other day care	0.00	0.00	0.08	0.06		46.00	Assumption same as community day care
Total costs	**£84**		**£94**		-	-	-
Hospital out-patients visits							

Obstetrics	0.15	0.07	0.13	0.04	(Petrou 2002)	191.43	NHS costs 2023/24 (NHS England. 2024)
A&E	0.16	0.04	0.20	0.04		272.97	
Other outpatient	1.05	0.25	0.75	0.22		191.43	
Total costs	**£273**		**£223**		-	-	-
Hospital in-patients visits							
Maternity ward (days)	2.95	0.30	2.69	0.20	(Petrou 2002)	461.33	Bowers 2016 (Bowers 2016)
Mother and baby unit (days)	0.10	0.10	0.00	0.00		1,171.00	NHS costs 2023/24 (NHS England. 2024)
Medical/surgical ward (days)	0.21	0.12	0.26	0.13		461.33	Assumption same as maternity ward cost
Other inpatient (days)	0.00	0.00	0.02	0.02		461.33	Assumption same as maternity ward cost
Total costs	**£2,030**		**£1,798**		**-**	**-**	**-**
Annualised cost over first 18 months	**£3,507**		**£3,164**		-	-	

Source: Cited within table.
Abbreviations: A&E: Accident and Emergency; GP: General Practitioner; NHS: National Health Service; PND, Postnatal Depression; PSSRU: Personal Social Services Research Unit; SE: Standard Error.

Table 45: Adjustments to base case HRU following clinical feedback for scenario analysis

Service	PND (EPDS ≥13)	Remission (EPDS <13)			Rationale	Source
	Mean	SE	Mean	SE		
Health visitor	[redacted]	-	[redacted]	-	[redacted]	(Biogen 2024e)
Home help	[redacted]	-	[redacted]	-	[redacted]	
Perinatal specialist practitioner[a]	[redacted]	-	[redacted]	-	[redacted]	
Perinatal psychologist[b]	[redacted]	-	[redacted]	-		
Perinatal psychiatrist[c]	[redacted]	-	[redacted]	-		
Community day care	[redacted]	-	[redacted]	-	[redacted]	
Mother and baby unit (days)	[redacted]	-	[redacted]	-	[redacted]	

[a] Replaces 'community psychiatric nurse' in Table 44.
[b] Replaces 'community psychologist' in Table 44.
[c] New line item for adjusted costs – unit cost £127.0062.
Abbreviations: EPDS: Edinburgh Postnatal Depression Scale; HRU: Healthcare Resource Utilisation; PND: Postnatal Depression; SE: Standard Error.

3.5.3.2 *HRU and costs (beyond 18 months post-birth)*

HRU and costs beyond 18 months post-birth are sourced from Byford et al., (2011) (Byford 2011) (Table 46). Byford et al., (2011) was identified through an ad hoc targeted search as the SLR did not identify studies exploring the HRU beyond 18 months post-birth.

Byford et al., (2011) was a naturalistic, longitudinal study investigating the economic impact of non-remission on health service use and costs over a 12 month period, among patients with depression. The study was carried out using data from a large primary care UK general practice research database between 2001 and 2006. A total of 88,935 patients aged ≥18 years, diagnosed with depression and in receipt of at least two ADT prescriptions, were included (Byford 2011).

There are several limitations that apply to the use of this data:

- HRU estimates come from a study of MDD patients, and therefore their use comes with the assumption that HRU for PND patients still experiencing symptoms of depression beyond 18 months post-birth are similar in terms of healthcare requirements to HRU for MDD patients.
- The study is relatively old, so may not represent current care.
- Unlike in the current model, in the study, depression is not assessed using EPDS.

In the model, HRU beyond 18 months post-birth is applied for the duration of the model time horizon. The resource use associated with non-remission has been applied to patients with EPDS score of >13, while the resource use associated with remission has been applied to patients with EPDS score of <13.

Table 46: HRU and costs, beyond 18 months post-birth

Resource	Non-remission N=35,281		Remission N=53,654		Source	Unit cost	Source
	Mean	SE	Mean	SE			
Service use							
GP visits	17.34	0.06	12.88	0.04	Byford 2011 (Byford 2011)	£49.00	PSSRU 2023 (Personal Social Services Research Unit. 2023)
GP phone calls	5.41	0.03	2.48	0.02		£20.27	
Psychological therapy contacts	0.00	0.00	0.00	0.00		£31.50	
Psychiatrist contacts	0.07	0.00	0.04	0.00		£136.40	NHS costs 2023/24 (NHS England. 2024)
Other specialist contacts	0.77	0.01	0.62	0.00		£136.40	Assumption same as mental health consultant
Inpatient days	0.08	0.00	0.07	0.00		£277.00	PSSRU 2023 (Personal Social Services Research Unit. 2023)
Emergency department attendance	0.05	0.00	0.04	0.00		£272.97	NHS costs 2023/24 (NHS England. 2024)
Annual costs by remission status	£1,169		£844		-	-	-

Source: Cited within table.
Abbreviations: GP: General Practitioner; HRU: Healthcare Resource Utilisation; NHS: National Health Service; PSSRU: Personal Social Services Research Unit; SE: Standard Error.

3.5.4 Adverse reaction unit costs and resource use

The AEs which occurred in ≥5% of patients during SKYLARK are shown in Table 38; costs were applied to the patient on a one-off basis for each incident AE and are shown in Table 47. It was assumed that urinary tract infection would incur the cost associated with immediate antibiotic management, while the other AEs were assumed to incur the cost associated with a GP visit.

Table 47: Summary of costs for treatment of AEs

Disutility	Mean	SE	Source
Somnolence	£49.00	9.80	GP visit - PSSRU 2023 (Personal Social Services Research Unit. 2023)
Dizziness			
Diarrhoea			
Fatigue			
Urinary tract infection	£30.60	4.02	Immediate antibiotic management – results from a RCT (£2005 price)(Turner 2010)

Source: Cited within table.
Abbreviations: AE: Adverse Event; GP: General Practitioner; PSSRU: Personal Social Services Research Unit; RCT: Randomised Control Trial; SE: Standard Error.

3.6 Severity

This submission does not meet the criteria for a severity weight.

3.7 Uncertainty

Not applicable.

3.8 Managed access proposal

A managed access agreement is not being considered at this stage.

3.9 Summary of base-case analysis inputs and assumptions

3.9.1 Summary of base-case analysis inputs

A summary of the base-case inputs is presented in Table 48.

Table 48: Summary of variables applied in the economic model

Variable	Value (reference to appropriate table or figure in submission)	Measurement of uncertainty and distribution: confidence interval (distribution)	Reference to section in submission
Population characteristics			
Baseline EPDS score	[redacted]	SE: [redacted] Distribution: Normal	Section 3.3.1

Age	[redacted]	SE: [redacted] Distribution: Normal	
ADT at baseline	[redacted]	SE: [redacted] Distribution: Beta (α=[redacted], β=[redacted])	
Short-term EPDS estimates			
SKYLARK MMRM: Intercept	[redacted]	SE: [redacted] Distribution: multivariate normal	Section 3.3.2.4
SKYLARK MMRM: EPDS0	[redacted]	SE: [redacted] Distribution: multivariate normal	
SKYLARK MMRM: Time	[redacted]	SE: [redacted] Distribution: multivariate normal	
SKYLARK MMRM: ADT use at baseline	[redacted]	SE: [redacted] Distribution: multivariate normal	
SKYLARK MMRM: Age	[redacted]	SE: [redacted] Distribution: multivariate normal	
SKYLARK MMRM: On zuranolone	[redacted]	SE: [redacted] Distribution: multivariate normal	
SKYLARK MMRM: On treatment (zur or pbo)	[redacted]	SE: [redacted] Distribution: multivariate normal	
SKYLARK MMRM: Time:On treatment	[redacted]	SE: [redacted] Distribution: multivariate normal	
Long-term EPDS estimates			
Random effects	[redacted]	N/A – not varied in the PSA	Section 3.3.3
ALSPAC MMRM: Intercept	[redacted]	SE: [redacted] Distribution: multivariate normal	
ALSPAC MMRM: EPDS8	[redacted]	SE: [redacted] Distribution: multivariate normal	
ALSPAC MMRM: Time	[redacted]	SE: [redacted] Distribution: multivariate normal	
ALSPAC MMRM: Time^2	[redacted]	SE: [redacted] Distribution: multivariate normal	
Zuranolone adverse event incidence rate			
Somnolence	36.00%	SE: 0.05 Distribution: Beta (α=35.28, β=62.72)	Section 3.3.5
Dizziness	13.00%	SE: 0.03 Distribution: Beta (α=12.74, β=85.26)	
Diarrhoea	6.00%	SE: 0.02 Distribution: Beta (α=5.88, β=92.12)	
Fatigue	5.00%	SE: 0.02 Distribution: Beta (α=4.9, β=93.1)	
Urinary tract infection	5.00%	SE: 0.02 Distribution: Beta (α=4.9, β=93.1)	
ECM Adverse event incidence rate			
Somnolence	6.00%	SE: 0.02 Distribution: Beta (α=5.88, β=92.12)	Section 3.3.5

Dizziness	9.00%	SE: 0.03 Distribution: Beta (α=8.82, β=89.18)	
Diarrhoea	2.00%	SE: 0.01 Distribution: Beta (α=1.96, β=96.04)	
Fatigue	2.00%	SE: 0.01 Distribution: Beta (α=1.96, β=96.04)	
Urinary tract infection	4.00%	SE: 0.02 Distribution: Beta (α=3.92, β=94.08)	
ROBIN SF-6D utility			
EPDS 0	0.93	SE: 0.02 Distribution: Beta (α=138.67, β=9.72)	Section 3.4.2
EPDS 1 to 6	0.79	SE: 0.01 Distribution: Beta (α=849.09, β=219.61)	
EPDS 7 to 13	0.66	SE: 0.01 Distribution: Beta (α=2472.03, β=1277.9)	
EPDS 14 to 18	0.57	SE: 0.01 Distribution: Beta (α=2085.26, β=1553.3)	
EPDS 19 to 30	0.52	SE: 0.01 Distribution: Beta (α=4053.85, β=3737.07)	
Adverse event disutility			
Somnolence	0.09	SE: 0.01 Distribution: Beta (α=87.88, β=945.97)	Section 3.4.4
Dizziness	0.03	SE: 0.01 Distribution: Beta (α=24.38, β=950.63)	
Diarrhoea	0.07	SE: 0.02 Distribution: Beta (α=17.48, β=223.63)	
Fatigue	0.04	SE: 0.01 Distribution: Beta (α=23.98, β=560.78)	
Urinary tract infection	0.01	SE: 0 Distribution: Beta (α=15.27, β=2530.43)	
Zuranolone price			
Zuranolone 50 mg - price per mg	[redacted]	N/A – not included in PSA	Section 3.5.2
Zuranolone 40 mg - price per mg	[redacted]	N/A – not included in PSA	
Zuranolone dose distribution			
Percent receiving 50 mg	[redacted]	N/A – not included in PSA	Section 3.5.2
Percent receiving 40 mg	[redacted]	N/A – not included in PSA	
Treatment duration			
Zuranolone (days)	14.00	SE: 2.80 Distribution: Log Normal	Section 3.5.2
Sertraline (months)	6.00	SE: 1.20	

		Distribution: Log Normal	
Fluoxetine (months)	6.00	SE: 1.20 Distribution: Log Normal	
Zuranolone dose reduction			
Cost of additional zuranolone pack	[redacted]	N/A – not included in PSA	Section 3.5.2
% of patients with dose reduction	[redacted]	SE: [redacted] Distribution: Beta (α=[redacted], β=[redacted])	
Additional perinatal psychiatrist visits	1.00	SE: 0.20 Distribution: Log Normal	
Cost of perinatal psychiatrist visits	£136.40	SE: 27.26 Distribution: Gamma (α=25.00, β=5.46)	
Concomitant SSRI distribution			
Sertraline	[redacted]	N/A – not included in PSA	Section 3.5.2
Fluoxetine	[redacted]	N/A – not included in PSA	
Concomitant SSRI dose			
Sertraline (mg/day)	67.50	SE: 21.50 Distribution: Normal	Section 3.5.2
Fluoxetine (mg/day)	20.00	SE: 4.00 Distribution: Normal	
Concomitant SSRI price			
Sertraline (per pack)	£0.56	SE: 0.00 Distribution: Log Normal	Section 3.5.2
Fluoxetine (per pack)	£0.40	SE: 0.00 Distribution: Log Normal	
HRU			
Resource use - 18 months post birth	Table 44	Log normal	Section 3.5.3.1
Costs - 18 months post birth		Gamma	
Resource use – 18+ months post birth	Table 46	Log normal	Section 3.5.3.2
Costs – 18+ months post birth		Gamma	
Adverse event costs			
Somnolence	£51.95	SE: 9.8 Distribution: Gamma (α=25, β=1.96)	Section 3.5.4
Dizziness	£51.95	SE: 9.8 Distribution: Gamma (α=25, β=1.96)	
Diarrhoea	£51.95	SE: 9.8 Distribution: Gamma (α=25, β=1.96)	
Fatigue	£51.95	SE: 9.8 Distribution: Gamma (α=25, β=1.96)	
Urinary tract infection	£54.17	SE: 4.02 Distribution: Gamma (α=57.86, β=0.53)	

Abbreviations: ADT; Antidepressant; ALSPAC: Avon Longitudinal Study of Parents and Children; ECM: Established Clinical Management; EPDS: Edinburgh Postnatal Depression Scale; HRU: Healthcare Resource Utilisation; MMRM: Mixed Model for

Repeated Measures; N/A: Not/Applicable; PSA: Probabilistic Sensitivity Analysis; SE: Standard Error; SF-6D: Short Form 6-Dimension; SSRI: Selective Serotonin Reuptake Inhibitor.

3.9.2 Assumptions

A summary of the base-case assumptions is presented in Table 49.

Table 49: Summary of assumptions applied in the economic model

Area	Assumption	Rationale
Model structure	Remission from depression is indicated by a patient reaching an EPDS threshold <13.	Thresholds for EPDS in the literature vary from 10 to 13, but 13 is commonly used as threshold and was validated by clinical feedback.
Comparator	The placebo arm of SKYLARK represents ECM since both ADTs and psychotherapy were allowed.	The primary comparator evaluated in this analysis is ECM, which is captured by the control arm of SKYLARK. Reflecting the zuranolone arm of the study, concomitant ADTs and psychotherapy were permitted, alongside placebo zuranolone treatment, to maintain treatment blinding.
Short-term modelling	MMRM modelling of the SKYLARK data provides both the trajectory of ECM and the treatment effectiveness of zuranolone with ECM up to 3 months post treatment initiation (8 months post-birth).	MMRM modelling trajectory is aligned with trial data from SKYLARK.
Efficacy / treatment effects (long-term)	**ALSPAC data** can be leveraged for the long-term trajectory of PND in the economic model. Efficacy at 8 months post-birth for both treatment arms is reflective of long-term trajectory, as EPDS at this time point is a good indicator for the long-term.	The ALSPAC dataset allows for the inclusion of covariates predictive of long-term EPDS score, and it uses the full follow up from the ALSPAC cohort (up to year 29), facilitating longer-term extrapolation of EPDS scores over time.
Safety	Only the AEs occurring with 5 % or more frequency in SKYLARK and are greater with zuranolone than placebo are included in the analysis.	Inclusion of events that have a greater incidence in the zuranolone arm is a conservative approach.
Costs	**Base case** That costs can be modelled using data from early 2000s (with inflation to 2024 prices). **Scenario** The HRU were validated and updated through feedback two UK clinicians to reflect current clinical practice	Lack of more recent data. A scenario was created to try to assess the impact of changes in provision.
	That costs can be modelled using data from PND patients for the first 18 months and by MDD patients thereafter.	These patients would be considered as MDD patients at this point – validated by clinical opinion.

Area	Assumption	Rationale
Utility values	**SF-6D from ROBIN:** That the SF-6D instrument is appropriate for HTA decision making.	Trial- based utility is preferred over other sources of utility, and the population in ROBIN closely matches the population in SKYLARK
Mortality	**Base case** There is no differential mortality between depressed and non-depressed patients.	A conservative assumption has been made that there is no increased mortality for patients with PND who are being treated.
	Scenario A standardised mortality ratio representing an increased rate of all-cause mortality can be applied to depressed patients.	There is some evidence that depressed patients have higher mortality risk, and clinical feedback confirms this.

Abbreviations: ADT: Antidepressant treatment; AE: Adverse Event; ALSPAC: Avon Longitudinal Study of Parents and Children; ECM: Established Clinical Management; EPDS: Edinburgh Postnatal Depression Scale; EQ-5D: EuroQol 5-Dimension; HRU: Healthcare Resource Utilisation; HTA: Health Technology Assessment; MDD: Major Depressive Disorder; MMRM: Mixed Model for Repeated Measures; PND: Postnatal Depression; SF-6D: Short Form 6-Dimension.

3.10 Base-case results

3.10.1 Base-case incremental cost-effectiveness analysis results

In the base case, the deterministic model was run with 1,000 patients to ensure model convergence (Figure 33). The base case costs, QALYs and ICER are shown in Table 50, with the net monetary benefit at a WTP threshold of £20,000 and £30,000 shown in Table 51. Zuranolone treatment with ECM was associated with higher per-patient costs compared to ECM alone (£[redacted] vs £[redacted], respectively), with the difference largely driven by treatment acquisition costs (Table 52). Patients receiving zuranolone with ECM experienced improved health outcomes compared to ECM in terms of QALYs gained ([redacted] vs [redacted], respectively). Under base case assumptions, the ICER for zuranolone with ECM compared to ECM is £ £12,704 per QALY, indicating that zuranolone with ECM can be considered a cost-effective treatment option for PND in the UK. The mean probabilistic results agree with those from the deterministic analysis (ICER £12,704 vs £12,706 for deterministic and mean probabilistic results respectively).

Table 50: Base-case results

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALY gain	ICER (£ per QALY)
Deterministic (1,000 patients)							
ECM alone	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	12,704
Zuranolone with ECM	[redacted]	[redacted]	[redacted]				

Costs expressed in 2024 GBP (£).
Abbreviations: ECM: Established Clinical Management; GBP: Great British Pounds; ICER: Incremental Cost-Effectiveness Ratio; LYG: Life Years Gained; QALY: Quality-Adjusted Life Year.

Figure 33: Model convergence plot

Abbreviations: ECM: Established Clinical Management; ICER: Incremental Cost-Effectiveness Ratio.

Table 51: Incremental net monetary benefit

Technologies	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	INMB at £20,000	INMB at £30,000
Deterministic (500 patients)						
ECM alone	[redacted]	[redacted]	[redacted]	[redacted]	3,744	8,876
Zuranolone with ECM	[redacted]	[redacted]				

Results based on an average of 1,000 probabilistic iterations. Costs expressed in 2024 GBP (£).
Abbreviations: ECM: Established Clinical Management; GBP: Great British Pounds; INMB: Incremental Net Monetary Benefit; QALY: Quality-Adjusted Life Year.

Table 52: Costs breakdown

Component	Zuranolone with ECM	ECM alone	Incremental
Treatment acquisition costs	[redacted]	[redacted]	[redacted]
Healthcare resource use	[redacted]	[redacted]	[redacted]
Remission (EPDS < 13)	[redacted]	[redacted]	[redacted]
Depression (EPDS ≥ 13)	[redacted]	[redacted]	[redacted]
Adverse events	[redacted]	[redacted]	[redacted]
Total	[redacted]	[redacted]	[redacted]

Costs expressed in 2024 GBP (£).
Abbreviations: ECM: Established Clinical Management; EPDS: Edinburgh Postnatal Depression Scale.

The difference in QALYs, and to a lesser extent costs, is being driven by the time spent in an unresolved PND state (Table 53). A larger proportion of patients who received zuranolone with ECM remained in remission compared with ECM ([redacted] vs [redacted], respectively), and had longer duration of treatment response (defined as 50% reduction from baseline EPDS; [redacted] vs [redacted], respectively) (See Appendix H). Figure 34 and Figure 35 presents the mean EPDS scores over time for all patients.

Table 53: QALY breakdown

Component	Zuranolone with ECM	ECM alone	Incremental
Remission (EPDS < 13)	[redacted]	[redacted]	[redacted]
Depression (EPDS >= 13)	[redacted]	[redacted]	[redacted]
Total	[redacted]	[redacted]	[redacted]

Abbreviations: ECM: Established Clinical Management; EPDS: Edinburgh Postnatal Depression Scale.

Figure 34: Mean EPDS score for all patients from 0-12 months

Note: The zuranolone arm corresponds to zuranolone + ECM.
Abbreviations: ECM: Established Clinical Management; EPDS: Edinburgh Postnatal Depression Scale.

Figure 35: Mean EPDS score for all patients from 0-29 years

Note: The zuranolone arm corresponds to zuranolone + ECM.
Abbreviations: ECM: Established Clinical Management; EPDS: Edinburgh Postnatal Depression Scale.

3.11 Exploring uncertainty

3.11.1 Probabilistic sensitivity analysis

The PSA was run with 1,000 patients for 1,000 iterations to ensure model convergence (Figure 36). The PSA results align with the deterministic results with an ICER of £12,706 (Table 54).

Table 54: Probabilistic sensitivity analysis results

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALY gain	ICER (£ per QALY)
PSA (1,000 patients, 1,000 iterations)							
ECM alone	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	12,706
Zuranolone with ECM	[redacted]	[redacted]	[redacted]				
95% CI (lower, upper)							
ECM alone	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	12,764, 12,650
Zuranolone with ECM	[redacted]	[redacted]	[redacted]				

Costs expressed in 2024 GBP (£).
Abbreviations: CI: Confidence Interval; ECM: Established Clinical Management; ICER: Incremental Cost-Effectiveness Ratio; LYG: Life Years Gained; PSA: Probabilistic Sensitivity Analysis; QALY: Quality-Adjusted Life Year.

Figure 36: PSA convergence plot

Abbreviations: ECM: Established Clinical Management; ICER: Incremental Cost-Effectiveness Ratio; PSA: Probabilistic Sensitivity Analysis.

At a willingness to pay threshold of £20,000 per QALY, zuranolone with ECM has an [redacted]% chance of being cost-effective and at £30,000 per QALY this rises to [redacted]% (Figure 37).

Figure 37: PSA - cost-effectiveness acceptability curve for the base case zuranolone vs ECM

The zuranolone arm corresponds to zuranolone + ECM.
Abbreviations: ECM: Established Clinical Management; PSA: Probabilistic Sensitivity Analysis.

The distribution of individual simulations in presented in Figure 38. It is evident that the uncertainty in parameters affecting the incremental QALYs is greater than that for the

incremental costs, which is reflected in the high degree of horizontal spread in the cost-effectiveness plane. As can be seen from the results of the deterministic sensitivity analysis (DSA), the biggest driver of the cost-effectiveness is the zuranolone with ECM treatment effect estimate (see Section 3.11.2), and therefore the uncertainty in this parameter is causing the horizontal spread.

Figure 38: PSA - cost-effectiveness plane the base case zuranolone vs ECM

Abbreviations: ECM: Established Clinical Management; ICER: Incremental Cost-effectiveness Ratio; PSA: Probabilistic Sensitivity Analysis; QALY: Quality Adjusted Life-Year; WTP: Willingness To Pay.

3.11.2 Deterministic sensitivity analysis

Figure 39 and Table 55 present the ten most impactful model parameters in terms of their impact on the CEA of zuranolone with ECM versus ECM alone in the base case. The parameter with the largest impact on the results is the treatment effect for zuranolone with ECM. Although the higher variation leads to a large increase in the ICER, in the DSA, this parameter has been varied individually while there may be correlation between this parameter and the other covariates in the SKYLARK MMRM model. Therefore, this parameter was also included in the PSA to capture the impact of uncertainty using the variance-covariance matrix of the SKYLARK MMRM model. The PSA results indicate a very narrow confidence interval, suggesting low uncertainty in the model's estimates. The deterministic ICER is £12,706 per QALY gained, while the probabilistic ICER ranges from £12,764 to £12,650. Similarly, the incremental costs in the PSA vary minimally between [redacted] and [redacted], and the incremental QALYs range only slightly from [redacted] to [redacted]. These tight intervals demonstrate

that the model is highly stable, with little variation in outcomes even when accounting for parameter uncertainty for the covariates in the SKYLARK MMRM model, including zuranolone treatment effect.

Figure 39: Top 10 parameters varied using 1,000 patients per simulation in base case DSA

Abbreviations: ADT: Antidepressant; ALSPAC: Avon Longitudinal Study of Parents and Children; DSA: Deterministic Sensitivity Analysis; EPDS: Edinburgh Postnatal Depression Scale; GP: General Practitioner; ICER: Incremental Cost-effectiveness Ratio; MMRM: Mixed Model for Repeated Measures; QALY: Quality Adjusted Life-Year; SF-6D: Short Form Health Survey – 6 Dimensions.

Table 55: Top 10 DSA results

#	Parameter	Variation	Inc. costs (£)	Inc. QALYs	ICER (£/QALY)	INMB at £20k (£)
Base case					**12,704**	**3,744**
1	SKYLARK MMRM: On zuranolone (,)	Lower			6,544	12,890
		Upper			77,729	-5,158
2	SF-6D EPDS 1 to 6 (,)	Lower			13,732	2,976
		Upper			11,854	4,481
3	ALSPAC MMRM: Time ,)	Lower			11,904	4,476
		Upper			13,693	2,994
4	SKYLARK MMRM: Age (,)	Lower			12,069	4,433
		Upper			13,345	3,229
5	ALSPAC MMRM: Time^2 (,)	Lower			12,063	4,307
		Upper			13,106	3,420
6	SF-6D EPDS 14 to 18 (,)	Lower			12,215	4,156
		Upper			13,232	3,335
7		Lower			12,942	3,622

	Unit costs: GP visits (£34, £74)	Upper	[redacted]	[redacted]	12,415	3,892
8	SF-6D EPDS 7 to 13 ([redacted], [redacted])	Lower	[redacted]	[redacted]	12,460	3,946
		Upper	[redacted]	[redacted]	12,956	3,545
9	Baseline EPDS score (20.58, 21.62)	Lower	[redacted]	[redacted]	12,634	3,811
		Upper	[redacted]	[redacted]	13,115	3,417
10	Age (29.67, 31.33)	Lower	[redacted]	[redacted]	13,026	3,498
		Upper	[redacted]	[redacted]	12,741	3,711

Abbreviations: ALSPAC: Avon Longitudinal Study of Parents and Children; DSA: Deterministic Sensitivity Analysis; EPDS: Edinburgh Postnatal Depression Scale; GP: General Practitioner; ICER: Incremental Cost-effectiveness Ratio; INMB: Incremental Net Monetary Benefit; MMRM: Mixed Model for Repeated Measures; QALY: Quality-Adjusted Life-Year; SF-6D: 36-item Short Form Health Survey.

3.11.3 Scenario analysis

Scenario analyses were performed to test the robustness of the analysis to variations in underlying model assumptions and to the use of alternative input parameters (e.g., a different EPDS threshold for remission). The incremental cost-effectiveness ratios were generated for each scenario and then compared against the base case results.

The full list of scenarios explored is presented in Table 56.

Table 56: Description of the scenario analysis

#	Parameter	Base case	Scenario
S1	**Mortality**	No additional mortality for patients with PND i.e., general population mortality applied for all patients	Additional mortality is applied for patients who have PND (Section 3.3.6)
S2	**ADT use at baseline**	The short-term EPDS model from SKYLARK forces the inclusion of ADT use at baseline	The short-term EPDS model with no forced ADT at baseline is applied, and the costs associated with ADT are removed (Section 3.3.2).
S3	**HRU inputs**	HRU resource use is sourced from Petrou et al. (2002)	The data from Petrou et al. (2002) has been augmented with feedback from clinicians to reflect current clinical practice (Section 3.5.3)
S4	**PND remission threshold (EPDS)**	An EPDS of <13 indicates PND remission	The remission threshold is changed so that an EPDS of <10 indicates PND remission (Section 3.2.2).
S5	**Model time horizon**	29 years to align with the long-term data from ALSPAC	Lifetime time horizon of 60 years (Section 3.2.2). The EPDS score at year 29 is carried forward for the duration of the time horizon.
S6	**Dose reductions**	Additional pack of zuranolone is applied to [redacted]% of patients who require a dose reduction	No dose reduction is applied (Section 3.5.2)

Abbreviations: ADT: Antidepressant Treatment; ALSPAC: Avon Longitudinal Study of Parents and Children; EPDS: Edinburgh Postnatal Depression Scale; HRU: Healthcare Resource Utilisation; ICER: Incremental Cost-Effectiveness Ratio; PND: Postnatal Depression;

Table 57 presents the scenario analyses results. While one scenarios lead to an increase in the ICER (S2), none of the scenarios result in an ICER above £20,000 per QALY, and the remaining five scenarios lead to a decrease in the ICER.

Table 57: Scenario analysis results, zuranolone vs comparator

#	Scenario	Inc. costs (£)	Inc. QALYs	ICER (£/QALY)	INMB at £20k (£)
Zuranolone with ECM vs ECM alone (SKYLARK)					
Base case		[redacted]	[redacted]	**12,704**	**3,744**
S1	Excess mortality for patients with PND (Section 3.3.6)	[redacted]	[redacted]	12,415	4,003
S2	0% ADT at baseline (alternative short-term model) (Section 3.3.2)	[redacted]	[redacted]	12,736	3,712
S3	HRU from Petrou + expert opinion (Section 3.5.3)	[redacted]	[redacted]	12,566	3,815
S4	EPDS < 10 for classifying depression (Section 3.2.2)	[redacted]	[redacted]	12,382	3,910
S5	Lifetime time horizon (60 years) (Section 3.2.2)	[redacted]	[redacted]	10,282	6,019
S6	No dose reduction (Section 3.5.2)	[redacted]	[redacted]	12,272	3,966

Abbreviations: ADT: Antidepressant Treatment; ECM: Established Clinical Management; EPDS: Edinburgh Postnatal Depression Scale; HRU: Healthcare Resource Utilisation; ICER: Incremental Cost-effectiveness Ratio; INMB: Incremental Net Monetary Benefit; PND: Postnatal Depression; QALY: Quality-Adjusted Life-Year.

3.12 Subgroup analysis

No subgroup analyses were conducted.

3.13 Benefits not captured in the QALY calculation

Introduction of zuranolone as a treatment option for PND is associated with additional benefits, which have not been captured in the QALY calculation:

- Beyond the adverse psychological effects, PND symptoms can also negatively impact the physical and overall functioning of patients (Posmontier 2008). This can lead to additional costs and disutilities which are not captured in the model.
- The QALY calculation does not capture any child outcomes, however, PND can negatively affect infants' HRQoL from as early as 1-month of age (Subbiah 2025). Additionally the symptomatic burden of PND can significantly impair mother-infant bonding (Section 1.3.3.1) and can lead to negative short- and long-term consequences for a child's development including behavioural problems, delayed cognitive development, significantly impaired language development, a higher degree of emotional disorders, and lower social engagement (Noorlander 2008, Netsi 2018, Slomian 2019). PND has been shown to substantially increase the risk of behavioural problems, academic performance, and depression, among children (Netsi 2018).

- PND may also significantly impair maternal function, including caring for the child, and disruptions arising due to maternal depression can also translate into unfavourable outcomes for siblings of the newborn (Canadian Paediatric Society 2004, Netsi 2018). However, the QALY calculation does not capture the impact of PND on other children, or if the woman has another child and has an additional PND event.
- The QALY calculation also does not capture any partner outcomes. PND can negatively affect the partner's own mental health and QoL, and the families of patients experiencing PND are subject to a higher level of disruption due to parenting behaviours and relationship separation (Section 1.3.3.1) (Moore Simas 2019). Partners of patients with PND are at an increased risk of depression, and any associated utility decrements (Nishimura 2015).
- Although the societal impact cannot be fully captured within the QALY calculation, the average cost to society of one case of PND in the UK is approximately £75,000 (2012/13 prices). The cost impact of PND on women is estimated to be £23,151 with productivity losses accounting for approximately 13% of these costs (£3,028 per case) (Section 1.3.3.2) (Bauer 2016b).

3.14 Validation

3.14.1 Expert engagement

3.14.1.1 Clinician validation

Two UK-based clinicians were engaged in an advisory board, with an email follow-up. The objective of the interviews was to gather clinician feedback on the treatment pathway for PND and the potential role of zuranolone within it. This included insights into current clinical practices, resource use, and outcome measures for diagnosing and treating PND, as well as understanding disease burden, remission criteria, and relapse patterns. The interviews also explored how zuranolone might impact patient and family quality of life, healthcare resource use, and cost-effectiveness. The feedback was used to refine the health economic modelling framework, as it stood at the time of the interviews (July 2024). An email questionnaire was sent in follow-up that tried to elicit more specific details around the level of HRU in England.

3.14.1.2 Health economic expert validation

A separate strategic validation was conducted via virtual interviews with an expert in HTA and a health economist who had not been involved in the development of the model (July 2024). The objective was to obtain independent and constructive advice to refine the model

development. A pre-read document was shared to facilitate feedback, which included an overview of the studies considered and an outline of the model concept. The interviews advised on aspects related to the model structure, namely ensuring that the use of patient level simulation was justified, and the scenario analyses to be included. Additionally, further validation conducted with two HTA and health economics experts in February 2025 to validate the key assumptions applied in this analysis.

3.14.2 Alignment of model and SKYLARK short-term EPDS predictions

The EPDS predictions of the cost-effectiveness model can be internally validated against the data they were informed by Figure 40 and Figure 41 show the predicted EPDS from the cost effectiveness model compared to the data from which the MMRM model was derived. For both zuranolone with ECM and ECM, agreement is good, except for the datapoint at day 8, which the cost effectiveness model overpredicts EPDS and underpredicts efficacy compared to the observed data

Figure 40: Comparison of predicted EPDS for patients on zuranolone with ECM vs observed data from SKYLARK

Abbreviations: ECM: Established Clinical Management; EPDS: Edinburgh Postnatal Depression Scale.

Figure 41: Comparison of predicted EPDS for patients on ECM vs observed data from SKYLARK

Abbreviations: ECM: Established Clinical Management; EPDS: Edinburgh Postnatal Depression Scale.

3.14.3 Model verification

The model was quality checked using the Tech-VER checklist, with the results of the tests presented in Appendix K.4.

3.15 Interpretation and conclusions of economic evidence

The base case analysis indicates that zuranolone with ECM can be considered cost-effective compared to ECM alone as a treatment for patients with PND. Zuranolone with ECM treatment was associated with higher per-patient lifetime costs compared to ECM (£ vs £) and higher per-patient lifetime QALYs (vs), with a resulting ICER of £12,704 per QALY.

Effort has been taken to ensure that this analysis aligns with the NICE scope (Section 1.1) and NICE reference case. Additionally, to ensure that the model reflects clinical practice as much as possible, the short-term SKYLARK regression model forced the inclusion of ADT use at baseline to allow ECM to better reflect the current standard of care. Furthermore, the *de novo* analysis of the ALSPAC data allows for an innovative modelling approach with a granular estimation of a patient's long-term EPDS scores for up to 29 years based on their baseline characteristics.

The key strengths of this analysis include a robust clinical evidence package with well-designed studies, and the novel use of UK based real-world long-term data to link the short-term EPDS outcomes from SKYLARK with the long-term EPDS projections.

While the DSA indicated that the zuranolone treatment effect has a large impact on the ICER, this does not result in uncertainty within the PSA when incorporated using the variance-covariance matrix, and the very close alignment between the PSA and deterministic results demonstrates that the base case analysis is robust with minimal uncertainty in the parameter estimates. Additionally, scenario analyses were conducted to test the robustness of the results to changes in key inputs and structural assumptions, with no scenarios resulting in changes to the conclusions about the cost-effectiveness of zuranolone. The only scenario which increased the ICER was the alternative short-term statistical model from SKYLARK with no ADT use (scenario two), which only resulted in an increase in the ICER of £32.31. The remaining scenarios all reduced the ICER, with the largest reduction in the ICER (reduction of £2,422) due to the scenario which explored a lifetime time horizon. Additionally, there are several potential benefits for zuranolone which were not captured in the QALY calculations, including the impact of PND on partners and offspring of patients, and the wider societal impact of PND, and therefore this analysis is likely to be a conservative estimate of the benefits of zuranolone.

One key limitation of the analysis, which could not be overcome, is the use of Petrou et al., (2002) for HRU, which may not reflect current resource use in clinical practice due to the age of the study. Nowadays, there is potentially greater involvement in the healthcare system to treat patients with PND, resulting in greater HRU for the affected patients, and therefore the use of historic data is likely to be a conservative assumption.

In summary, the burden of PND extends well beyond the suffering patient as it impacts patient's families and societies as a whole, both in terms of a humanistic burden and costs to the society. With the recent policy focus on improving maternal health services and raising awareness and improving access to mental health services, zuranolone is uniquely positioned to address these areas of unmet need as it offers a potential cost-effective and innovative treatment option for PND.

4 References

Agrawal, I., Mehendale, A. M. and Malhotra, R. (2022). Risk Factors of Postpartum Depression. Cureus 14(10): e30898.

Ahmed, H. M., Alalaf, S. K. and Al-Tawil, N. G. (2012). Screening for postpartum depression using Kurdish version of Edinburgh postnatal depression scale. Arch Gynecol Obstet 285(5): 1249-1255.

Ahmed, Q., Aga, Q., Bataineh, Y., Saadeh, M., Ismael, N., et al. (2020). Perinatal depression. Prevalence, suicidal idea, and associated factors. J Pharm Sci. 21: 112-118.

Alasoom, L. I. and Koura, M. R. (2014). Predictors of postpartum depression in the eastern province capital of saudi arabia. J Family Med Prim Care 3(2): 146-150.

Alhusaini, N. A., Zarban, N. A., Shoukry, S. T., Alahmadi, M., Gharawi, N. K., et al. (2022). Prevalence of Postpartum Depression Among Mothers Giving Birth at King Abdulaziz University Hospital, Jeddah, Kingdom of Saudi Arabia From 2020 Until 2022. Cureus 14(11): e31365.

Alliance, M. M. H. (2023). "Over 30,000 new or expectant mothers are on waiting lists for mental health support." Retrieved 29 August, 2024, from https://maternalmentalhealthalliance.org/news/labour-party-waiting-lists-perinatal-mental-health-support/.

Alonazi, H. G. and Jahan, S. (2022). Prevalence of Postpartum Depression among Women in Childbearing Age Attending Primary Health Care Centres, Qassim Region. Health Psychol Res 10(3): 35642.

Althaus, A. L., Ackley, M. A., Belfort, G. M., Gee, S. M., Dai, J., et al. (2020). Preclinical characterization of zuranolone (SAGE-217), a selective neuroactive steroid GABA(A) receptor positive allosteric modulator. Neuropharmacology 181: 108333.

American College of Obstetricians and Gynecologists (ACOG). "Treatment and management of mental health conditions during pregnancy and postpartum." Retrieved June 16, 2023, from https://www.acog.org/clinical/clinical-guidance/clinical-practice-guideline/articles/2023/06/treatment-and-management-of-mental-health-conditions-during-pregnancy-and-postpartum

American Psychiatric Association (APA) (2022). Diagnostic and Statistical Manual of Mental Disorders. 5th Edition, Text Revision. American Psychiatric Association; 2022. .

American Psychological Association (2022). "Postpartum depression: Causes, symptoms, risk factors, and treatment options." Retrieved September 30, 2024, from https://www.apa.org/topics/women-girls/postpartum-depression.

Angst, J., Angst, F. and Stassen, H. H. (1999). Suicide risk in patients with major depressive disorder. Journal of clinical psychiatry 60(2): 57-62.

Antonoudiou, P., Colmers, P. L. W., Walton, N. L., Weiss, G. L., Smith, A. C., et al. (2022). Allopregnanolone Mediates Affective Switching Through Modulation of Oscillatory States in the Basolateral Amygdala. Biol Psychiatry 91(3): 283-293.

Appleby, L., Warner, R., Whitton, A. and Faragher, B. (1997). A controlled study of fluoxetine and cognitive-behavioural counselling in the treatment of postnatal depression. BMJ 314(7085): 932-936.

Atuhaire, C., Taseera, K., Atukunda, E. C., Atwine, D., Matthews, L. T., et al. (2024). Prevalence of postpartum depression and its association with diabetes mellitus among mothers in public health facilities in Mbarara, Southwestern Uganda. Psychol Health Med: 1-17.

Auger, N., Potter, B. J., Healy-Profitós, J., He, S., Schnitzer, M. E., et al. (2020). Mood disorders in pregnant women and future cardiovascular risk. J Affect Disord 266: 128-134.

Batt, M. M., Duffy, K. A., Novick, A. M., Metcalf, C. A. and Epperson, C. N. (2020). Is Postpartum Depression Different From Depression Occurring Outside of the Perinatal Period? A Review of the Evidence. Focus 18(2): 106-119.
Bauer, A., , et al., (2016a). Best practice for perinatal mental health care: the economic case. PSSRU Discussion Paper DP2913, November 2016.
Bauer, A., et al., (2014). The costs of perinatal mental health problems., London School of Economics & Centre for Mental Health.
Bauer, A., Knapp, M. and Parsonage, M. (2016b). Lifetime costs of perinatal anxiety and depression. J Affect Disord 192: 83-90.
Bauman, B. L., Ko, J. Y., Cox, S., D'Angelo Mph, D. V., Warner, L., et al. (2020). Vital Signs: Postpartum Depressive Symptoms and Provider Discussions About Perinatal Depression - United States, 2018. MMWR Morb Mortal Wkly Rep 69(19): 575-581.
Bell, L., Feeley, N., Hayton, B., Zelkowitz, P., Tait, M., et al. (2016). Barriers and Facilitators to the Use of Mental Health Services by Women With Elevated Symptoms of Depression and Their Partners. Issues Ment Health Nurs 37(9): 651-659.
Biogen (2023). Biogen and Sage Therapeutics Announce FDA Accepts Filing of New Drug Application and Grants Priority Review of Zuranolone in the Treatment of Major Depressive Disorder and Postpartum Depression.
Biogen (2024a). ALSPAC Analysis Report. Version 1. Data on file.
Biogen (2024b). Meaningful Change Analysis of the HAMD-17 in Postpartum Depression. Data on file.
Biogen (2024c). PPD Country Landscaping 2: Country-specific reports for Brazil, Canada, France, UK. Data on file. .
Biogen (2024d). Therapy Watch 2023 Postpartum Depression Report. Version 2. Data on file.
Biogen (2024e). UK PND clinician interview July 2024. Data on file. .
Biogen (2024f). Zuranolone US IQVIA prescription data. Data on file.
Biogen (2024g). ZURZUVAE™ (zuranolone) Prescribing Information. Available at: https://www.accessdata.fda.gov/drugsatfda_docs/label/2024/217369s001lbl.pdf. [Accessed: September 2024].
Biogen (2025a). Analysis of Clinical Practice Research Datalink Aurum. Data on file.
Biogen (2025b). Analysis of IQVIA Longitudinal Access and Adjudication Data. Data on file.
Biogen (2025c). Biogen post-hoc analysis April 2025 [Data on file].
Biogen (2025d). SKYLARK and ROBIN Post-hoc analysis January 2025. Data on file.
Biogen (2025e). ZURZUVAE™ (zuranolone) MHRA Draft SmPC. Data on file.
Bloch, M., Meiboom, H., Lorberblatt, M., Bluvstein, I., Aharonov, I., et al. (2012). The effect of sertraline add-on to brief dynamic psychotherapy for the treatment of postpartum depression: a randomized, double-blind, placebo-controlled study. J Clin Psychiatry 73(2): 235-241.
Bodkin, J. A. and Dunlop, B. W. (2021). Moving on With Monoamine Oxidase Inhibitors. Focus (Am Psychiatr Publ) 19(1): 50-52.
Bowers, J. and Cheyne, H. (2016). Reducing the length of postnatal hospital stay: implications for cost and quality of care. BMC Health Serv Res 16: 16.
Boyd, A., Golding, J., Macleod, J., Lawlor, D. A., Fraser, A., et al. (2013). Cohort Profile: the 'children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children. Int J Epidemiol 42(1): 111-127.
Braund, T. A., Tillman, G., Palmer, D. M., Gordon, E., Rush, A. J., et al. (2021). Antidepressant side effects and their impact on treatment outcome in people with major depressive disorder: an iSPOT-D report. Transl Psychiatry 11(1): 417.
Brazier, J., Roberts, J. and Deverill, M. (2002). The estimation of a preference-based measure of health from the SF-36. J Health Econ 21(2): 271-292.
Brodey, B. B., Goodman, S. H., Baldasaro, R. E., Brooks-DeWeese, A., Wilson, M. E., et al. (2016). Development of the Perinatal Depression Inventory (PDI)-14 using item

response theory: a comparison of the BDI-II, EPDS, PDI, and PHQ-9. Arch Womens Ment Health 19(2): 307-316.
Brown, J. V. E., Wilson, C. A., Ayre, K., Robertson, L., South, E., et al. (2021). Antidepressant treatment for postnatal depression. Cochrane Database Syst Rev 2(2): Cd013560.
Bryson, A., Hatch, R. J., Zandt, B. J., Rossert, C., Berkovic, S. F., et al. (2020). GABA-mediated tonic inhibition differentially modulates gain in functional subtypes of cortical interneurons. Proc Natl Acad Sci U S A 117(6): 3192-3202.
Busner, J. and Targum, S. D. (2007). The clinical global impressions scale: applying a research tool in clinical practice. Psychiatry (Edgmont) 4(7): 28-37.
Byford, S., Barrett, B., Despiégel, N. and Wade, A. (2011). Impact of treatment success on health service use and cost in depression: longitudinal database analysis. Pharmacoeconomics 29(2): 157-170.
Callander, E. J., Gamble, J. and Creedy, D. K. (2021). Postnatal Major Depressive Disorder in Australia: Inequalities and Costs of Healthcare to Individuals, Governments and Insurers. Pharmacoeconomics 39(6): 731-739.
Canadian Paediatric Society (2004). Maternal depression and child development. Paediatr Child Health 9(8): 575-598.
Carlson, M. S., Azhar Y., et al.,, (2024). Postpartum Depression. [Updated 2024 Aug 12]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK519070/.
Cartwright, C., Gibson, K., Read, J., Cowan, O. and Dehar, T. (2016). Long-term antidepressant use: patient perspectives of benefits and adverse effects. Patient Prefer Adherence 10: 1401-1407.
Carvalho, A. F., Sharma, M. S., Brunoni, A. R., Vieta, E. and Fava, G. A. (2016). The Safety, Tolerability and Risks Associated with the Use of Newer Generation Antidepressant Drugs: A Critical Review of the Literature. Psychother Psychosom 85(5): 270-288.
Chibanda, D., Shetty, A. K., Tshimanga, M., Woelk, G., Stranix-Chibanda, L., et al. (2014). Group problem-solving therapy for postnatal depression among HIV-positive and HIV-negative mothers in Zimbabwe. J Int Assoc Provid AIDS Care 13(4): 335-341.
Chrousos, G. P., Torpy, D. J. and Gold, P. W. (1998). Interactions between the hypothalamic-pituitary-adrenal axis and the female reproductive system: clinical implications. Ann Intern Med 129(3): 229-240.
Clement, S., Schauman, O., Graham, T., Maggioni, F., Evans-Lacko, S., et al. (2015). What is the impact of mental health-related stigma on help-seeking? A systematic review of quantitative and qualitative studies. Psychol Med 45(1): 11-27.
Clinical Trials GOV (2023a). "A Study to Evaluate SAGE-217 in Participants With Severe Postpartum Depression." Retrieved 24 October, 2024, from https://clinicaltrials.gov/study/NCT02978326.
Clinical Trials GOV (2023b). "A Study to Evaluate the Efficacy and Safety of SAGE-217 in Participants With Severe Postpartum Depression (PPD)." Retrieved 24 October, 2024, from https://clinicaltrials.gov/study/NCT04442503.
Collett, D. (2023). Modelling survival data in medical research, Chapman and Hall/CRC.
Cooper, P. J., Murray, L., Wilson, A. and Romaniuk, H. (2003). Controlled trial of the short- and long-term effect of psychological treatment of post-partum depression: I. Impact on maternal mood. British Journal of Psychiatry 182(5): 412-419.
Cosci, F. and Chouinard, G. (2019). Chapter 7 - The Monoamine Hypothesis of Depression Revisited: Could It Mechanistically Novel Antidepressant Strategies? Neurobiology of Depression. Quevedo, J., Carvalho, A. F. and Zarate, C. A., Academic Press: 63-73.
Davies, J. and Read, J. (2019). A systematic review into the incidence, severity and duration of antidepressant withdrawal effects: Are guidelines evidence-based? Addictive Behaviors 97: 111-121.

De Crescenzo, F., Perelli, F., Armando, M. and Vicari, S. (2014). Selective serotonin reuptake inhibitors (SSRIs) for post-partum depression (PPD): a systematic review of randomized clinical trials. J Affect Disord 152-154: 39-44.
Decision Support Unit. (2022). "Estimating EQ-5D by Age and Sex for the UK (January 2022)." Retrieved 3 December 2024, from https://www.sheffield.ac.uk/media/34059/download?attachment.
Dekel, S., Ein-Dor, T., Ruohomäki, A., Lampi, J., Voutilainen, S., et al. (2019). The dynamic course of peripartum depression across pregnancy and childbirth. J Psychiatr Res 113: 72-78.
Deligiannidis, K. M., Citrome, L., Huang, M. Y., Acaster, S., Fridman, M., et al. (2023a). Effect of Zuranolone on Concurrent Anxiety and Insomnia Symptoms in Women With Postpartum Depression. J Clin Psychiatry 84(1).
Deligiannidis, K. M., Fales, C. L., Kroll-Desrosiers, A. R., Shaffer, S. A., Villamarin, V., et al. (2019). Resting-state functional connectivity, cortical GABA, and neuroactive steroids in peripartum and peripartum depressed women: a functional magnetic resonance imaging and spectroscopy study. Neuropsychopharmacology 44(3): 546-554.
Deligiannidis, K. M., Meltzer-Brody, S., Gunduz-Bruce, H., Doherty, J., Jonas, J., et al. (2021). Effect of Zuranolone vs Placebo in Postpartum Depression: A Randomized Clinical Trial. JAMA Psychiatry 78(9): 951-959.
Deligiannidis, K. M., Meltzer-Brody, S., Maximos, B., Peeper, E. Q., Freeman, M., et al. (2023b). Zuranolone for the Treatment of Postpartum Depression. Am J Psychiatry 180(9): 668-675.
Della Corte, L., La Rosa, V. L., Cassinese, E., Ciebiera, M., Zaręba, K., et al. (2022). Prevalence and associated psychological risk factors of postpartum depression: a cross-sectional study. J Obstet Gynaecol 42(5): 976-980.
Department of Health & Social Care (2022). "Women's Health Strategy for England (Policy Paper Updated 30 August 2022)." from https://www.gov.uk/government/publications/womens-health-strategy-for-england/womens-health-strategy-for-england.
Dikmen-Yildiz, P., Ayers, S. and Phillips, L. (2017). Depression, anxiety, PTSD and comorbidity in perinatal women in Turkey: A longitudinal population-based study. Midwifery 55: 29-37.
Du Preez, A., Conroy, S., Pawlby, S., Moran, P. and Pariante, C. M. (2016). Differential effects of ethnic density on the risk of postnatal depression and personality dysfunction. British Journal of Psychiatry 208(1): 49-55.
Duman, R. S., Sanacora, G. and Krystal, J. H. (2019). Altered Connectivity in Depression: GABA and Glutamate Neurotransmitter Deficits and Reversal by Novel Treatments. Neuron 102(1): 75-90.
Eldar-Lissai, A., Cohen, J. T., Meltzer-Brody, S., Gerbasi, M. E., Chertavian, E., et al. (2020). Cost-effectiveness of brexanolone versus selective serotonin reuptake inhibitors for the treatment of postpartum depression in the United States. Journal of Managed Care & Specialty Pharmacy 26(5): 627-638.
Epperson, C. N., Huang, M.-Y., Cook, K., Gupta, D., Chawla, A., et al. (2020). Healthcare resource utilization and costs associated with postpartum depression among commercially insured households. Current Medical Research and Opinion 36(10): 1707-1716.
European Medicines Agency (2023). Guideline on clinical investigation of medicinal products in the treatment of depression.
Fathi, F., Mohammad-Alizadeh-Charandabi, S. and Mirghafourvand, M. (2018). Maternal self-efficacy, postpartum depression, and their relationship with functional status in Iranian mothers. Women Health 58(2): 188-203.

FDA (2018). "Priority Review." Retrieved 23 October, 2024, from https://www.fda.gov/patients/fast-track-breakthrough-therapy-accelerated-approval-priority-review/priority-review.
Fee, C., Banasr, M. and Sibille, E. (2017). Somatostatin-Positive Gamma-Aminobutyric Acid Interneuron Deficits in Depression: Cortical Microcircuit and Therapeutic Perspectives. Biol Psychiatry 82(8): 549-559.
Feng, Y. F., Zhou, Y. Y. and Duan, K. M. (2024). The Role of Extrasynaptic GABA Receptors in Postpartum Depression. Mol Neurobiol 61(1): 385-396.
Fogaça, M. V. and Duman, R. S. (2019). Cortical GABAergic Dysfunction in Stress and Depression: New Insights for Therapeutic Interventions. Front Cell Neurosci 13: 87.
Food and Drug Administration (FDA) (2022). ZURZUVAE™ (zuranolone) NDA Application Integrated Review.
Food and Drug Administration (FDA) (June 2018). "Major depressive disorder: developing drugs for treatment." Retrieved March 16, 2022, from https://www.fda.gov/media/113988/download
Fraser, A., Macdonald-Wallis, C., Tilling, K., Boyd, A., Golding, J., et al. (2013). Cohort profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. International journal of epidemiology 42(1): 97-110.
Frieder, A., Fersh, M., Hainline, R. and Deligiannidis, K. M. (2019). Pharmacotherapy of Postpartum Depression: Current Approaches and Novel Drug Development. CNS Drugs 33(3): 265-282.
Gartlehner et al. (2016). Comparative Benefits and Harms of Antidepressant, Psychological, Complementary, and Exercise Treatments for Major Depression: An Evidence Report for a Clinical Practice Guideline From the American College of Physicians. Annals of Internal Medicine 164(5): 331-341.
Gerbasi, M. E., Eldar-Lissai, A., Acaster, S., Fridman, M., Bonthapally, V., et al. (2020). Associations between commonly used patient-reported outcome tools in postpartum depression clinical practice and the Hamilton Rating Scale for Depression. Arch Womens Ment Health 23(5): 727-735.
Gerbasi, M. E., Kosinski, M., Meltzer-Brody, S., Acaster, S., Fridman, M., et al. (2021). Achieving clinical response in postpartum depression leads to improvement in health-related quality of life. Curr Med Res Opin 37(7): 1221-1231.
Gheorghe, M., Varin, M., Wong, S. L., Baker, M., Grywacheski, V., et al. (2021). Symptoms of postpartum anxiety and depression among women in Canada: findings from a national cross-sectional survey. Can J Public Health 112(2): 244-252.
Goodman, J. H. (2004). Paternal postpartum depression, its relationship to maternal postpartum depression, and implications for family health. J Adv Nurs 45(1): 26-35.
GOV.UK (2021, 5 April 2024). "Innovative Licensing and Access Pathway." Retrieved 23 October, 2024, from https://www.gov.uk/guidance/innovative-licensing-and-access-pathway.
GOV.UK. (2024). "Drugs and pharmaceutical electronic market information tool (eMIT) Pharmex data for the period 1 July 2023 to 30 June 2023." Retrieved 23 July 2024, from https://www.gov.uk/government/publications/drugs-and-pharmaceutical-electronic-market-information-emit.
Grady, M. M. and Stahl, S. M. (2012). Practical guide for prescribing MAOIs: debunking myths and removing barriers. CNS Spectr 17(1): 2-10.
Hamilton, M. (1960). A rating scale for depression. J Neurol Neurosurg Psychiatry 23(1): 56-62.
Hanach, N., Radwan, H., Fakhry, R., Dennis, C.-L., issa, W. B., et al. (2023). Prevalence and risk factors of postpartum depression among women living in the United Arab Emirates. Social Psychiatry and Psychiatric Epidemiology 58(3): 395-407.
Hantsoo, L., Ward-O'Brien, D., Czarkowski, K. A., Gueorguieva, R., Price, L. H., et al. (2014). A randomized, placebo-controlled, double-blind trial of sertraline for postpartum depression. Psychopharmacology (Berl) 231(5): 939-948.

Hanusa, B. H., Scholle, S. H., Haskett, R. F., Spadaro, K. and Wisner, K. L. (2008). Screening for depression in the postpartum period: a comparison of three instruments. J Womens Health (Larchmt) 17(4): 585-596.
Harrison, S., Quigley, M. A., Fellmeth, G., Stein, A. and Alderdice, F. (2023). The impact of the Covid-19 pandemic on postnatal depression: analysis of three population-based national maternity surveys in England (2014-2020). Lancet Reg Health Eur 30: 100654.
Henshaw, E. J., Durkin, K. M. and Snell, R. J. (2016). First-time parents' shared representation of postpartum depressive symptoms: A qualitative analysis. Soc Sci Med 160: 102-110.
Hobden, B., Schwandt, M. L., Carey, M., Lee, M. R., Farokhnia, M., et al. (2017). The Validity of the Montgomery-Asberg Depression Rating Scale in an Inpatient Sample with Alcohol Dependence. Alcohol Clin Exp Res 41(6): 1220-1227.
Hoffmann, E., Nomikos, G. G., Kaul, I., Raines, S., Wald, J., et al. (2020). SAGE-217, A Novel GABA(A) Receptor Positive Allosteric Modulator: Clinical Pharmacology and Tolerability in Randomized Phase I Dose-Finding Studies. Clin Pharmacokinet 59(1): 111-120.
Horáková, A., Němcová, H., Mohr, P. and Sebela, A. (2022). Structural, functional, and metabolic signatures of postpartum depression: A systematic review. Front Psychiatry 13: 1044995.
Impact on Urban Health. "The Motherhood Group - Children's Mental Health " Retrieved 31 October, 2024, from https://urbanhealth.org.uk/partnerships/current-partnerships/the-motherhood-group.
Ioannidis, J. P. (2008). Effectiveness of antidepressants: an evidence myth constructed from a thousand randomized trials? Philos Ethics Humanit Med 3: 14.
Jankovic, J., Parsons, J., Jovanović, N., Berrisford, G., Copello, A., et al. (2020). Differences in access and utilisation of mental health services in the perinatal period for women from ethnic minorities—a population-based study. BMC Medicine 18(1): 245.
Jiang, Q., Qi, Y., Zhou, M., Dong, Y., Zheng, W., et al. (2024). Effect of esketamine on serum neurotransmitters in patients with postpartum depression: a randomized controlled trial. BMC Anesthesiol 24(1): 293.
Jones, I. and Shakespeare, J. (2014). Postnatal depression. Bmj 349: g4500.
Jones, R. M., Patel, K., Moscovici, M., McMaster, R., Glancy, G., et al. (2019). Adaptation of the Clinical Global Impression for Use in Correctional Settings: The CGI-C. Front Psychiatry 10: 687.
Kanes, S., Colquhoun, H., Gunduz-Bruce, H., Raines, S., Arnold, R., et al. (2017). Brexanolone (SAGE-547 injection) in post-partum depression: a randomised controlled trial. The Lancet 390(10093): 480-489.
Kaufman, Y., Carlini, S. V. and Deligiannidis, K. M. (2022). Advances in pharmacotherapy for postpartum depression: a structured review of standard-of-care antidepressants and novel neuroactive steroid antidepressants. Ther Adv Psychopharmacol 12: 20451253211065859.
Kerstis, B., Engström, G., Sundquist, K., Widarsson, M. and Rosenblad, A. (2012). The association between perceived relationship discord at childbirth and parental postpartum depressive symptoms: a comparison of mothers and fathers in Sweden. Ups J Med Sci 117(4): 430-438.
Knight, B. K., Patel R, Shakespeare J, Kotnis R, Kenyon S, Kurinczuk JJ (Eds.) on behalf of MBRRACE-UK. (2022). Saving Lives, Improving Mothers' Care Core Report - Lessons learned to inform maternity care from the UK and Ireland Confidential Enquiries into Maternal Deaths and Morbidity 2018-20. Oxford: National Perinatal Epidemiology Unit, University of Oxford 2022.
Kroenke, K., Spitzer, R. L. and Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 16(9): 606-613.

Lara, M. A., Navarrete, L., Nieto, L., Martín, J. P., Navarro, J. L., et al. (2015). Prevalence and incidence of perinatal depression and depressive symptoms among Mexican women. J Affect Disord 175: 18-24.
Lee, V. and Maguire, J. (2014). The impact of tonic GABAA receptor-mediated inhibition on neuronal excitability varies across brain region and cell type. Front Neural Circuits 8: 3.
Letourneau, N. L., Dennis, C. L., Benzies, K., Duffett-Leger, L., Stewart, M., et al. (2012). Postpartum depression is a family affair: addressing the impact on mothers, fathers, and children. Issues Ment Health Nurs 33(7): 445-457.
Leuchter, A. F., Cook, I. A., Hunter, A. M. and Korb, A. S. (2009). A new paradigm for the prediction of antidepressant treatment response. Dialogues Clin Neurosci 11(4): 435-446.
Levis, B., Negeri, Z., Sun, Y., Benedetti, A. and Thombs, B. D. (2020). Accuracy of the Edinburgh Postnatal Depression Scale (EPDS) for screening to detect major depression among pregnant and postpartum women: systematic review and meta-analysis of individual participant data. BMJ 371: m4022.
Li, Y. F. (2020). A hypothesis of monoamine (5-HT) - Glutamate/GABA long neural circuit: Aiming for fast-onset antidepressant discovery. Pharmacol Ther 208: 107494.
Luscher, B., Shen, Q. and Sahir, N. (2011). The GABAergic deficit hypothesis of major depressive disorder. Mol Psychiatry 16(4): 383-406.
Ma, H., Cai, M. and Wang, H. (2021). Emotional Blunting in Patients With Major Depressive Disorder: A Brief Non-systematic Review of Current Research. Front Psychiatry 12: 792960.
Maguire, J. (2019). Neuroactive Steroids and GABAergic Involvement in the Neuroendocrine Dysfunction Associated With Major Depressive Disorder and Postpartum Depression. Front Cell Neurosci 13: 83.
Maguire, J., Ferando, I., Simonsen, C. and Mody, I. (2009). Excitability changes related to GABAA receptor plasticity during pregnancy. J Neurosci 29(30): 9592-9601.
Maguire, J. and Mody, I. (2008). GABA(A)R plasticity during pregnancy: relevance to postpartum depression. Neuron 59(2): 207-213.
Marinus Pharmaceuticals (2017). A clinical trial of intravenous (IV) ganaxolone in women with postpartum depression. ClinicalTrials.gov. Bethesda, US National Library of Medicine.
Marinus Pharmaceuticals (2018). A clinical trial of oral ganaxolone in women with postpartum depression. ClinicalTrials.gov. Bethesda, US National Library of Medicine.
Martinez Botella, G., Salituro, F. G., Harrison, B. L., Beresis, R. T., Bai, Z., et al. (2017). Neuroactive Steroids. 2. 3α-Hydroxy-3β-methyl-21-(4-cyano-1H-pyrazol-1'-yl)-19-nor-5β-pregnan-20-one (SAGE-217): A Clinical Next Generation Neuroactive Steroid Positive Allosteric Modulator of the (γ-Aminobutyric Acid)(A) Receptor. J Med Chem 60(18): 7810-7819.
Maternal Mental Health Alliance (2023). Specialist perinatal mental health care in the UK 2023.
Matsumura, K., Hamazaki, K., Tsuchida, A., Kasamatsu, H. and Inadera, H. (2019). Education level and risk of postpartum depression: results from the Japan Environment and Children's Study (JECS). BMC Psychiatry 19(1): 419.
Maurer, D. M. (2012). Screening for depression. Am Fam Physician 85(2): 139-144.
Melo, E. F., Jr., Cecatti, J. G., Pacagnella, R. C., Leite, D. F., Vulcani, D. E., et al. (2012). The prevalence of perinatal depression and its associated factors in two different settings in Brazil. J Affect Disord 136(3): 1204-1208.
Meltzer-Brody, S., Colquhoun, H., Riesenberg, R., Epperson, C. N., Deligiannidis, K. M., et al. (2018). Brexanolone injection in post-partum depression: Two multicentre, double-blind, randomised, placebo-controlled, phase 3 trials. Lancet 392(10152): 1058-1070.

Meltzer-Brody, S., Gerbasi, M. E., Mak, C., Toubouti, Y., Smith, S., et al. (2024). Indirect comparisons of relative efficacy estimates of zuranolone and selective serotonin reuptake inhibitors for postpartum depression. Journal of Medical Economics 27(1): 582-595.
Meltzer-Brody, S. and Kanes, S. J. (2020). Allopregnanolone in postpartum depression: Role in pathophysiology and treatment. Neurobiol Stress 12: 100212.
Meltzer-Brody, S., Shirikjian, L., Laguerre, K., Moseley, M. K., Li, S., et al. (2023). Zuranolone Safety and Tolerability in Adults With Postpartum Depression: Analyses From SKYLARK, a 50 Mg Placebo-Controlled Study. Neuropsychopharmacology, SPRINGERNATURE CAMPUS, 4 CRINAN ST, LONDON, N1 9XW, ENGLAND.
Milgrom, J., Gemmill, A. W., Ericksen, J., Burrows, G., Buist, A., et al. (2015). Treatment of postnatal depression with cognitive behavioural therapy, sertraline and combination therapy: A randomised controlled trial. Australian & New Zealand Journal of Psychiatry 49(3): 236-245.
Misri, S., Reebye, P., Corral, M. and Milis, L. (2004). The use of paroxetine and cognitive-behavioral therapy in postpartum depression and anxiety: a randomized controlled trial. J Clin Psychiatry 65(9): 1236-1241.
Montano, C. B., Jackson, W. C., Vanacore, D. and Weisler, R. (2023). Considerations when selecting an antidepressant: a narrative review for primary care providers treating adults with depression. Postgraduate Medicine 135(5): 449-465.
Moore Simas, T. A., Huang, M. Y., Packnett, E. R., Zimmerman, N. M., Moynihan, M., et al. (2020). Matched cohort study of healthcare resource utilization and costs in young children of mothers with postpartum depression in the United States. J Med Econ 23(2): 174-183.
Moore Simas, T. A., Huang, M. Y., Patton, C., Reinhart, M., Chawla, A. J., et al. (2019). The humanistic burden of postpartum depression: a systematic literature review. Curr Med Res Opin 35(3): 383-393.
Morgan, J. K., Santosa, H., Fridley, R. M., Conner, K. K., Hipwell, A. E., et al. (2021). Postpartum Depression Is Associated With Altered Neural Connectivity Between Affective and Mentalizing Regions During Mother-Infant Interactions. Front Glob Womens Health 2: 744649.
Müller, M. J., Szegedi, A., Wetzel, H. and Benkert, O. (2000). Moderate and severe depression. Gradations for the Montgomery-Asberg Depression Rating Scale. J Affect Disord 60(2): 137-140.
Myczkowski, M. L., Dias, Á. M., Luvisotto, T., Arnaut, D., Bellini, B. B., et al. (2012). Effects of repetitive transcranial magnetic stimulation on clinical, social, and cognitive performance in postpartum depression. Neuropsychiatric Disease and Treatment 8.
Nafees, B., Lloyd, A. J., Dewilde, S., Rajan, N. and Lorenzo, M. (2017). Health state utilities in non-small cell lung cancer: An international study. Asia Pac J Clin Oncol 13(5): e195-e203.
National Health Service (2018). "The Perinatal Mental Health Care Pathways. NHS England and NHS Improvement." from https://www.england.nhs.uk/wp-content/uploads/2018/05/perinatal-mental-health-care-pathway.pdf.
National Health Service (2019). "The NHS Long Term Plan." from https://www.longtermplan.nhs.uk/wp-content/uploads/2019/08/nhs-long-term-plan-version-1.2.pdf.
National Health Service England (2023). "Three year delivery plan for maternity and neonatal services." from https://www.england.nhs.uk/publication/three-year-delivery-plan-for-maternity-and-neonatal-services/.
National Institute for Health and Care Excellence (2020). "Clinical guideline [CG129]. Antenatal and postnatal mental health: clinical management and service guidance." Retrieved August 20, 2024, from https://www.nice.org.uk/guidance/cg192.
National Institute for Health and Care Excellence (2022). Depression in adults: treatment and management: NICE guideline [NG222].
Company evidence submission for zuranolone for treating postnatal depression [ID6431]

National Institute for Health and Care Excellence (2023a). "Depression - antenatal and postnatal." Retrieved August 20, 2024, from https://cks.nice.org.uk/topics/depression-antenatal-postnatal/.
National Institute for Health and Care Excellence (2023b). Maribavir for treating refractory cytomegalovirus infection after transplant. Technology appraisal guidance. Reference number:TA860.
National Institute for Health and Care Excellence (2025). Health Technology Evaluation. Zuranolone for treating postnatal depression [ID6431]. Final scope.
National Institute for Health and Care Excellence. (2022). "NICE health technology evaluations: the manual. NICE process and methods [PMG36; 31 January 2022]." Retrieved 23 July 2024, from https://www.nice.org.uk/process/pmg36.
National Institute of Mental Health (NIMH) (2022). "Perinatal depression." Retrieved February 25, 2022, from https://www.nimh.nih.gov/health/publications/perinatal-depression.
Netsi, E., Pearson, R. M., Murray, L., Cooper, P., Craske, M. G., et al. (2018). Association of Persistent and Severe Postnatal Depression With Child Outcomes. JAMA Psychiatry 75(3): 247-253.
NHS Devon Partnership (2021). "Black Maternal Mental Health Awareness Week " Retrieved 31 October, 2024, from https://www.dpt.nhs.uk/news/black-maternal-mental-health-awareness-week-2021.
NHS England (2021). "Overview - Antidepressants." Retrieved 21 October, 2024, from https://www.nhs.uk/mental-health/talking-therapies-medicine-treatments/medicines-and-psychiatry/antidepressants/overview/.
NHS England. (2022). "How and when to take fluoxetine." Retrieved 2025, January, from https://www.nhs.uk/medicines/fluoxetine-prozac/how-and-when-to-take-fluoxetine/.
NHS England. (2024). "2022/23 National Cost Collection Data Publication. National Schedule of NHS Costs 2022/23." Retrieved 3 December 2024, from https://www.england.nhs.uk/wp-content/uploads/2024/07/national-schedule-of-NHS-costs-22-23-lns.xlsx.
Niarchou, E., Roberts, L. H. and Naughton, B. D. (2024). What is the impact of antidepressant side effects on medication adherence among adult patients diagnosed with depressive disorder: A systematic review. J Psychopharmacol 38(2): 127-136.
Nishimura, A., Fujita, Y., Katsuta, M., Ishihara, A. and Ohashi, K. (2015). Paternal postnatal depression in Japan: an investigation of correlated factors including relationship with a partner. BMC Pregnancy and Childbirth 15: 1-8.
Noorlander, Y., Bergink, V. and van den Berg, M. P. (2008). Perceived and observed mother-child interaction at time of hospitalization and release in postpartum depression and psychosis. Arch Womens Ment Health 11(1): 49-56.
Northoff, G. and Sibille, E. (2014). Why are cortical GABA neurons relevant to internal focus in depression? A cross-level model linking cellular, biochemical and neural network findings. Mol Psychiatry 19(9): 966-977.
O'Callaghan, L., Chertavian, E., Johnson, S. J., Ferries, E. and Deligiannidis, K. M. (2024). The cost-effectiveness of zuranolone versus selective serotonin reuptake inhibitors for the treatment of postpartum depression in the United States. J Med Econ 27(1): 492-505.
O'Hara, M. W., Pearlstein, T., Stuart, S., Long, J. D., Mills, J. A., et al. (2019). A placebo controlled treatment trial of sertraline and interpersonal psychotherapy for postpartum depression. Journal of Affective Disorders 245: 524-532.
Office for National Statistics. (2024). "National life tables: UK." Retrieved 29 November 2024, from https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/datasets/nationallifetablesunitedkingdomreferencetables.
Office for National Statistics. (2025). Consumer price inflation time series.

Parsa, B., Parsa, P., Boojarzadeh, B. and Mohammadi, Y. (2019). Prediction of postpartum depression based on women's quality of life. Family Medicine & Primary Care Review 21(4): 343-348.
Paulson, J. F. and Bazemore, S. D. (2010). Prenatal and postpartum depression in fathers and its association with maternal depression: a meta-analysis. Jama 303(19): 1961-1969.
Peeper, E. Q., Moseley, M. K., LaGuerre, K., Huang, M. Y., Mak, C., et al. (2023). Impact of time to treatment with zuranolone on efficacy stratified by postpartum depression onset in the phase 3 SKYLARK study. Neuroscience Applied 2: 102969.
Pennington, E. L., Barner, J. C., Brown, C. M., Moczygemba, L. R., Patel, D. A., et al. (2025). Antidepressant adherence using group-based trajectory modeling among postpartum women with Texas Medicaid. J Manag Care Spec Pharm 31(2): 167-178.
Personal Social Services Research Unit. (2023). "Unit Costs of Health and Social Care 2023 Manual." Retrieved 23 July 2024, from https://kar.kent.ac.uk/105685/1/The%20unit%20costs%20of%20health%20and%20social%20care_Final3.pdf.
Petersen, I., Peltola, T., Kaski, S., Walters, K. R. and Hardoon, S. (2018). Depression, depressive symptoms and treatments in women who have recently given birth: UK cohort study. BMJ Open 8(10): e022152.
Petrou, S., Cooper, P., Murray, L. and Davidson, L. L. (2002). Economic costs of post-natal depression in a high-risk British cohort. Br J Psychiatry 181: 505-512.
Pickard, R., Lam, T., Maclennan, G., Starr, K., Kilonzo, M., et al. (2012). Types of urethral catheter for reducing symptomatic urinary tract infections in hospitalised adults requiring short-term catheterisation: multicentre randomised controlled trial and economic evaluation of antimicrobial- and antiseptic-impregnated urethral catheters (the CATHETER trial). Health Technol Assess 16(47): 1-197.
Posmontier, B. (2008). Functional status outcomes in mothers with and without postpartum depression. J Midwifery Womens Health 53(4): 310-318.
Putnick, D. L., Sundaram, R., Bell, E. M., Ghassabian, A., Goldstein, R. B., et al. (2020). Trajectories of Maternal Postpartum Depressive Symptoms. Pediatrics 146(5).
Quilty, L. C., Robinson, J. J., Rolland, J. P., Fruyt, F. D., Rouillon, F., et al. (2013). The structure of the Montgomery-Åsberg depression rating scale over the course of treatment for depression. Int J Methods Psychiatr Res 22(3): 175-184.
Raja, A., Ahmed, S., Basit Ali Siddiqui, M., Lamiya Mir, S., Kumar, R., et al. (2024). Evaluating the safety and efficacy of zuranolone in the management of major depressive disorder and postpartum depression, with or without concurrent insomnia: a rigorous systematic review and meta-analysis. Front Psychiatry 15: 1425295.
Reddy, D. S. and Estes, W. A. (2016). Clinical Potential of Neurosteroids for CNS Disorders. Trends Pharmacol Sci 37(7): 543-561.
Richards, J. B., Papaioannou, A., Adachi, J. D., Joseph, L., Whitson, H. E., et al. (2007). Effect of selective serotonin reuptake inhibitors on the risk of fracture. Arch Intern Med 167(2): 188-194.
Roumieh, M., Bashour, H., Kharouf, M. and Chaikha, S. (2019). Prevalence and risk factors for postpartum depression among women seen at Primary Health Care Centres in Damascus. BMC Pregnancy and Childbirth 19(1): 519.
Royal College of Obstetricians & Gynaecologists. "Maternal mental health - women's voices." from https://www.rcog.org.uk/for-the-public/rcog-engagement-listening-to-patients/maternal-mental-health-womens-voices/.
Royal College of Psychiatrists (2018, November 2021). "Mother and Baby Units (MBUs)." Retrieved 05/02/2025, from https://www.rcpsych.ac.uk/mental-health/treatments-and-wellbeing/mother-and-baby-units-(mbus).
Royal College of Psychiatrists (2021). Perinatal mental health services: Recommendations for the provision of services for childbearing women.
Sage Therapeutics (2019). ROBIN Clinical Study Report.

Sage Therapeutics (2022a). SKYLARK Clinical Study Report.
Sage Therapeutics (2022b). SKYLARK Clinical Study Report. Tables.
Salahudeen, M., Wright, C. and Peterson, G. (2020). Esketamine: new hope for the treatment of treatment-resistant depression? A narrative review. Ther Adv Drug Saf 11: 2042098620937899.
Sambrook Smith, M., Lawrence, V., Sadler, E. and Easter, A. (2019). Barriers to accessing mental health services for women with perinatal mental illness: systematic review and meta-synthesis of qualitative studies in the UK. BMJ Open 9(1): e024803.
Sansone, R. A. and Sansone, L. A. (2010). SSRI-induced indifference. Psychiatry (Edgmont) 7(10): 14.
Santarsieri, D. and Schwartz, T. L. (2015). Antidepressant efficacy and side-effect burden: a quick guide for clinicians. Drugs Context 4: 212290.
Sapra, A., Bhandari, P., Sharma, S., Chanpura, T. and Lopp, L. (2020). Using Generalized Anxiety Disorder-2 (GAD-2) and GAD-7 in a Primary Care Setting. Cureus 12(5): e8224.
Sarawagi, A., Soni, N. D. and Patel, A. B. (2021). Glutamate and GABA Homeostasis and Neurometabolism in Major Depressive Disorder. Front Psychiatry 12: 637863.
SAS (2024). "SAS® 9.4 Macro Language: Reference, Fifth Edition." Retrieved 05/02/2025, 2025, from https://documentation.sas.com/doc/en/pgmsascdc/9.4_3.5/mcrolref/titlepage.htm.
Schmidt, R. M., Wiemann, C. M., Rickert, V. I. and Smith, E. O. (2006). Moderate to severe depressive symptoms among adolescent mothers followed four years postpartum. J Adolesc Health 38(6): 712-718.
Sharp, D. J., Chew-Graham, C., Tylee, A., Lewis, G., Howard, L., et al. (2010). A pragmatic randomised controlled trial to compare antidepressants with a community-based psychosocial intervention for the treatment of women with postnatal depression: the RESPOND trial. Health Technol Assess 14(43): iii-iv, ix-xi, 1-153.
Shi, X., Ying, Y., Yu, Z., Xing, M., Zhu, J., et al. (2021). Risk factors for postpartum depression in Chinese women: A cross-sectional study at 6 weeks postpartum. J Psychosom Res 140: 110295.
Sit, D., Luther, J., Buysse, D., Dills, J. L., Eng, H., et al. (2015). Suicidal ideation in depressed postpartum women: Associations with childhood trauma, sleep disturbance and anxiety. J Psychiatr Res 66-67: 95-104.
Siu, A. L., Bibbins-Domingo, K., Grossman, D. C., Baumann, L. C., Davidson, K. W., et al. (2016). Screening for Depression in Adults: US Preventive Services Task Force Recommendation Statement. Jama 315(4): 380-387.
Slomian, J., Honvo, G., Emonts, P., Reginster, J. Y. and Bruyère, O. (2019). Consequences of maternal postpartum depression: A systematic review of maternal and infant outcomes. Womens Health (Lond) 15: 1745506519844044.
Snaith, R. P., Harrop, F. M., Newby, D. A. and Teale, C. (1986). Grade scores of the Montgomery-Asberg Depression and the Clinical Anxiety Scales. Br J Psychiatry 148: 599-601.
Stewart, D. E. and Vigod, S. (2016). Postpartum Depression. New England Journal of Medicine 375(22): 2177-2186.
Subbiah, G. K., de Kroon, M. L. A., Boere-Boonekamp, M. M., van der Zee-van den Berg, A. I., Hartman, C. A., et al. (2025). Maternal postpartum mental health negatively affects infants' health related quality of life. Journal of Affective Disorders 374: 381-389.
Suchting, R., Tirumalajaru, V., Gareeb, R., Bockmann, T., de Dios, C., et al. (2021). Revisiting monoamine oxidase inhibitors for the treatment of depressive disorders: A systematic review and network meta-analysis. J Affect Disord 282: 1153-1160.
Sullivan, P. W., Slejko, J. F., Sculpher, M. J. and Ghushchyan, V. (2011). Catalogue of EQ-5D scores for the United Kingdom. Med Decis Making 31(6): 800-804.

Taylor, S., Annand, F., Burkinshaw, P., Greaves, F., Kelleher, M., et al. (2019). Dependence and withdrawal associated with some prescribed medicines. London: Public Health England.
Terrone, G., Bianciardi, E., Fontana, A., Pinci, C., Castellani, G., et al. (2023). Psychological Characteristics of Women with Perinatal Depression Who Require Psychiatric Support during Pregnancy or Postpartum: A Cross-Sectional Study. Int J Environ Res Public Health 20(8).
Thase, M. E. (2023). A New Option for Depressed Patients Who Do Not Respond to Antidepressant Medications. American Journal of Psychiatry 180(3): 188-189.
Thorsteinsson, E. B., Loi, N. M. and Farr, K. (2018). Changes in stigma and help-seeking in relation to postpartum depression: non-clinical parenting intervention sample. PeerJ 6: e5893.
Trivedi, M. H., Rush, A. J., Wisniewski, S. R., Nierenberg, A. A., Warden, D., et al. (2006). Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. Am J Psychiatry 163(1): 28-40.
Turner, D., Little, P., Raftery, J., Turner, S., Smith, H., et al. (2010). Cost effectiveness of management strategies for urinary tract infections: results from randomised controlled trial. Bmj 340.
Vliegen, N., Casalin, S. and Luyten, P. (2014). The course of postpartum depression: a review of longitudinal studies. Harv Rev Psychiatry 22(1): 1-22.
Watson, H., Harrop, D., Walton, E., Young, A. and Soltani, H. (2019). A systematic review of ethnic minority women's experiences of perinatal mental health conditions and services in Europe. PLOS ONE 14(1): e0210587.
Winslow, M., White, E., Rose, S. J., Salzer, E. and Nemec, E. C., 2nd (2024). The efficacy of zuranolone versus placebo in postpartum depression and major depressive disorder: a systematic review and meta-analysis. Int J Clin Pharm 46(3): 590-601.
Wisner, K. L., Hanusa, B. H., Perel, J. M., Peindl, K. S., Piontek, C. M., et al. (2006). Postpartum Depression: A Randomized Trial of Sertraline Versus Nortriptyline. Journal of Clinical Psychopharmacology 26(4).
Wisner, K. L., Sit, D. K., Moses-Kolko, E. L., Driscoll, K. E., Prairie, B. A., et al. (2015). Transdermal Estradiol Treatment for Postpartum Depression: A Pilot, Randomized Trial. J Clin Psychopharmacol 35(4): 389-395.
Yonkers, K. A., Lin, H., Howell, H. B., Heath, A. C. and Cohen, L. S. (2008). Pharmacologic treatment of postpartum women with new-onset major depressive disorder: a randomized controlled trial with paroxetine. J Clin Psychiatry 69(4): 659-665.
Yu, H., Shen, Q., Bränn, E., Yang, Y., Oberg, A. S., et al. (2024). Perinatal Depression and Risk of Suicidal Behavior. JAMA Network Open 7(1): e2350897-e2350897.
Zhang, Y., Cheng, Y., Carrillo-Larco, R. M., Zhou, Y., Wang, H., et al. (2025). Postpartum depression in relation to chronic diseases and multimorbidity in women's mid-late life: a prospective cohort study of UK Biobank. BMC Med 23(1): 24.
Zorumski, C. F., Paul, S. M., Covey, D. F. and Mennerick, S. (2019). Neurosteroids as novel antidepressants and anxiolytics: GABA-A receptors and beyond. Neurobiol Stress 11: 100196.

n-NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Single technology appraisal

Zuranolone for treating postnatal depression [ID6431]

Summary of Information for Patients (SIP)

March 11, 2025

Template version	Date amended	Changes since previous version
2.0	Dec 2023	Clarifications made to guidance notes in section 3i regarding inclusion of statements on cost effectiveness.

File name	Version	Contains confidential information	Date
ID6431_Zuranolone for the treatment of PND_Biogen_SIP_Final_v1.1	**Version 1.1**	**Yes**	**March 25, 2025**

Summary of Information for Patients (SIP):

The pharmaceutical company perspective

What is the SIP?

The Summary of Information for Patients (SIP) is written by the company who is seeking approval from NICE for their treatment to be sold to the NHS for use in England. It is a plain English summary of their submission written for patients participating in the evaluation. It is not independently checked, although members of the public involvement team at NICE will have read it to double-check for marketing and promotional content before it is sent to you.

The **Summary of Information for Patients** template has been adapted for use at NICE from the Health Technology Assessment International – Patient & Citizens Involvement Group (HTAi PCIG). Information about the development is available in an open-access IJTAHC journal article.

SECTION 1: Submission summary

1a) Name of the medicine (generic and brand name):

Generic name: zuranolone Brand name: Zurzuvae®

1b) Population this treatment will be used by. Please outline the main patient population that is being appraised by NICE:

If the marketing authorisation is granted, zuranolone will be used for the treatment of postnatal depression (PND) in adults. It should be taken after giving birth by adults aged 18 years or older. The proposed patient population is the same as the patient population being appraised by the National Institute for Health and Care Excellence (NICE).

1c) Authorisation: Please provide marketing authorisation information, date of approval and link to the regulatory agency approval. If the marketing authorisation is pending, please state this, and reference the section of the company submission with the anticipated dates for approval.

The proposed indication for zuranolone is the treatment of PND in adults, following childbirth. The marketing authorization is pending. The anticipated date of approval can be found in table 2 in section 1.2 (description of the technology being evaluated) of the company submission.

1d) Disclosures. Please be transparent about any existing collaborations (or broader conflicts of interest) between the pharmaceutical company and patient groups relevant to the medicine. Please outline the reason and purpose for the engagement/activity and any financial support provided:

Nothing to disclose.

SECTION 2: Current landscape

2a) The condition – clinical presentation and impact

Please provide a few sentences to describe the condition that is being assessed by NICE and the number of people who are currently living with this condition in England.

Please outline in general terms how the condition affects the quality of life of patients and their families/caregivers. Please highlight any mortality/morbidity data relating to the condition if available. If the company is making a case for the impact of the treatment on carers this should be clearly stated and explained.

PND is a common and serious mental health event happening between the final trimester of pregnancy and 1 year following childbirth [1, 2]. PND is a potentially debilitating condition that can lead to severe short- and long-term consequences for affected patients, children, and their families [3]. In the UK, it is estimated that up to 16% of individuals experience PND symptoms after giving birth [4-6]. However, published literature suggest that PND is often under-detected with only 4%-15% of women suffering from PND receiving a formal diagnosis, suggesting that many patients remain undiagnosed [7, 8].

Patients with PND may feel low or sad, lose interest in things they used to enjoy, have trouble sleeping, experience feelings of guilt or worthlessness, experience changes in weight, have difficulty concentrating, or feel slowed down or anxious [9, 10]. These symptoms can be different for each person, and they can change from day to day. Patients with PND often struggle with personal, social, and household tasks and may have a harder time taking care of themselves and their baby [11]. As a result, PND can have a substantial negative impact on daily functioning and quality of life from both a physical and mental health perspective, with these negative impacts being potentially long-lasting [3, 12].

PND can have very serious consequences and in the most severe of cases can lead to suicide. Maternal suicide was the leading cause of direct death within a year after pregnancy from 2017 to 2019. In 2020, the risk of suicide during pregnancy or up to 6 weeks postpartum tripled, and psychiatric causes accounted for nearly 40% of deaths within a year after birth [13].

Importantly, the impact of PND extends beyond the patient and can affect the whole family, including the child/children and the partner of a patient. Lack of or disrupted bonding of a PND patient with the child can have long-term effects on the child's development, including their language skills, behaviour, and emotional well-being [5]. Children whose mothers had PND might have behavioural problems, do worse in school, and are more likely to feel depressed when they are older [14]. Partners of patients with PND are also impacted. Compared to partners of unaffected individuals, they are more likely to feel stressed, anxious, or depressed themselves [3]. As a result, PND can strain relationships and may lead to lower satisfaction in relationships [15, 16].

2b) Diagnosis of the condition (in relation to the medicine being evaluated)

Please briefly explain how the condition is currently diagnosed and how this impacts patients. Are there any additional diagnostic tests required with the new treatment?

PND is often under-detected and may be overlooked with normal changes in emotions and mental state that happen after having a baby [9]. These changes can hide signs of depression or be confused with other conditions like "baby blues" (which only lasts up to 2 weeks) [2, 17]. Additionally, some patients with PND may be unwilling to discuss their problem due to fear of stigma and taboo associated with mental health conditions.

Multiple healthcare professionals may be involved in screening and diagnosing PND. Health visitors may refer a patient to a general practitioner (GP) or community mental health team if they suspect PND. Patients may also be screened by midwives at the hospital or during their GP check-ups.

If a healthcare professional thinks a patient might have PND, they will follow recommendations and guidelines set by health organisations like National Health Service (NHS) England and NICE to assess the patient [18, 19]. These guidelines help to ensure proper identification and care are offered to those who need it. The guidelines outline a number of tools that healthcare professionals (midwives, GPs, psychiatrists) can use to check for PND. The Edinburgh Postnatal Depression Scale (EPDS) is the most commonly used test in a form of a short questionnaire used by doctors to screen for PND after childbirth.

To diagnose PND, healthcare professionals look at when symptoms started, how long they have lasted, and how they affect the patient's daily life. They also consider the results from the screening tools. However, under-detection of PND and under-diagnosis remain a significant challenge. Treatment with zuranolone is not anticipated to require additional diagnostic tests.

2c) Current treatment options:

The purpose of this section is to set the scene on how the condition is currently managed:

- What is the treatment pathway for this condition and where in this pathway the medicine is likely to be used? Please use diagrams to accompany text where possible. Please give emphasis to the specific setting and condition being considered by NICE in this review. For example, by referencing current treatment guidelines. It may be relevant to show the treatments people may have before and after the treatment under consideration in this SIP.
- Please also consider:
    - if there are multiple treatment options, and data suggest that some are more commonly used than others in the setting and condition being considered in this SIP, please report these data.
    - are there any drug–drug interactions and/or contraindications that commonly cause challenges for patient populations? If so, please explain what these are.

In the UK, at the present time there are no approved treatments indicated for PND. Many of the expected cases of PND go undiagnosed and thus untreated. The available treatment options are limited to psychological therapies and off-label use of antidepressants (i.e., antidepressants that do not have a marketing authorisation in PND) – together these treatment options are part of the established clinical management.

Zuranolone is expected to be used as an alternative or as a complementary option to established clinical management. It is expected to be a short-term treatment with rapid effect, which may reduce the need for long-term follow-up.

NICE currently recommends that patients with mild to moderate PND try psychological therapies first (for example, a guided self-help, which focuses on structured activities to help change thoughts and behaviours, with support from a trained helper). For moderate to severe PND, more intensive psychological therapies or antidepressants might be recommended if the risks associated with the medication and the mental health problem in pregnancy and the postnatal period are understood. The three main classes of antidepressants currently used for PND treatment include selective serotonin reuptake inhibitors (SSRIs), serotonin and norepinephrine reuptake inhibitors (SNRIs) and tricyclic antidepressants (TCAs) [2]. Sometimes, a mix of both a psychological therapy and antidepressants can be used if one intervention alone does not offer sufficient help.

Despite their lack of marketing authorisation in PND, antidepressants are often prescribed. However, they present substantial drawbacks:

- It takes a long time to assess if they are working or not (on average 4 to 6 weeks, but can take up to 12 weeks) [20]
- They need to be taken long-term (6-12 months) [21, 22]
- They can cause side effects (like weight gain, feeling sick, trouble sleeping, or anxiety) [23, 24], which can make them challenging to take regularly (46% to 83% of individuals don't take them as prescribed) [25]
- Patients can experience unpleasant withdrawal symptoms (27% to 86%) [26]

Given these limitations, there is a clear need for better PND treatments that offer faster relief of depressive symptoms and have fewer side effects.

2d) Patient-based evidence (PBE) about living with the condition

Context:

- **Patient-based evidence (PBE)** is when patients input into scientific research, specifically to provide experiences of their symptoms, needs, perceptions, quality of life issues or experiences of the medicine they are currently taking. PBE might also include carer burden and outputs from patient preference studies, when conducted in order to show what matters most to patients and carers

and where their greatest needs are. Such research can inform the selection of patient-relevant endpoints in clinical trials.

In this section, please provide a summary of any PBE that has been collected or published to demonstrate what is understood about **patient needs and disease experiences**. Please include the methods used for collecting this evidence. Any such evidence included in the SIP should be formally referenced wherever possible and references included.

Patient-based evidence highlights that those affected by PND experience profound impacts on various crucial aspects of their quality of life. Overall, PND patients struggle more with daily functioning, including personal, social, and household tasks, compared to unaffected individuals [11].

The following findings describe key impacts of PND as reported by PND patients in studies leveraging clinical interviews [11] and online surveys [27], systematic literature reviews evaluating humanistic burden on PND patients [3] and qualitative research [28] as well as UK-specific semi-structured interviews [29, 30]:

- Physical and mental well-being: A majority of PND patients (94%) report mood disorder symptoms affecting daily life, with fewer feeling physically well compared to the general population [3, 27]
- Social and family impact: PND disrupts family life, relationships, and work/school performance. Many patients report being judged in social settings due to their condition [27]
- Psychological effects: Patients with PND often feel a loss of control and sense of self, experiencing low resilience in handling their new situation [28]
- Conflicting feelings and relationships: Patients with PND experience conflicting feelings towards their babies, partners, and in-laws, causing distress and suffering. This can potentially lead to negative bonding and difficulties in day-to-day interactions. The core symptoms of depression (loss of interest and low mood) can interfere with parenting tasks and affect relationships with both child and partner [28]
- Treatment perceptions and access: While most believe in the effectiveness of proper treatment, many PND sufferers report difficulties accessing quality mental health care [27, 29, 30]

This evidence underscores the comprehensive impact of PND on the lives of impacted patients, their partners and children, highlighting the need for accessible interventions for those affected.

SECTION 3: The treatment

3a) How does the new treatment work?

What are the important features of this treatment?

Please outline as clearly as possible important details that you consider relevant to patients relating to the mechanism of action and how the medicine interacts with the body

Where possible, please describe how you feel the medicine is innovative or novel, and how this might be important to patients and their communities.

If there are relevant documents which have been produced to support your regulatory submission such as a summary of product characteristics or patient information leaflet, please provide a link to these.

Zuranolone is a novel antidepressant medicine used to treat PND. It is a fast-acting treatment, which is taken orally at home once daily (two 25 mg capsules), every day for 14 days as a single course of treatment. It can be taken after giving birth by adults aged 18 years or older.

The exact biological reasons why some individuals get PND are unclear, however, it is likely due to a mix of factors such as changes in brain chemicals, hormones, life stress, and a lack of support [4]. It is thought that one brain chemical called gamma-aminobutyric acid, referred to as GABA, plays a particularly important role in the regulation of mood and stress, and the disruption of GABA-related system contributes to development of PND in some individuals [31].

Zuranolone works by targeting the brain's GABA system, which helps control mood and anxiety. Zuranolone is similar to allopregnanolone, a natural substance that is produced by the brain, which decreases rapidly in patients after giving birth. By boosting GABA activity in the brain, zuranolone helps to restore balance in brain function that is thought to be disrupted in those suffering from PND [32-34].

Another key benefit associated with zuranolone treatment is how quickly it works. Many individuals start feeling better as soon as Day 3 after starting the treatment. Because the treatment course is short (14 days) and is well tolerated, more patients may be willing to start and continue the treatment. In the zuranolone clinical trial program, 91% of patients completed the full treatment course [35, 36].

In addition, zuranolone treatment is convenient for patients as it is taken at home. This could be particularly helpful for patients living far from healthcare centres or those facing economic challenges and for whom healthcare visits could be burdensome.

Zuranolone is anticipated to be the first treatment in the UK designed specifically for PND, making it different from current antidepressants. Zuranolone can help address the current unmet needs in PND due to the following:

Zuranolone has been developed specifically for treatment of PND, and was designed to target the changes which are believed to contribute to development of PND

- Zuranolone is fast-acting. Treated patients may start feeling better as early as Day 3 of the treatment [35, 36]. Since PND is known to have a negative impact on child's development, this quick relief is important for both the patient and the child as it can support their bonding, which is crucial for development of the child
- Zuranolone is well tolerated and unlike daily antidepressants, which may need to be taken over a long period of time, zuranolone is a short-course treatment that needs to be taken for just two weeks, which may improve adherence to treatment

Zuranolone represents an important development in the treatment of PND. If granted marketing authorisation, it will be the first UK-approved PND-specific medication, with the potential to address certain limitations of existing treatment options. This new approach offers benefits for patients with PND, allowing for fast resolution of symptoms while having a well-defined safety and tolerability profile, improving patient outcomes, and treatment adherence.

3b) Combinations with other medicines

Is the medicine intended to be used in combination with any other medicines?

- Yes / No

If yes, please explain why and how the medicines work together. Please outline the mechanism of action of those other medicines so it is clear to patients why they are used together.

If yes, please also provide information on the availability of the other medicine(s) as well as the main side effects.

If this submission is for a combination treatment, please ensure the sections on efficacy (3e), quality of life (3f) and safety/side effects (3g) focus on data that relate to the combination, rather than the individual treatments.

In the zuranolone clinical trial program, patients were allowed concomitant use of antidepressants and/or psychological therapies. As a result, the data support the use of zuranolone to be taken alone or in combination with currently used treatments.

Zuranolone was demonstrated to be well-tolerated by all patients, regardless of concomitant use with other treatments [37].

3c) Administration and dosing

How and where is the treatment given or taken? Please include the dose, how often the treatment should be given/taken, and how long the treatment should be given/taken for.

How will this administration method or dosing potentially affect patients and caregivers? How does this differ to existing treatments?

Zuranolone is recommended to be taken at a dose of 50 mg (two 25 mg capsules) that are taken orally, once a day, in the evening (preferably with fat-containing food), over a period of 14 days. Zuranolone must be taken for the whole treatment course (i.e., 14 days), even if patients start to feel better early on during the treatment course. Healthcare practitioners may reduce the dose to 40 mg (two 20 mg capsules) taken once daily if a patient has trouble with side effects.

The convenient, ready-to-use (fixed dosing) oral formulation means that patients can take zuranolone at home.

Unlike daily antidepressants which may need to be taken over a long period of time, zuranolone is a short-course treatment that needs to be taken for just two weeks, which may improve adherence to treatment.

3d) Current clinical trials

Please provide a list of completed or ongoing clinical trials for the treatment. Please provide a brief top-level summary for each trial, such as title/name, location, population, patient group size, comparators, key inclusion and exclusion criteria and completion dates etc. Please provide references to further information about the trials or publications from the trials.

Completed Trials in PND [35, 36]:

SKYLARK:

- SKYLARK was a randomised, double-blinded, placebo-controlled study to evaluate the efficacy and safety of zuranolone 50 mg once daily for 14-days compared to placebo in adult patients with PND.
- The study included 200 patients and was conducted across 78 sites in the United States, 9 sites in Spain, and 5 sites in the United Kingdom
- Study follow-up: 45 days
- SKYLARK was completed on 12/04/2022
- ClinicalTrials.gov link: SKYLARK (NCT04442503)
- Publication with the trials results: Zuranolone for the Treatment of Postpartum Depression

ROBIN:

- ROBIN was a randomised, double-blinded, placebo-controlled study to evaluate the efficacy and safety of zuranolone 30 mg once daily for 14-days compared to placebo in the treatment of adult patients with PND.
- The study included 153 patients and was conducted across 33 centres in the United States
- Study follow-up: 45 days

- ROBIN was completed on 11/12/2018
- ClinicalTrials.gov link: ROBIN (NCT03238963)
- Publication with the trials results: Effect of Zuranolone vs Placebo in Postpartum Depression

In both studies, patients were allowed to receive established clinical management (off-label antidepressants and psychological therapies), both in the study arm receiving placebo and in the study arm receiving zuranolone treatment.

Key inclusion criteria for both trials are as follows:

- Adult female patients
- Aged 18 to 45 years old
- 6 months (ROBIN)/12 months (SKYLARK) or less after birth
- Major depressive episode that began no earlier than the third trimester and no later than the first 4 weeks following delivery
- A baseline score of 26 or higher on the Hamilton Depression scale (HAMD-17), which is a questionnaire used by clinicians to evaluate depression symptoms in patients

Key exclusion criteria for both trials are as follows:

- At significant risk of suicide, or had attempted suicide associated with the current episode of PND
- Active psychosis & medical history of bipolar disorder, schizophrenia and/or schizoaffective disorder
- Using psychotropic medications to treat depressive symptoms (e.g., antidepressants) which have not been taken at the same dose for at least 30 days prior to Day 1 of treatment with zuranolone

There are no ongoing clinical trials of zuranolone in patients with PND.

3e) Efficacy

Efficacy is the measure of how well a treatment works in treating a specific condition.

In this section, please summarise all data that demonstrate how effective the treatment is compared with current treatments at treating the condition outlined in section 2a. Are any of the outcomes more important to patients than others and why? Are there any limitations to the data which may affect how to interpret the results? Please do not include academic or commercial in confidence information but where necessary reference the section of the company submission where this can be found.

The clinical trial programme for zuranolone examined effectiveness of zuranolone in reducing depression symptoms, its speed of action, the duration of effect, side effects and tolerability.

The SKYLARK and ROBIN studies showed that zuranolone met the primary outcome, i.e., compared with placebo, zuranolone provided significant improvement in depression symptoms by Day 15, as assessed by HAMD-17 score. Importantly, in both studies, significant improvement in HAMD-17 score was observed as early as Day 3 from zuranolone treatment initiation, and patients continued to have a significant benefit throughout the study duration – until Day 45 – despite zuranolone being only a 14-day treatment (see more information in Section 2.6 of the company submission) [35, 36].

Additionally, SKYLARK and ROBIN demonstrated that zuranolone was more effective than placebo across all secondary outcomes, which assessed changes in depression and anxiety symptoms at

different time points using different questionnaires and rating scales including Montgomery-Åsberg Depression Rating Scale (MADRS), Clinical Global Impression scale (CGI-S), Edinburgh Postnatal Depression Scale (EPDS), and Hamilton Rating Scale for Anxiety (HAM-A) (see more information in Section 2.6 of the company submission) [35, 36].

The tools used in SKYLARK and ROBIN may be considered of increased relevance to patients as they assessed benefits of zuranolone that may be especially meaningful for patients, including:

- Resolution of symptoms of depression (as demonstrated by HAMD-17, MADRS, CGI-S – clinician-rated tools, as well as EPDS – patient's self-reported assessment)
- Resolution of symptoms of anxiety (as demonstrated by EPDS, HAM-A)
- Impact on components of social functioning (as demonstrated by HAMD-17 and CGI-S)
- Impact on performance of everyday activities (as demonstrated by MADRS)
- Improvements in sleep quality (as demonstrated by MADRS – Reduced Sleep, HAMD-17 – Insomnia)

Zuranolone was also well tolerated in both studies (see more information in Section 3g of this document and Section 2.11 of the company submission) [35, 36].

The limitations of the SKYLARK and ROBIN studies, which are common to antidepressant trials, include a 45-day follow-up (antidepressant trials rarely last above 8 weeks [38]).

3f) Quality of life impact of the medicine and patient preference information

What is the clinical evidence for a potential impact of this medicine on the quality of life of patients and their families/caregivers? What quality of life instrument was used? If the EuroQol-5D (EQ-5D) was used, does it sufficiently capture quality of life for this condition? Is there other disease specific quality of life measures that should also be considered as supplementary information?

Please outline in plain language any quality-of-life related data such as **patient reported outcomes (PROs).**

Please include any **patient preference information (PPI)** relating to the drug profile, for instance research to understand willingness to accept the risk of side effects given the added benefit of treatment. Please include all references as required.

Quality of life data was collected in the ROBIN study via the SF-36v2 Health Survey (SF-36). This tool is a patient-reported instrument for measuring health-related quality of life, functional health and well-being across 8 components (physical functioning, role physical, bodily pain, general health, vitality, social functioning, role emotional, and mental health) [39]. These 8 components provide scores which are totalled into a physical component summary and a mental component summary, with higher SF-36 scores indicating a better state of health and quality of life [40].

By Day 45 (end of follow-up in the ROBIN trial), statistically significant differences favouring zuranolone (versus placebo with established clinical management) were observed for 5 of the 8 dimensions (3 reflecting physical health status–physical functioning, role physical, and bodily pain–and 2 reflecting mental health status–social functioning and mental health) as well as the mental component summary [40].

It is important to consider that there is a relationship between depression severity and quality of life. Severity of depression and anxiety symptoms are key factors for poor quality of life and as such, when depression severity decreases, quality of life often improves [41].

3g) Safety of the medicine and side effects

When NICE appraises a treatment, it will pay close attention to the balance of the benefits of the treatment in relation to its potential risks and any side effects. Therefore, please outline the main side effects (as opposed to a complete list) of this treatment and include details of a benefit/risk assessment where possible. This will support patient reviewers to consider the potential overall benefits and side effects that the medicine can offer.

Based on available data, please outline the most common side effects, how frequently they happen compared with standard treatment, how they could potentially be managed and how many people had treatment adjustments or stopped treatment. Where it will add value or context for patient readers, please include references to the Summary of Product Characteristics from regulatory agencies etc.

Overall, zuranolone was well-tolerated, with side effects that are usually mild or moderate and can be managed [35, 36, 40]. There were no significant changes in vital signs, heart tests, or laboratory results. There were also no signs of increased thoughts of suicide or withdrawal symptoms after stopping the medicine.

Like all medicines, this medicine can cause side effects, although not everybody gets them. The following side effects may happen with this medicine.

Very common (may affect more than 1 in 10 people):

- Drowsiness or sleepiness
- Dizziness

Common (may affect up to 1 in 10 people):

- Loose stools (diarrhoea)
- Lack of energy
- Trouble remembering information
- Trembling or shaking
- Feeling confused

In the zuranolone clinical trial program, most side effects were mild or moderate. In the SKYLARK study (50 mg dose), about one-third of patients had mild side effects, and slightly less than one-third had moderate side effects [36]. In the ROBIN study (30 mg dose), about four in ten patients had mild side effects, and less than two in ten had moderate side effects [35, 40].

Serious side effects were rare. In the SKYLARK study, two serious events happened in the zuranolone group, but doctors thought these were not related to the medicine. In the ROBIN study, one patient became confused for a short time, but this cleared up within 7 hours.

If side effects are bothersome, doctors might lower the dose. In the SKYLARK study, about one in six patients had their dose reduced due to side effects, but most of these patients finished the study (see more information in Section 2.11 of the company submission) [36].

When compared to placebo, zuranolone had slightly more side effects, but there were no large differences. For patients already taking antidepressants, taking zuranolone did not cause more side effects than in those not taking antidepressants.

Zuranolone can reduce awareness and alertness. It is important that patients discuss these possible effects with their healthcare provider before taking zuranolone. Patients are advised not to drive for at least 12 hours after taking each zuranolone dose. Patients may not be able to tell on their own how much zuranolone is affecting them and whether it is safe to drive.

3h) Summary of key benefits of treatment for patients

Issues to consider in your response:

- Please outline what you feel are the key benefits of the treatment for patients, caregivers and their communities when compared with current treatments.
- Please include benefits related to the mode of action, effectiveness, safety and mode of administration

Zuranolone as a treatment for PND could provide the following benefits to patients, caregivers, and communities:

- Be the first PND-specific treatment option, designed to target the brain changes seen in PND
- Rapid onset of action, with many patients experience improvement early after initiating the 14-day treatment with significant improvement in depression symptoms compared to placebo as early as Day 3 [35, 36]. This fast relief is crucial as it allows the patients to regain a good quality of life, improve their functioning, including taking care of themselves and their child, potentially supporting bonding with the child, which has positive long-term impact on development of the child.
- Good treatment adherence due to its short duration and favourable safety profile. Zuranolone was generally well-tolerated in both ROBIN and SKYLARK, with 91% of patients completing the full course of zuranolone in the clinical trial programme [35, 36]
- Potential to reduce treatment barriers due to its short-treatment course and PND specificity, which may encourage more patients to seek treatment

3i) Summary of key disadvantages of treatment for patients

Issues to consider in your response:

- Please outline what you feel are the key disadvantages of the treatment for patients, caregivers and their communities when compared with current treatments. Which disadvantages are most important to patients and carers?
- Please include disadvantages related to the mode of action, effectiveness, side effects and mode of administration
- What is the impact of any disadvantages highlighted compared with current treatments

Zuranolone passes into breast milk. Patients should discuss the risks and benefits of zuranolone while breast-feeding.

No disadvantages are deemed to be associated with the mode of action, effectiveness or the mode of administration of zuranolone compared to current treatments. This is also the case for disadvantages linked with communities and caregivers.

Concerning side effects, zuranolone was well tolerated as demonstrated by low discontinuation rates in both SKYLARK and ROBIN. The most common side effects which may affect more than 1 in 10 people included feeling drowsy, sleepy, dizzy. Other common side effects which may affect up to 1 in 10 people included loose stools (diarrhoea), lack of energy, trouble remembering information, trembling or shaking, and feeling confused (see Section 3g).

3i) Value and economic considerations

Introduction for patients:

Health services want to get the most value from their budget and therefore need to decide whether a new treatment provides good value compared with other treatments. To do this they consider the costs of treating patients and how patients' health will improve, from feeling better and/or living longer, compared with the treatments already in use. The drug manufacturer provides this information, often presented using a health economic model.

In completing your input to the NICE appraisal process for the medicine, you may wish to reflect on:

- The extent to which you agree/disagree with the value arguments presented below (e.g., whether you feel these are the relevant health outcomes, addressing the unmet needs and issues faced by patients; were any improvements that would be important to you missed out, not tested or not proven?)

- If you feel the benefits or side effects of the medicine, including how and when it is given or taken, would have positive or negative financial implications for patients or their families (e.g., travel costs, time-off work)?
- How the condition, taking the new treatment compared with current treatments affects your quality of life.

As part of the submission to NICE, the manufacturer built a cost-effectiveness model to assess whether the benefits of treating PND patients with zuranolone outweighed the associated costs to the NHS in comparison to established clinical management.

The cost-effectiveness model developed for this submission was a patient-level simulation model, i.e., a type of model that allows to capture diversity of PND by tracking disease course of individual patients (based on EPDS scores), health outcomes, and costs. EPDS was selected as the measure of PND symptoms as it is patient-relevant (self-reported scale), it was collected in the SKYLARK study, and it used in clinical practice and therefore collected in real-world data.

The cost-effectiveness analysis was designed to capture potential long-term effects of PND on patients by modelling the disease course over extended period of time. This was achieved by using and linking short-term and long-term EPDS scores from two different sources: clinical data from the SKYLARK study to provide information on the short-term outcomes, and real-world, natural history data from the Avon longitudinal study of parents and children (ALSPAC) study to support modelling of the long-term outcomes and course of PND in patients over a time horizon of 29 years (the length of the available follow-up in the ALSPAC study).

The analysis demonstrated that compared with established clinical management, treatment with zuranolone resulted in improved health outcomes in form of faster resolution of PND symptoms among a larger proportion of patients. This resulted in a larger health-related quality of life benefit, on average, when compared with established clinical management. Compared to established clinical management, treatment with zuranolone was associated with higher per-patient costs for the health service (with the difference largely driven by treatment acquisition [drug] costs).

It is important to note that there are many benefits of zuranolone treatment, which have not been captured in the cost-effectiveness analysis:

- Faster resolution of symptoms among PND patients, who were treated with zuranolone may have an indirect benefit in a form of avoided negative consequences of PND on the children
    - PND negatively affects infants' health-related quality of life from as early as 1-month of age [42]. Additionally, PND can significantly impair mother-infant bonding and can lead to negative short- and long-term consequences for a child's development including behavioural problems, delayed cognitive development, significantly impaired language development, a higher degree of emotional disorders, and lower social engagement [5, 14, 43].
    - PND can significantly impair maternal function, including caring for the child, and disruptions arising due to maternal depression can also translate into unfavourable outcomes for siblings of the newborn [14, 44].
- The cost-effectiveness analysis also does not capture any partner outcomes. PND can negatively affect the partner's own mental health and quality of life, and the families of patients experiencing PND are subject to a higher level of disruption due to parenting behaviours and relationship separation [3]. Partners of patients with PND are at an increased risk of depression, and any associated health-related quality of life loss [45]. These negative outcomes of partners of PND patients may potentially be avoided by introducing zuranolone as a treatment option for PND

- Although the societal impact cannot be fully captured within the cost-effectiveness analysis, the average cost to society of one case of PND in the UK is approximately £75,000 (2012/13 prices). The cost impact of PND on women is estimated to be £23,151 with productivity losses accounting for approximately 13% of these costs (£3,028 per case) [46].

3j) Innovation

NICE considers how innovative a new treatment is when making its recommendations.
If the company considers the new treatment to be innovative, please explain how it represents a 'step change' in treatment and/ or effectiveness compared with current treatments. Are there any QALY benefits that have not been captured in the economic model that also need to be considered (see section 3f)

Zuranolone represents a significant innovation in the treatment of PND, as recognized by both international (FDA-approved with priority review and fast-track designation) [47] and national (granted an Innovative Licensing and Access Pathway designation by the MHRA) regulators. This is a testament to how zuranolone addresses a long-standing unmet need in PND treatment, where no specific pharmacological treatment options were previously available in the UK, and to its potential benefit to patients.

The key benefits of zuranolone are as follows:

- New mechanism of action: It is the first treatment in the UK that works by targeting the brain's GABA system, which helps control mood and anxiety in patients with PND [36]
- Rapid action: Many patients start feeling better after just 3 days of taking zuranolone [48]
- Short treatment time: Zuranolone has a short treatment course of 14 days, unlike other antidepressants that can be taken for months or years. This could mean fewer side effects and make it easier for patients to complete the treatment (91% of patients persisted with zuranolone treatment in the clinical program)[35, 36]

3k) Equalities

Are there any potential equality issues that should be taken into account when considering this condition and this treatment? Please explain if you think any groups of people with this condition are particularly disadvantaged.
Equality legislation includes people of a particular age, disability, gender reassignment, marriage and civil partnership, pregnancy and maternity, race, religion or belief, sex, and sexual orientation or people with any other shared characteristics

More information on how NICE deals with equalities issues can be found in the NICE equality scheme
Find more general information about the Equality Act and equalities issues here

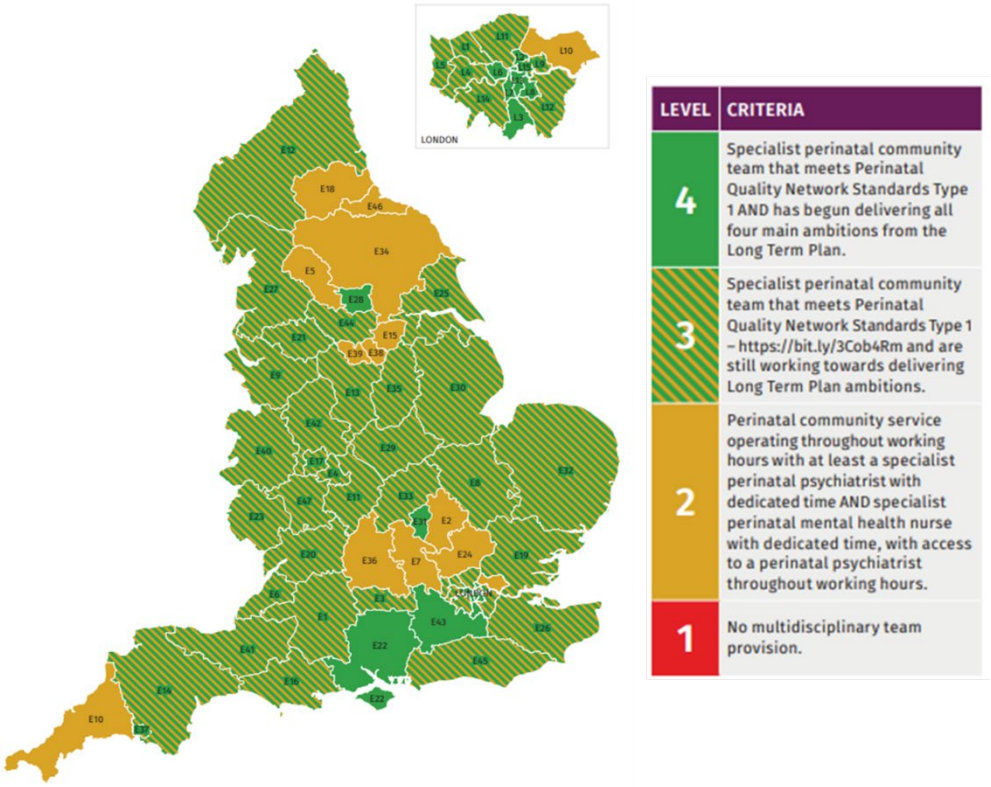
Despite recent progress in expanding specialist perinatal mental health services in the UK, significant inequalities in access to high-quality care remain [49]. These disparities particularly affect ethnic minority groups, younger women, and those living in rural areas [50-52].

Ethnic minorities face a higher burden of mental health disorders. Black women are 13% more likely to develop PND, experience lower access to community mental health services and have higher rates of involuntary admissions versus White British patients [52, 53]. Younger women are also particularly vulnerable to PND, often being reluctant to seek help due to stigma and fear of judgment [54].

Audit by Maternal Mental Health Alliance highlights variation in service provision. The extent of these challenges is evident in the fact that only 16% of specialist perinatal mental health services

community teams in England have met full standards and NHS Long Term Plan ambitions (Figure 1) [49].

Figure 1: Specialist Perinatal Mental Health Community Teams in England



SECTION 4: Further information, glossary and references

4a) Further information

Feedback suggests that patients would appreciate links to other information sources and tools that can help them easily locate relevant background information and facilitate their effective contribution to the NICE assessment process. Therefore, please provide links to any relevant online information that would be useful, for example, published clinical trial data, factual web content, educational materials etc.
Where possible, please provide open access materials or provide copies that patients can access.

Further information on NICE and the role of patients**:**

- Public Involvement at NICE Public involvement | NICE and the public | NICE Communities | About | NICE
- NICE's guides and templates for patient involvement in HTAs Guides to developing our guidance | Help us develop guidance | Support for voluntary and community sector (VCS) organisations | Public involvement | NICE and the public | NICE Communities | About | NICE
- EUPATI guidance on patient involvement in NICE: https://www.eupati.eu/guidance-patient-involvement/
- EFPIA – Working together with patient groups: https://www.efpia.eu/media/288492/working-together-with-patient-groups-23102017.pdf
- National Health Council Value Initiative. https://nationalhealthcouncil.org/issue/value/
- INAHTA: http://www.inahta.org/

- European Observatory on Health Systems and Policies. Health technology assessment - an introduction to objectives, role of evidence, and structure in Europe: http://www.inahta.org/wp-content/themes/inahta/img/AboutHTA_Policy_brief_on_HTA_Introduction_to_Objectives_Role_of_Evidence_Structure_in_Europe.pdf

4b) Glossary of terms

- ALSPAC – Avon Longitudinal Study of Parents and Children
- BIMF - Barkin Index of Maternal Functioning
- CBT - Cognitive Behavioural Therapy
- CGI - Clinical Global Impression
- EPDS - Edinburgh Postnatal Depression Scale
- FDA – Food and Drug Administration
- GABA - Gamma-aminobutyric Acid
- Hamilton Rating Scale for Depression-17 – HAMD-17
- HTA – Health Technology Assessment
- ILAP - Innovative Licensing and Access Pathway
- MHRA - Medicines and Healthcare products Regulatory Agency
- MADRS - Montgomery–Åsberg Depression Rating Scale
- NHS – National Health Service
- NICE – National Institute for Health and Care Excellence
- PND – Postnatal Depression
- SNRI - Serotonin and Norepinephrine Reuptake Inhibitors
- SSRI - Selective Serotonin Reuptake Inhibitors
- TCA - Tricyclic Antidepressants
- UK - United Kingdom

4c) References

Please provide a list of all references in the Vancouver style, numbered and ordered strictly in accordance with their numbering in the text:

1. American College of Obstetricians and Gyneoclogists, *Treatment and Management of Mental Health Conditions During Pregnancy and Postpartum*. 2023.
2. National Institute of Health and Care Excellence, *Depression - antenatal and postnatal*. 2023.
3. Moore Simas, T.A., et al., *The humanistic burden of postpartum depression: a systematic literature review.* Curr Med Res Opin, 2019. **35**(3): p. 383-393.
4. Duman, R.S., G. Sanacora, and J.H. Krystal, *Altered Connectivity in Depression: GABA and Glutamate Neurotransmitter Deficits and Reversal by Novel Treatments.* Neuron, 2019. **102**(1): p. 75-90.
5. Slomian, J., et al., *Consequences of maternal postpartum depression: A systematic review of maternal and infant outcomes.* Womens Health (Lond), 2019. **15**: p. 1745506519844044.
6. Harrison, S., et al., *The impact of the Covid-19 pandemic on postnatal depression: analysis of three population-based national maternity surveys in England (2014-2020).* Lancet Reg Health Eur, 2023. **30**: p. 100654.
7. Jones, I. and J. Shakespeare, *Postnatal depression.* Bmj, 2014. **349**: p. g4500.
8. Petersen, I., et al., *Depression, depressive symptoms and treatments in women who have recently given birth: UK cohort study.* BMJ Open, 2018. **8**(10): p. e022152.

9. National Institute of Mental Health, *Perinatal Depression*. 2022.
10. FDA. *Zurzuvae (zuranolone) Prescribing Information*. 2024 [cited 2025 21/01/2025]; Available from: https://www.accessdata.fda.gov/drugsatfda_docs/label/2023/217369Orig2s000Corrected_lbl.pdf.
11. Posmontier, B., *Functional status outcomes in mothers with and without postpartum depression.* J Midwifery Womens Health, 2008. **53**(4): p. 310-8.
12. Gerbasi, M.E., et al., *Achieving clinical response in postpartum depression leads to improvement in health-related quality of life.* Current Medical Research and Opinion, 2021. **37**(7): p. 1221-1231.
13. MBRRACE-UK, *Lessons learned to inform maternity care from the UK and Ireland Confidential Enquiries into Maternal Deaths and Morbidity 2018-20*. 2022.
14. Netsi, E., et al., *Association of Persistent and Severe Postnatal Depression With Child Outcomes.* JAMA Psychiatry, 2018. **75**(3): p. 247-253.
15. Kerstis, B., et al., *The association between perceived relationship discord at childbirth and parental postpartum depressive symptoms: a comparison of mothers and fathers in Sweden.* Upsala Journal of Medical Sciences, 2012. **117**: p. 430 - 438.
16. Letourneau, N.L., et al., *Postpartum depression is a family affair: addressing the impact on mothers, fathers, and children.* Issues Ment Health Nurs, 2012. **33**(7): p. 445-57.
17. Carlson K, M.S., Azhar Y, et al.. , *Postpartum Depression*. 2024: StatPearls [Internet].
18. NHS England, *The Perinatal Mental Health Care Pathways*. 2018.
19. National Institute of Health and Care Excellence, *Antenatal and postnatal mental health: clinical management and service guidance: Clinical guideline [CG192]*. 2020.
20. Trivedi, M.H., et al., *Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice.* Am J Psychiatry, 2006. **163**(1): p. 28-40.
21. Thase, M.E., *A New Option for Depressed Patients Who Do Not Respond to Antidepressant Medications.* American Journal of Psychiatry, 2023. **180**(3): p. 188-189.
22. Read, J., et al., *Staying on, and coming off, antidepressants: The experiences of 752 UK adults.* Addict Behav, 2019. **88**: p. 82-85.
23. Richards, J.B., et al., *Effect of selective serotonin reuptake inhibitors on the risk of fracture.* Arch Intern Med, 2007. **167**(2): p. 188-94.
24. Cosci, F. and G. Chouinard, *The Monoamine Hypothesis of Depression Revisited: Could It Mechanistically Novel Antidepressant Strategies?* Neurobiology of Depression, 2019.
25. Niarchou, E., L.H. Roberts, and B.D. Naughton, *What is the impact of antidepressant side effects on medication adherence among adult patients diagnosed with depressive disorder: A systematic review.* J Psychopharmacol, 2024. **38**(2): p. 127-136.
26. Marsden, J., et al., *Medicines associated with dependence or withdrawal: a mixed-methods public health review and national database study in England.* The Lancet Psychiatry, 2019. **6**(11): p. 935-950.
27. Elizabeth Stafford, Z.C., Jessica Walthall, Theresa Vera, Mamta Parikh, Catherine Mak, Ellison Suthoff, Robin Tan,. *The Burden of Postpartum Depression, Barriers to Treatment, and Disparities by Race Among Individuals in the United States: A Cross-Sectional Study*. in *Psych Congress*. 2024. Boston, United States.
28. Holopainen, A. and T. Hakulinen, *New parents' experiences of postpartum depression: a systematic review of qualitative evidence.* JBI Database System Rev Implement Rep, 2019. **17**(9): p. 1731-1769.
29. Gardner, P.L., et al., *The experience of postnatal depression in West African mothers living in the United Kingdom: a qualitative study.* Midwifery, 2014. **30**(6): p. 756-63.

30. Ling, L., Y. Eraso, and V.D. Mascio, *First-generation Nigerian mothers living in the UK and their experience of postnatal depression: an interpretative phenomenological analysis.* Ethn Health, 2023. **28**(5): p. 738-756.
31. Gunduz-Bruce, H., K. Takahashi, and M.-Y. Huang, *Development of neuroactive steroids for the treatment of postpartum depression.* Journal of Neuroendocrinology, 2022. **34**(2): p. e13019.
32. Althaus, A.L., et al., *Preclinical characterization of zuranolone (SAGE-217), a selective neuroactive steroid GABA(A) receptor positive allosteric modulator.* Neuropharmacology, 2020. **181**: p. 108333.
33. Hoffmann, E., et al., *SAGE-217, A Novel GABA(A) Receptor Positive Allosteric Modulator: Clinical Pharmacology and Tolerability in Randomized Phase I Dose-Finding Studies.* Clin Pharmacokinet, 2020. **59**(1): p. 111-120.
34. Gunduz-Bruce, H., K. Takahashi, and M.Y. Huang, *Development of neuroactive steroids for the treatment of postpartum depression.* J Neuroendocrinol, 2022. **34**(2): p. e13019.
35. Deligiannidis, K.M., et al., *Effect of Zuranolone vs Placebo in Postpartum Depression: A Randomized Clinical Trial.* JAMA Psychiatry, 2021. **78**(9): p. 951-959.
36. Deligiannidis, K.M., et al., *Zuranolone for the Treatment of Postpartum Depression.* Am J Psychiatry, 2023. **180**(9): p. 668-675.
37. Therapeutics, S., *SKYLARK Clinical Study Report [Data on file]*. 2022.
38. Ioannidis, J.P., *Effectiveness of antidepressants: an evidence myth constructed from a thousand randomized trials?* Philos Ethics Humanit Med, 2008. **3**: p. 14.
39. Ware, J.E. and B. Gandek, *Overview of the SF-36 Health Survey and the International Quality of Life Assessment (IQOLA) Project.* Journal of Clinical Epidemiology, 1998. **51**(11): p. 903-912.
40. Therapeutics, S., *ROBIN Clinical Study Report [Data on file]*. 2019.
41. Gao, K. and J. Zhang, *Chapter 39 - Depression, anxiety, and quality of life*, in *The Neuroscience of Depression*, C.R. Martin, et al., Editors. 2021, Academic Press. p. 415-424.
42. Subbiah, G.K., et al., *Maternal postpartum mental health negatively affects infants' health related quality of life.* Journal of Affective Disorders, 2025. **374**: p. 381-389.
43. Noorlander, Y., V. Bergink, and M.P. van den Berg, *Perceived and observed mother-child interaction at time of hospitalization and release in postpartum depression and psychosis.* Arch Womens Ment Health, 2008. **11**(1): p. 49-56.
44. Canadian Paediatric Society, *Maternal depression and child development.* Paediatr Child Health, 2004. **9**(8): p. 575-598.
45. Nishimura, A., et al., *Paternal postnatal depression in Japan: an investigation of correlated factors including relationship with a partner.* BMC Pregnancy and Childbirth, 2015. **15**: p. 1-8.
46. Bauer, A., M. Knapp, and M. Parsonage, *Lifetime costs of perinatal anxiety and depression.* J Affect Disord, 2016. **192**: p. 83-90.
47. FDA, *FDA Approves First Oral Treatment for Postpartum Depression*. 2023.
48. Kaufman, Y., S.V. Carlini, and K.M. Deligiannidis, *Advances in pharmacotherapy for postpartum depression: a structured review of standard-of-care antidepressants and novel neuroactive steroid antidepressants.* Ther Adv Psychopharmacol, 2022. **12**: p. 20451253211065859.
49. Maternal Mental Health Alliance, *Specialist perinatal mental health care in the UK 2023*. 2023.
50. Du Preez, A., et al., *Differential effects of ethnic density on the risk of postnatal depression and personality dysfunction.* British Journal of Psychiatry, 2016. **208**(1): p. 49-55.
51. Watson, H., et al., *A systematic review of ethnic minority women's experiences of perinatal mental health conditions and services in Europe.* PLOS ONE, 2019. **14**(1): p. e0210587.

52. Impact on Urban Health. *The Motherhood Group - Children's Mental Health* [cited 2024 31 October]; Available from: https://urbanhealth.org.uk/partnerships/current-partnerships/the-motherhood-group.
53. Jankovic, J., et al., *Differences in access and utilisation of mental health services in the perinatal period for women from ethnic minorities—a population-based study.* BMC Medicine, 2020. **18**(1): p. 245.
54. Schmidt, R.M., et al., *Moderate to severe depressive symptoms among adolescent mothers followed four years postpartum.* J Adolesc Health, 2006. **38**(6): p. 712-8.

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Single Technology Appraisal

Zuranolone for treating severe postnatal depression (ID6431)

Clarification questions

April 2025

File name	Version	Contains confidential information	Date
ID6431 Zuranolone Clarification Questions _Biogen response [Redacted].docx	1	Yes	17/04/25

Section A: Clarification on effectiveness data

Company's clinical trials

A1. Please can the company provide the clinical study reports (including all appendices), the protocols, and the statistical analysis plans for the SKYLARK and ROBIN trials?

Following the External Assessment Group (EAG) request made on April 03, 2025, and in response to this question, Biogen confirms that requested materials supporting SKYLARK and ROBIN studies will be shared by Biogen alongside the responses to the clarification questions. The list of uploaded materials and the file names is provided below.

SKYLARK:

- Clinical study report with appendices (file name: *SKYLARK CSR*)
- Study protocol (file name: *SKYLARK protocol*)
- Statistical analysis plan (file name: *SKYLARK SAP*)

ROBIN materials:

- Clinical study report with appendices (file name: *ROBIN CSR*)
- Study protocol (file name: *ROBIN protocol*)
- Statistical analysis plan (file name: *ROBIN SAP*)

If any additional information is needed, Biogen are happy to share additional documents at EAG request.

A2. Please can the company provide details of which antidepressants and psychological therapies patients in the SKYLARK and ROBIN trials received and how many patients received them by treatment arm?

Details on which antidepressant treatments (ADTs) were used by patients at baseline in both ROBIN and SKYLARK can be found in Table 1 and Table 2, respectively.

Table 1: ROBIN: Summary of ADT Use at Treatment Start – Efficacy Set (Part B)

Preferred term	Placebo N=74 (%)	Zuranolone N=76 (%)	Overall N=150 (%)
Subjects with ≥1 ADT at treatment start	[redacted]	[redacted]	[redacted]
SERTRALINE/SERTRALINE HYDROCHLORIDE	[redacted]	[redacted]	[redacted]
BUPROPION/BUPROPION HYDROCHLORIDE	[redacted]	[redacted]	[redacted]
CITALOPRAM HYDROBROMIDE	[redacted]	[redacted]	[redacted]
FLUOXETINE HYDROCHLORIDE	[redacted]	[redacted]	[redacted]
PAROXETINE/PAROXETINE HYDROCHLORIDE	[redacted]	[redacted]	[redacted]
DULOXETINE HYDROCHLORIDE	[redacted]	[redacted]	[redacted]
ESCITALOPRAM OXALATE	[redacted]	[redacted]	[redacted]

SERTRALINE/SERTRALINE HYDROCHLORIDE, BUPROPION/BUPROPION HYDROCHLORIDE and PAROXETINE/PAROXETINE HYDROCHLORIDE have each been combined with their two available formulations.
Source: (Sage Therapeutics 2019)
Abbreviations: ADT: Antidepressant treatment.

Table 2: SKYLARK: Summary of ADT Use at Treatment Start – Full Analysis Set

Preferred term	Placebo N=97 (%)	Zuranolone N=98 (%)	Overall N=195 (%)
Subjects with ≥1 ADT at treatment start	[redacted]	[redacted]	[redacted]
SERTRALINE/SERTRALINE HYDROCHLORIDE	[redacted]	[redacted]	[redacted]
ESCITALOPRAM/ESCITALOPRAM OXALATE	[redacted]	[redacted]	[redacted]
BUPROPION/BUPROPION HYDROCHLORIDE	[redacted]	[redacted]	[redacted]
FLUOXETINE/FLUOXETINE HYDROCHLORIDE	[redacted]	[redacted]	[redacted]
PAROXETINE	[redacted]	[redacted]	[redacted]
VILAZODONE HYDROCHLORIDE	[redacted]	[redacted]	[redacted]

SERTRALINE/SERTRALINE HYDROCHLORIDE, ESCITALOPRAM/ESCITALOPRAM OXALATE and BUPROPION/BUPROPION HYDROCHLORIDE have each been combined with their two available formulations.
Source: (Biogen 2025b)
Abbreviations: ADT: Antidepressant treatment.

In SKYLARK, the study protocol allowed background use of ADTs and psychological therapies. However, psychological therapies were not captured as standalone variables at baseline, and therefore this information is not available for the SKYLARK and ROBIN trials/studies.

A3. Could the company please provide information on parity, previous history of MDD, and baseline psychological therapies for the participants of SKYLARK and ROBIN trials by treatment arm? Please provide this information in updated versions of Tables 9 and 10.

Revised baseline patient demographic tables with inclusion of prior Major Depressive Disorder (MDD) history by treatment arm are presented for both ROBIN and SKYLARK in Table 3 and Table 4, respectively. Please note, in order to be consistent with the publicly available information (i.e., the 2023 publication by Deligiannidis et

al.), the patient demographics and baseline characteristics presented in the Company Submission (CS) (Table 9, page 51 of the CS) were from the Safety Set. As the analysis of the MDD history was conducted post hoc, only data from the Full Analysis Set is available. Therefore, in Table 4 below, two analysis sets are included – the necessary footnotes were added for clarity. All necessary changes were introduced in the CS document. For a summary of all changes, please see Table 12 in the Appendix of this response document.

Concerning usage of psychological therapies, as these were not captured as a standalone variable, no information is available to be shared for neither SKYLARK nor ROBIN. Parity (and gravidity) data were not collected in ROBIN and SKYLARK.

Table 3: ROBIN: Patient Demographics and Baseline Characteristics (Efficacy Set)

Characteristics	Zuranolone 30 mg (N=76)	Placebo (N=74)
Age, mean (SD), years	29.3 (5.4)	27.4 (5.3)
Race, n (%)		
White	44 (58)	40 (54)
African American	31 (41)	31 (42)
Other[a]	1 (1)	3 (4)
Hispanic/Latino ethnicity, n (%)	16 (21)	18 (24)
Weight, mean (SD), kg	85.1 (19)	80.2 (24)
HAMD-17 total score, mean (SD)	28.4 (2)	28.8 (2)
Baseline ADT use, n (%)	16 (21)	13 (18)
Family history of PND, n (%)	10 (13)	10 (14)
Onset of PND, n (%)		
Third trimester	32 (42)	31 (42)
≤4 weeks after delivery	44 (58)	43 (58)
History of MDD, n (%)	[redacted]	[redacted]

[a]Other included Asian, Native Hawaiian/Pacific Islander, and more than 1 race.
Source: (Deligiannidis 2021)(Biogen 2025a)
Abbreviations: ADT: Antidepressant treatment; HAMD-17: 17-item Hamilton Rating Scale for Depression; PND: Postnatal Depression; SD: Standard deviation.

Table 4: SKYLARK: Patient Demographics and Baseline Characteristics (Safety Set / Full Analysis Set)

Characteristics	Safety Set	
	Zuranolone 50 mg (N=98)	Placebo (N=98)
Age, mean (SD), years	30.0 (5.9)	31.0 (6.0)
Age group, years, n (%)		
18-24	[redacted]	[redacted]
25-45	[redacted]	[redacted]
Race, n (%)		
White	68 (69.4)	69 (70.4)
Black/African American	25 (25.5)	18 (18.4)
Other[a]	5 (5.1)	11 (11.2)
Hispanic/Latina ethnicity, n (%)	33 (33.7)	42 (42.9)
BMI, mean (SD), kg/m2	30.9 (6.3)	29.6 (6.3)
HAMD-17 total score, mean (SD)	28.6 (2.5)	28.8 (2.3)
CGI-S score, mean (SD)	5.0 (0.7)	4.9 (0.6)
Country, n (%)		
USA	95 (96.9)	96 (98.0)
Rest of the world	3 (3.1)	2 (2.0)
Baseline ADT use, n (%)	15 (15.3)	15 (15.3)
History of PND, n (%)		
First episode	81 (82.7)	87 (88.8)
Recurrent episode	17 (17.3)	11 (11.2)
Onset of PND, n (%)		
Third trimester	34 (34.7)	31 (31.6)
≤4 weeks postnatal	64 (65.3)	67 (68.4)
Characteristics	**Full Analysis Set**	
	Zuranolone 50 mg (N=98)	**Placebo (N=98)**
History of MDD[b], n (%)	[redacted]	[redacted]

[a]Other included Asian, American Indian, Alaska Native, Native Hawaiian/Pacific Islander, multiple, other race, and/or not reported.

[b][redacted]

Source: (Sage Therapeutics 2022, Deligiannidis 2023)(Biogen 2025a)
Abbreviations: ADT: Antidepressant treatment; BMI: Body mass index; CGI-S: Clinical Global Impressions – Severity; HAMD-17: 17-item Hamilton Rating Scale for Depression; MDD: Major Depressive Disorder; PND: Postnatal Depression; SD: Standard deviation; USA: United States of America.

A4. In the SKYLARK and ROBIN trials, how many patients who were on baseline ADT would be considered non-responders to ADT, by treatment arm? Was non-response to ADT an eligibility criterion for inclusion in the trials?

Non-response was not a criterion for inclusion in the trial. Patients were allowed concomitant ADT use as long as they were on a stable dose for at least 30 days prior to Day 1 and agreed to continue on a stable dose through completion of the Day 45 assessments.

Biogen would like to reiterate that zuranolone is expected to fit into the existing clinical pathway as a treatment option to be used alone or as an adjunct to therapeutic alternatives used as part of the established clinical management (ECM), regardless of the history of prior ADT use. Unlike slow-acting ADTs (Posmontier 2008, Moore Simas 2019, Slomian 2019), treatment with zuranolone can offer statistically significant improvements in depressive symptoms (HAMD-17) as early as Day 3; therefore, it is important that patients in need can access zuranolone treatment in a timely manner and benefit from a rapid resolution of symptoms.

A5. Could you provide a subgroup analysis for history of major depressive disorder or not, as per the scope? We acknowledge that the sample size is small (as noted in the CS, Doc B) and results will be uncertain, but this subgroup was listed in the NICE scope and the company do report analyses for other subgroups that have small sample sizes.

Biogen would like to emphasize that neither ROBIN nor SKYLARK studies were powered to demonstrate differences in efficacy of zuranolone in subgroups of patients based on their history of mental health problems or concomitant use of ADTs.

Additionally, Biogen would like to reiterate that Postnatal Depression (PND) and MDD are distinct disorders. As summarised in the CS, PND is defined as a depressive episode characterised by onset in the peripartum period and is driven by pathophysiological mechanisms that differentiate it from MDD (Slomian 2019). While the underlying pathophysiology of PND is not completely understood, it is thought that genetic, metabolic, endocrine, environmental, and neurobiological factors are associated with the manifestation of depressive symptoms in PND (Duman 2019, Batt 2020). Specifically, the perinatal period is characterised by rapid physiological changes in levels of endocrine hormones, peptides and neuroactive steroids, which may contribute to the vulnerability to depressive symptoms in some patients (Maguire 2009, Zorumski 2019). It is also important to stress that any woman, irrespective of the history of mental health problems (and/or situational factors) can experience symptoms of PND (American Psychological Association 2022) and for many patients, the onset of PND can be the first time they experience any mental health problems.

Nevertheless, upon request from EAG, Biogen conducted additional post hoc analyses for subgroup of patients with history of MDD for ROBIN and SKYLARK (Figure 1 and

Figure 2, respectively). In ROBIN, [redacted] had history of MDD, which is aligned with the literature which suggests 25-40% of PND patients have prior MDD history (Kroh 2021). In ROBIN, [redacted] least-square means (LSM) change from baseline of 17-item Hamilton Rating Scale for Depression (HAMD-17) total score [redacted].

Figure 1: ROBIN: Forest Plot of LSM (95% CI) Treatment Difference in Change from Baseline in HAMD-17 Total Score by History of MDD Subgroups at Day 15 and Day 45 - Efficacy Set (Part B)

For subgroup analyses, results are from a mixed model for repeated measures (MMRM) with treatment, baseline result, assessment time point, and time point-by-treatment interaction as fixed effects. For Overall analysis only, an additional fixed effect term for antidepressant use at baseline was included.
[1] Post hoc subgroup analysis.
Source: (Biogen 2025a)
Abbreviations: CI: Confidence interval; HAMD-17: 17-item Hamilton Rating Scale for Depression; LSM: Least-squares mean; MDD: Major Depressive Disorder.

Biogen would like to emphasize that SKYLARK was not designed or powered to demonstrate a treatment difference in these subgroups, and the number of enrolled patients with MDD history was very small [redacted], and MDD was not a standalone variable for collection in the study. As a result, the analysis of history of MDD was conducted as part of the post hoc analyses and was extracted from across 'Other' fields in data collection sheets ([redacted]

[redacted]). In SKYLARK, LSM change from baseline of HAMD-17 total score [redacted]. The heterogeneity between subgroup results observed within SKYLARK and compared to ROBIN, is likely due to the limitations of the sample size and is likely further impacted by the way the data were collected in SKYLARK (i.e., MDD was not a standalone variable for collection).

Figure 2: SKYLARK: Forest Plot of LSM (95% CI) Treatment Difference in Change from Baseline in HAMD-17 Total Score by History of MDD Subgroups at Day 15 and Day 45 – Full Analysis Set

Results are from a mixed model for repeated measures (MMRM) with treatment, baseline HAMD-17 total score, antidepressant use at baseline, assessment time point, and time point-by-treatment interaction as fixed effects.
[1] Post hoc subgroup analysis.
Source: (Biogen 2025a)
Abbreviations: CI: Confidence interval; HAMD-17: 17-item Hamilton Rating Scale for Depression; LSM: Least-squares mean; MDD: Major Depressive Disorder.

A6. Priority question: The proportion of patients on baseline ADT in SKYLARK was 15%. However, our clinical experts believe that zuranolone would likely be offered to those referred to a perinatal psychiatrist (secondary care), most of whom would already be taking an antidepressant (and not responding to ADT). Could you provide justification why the proportion of patients on ADT in SKYLARK is representative of the likely use in the NHS?

Biogen would like to emphasise that PND is a complex, multifactorial disease and that being referred to secondary care may not only be related to response or lack of

response to ADTs. Patients who are likely to be referred to secondary care would include patients with moderate-severe disease, complex PND patients with other psychiatric comorbidities, or those who may not be responding to or with in adequate response to ADTs or other treatments. As outlined in the CS, the United Kingdom (UK) -specific information on the demographics and other baseline characteristics of PND patients is limited, but based on the limited information available, Biogen believes that the SKYLARK population is generally expected to be similar to the population of PND patients in the UK, with majority of patients to be aged 25-45 years (Harrison 2023, Biogen 2024), majority of patients (>80%) experiencing their first PND episode. Finally, it is not anticipated that the pathophysiology of PND would differ by region (Stewart 2016).

Given the rapid onset of action of zuranolone – resolution of symptoms as early as Day 3 (as demonstrated in the SKYLARK and ROBIN studies) – zuranolone is uniquely positioned to offer rapid improvements across the symptoms of PND. This is particularly important as zuranolone treatment has a potential to contribute to improving patients' functioning, quality of life (QoL), and allow for development of the bond with the child, more rapidly than existing ADTs (Posmontier 2008, Moore Simas 2019, Slomian 2019). Unresolved symptoms can have long-lasting effects on the patient and child (Netsi 2018). Early intervention and timely symptom resolution may also help to prevent deterioration of the patient's state and admission to mother and baby units, which are associated with substantial economic burden.

As outlined in the CS and in the response to question A4, Biogen would like to reiterate that zuranolone is expected to fit into the existing clinical pathway as a treatment option to be used alone or as an adjunct to therapeutic alternatives used as part of the ECM, regardless of the history of prior ADT use. Additionally, while Biogen agrees that zuranolone is most likely to be prescribed in a secondary care setting, there is an expectation that introduction of zuranolone will initiate a paradigm shift towards a faster referral to specialists. Biogen is committed to working with the relevant stakeholders on improving timely access to specialist care for patients across England and Wales.

A7. Please can the company provide the number of patients who had data available for EPDS at each timepoint

The number of patients with available data for Edinburgh Postnatal Depression Scale (EPDS) at each timepoint are presented in Table 5 and Table 6 for ROBIN and SKYLARK, respectively.

Table 5: ROBIN: Number of patients by study visit and treatment group who had data for EPDS – Efficacy set (Part B)

Study Visit	Placebo, n (%)	Zuranolone, n (%)
Baseline	[redacted]	[redacted]
Day 3	[redacted]	[redacted]
Day 8	[redacted]	[redacted]
Day 15	[redacted]	[redacted]
Day 21	[redacted]	[redacted]
Day 45	[redacted]	[redacted]

Source: (Sage Therapeutics 2019)
Abbreviations: EPDS: Edinburgh Postnatal Depression Scale; SD: Standard deviation.

Table 6: SKYLARK: Summary and mean change from baseline in EPDS total score by study visit and treatment group

Study Visit	Placebo, n (%)	Zuranolone, n (%)
Baseline	[redacted]	[redacted]
Day 3	[redacted]	[redacted]
Day 8	[redacted]	[redacted]
Day 15	[redacted]	[redacted]
Day 21	[redacted]	[redacted]
Day 28	[redacted]	[redacted]
Day 45	[redacted]	[redacted]

Source: (Sage Therapeutics 2022)
Abbreviations: EPDS: Edinburgh Postnatal Depression Scale; SD: Standard deviation.

A9. Please can the company provide mean MADRS, EPDS, and HAMD-17 baseline scores and SDs for the moderate and severe post-hoc subgroups.

Mean Montgomery–Åsberg Depression Rating Scale (MADRS), EPDS and HAMD-17 baseline scores alongside their standard deviations for post hoc subgroups of patients categorised as moderate PND and severe PND (based on baseline MADRS score) are presented in Table 7 and Table 8 for ROBIN and SKYLARK, respectively. For more details on the post-hoc analysis by PND severity, please refer to Section 2.8.2.2 of the CS.

Table 7: ROBIN: Mean MADRS, EPDS and HAMD-17 score at Baseline by MADRS Severity – Efficacy Set (Part B)

Baseline score (SD)	Zuranolone 30 mg Moderate PND			Zuranolone 30 mg Severe PND		
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
HAMD-17	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
MADRS	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
EPDS	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Source: (Biogen 2025a)
Abbreviations: EPDS: Edinburgh Postnatal Depression Scale; HAMD-17: HAMD-17: 17-item Hamilton Rating Scale for Depression; MADRS: Montgomery–Åsberg Depression Rating Scale; PND: Postnatal Depression; SD: Standard deviation.

Table 8: SKYLARK: Mean MADRS, EPDS and HAMD-17 score at Baseline by MADRS Severity – Full Analysis Set

Baseline score (SD)	Zuranolone 50 mg Moderate PND			Zuranolone 50 mg Severe PND		
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
HAMD17	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
MADRS	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
EPDS	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Source: (Biogen 2025a)
Abbreviations: EPDS: Edinburgh Postnatal Depression Scale; HAMD-17: HAMD-17: 17-item Hamilton Rating Scale for Depression; MADRS: Montgomery–Åsberg Depression Rating Scale; PND: Postnatal Depression; SD: Standard deviation.

A10. How was 'recurrent episode of PND' defined (in Tables 9 and 10)? Does this refer to an episode of PND experienced in a previous pregnancy?

'Recurrent episode of PND' refers to a patient who previously had a PND episode experienced during/after previous pregnancy, and thus this current episode is an recurrent episode of PND. As such, in SKYLARK, 28 patients (14.3%) had a recurrent episode of PND, whereas for 168 patients (85.7%) this was their first experience of PND. In ROBIN, 25 patients (16.7%) had a recurrent episode of PND, whereas 125 patients (83.3%) had their first episode of PND.

Section B: Clarification on cost-effectiveness data

Treatment effects

B1. Priority question: Can you provide scenarios to explore the impact of a waning of treatment effect over time, where EPDS scores converge after different lengths of time, and provide functionality in the model to adjust this? For example, where EPDS curves converge after: 3 months, 6 months, 1 year, 2 years, 5 years, 10 years.

The convergence of EPDS scores for patients who received zuranolone + ECM with those who receive ECM only in the model could be implemented in two ways:

1. The EPDS scores for patients treated with zuranolone + ECM scores worsen with time and eventually converges with the ECM curve

2. The EPDS scores for patients treated with ECM improve with time and eventually converges with the zuranolone + ECM curve

The first approach lacks clinical plausibility as it is unclear why depression in patients treated with zuranolone would worsen, and why any justification made for this change would not also apply to the ECM arm too. For approach 2, the assumption that only patients treated with ECM would improve also lacks clinical plausibility. It is unclear why patients treated with ECM could only improve yet the same assumption could not be made for the zuranolone + ECM patients. The principle of worsening EPDS scores with time which underpins approach 1 lacks face validity, and therefore this has not been explored further. However, we do provide an alternative to approach 2 with refinements to improve clinical plausibility.

In its current form, approach 2 assumes all patients in the ECM arm would have an improvement in EPDS, while no changes are made to patients treated with zuranolone + ECM. An alternative approach would be to focus on a patient, in both treatment arms (zuranolone + ECM and ECM alone), with persistent PND that results in an EPDS ≥13 i.e., patients who have not achieved remission. In this approach, an assumption has been made that the NHS would intervene to ensure that patients are able to achieve remission in the long term (i.e., an EPDS <13). To implement this approach in the

model, scenarios have been included where after a defined time period, say 10 years, patients with an EPDS ≥13, in both treatment arms, would have their EPDS scores gradually decreased to 12.99 (classified as ‘remission’) over a period of 24 months. This results in a convergence of EPDS scores between zuranolone + ECM and ECM for patients who have persistent PND. Biogen provides the results of these scenarios in Table 9 as an alternative to the treatment waning approaches (approach 1 and 2) proposed by the EAG. For more detail on the scenarios conducted in response to the question, please see file named ”ID6431_Zuranolone for the treatment of PND_Biogen_CS_Final_v2.0 [CIC]”, which will be shared by Biogen alongside the responses to the clarification questions.

Finally, Biogen still maintains that the approach in the base case is the most robust approach because the EPDS trajectories are based on long-term natural history data from the Avon Longitudinal Study of Parents and Children (ALSPAC). As the only empirical source for this long-term EPDS prediction, its use in the model in its current form would be our preferred approach.

Table 9: Scenario analysis exploring the impact of reducing a patient's EPDS level if they have not achieved remission after 10, 5, and 2 years.

Treatment	Total costs (£)	Total QALYs	Total LYs	Incr. cost	Incr. QALYs	Incr. LYs	ICER (£/QALY)
Biogen original submitted base case							
ECM	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	12,704.08
Zuranolone	[redacted]	[redacted]	[redacted]				
Scenario 1: Biogen original base case with corrections noted in B17 and B18							
ECM	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	12,704.08
Zuranolone	[redacted]	[redacted]	[redacted]				
Scenario 2: EPDS is decreased to 12.99 over 24 months for patients who have not achieved remission after 10 years – applied to Scenario 1. (Scenario 7 in revised model)							
ECM	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	15,227.44
Zuranolone	[redacted]	[redacted]	[redacted]				
Scenario 3: EPDS is decreased to 12.99 over 24 months for patients who have not achieved remission after 5 years – applied to Scenario 1. (Scenario 8 in revised model)							
ECM	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	16,526.33
Zuranolone	[redacted]	[redacted]	[redacted]				
Scenario 4: EPDS is decreased to 12.99 over 24 months for patients who have not achieved remission after 2 years – applied to Scenario 1. (Scenario 9 in revised model)							
ECM	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	17,567.74
Zuranolone	[redacted]	[redacted]	[redacted]				

Source: Cost-effectiveness model
Abbreviations: ADTs: Antidepressant treatment; ECM: Established clinical management; ICER: Incremental cost-effectiveness ratio; Incr.: Incremental; LY: Life year; QALY: Quality-adjusted life year.

B2. Priority question: In the MMRM analysis for short-term EPDS that informs the economic model, which interaction terms did you consider for inclusion? Did you consider an interaction between Treatment Arm x ADT? If not, could you explore a potential scenario that includes this interaction?

Scenario analysis on the mixed model for repeated measures (MMRM) model to include an interaction between treatment arm and ADT has not been conducted as ADT was not a statistically significant covariate (Table 35 in the CS). As the main variable was not significant, any statistically significant effects of the interaction term could be spurious and are, therefore, unlikely to be meaningful. When conducting the

short-term MMRM analysis, the only interaction term considered was Time x On-treatment Period. In the base case model, the covariate for ADT use was forced to better reflect clinical practice where ADTs are used, however scenario analysis (Scenario 2 in the economic model, file name: "*ID6431_Zuranolone for the treatment of PND_Biogen_CS_Final_v2.0 [CIC]*") was also conducted without the forced inclusion of ADT in the MMRM model, and did not lead to a significant change in the incremental cost-effectiveness ratio (ICER).

B3. Priority question: Could you provide a scenario to explore the cost-effectiveness in the subgroup that are receiving ADT (taking into account question B2 above).

Due to the small number of patients receiving ADTs in the zuranolone studies, no scenario analysis has been conducted based on a MMRM model for the subgroup of patients only receiving ADTs, and therefore this scenario is not possible to conduct. Additionally, as Biogen do not hold the data for the ALSPAC analysis, it is not possible to estimate the long-term EPDS projection for the subgroup of patients receiving ADTs. Alternatively, the percentage of patients on ADTs in the current cost-effectiveness model can be set to 100% to run a scenario. Note that as the base case MMRM model forces the use of ADT as a covariate, the scenario presented in Table 10 results in changes to total and incremental quality-adjusted life years (QALY).

For more detail on the scenarios conducted in response to the question, please see file named "*ID6431_Zuranolone for the treatment of PND_Biogen_CS_Final_v2.0 [CIC]*", which will be shared by Biogen alongside the responses to the clarification questions.

Table 10: Scenario analysis exploring the impact of 100% of patients on ADTs

Treatment	Total costs (£)	Total QALYs	Total LYs	Incr. cost	Incr. QALYs	Incr. LYs	ICER (£/QALY)
Biogen original submitted base case							
ECM	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	12,704.08
Zuranolone	[redacted]	[redacted]	[redacted]				
Scenario 1: Biogen original base case with corrections noted in B17 and B18							
ECM	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	12,704.08
Zuranolone	[redacted]	[redacted]	[redacted]				
Scenario 2: 100% of patients on ADT with ADT implemented as a covariate in the MMRM applied to Scenario 1 (Scenario 10 in revised model)							
ECM	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	13,399.77
Zuranolone	[redacted]	[redacted]	[redacted]				

Source: Cost-effectiveness model
Abbreviations: ADTs: Antidepressant treatment; ECM: Established clinical management; ICER: Incremental cost-effectiveness ratio; Incr.: Incremental; LY: Life year; QALY: Quality-adjusted life year.

B4. Priority question: Could you provide a scenario to explore the cost-effectiveness in the moderate and severe subgroups.

To conduct this scenario, additional analysis on the ALSPAC database would be required to generate long-term EPDS trajectories specifically for the moderate and severe subgroups. As Biogen do not have the relevant data from the ALSPAC dataset, it is not possible to conduct this scenario. As explained in the CS (page 100, Table 29), the long-term EPDS estimations used in the model, were sourced from a de novo analysis of the ALSPAC data commissioned by Biogen and conducted by the University of Bristol. Biogen are unable to obtain additional data within the timeframe of this consultation.

B5. In the MMRM analysis for short-term EPDS that informs the economic model you assume a linear trend with time, but from Fig 14 this relationship looks non-linear, which would impact on the extrapolation of this model up-to 8 months. Could you explore an analysis that accounts for non-linearity?

The model that Biogen has fitted already accounts for non-linearity by including the interaction term between Time and On-treatment period. The On-treatment period

covariate flags whether patients are still receiving treatment (i.e., days 1-14) and, therefore, the interaction of time with this covariate allows for a change in the slope of the predicted EPDS curve after treatment is stopped. The model is, therefore, not a linear model across all time points but a “broken stick” model, incorporating separate linear relationships whilst on and off treatment. This, therefore, captures the steeper gradient in observed EPDS values during the first 14 days whilst on treatment and the shallower gradient in observed EPDS values from day 15, which is then reflected in the extrapolations.

B6. Please provide the Statistical Analysis Plan of the ALSPAC data (it says available on request in Appendix K.3.1)?

Following the EAG request made on April 03, 2025, and in response to this question, Biogen confirms that the requested statistical analysis plan (SAP) of the ALSPAC data (file name: *ALSPAC_SAP_Aim1_v1.0*) will be shared by Biogen alongside the responses to the clarification questions.

Utilities

B7. Priority question: Could you provide the completion rates of the SF36 questionnaire at baseline and at all follow-up points per group for the ROBIN trial?

Table 11 outlines the completion rates of the of the Short-Form Health Survey-36 items (SF-36) questionnaire at baseline and at all follow-up points per group for the ROBIN trial.

Table 11: ROBIN: SF-36 Completion Rates by Study Visit and Treatment Group – Efficacy Set (Part B)

Study Visit	Placebo N=74 (%)	Zuranolone N=76 (%)
Baseline	[redacted]	[redacted]
Day 15	[redacted]	[redacted]
Day 45	[redacted]	[redacted]

Source: (Sage Therapeutics 2019)
Abbreviations: SF-36: Short-Form Healthy Survey – 36 items.

B8. Priority question: Could you please provide the number of responses per group that the utilities estimated per EPDS bracket in the model are based on?

The EPDS to Short-Form-6 Dimensions (SF-6D) utility data was taken directly from the O'Callaghan publication and referenced as such in the economic model. Biogen does not have access to the information requested.

B9. Priority question: Could you please provide adjusted estimates of SF-6D utilities per group using regression methods, adjusting for baseline utility, intervention group, and EPDMS score. We would please ask these also to be provided by subgroup of patients on ADT at baseline and not on ADTs at baseline. Note that if there is imbalance of missing data between groups, could you provide a scenario of these estimates where you have used multiple imputation methods to adjust for missing data.

The EPDS to SF-6D utility data was taken directly from the O'Callaghan publication and referenced as such in the economic model. Biogen does not have access to the information requested.

B10. Would it be possible to provide a range of EQ-5D-3L utility estimates by mapping from collected disease-specific questionnaire responses, such as the PHQ-9 in the ROBIN and SKYLARK trials, using existing mapping models, such as Mukuria et al (2025), Franklin and Hernandez-Alava (2023), and Abdin et al (2021) (REFs). Please provide these by EPDS score and adjusted as per question above.

During the clarification call with EAG (April 11th), Biogen requested additional time to provide response to this question. EAG and NICE granted the extension, and as agreed via email, the response to B10 will be provided by April 28th, 5pm BST.

B11. Our clinical advisors suggest that important adverse events include decreased consciousness and addiction, the latter with longer-term effects. Could you please comment on these and provide a scenario where you include disutilities (and associated costs) in the economic model for these adverse events?

The adverse events included in the economic model have been sourced from the SKYLARK study. As addiction (in the SKYLARK study captured as withdrawal

symptoms assessed by 20-item Physician Withdrawal Checklist) was not observed, there were no incidence rates that could be included in the economic model. Due to a one-off, short-course nature of zuranolone treatment, and the fact that zuranolone is likely to be prescribed in a specialist setting, patients are not expected to have access to extra supply of zuranolone and therefore the long-term risk of addiction is limited. Decreased consciousness was not observed in the SKYLARK study. Instead, somnolence was captured in the study, the incidence rate and disutility/associated cost were included in the model.

B12. Could you provide a rationale for the choice of the SF-36 questionnaire rather than EQ-5D to capture general health-related quality of life in the ROBIN trial?

While both SF-36 and EuroQol-5 Dimensions (EQ-5D) capture aspects of mental health and their impact on the overall quality of life, studies suggest that the SF-36 may be better suited for capturing depressive symptoms and overall mental health, while the EQ-5D focuses more on broader health-related quality of life (LoMartire 2020, Stolz 2023). A recent study assessing the validity and responsiveness of the EQ-5D-3 Levels (EQ-5D-3L) and the SF-6D, both the convergent validity and responsiveness of the SF-6D are better than those of the EQ-5D-3L in patients with depressive symptoms (Stolz 2023). Therefore, SF-6D appears more recommendable for use in studies evaluating new interventions for this population, as was done in the ROBIN study.

Costs

B13. Cost in the 18 months after birth are based on older studies from 2002 and 2011 which may over-estimate the level of services used by today's population in the English NHS. Could you please comment on this and potentially revise or adjust your resource use accordingly?

Biogen accept that the treatment of PND has changed since 2002, and therefore a scenario analysis was conducted in the original submission (Scenario 3) by augmenting the data from Petrou et al. (2002) with feedback from clinicians to reflect current clinical practice, which resulted in a small decrease in the ICER. While some services used in 2002 may not be currently in use, clinicians indicated that for some

services there would be an increase in the resource use. For example, clinicians advised that health visitor contacts and mother and baby unit use has increased significantly since 2002, and any admission to a mother and baby unit is likely to result in considerable National Health Service (NHS) healthcare resource use given that admissions can result in stays of several weeks.

B14. Can you clarify the resource implications to support women who are breast-feeding? For example, supporting expressing, and supporting the decision as to future breast-feeding.

Biogen is still investigating the resource implications associated with support that may be required for women who are breast-feeding, and who would like to stop for the 14-day period while receiving zuranolone treatment. Biogen is committed to engage with relevant stakeholders to fully understand the support needs and to partner with relevant services to deliver effective support for patients without additional costs to the healthcare system.

B15. Can you clarify whether the costs of SSRIs were included in both groups?

The cost of selective serotonin reuptake inhibitors (SSRIs) was included in both the intervention and the comparator arm and was adjusted for the percentage of ADT use at baseline (as per baseline characteristics of the SKYLARK patient population). The cost per treatment course is calculated for zuranolone and ECM in cells O26 and T9 of the 'patient_simulations' sheet cell, respectively. For zuranolone, the value in cell O26 is then used in cell F62 to calculate the total cost per treatment course including dose adjustment.

Model

B16. Please provide your completed TECH-VER checklist. The link provided does not work.

Following the EAG request made on April 03, 2025, Biogen confirms that the TECH-VER (file name: *TechVer_Biogen_Zuranolone 2*) will be shared alongside the responses to the clarification questions.

B17. In the patient_simulations sheet of the Excel model, the formulas in cells AL90 to AL94 include reference to cells AMN90 to AMN94. Is this an error?

Biogen has checked the model and confirm that this was an error. However, this error did not impact the results of the analysis; the cells AMN90 to AMN94 did not contain any data and therefore the formula was subtracting 0 from the QALY value. This has now been corrected, and the updated model has been shared with NICE as part of the clarification response (file name: "*ID6431_Zuranolone for the treatment of PND_Biogen_CEA_Final_V2.0 [CIC]*").

B18. In the patient_simulations sheet of the Excel model the formula in cell T19 is =(O15*(O16*(O18/12)*365.25*O20*O23)+(O17*(O19/12)*365.25*O21*O24)) ,but we think the O15 should be before the first bracket, as O15 is the percentage on ADT so should be applied to both the weighted sertraline and fluoxetine costs. There is a similar issue in O26 (the cost of the cost of zuranolone treatment, as that includes ADT costs for those on baseline ADT). Can you please check?

Biogen has checked the model and can confirm that this was an error. Biogen would like to highlight that due to the low cost of ADTs the impact of this error on the ICER was minimal. This has now been corrected, and the deterministic sensitivity analysis (DSA), probabilistic sensitivity analysis (PSA), and scenarios have been re-run. The updated model has been shared with NICE as part of the clarification response (file name: "*ID6431_Zuranolone for the treatment of PND_Biogen_CEA_Final_V2.0 [CIC]*").

Section C: Textual clarification and additional points

Company systematic reviews

C1. Please can the company provide the protocols for the systematic reviews of clinical effectiveness, HRQoL, and cost-effectiveness. Please also provide details about any post-protocol amendments, if made.

Following the EAG request made on April 03, 2025, and in response to this question, Biogen confirms that all requested protocols for the systematic reviews of clinical effectiveness (file name: *Biogen PPD Clinical SLR Update Protocol_Final*), health-

related quality of life (HRQoL) (file name: *Biogen PPD Economic SLR Update Protocol_Final*), and cost-effectiveness (file name: *Biogen PPD Economic SLR Update Protocol_Final*) will be shared by Biogen alongside the responses to the clarification questions.

No post-protocol amendments were made.

C2. Please can the company clarify if atypical antidepressants (as specified in the NICE scope under "comparators"), other than mirtazapine, were included in the systematic review of clinical effectiveness? If not, please clarify why.

Biogen confirm that atypical antidepressants (tetracyclic antidepressant [mirtazapine] and aminoketone antidepressant [bupropion]) were included in the systematic review of clinical effectiveness. Other treatment classes either alone or in combination included in the eligibility criteria were: benzodiazepines, neurosteroids, *N*-methyl-D-aspartates (NMDA), SSRIs, selective serotonin norepinephrine reuptake inhibitors (SNRIs), tricyclic antidepressants, anxiolytics, mood stabilisers, and antipsychotics.

For more detail, please refer to Table 1 in the clinical systematic literature review (SLR) protocol (page 3, file name: *Biogen PPD Clinical SLR Update Protocol_Final*).

C3. Please can the company clarify if best supportive care (as specified in the NICE scope under "comparators") was included in the systematic review of clinical effectiveness?

Best supportive care (BSC) was not in the scope of the systematic review of clinical effectiveness. Current NICE guidelines (CG192) for the management of antenatal and postnatal mental health do not list BSC as a treatment option and do not provide a definition of what would be considered as BSC (National Institute for Health and Care Excellence 2020). There is no universal definition of BSC in the context of PND. The National Health Service (NHS) lists a number of self-help options that could be considered as BSC (NHS 2022):

- talking to your partner, friends and family – try to help them understand how you're feeling and what they can do to support you

- not trying to be a "supermum" – accept help from others when it's offered and ask your loved ones if they can help look after the baby and do tasks such as housework, cooking and shopping

- making time for yourself – try to do activities that you find relaxing and enjoyable, such as going for a walk, listening to music, reading a book or having a warm bath

- resting when you can – although it can be difficult when you're looking after a baby, try to sleep whenever you get the chance, follow good sleeping habits and ask your partner to help with the night-time work

- exercising regularly to boost your mood

- eating regular, healthy meals and not going for long periods without eating

- not drinking alcohol or taking drugs, as this can make you feel worse

However, guidance varies throughout clinical practice. Therefore, in the absence of a clear definition of BSC, incorporation of BSC in the clinical effectiveness SLR would be challenging.

For more detail, please refer to Table 1 in the clinical SLR protocol (page 3, file name: *Biogen PPD Clinical SLR Update Protocol_Final*).

C4. Please can the company clarify whether PsycINFO was searched in the systematic reviews? It is not listed as a resource search in Appendices, Table 1 (page 12-13), but syntax is provided in Appendices page 23 (NB: it also says 'to be searched').

Biogen would like to clarify that APA PsycINFO was searched in the SLR and is listed in Table 1 of the Appendix B (7th row under databases; page 12-13). PsycINFO was searched via OvidSP interface.

C5. The systematic reviews of clinical effectiveness and HRQoL state that a single reviewer assessed the search results and removed "obviously irrelevant

records such as those about ineligible diseases". Please clarify why this was not checked by a second reviewer, or conducted in duplicate.

The single reviewer process was only used to exclude publications at the initial stage of the review, with the Cochrane randomised controlled trial (RCT) classifier, which is an accepted and validated tool and was used in developing NICE guidelines.

Below the three tasks used a single reviewer approach are described:

1. The records were de-duplicated, both against earlier searches and within the new search yield.

2. One senior reviewer (highly experienced in conducting systematic reviews for the purposes of Health Technology Assessment [HTA]) performed a 'first pass' of the records in an Endnote library with the aim of removing obviously irrelevant records. Examples include: case reports, opinion articles, letters to the editor, studies conducted in children, studies conducted in completely different diseases or topics (such as cancer, diabetes and Coronavirus Disease [COVID]).

3. Following removal of these records, the remaining records were passed through the Cochrane RCT classifier. The Cochrane RCT classifier is an accepted and validated tool (Thomas 2021), with several publications documenting its performance, and is widely considered as appropriate for use in systematic reviews of RCTs. The RCT classifier assessed each record to determine whether or not it was an RCT. The records excluded at this stage were not considered to be RCTs. The Cochrane RCT classifier offers both sensitivity-maximising and precision-maximising versions. The sensitive version that we used prioritises finding all relevant trials (high recall) and is likely to include more irrelevant studies by erring on the side of inclusion. The sensitivity of this classifier was used as published by Thomas et al 2021 (Thomas 2021); it cannot be changed as it is 'baked' into the algorithm, and was used as provided (freely available algorithm).

A two-reviewer process was used for the remaining records to screen abstract and titles, full text and extraction of relevant information according to the predefined protocol.

C6. The systematic review of clinical effectiveness used the Downs and Black checklist to assess the quality of the randomised controlled trials (RCTs). The Risk of Bias 2 (RoB 2) tool is more robust and designed specifically for the assessment of risk of bias in RCTs at the outcome level. Please can the company repeat the risk of bias assessment using RoB 2 for the following outcomes at day 45: EPDS & HAM-D (SKYLARK); HAM-D & SF-36 (ROBIN)

A summary of the Downs and Black quality checklist is provided in the CS (page 55, Table 13; Section 2.5) and the full quality assessment results for SKYLARK and ROBIN can be found in Table 13 and Table 14, respectively, in Appendix B of the CS Appendices (page 38). The quality assessment for the other trials identified in the clinical SLR is also presented in Section B.4.3 (page 38) of the CS Appendix B in the embedded file, which will be provided alongside the responses to clarification questions (file name: *Biogen_zuranolone_clinical SLR_extractions*). Furthermore, the Risk of Bias 2 tool was employed for both ROBIN and SKYLARK and is presented in the Appendix of this document in Table 14.

C7. Please can the company clarify how many reviewers undertook risk of bias assessment in the systematic review of clinical effectiveness, whether each assessment was done independently by more than one reviewer, and how any disagreements were resolved?

Biogen would like to clarify that as described in the protocol of the clinical effectiveness SLR which will be provided alongside the responses to clarification questions (page 8, file name: *Biogen PPD Clinical SLR Update Protocol_Final*), the quality of all included RCTs was assessed using the Cochrane Risk of Bias 2 tool (Higgins JPT 2024). Furthermore, the quality assessment was completed by one individual and verified by a second independent reviewer. The number of disagreements resolved were not captured in the clinical effectiveness SLR report, however, a table outlining the risk of bias assessment for both ROBIN and SKYLARK and is presented in the Appendix of this document in Table 15.

C8. Please can the company explain why a formal tool was not used to critically appraise studies in the HRQoL review? Please also clarify how many independent reviewers undertook quality appraisal.

Biogen would like to clarify that the quality of all included economic evaluations was assessed using the assessment developed by Philips et al. 2004 (Philips 2004). Biogen considers that HTA bodies such as NICE typically do not require validated tools for critical appraisals of cost and resource use studies or utility studies as part of manufacturers' evidence submissions. Additionally, Biogen is not aware of any validated tool for critical appraisals of such studies. Therefore, formal quality assessments were not conducted and instead were completed by one individual and verified by a second independent reviewer.

For more detail on data selection and analysis of economic evaluations, please refer to Section 3.3 in the economic SLR protocol (page 8-10, file name: *Biogen PPD Economic SLR Update Protocol_Final*).

C9. A limit to English language was applied to the searches in 2023 (Appendices page 12) but not subsequent (update) searches. Please can you provide the code from Ovid that you used to limit searches. Please can you also clarify if you undertook any searches to identify studies in languages other than English reported prior to 2023? The concern here is that your post 2023 searches will identify non-English language publications, and your 2023 searches exclude non-English publications. This introduces an inconsistency into study identification and the potential for bias into your reviews.

Biogen would like to clarify that despite not applying an English language limit in the search updates from November 2023 onwards, any studies identified in languages other than English were not assessed for eligibility (as the PICO focussed exclusively on English language-based studies). In the January 2023 search cited 'module 1', the following code was used from Ovid to limit searches: "35: limit 34 to English language (2125)". There were 41 studies excluded by limiting the previous line to English only.

Target population

C10. Please can the company clarify if the target population for zuranolone in this submission is 'adults with postnatal depression' or 'adults with postnatal

depression, following childbirth'? Please can the company confirm if 'adults with postnatal depression, following childbirth' is a subset of the population defined by the NICE scope? If it is, please define that population subset.

Biogen would like to clarify that the term "following childbirth" does not apply to the symptom onset. The term "following childbirth" only applies to the timing of treatment initiation and is to emphasise that zuranolone is intended to be used after childbirth and not during pregnancy. Therefore, this distinction does not limit the target population to a specific subset. The target population for zuranolone in this submission being 'adults with postnatal depression" as defined in the NICE scope. [redacted]

References

American Psychological Association (2022). "Postpartum depression: Causes, symptoms, risk factors, and treatment options." Retrieved September 30, 2024, from https://www.apa.org/topics/women-girls/postpartum-depression.

Appleby, L., Warner, R., Whitton, A. and Faragher, B. (1997). A controlled study of fluoxetine and cognitive-behavioural counselling in the treatment of postnatal depression. BMJ 314(7085): 932.

Batt, M. M., Duffy, K. A., Novick, A. M., Metcalf, C. A. and Epperson, C. N. (2020). Is Postpartum Depression Different From Depression Occurring Outside of the Perinatal Period? A Review of the Evidence. Focus 18(2): 106-119.

Biogen (2024). Therapy Watch 2023 Postpartum Depression Report. Version 2. Data on file.

Biogen (2025a). Biogen post-hoc analysis April 2025 [Data on file].

Biogen (2025b). Biogen post-hoc analysis February 2025 [Data on file].

Bloch, M., Meiboom, H., Lorberblatt, M., Bluvstein, I., Aharonov, I., et al. (2012). The effect of sertraline add-on to brief dynamic psychotherapy for the treatment of postpartum depression: a randomized, double-blind, placebo-controlled study. J Clin Psychiatry 73(2): 235-241.

Chibanda, D., Shetty, A. K., Tshimanga, M., Woelk, G., Stranix-Chibanda, L., et al. (2014). Group problem-solving therapy for postnatal depression among HIV-positive and HIV-negative mothers in Zimbabwe. J Int Assoc Provid AIDS Care 13(4): 335-341.

Cooper, P. J., Murray, L., Wilson, A. and Romaniuk, H. (2003). Controlled trial of the short- and long-term effect of psychological treatment of post-partum depression: I. Impact on maternal mood. British Journal of Psychiatry 182(5): 412-419.

Deligiannidis, K. M., Citrome, L., Huang, M. Y., Acaster, S., Fridman, M., et al. (2023). Effect of Zuranolone on Concurrent Anxiety and Insomnia Symptoms in Women With Postpartum Depression. J Clin Psychiatry 84(1).

Deligiannidis, K. M., Meltzer-Brody, S., Gunduz-Bruce, H., Doherty, J., Jonas, J., et al. (2021). Effect of Zuranolone vs Placebo in Postpartum Depression: A Randomized Clinical Trial. JAMA Psychiatry 78(9): 951-959.

Duman, R. S., Sanacora, G. and Krystal, J. H. (2019). Altered Connectivity in Depression: GABA and Glutamate Neurotransmitter Deficits and Reversal by Novel Treatments. Neuron 102(1): 75-90.

Hantsoo, L., Ward-O'Brien, D., Czarkowski, K. A., Gueorguieva, R., Price, L. H., et al. (2014). A randomized, placebo-controlled, double-blind trial of sertraline for postpartum depression. Psychopharmacology (Berl) 231(5): 939-948.

Harrison, S., Quigley, M. A., Fellmeth, G., Stein, A. and Alderdice, F. (2023). The impact of the Covid-19 pandemic on postnatal depression: analysis of three population-based national maternity surveys in England (2014-2020). Lancet Reg Health Eur 30: 100654.

Higgins JPT, T. J., Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). (2024). Cochrane Handbook for Systematic Reviews of Interventions version 6.5 (updated August 2024).

Kanes, S., Colquhoun, H., Gunduz-Bruce, H., Raines, S., Arnold, R., et al. (2017). Brexanolone (SAGE-547 injection) in post-partum depression: a randomised controlled trial. The Lancet 390(10093): 480-489.

Kroh, S. J. and Lim, G. (2021). Perinatal depression. International Anesthesiology Clinics 59(3): 45-51.
LoMartire, R., Äng, B. O., Gerdle, B. and Vixner, L. (2020). Psychometric properties of Short Form-36 Health Survey, EuroQol 5-dimensions, and Hospital Anxiety and Depression Scale in patients with chronic pain. Pain 161(1): 83-95.
Maguire, J., Ferando, I., Simonsen, C. and Mody, I. (2009). Excitability changes related to GABAA receptor plasticity during pregnancy. J Neurosci 29(30): 9592-9601.
Marinus Pharmaceuticals (2017). A clinical trial of intravenous (IV) ganaxolone in women with postpartum depression. ClinicalTrials.gov. Bethesda, US National Library of Medicine.
Marinus Pharmaceuticals (2018). A clinical trial of oral ganaxolone in women with postpartum depression. ClinicalTrials.gov. Bethesda, US National Library of Medicine.
Meltzer-Brody, S., Colquhoun, H., Riesenberg, R., Epperson, C. N., Deligiannidis, K. M., et al. (2018). Brexanolone injection in post-partum depression: two multicentre, double-blind, randomised, placebo-controlled, phase 3 trials. The Lancet 392(10152): 1058-1070.
Milgrom, J., Gemmill, A. W., Ericksen, J., Burrows, G., Buist, A., et al. (2015). Treatment of postnatal depression with cognitive behavioural therapy, sertraline and combination therapy: A randomised controlled trial. Australian & New Zealand Journal of Psychiatry 49(3): 236-245.
Misri, S., Reebye, P., Corral, M. and Milis, L. (2004). The use of paroxetine and cognitive-behavioral therapy in postpartum depression and anxiety: a randomized controlled trial. J Clin Psychiatry 65(9): 1236-1241.
Moore Simas, T. A., Huang, M. Y., Patton, C., Reinhart, M., Chawla, A. J., et al. (2019). The humanistic burden of postpartum depression: a systematic literature review. Curr Med Res Opin 35(3): 383-393.
Myczkowski, M. L., Dias, Á. M., Luvisotto, T., Arnaut, D., Bellini, B. B., et al. (2012). Effects of repetitive transcranial magnetic stimulation on clinical, social, and cognitive performance in postpartum depression. Neuropsychiatric Disease and Treatment 8.
National Institute for Health and Care Excellence (2020). "Clinical guideline [CG129]. Antenatal and postnatal mental health: clinical management and service guidance." Retrieved August 20, 2024, from https://www.nice.org.uk/guidance/cg192.
Netsi, E., Pearson, R. M., Murray, L., Cooper, P., Craske, M. G., et al. (2018). Association of Persistent and Severe Postnatal Depression With Child Outcomes. JAMA Psychiatry 75(3): 247-253.
NHS (2022). "Treatment - Postnatal depression." from https://www.nhs.uk/mental-health/conditions/post-natal-depression/treatment/.
O'Hara, M. W., Pearlstein, T., Stuart, S., Long, J. D., Mills, J. A., et al. (2019). A placebo controlled treatment trial of sertraline and interpersonal psychotherapy for postpartum depression. Journal of Affective Disorders 245: 524-532.
Philips, Z., Ginnelly, L., Sculpher, M., Claxton, K., Golder, S., et al. (2004). Review of guidelines for good practice in decision-analytic modelling in health technology assessment. Health Technol Assess 8(36): iii-iv, ix-xi, 1-158.
Posmontier, B. (2008). Functional status outcomes in mothers with and without postpartum depression. J Midwifery Womens Health 53(4): 310-318.

Sage Therapeutics (2019). ROBIN Clinical Study Report.
Sage Therapeutics (2022). SKYLARK Clinical Study Report.
Sage Therapeutics (2023). "A randomized, double-blind, placebo-controlled study evaluating the efficacy and safety of Sage-217 in the treatment of adults with severe postpartum depression ", from https://www.clinicaltrialsregister.eu/ctr-search/trial/2020-001424-34/results.
Sharp, D. J., Chew-Graham, C., Tylee, A., Lewis, G., Howard, L., et al. (2010). A pragmatic randomised controlled trial to compare antidepressants with a community-based psychosocial intervention for the treatment of women with postnatal depression: the RESPOND trial. Health Technol Assess 14(43): iii-iv, ix-xi, 1-153.
Slomian, J., Honvo, G., Emonts, P., Reginster, J. Y. and Bruyère, O. (2019). Consequences of maternal postpartum depression: A systematic review of maternal and infant outcomes. Womens Health (Lond) 15: 1745506519844044.
Stewart, D. E. and Vigod, S. (2016). Postpartum Depression. New England Journal of Medicine 375(22): 2177-2186.
Stolz, M., Albus, C., Beutel, M. E., Deter, H.-C., Fritzsche, K., et al. (2023). Assessment of health-related quality of life in individuals with depressive symptoms: validity and responsiveness of the EQ-5D-3L and the SF-6D. The European Journal of Health Economics 24(8): 1297-1307.
Thomas, J., McDonald, S., Noel-Storr, A., Shemilt, I., Elliott, J., et al. (2021). Machine learning reduced workload with minimal risk of missing studies: development and evaluation of a randomized controlled trial classifier for Cochrane Reviews. Journal of Clinical Epidemiology 133: 140-151.
Wisner, K. L., Hanusa, B. H., Perel, J. M., Peindl, K. S., Piontek, C. M., et al. (2006). Postpartum Depression: A Randomized Trial of Sertraline Versus Nortriptyline. Journal of Clinical Psychopharmacology 26(4).
Wisner, K. L., Sit, D. K., Moses-Kolko, E. L., Driscoll, K. E., Prairie, B. A., et al. (2015). Transdermal Estradiol Treatment for Postpartum Depression: A Pilot, Randomized Trial. J Clin Psychopharmacol 35(4): 389-395.
Yonkers, K. A., Lin, H., Howell, H. B., Heath, A. C. and Cohen, L. S. (2008a). Pharmacologic treatment of postpartum women with new-onset major depressive disorder: A randomized controlled trial with paroxetine. Journal of Clinical Psychiatry 69(4): 659-665.
Yonkers, K. A., Lin, H., Howell, H. B., Heath, A. C. and Cohen, L. S. (2008b). Pharmacologic treatment of postpartum women with new-onset major depressive disorder: a randomized controlled trial with paroxetine. J Clin Psychiatry 69(4): 659-665.
Zorumski, C. F., Paul, S. M., Covey, D. F. and Mennerick, S. (2019). Neurosteroids as novel antidepressants and anxiolytics: GABA-A receptors and beyond. Neurobiol Stress 11: 100196.

Appendix

The CS documents were updated following the clarification procedure and discussions with the EAG. All changes were marked in the CS documents in green highlighting for visibility.

The summary of changes is provided in Table 13 below.

Table 12: Summary of changes across CS documents

EAG Question	Summary of changes to the CS document(s)	Document where the updates were made
A3. Could the company please provide information on parity, previous history of MDD, and baseline psychological therapies for the participants of SKYLARK and ROBIN trials by treatment arm? Please provide this information in updated versions of Tables 9 and 10.	Additional rows detailing the history of MDD for patients in the SKYLARK and ROBIN patient demographics and baseline characteristics tables have been incorporated: • Table 9, Section 2.3.4.1, page 51 and • Table 10, Section 2.3.4.2, page 52	Company submission (redacted and unredacted version)
A5. Could you provide a subgroup analysis for history of major depressive disorder or not, as per the scope? We acknowledge that the sample size is small (as noted in the CS, Doc B) and results will be uncertain, but this subgroup was listed in the NICE scope and the company do report analyses for other subgroups that have small sample sizes.	Additional post-hoc analyses were conducted for subgroup of patients with history of MDD for ROBIN and SKYLARK. The results from the post hoc analyses for both studies are now presented in Section 2.8.2.1: • Figure 24 (SKYLARK), page 75 • Figure 25 (ROBIN), page 76	Company submission (redacted and unredacted version)
B1. Priority question: Can you provide scenarios to explore the impact of a waning of treatment effect over time, where EPDS scores converge after different lengths of time, and provide functionality in the model to adjust this? For example, where EPDS curves converge after: 3 months, 6 months, 1 year, 2 years, 5 years, 10 years.	Additional scenario analysis was conducted to explore the impact of gradually decreasing EPDS scores over 24 months for patients with persistent PND (EPDS≥13) to 12.99 at pre-defined time points (10 years, 5 years and 2 years) . The settings associated with these new scenarios can be found in the "Clinical" tab under the heading "EPDS adjustment".	Excel cost-effectiveness model (redacted and unredacted version)
B2. Priority question: In the MMRM analysis for short-term EPDS that informs the economic model, which interaction terms did you consider for inclusion? Did you consider an interaction between Treatment Arm x ADT? If not, could you explore a potential scenario that includes this interaction?	Changed name of scenario 2 in the economic model from "0% ADT at baseline including in short-term model" to "No ADT use included in the short-term MMRM model". An additional scenario was added to the economic model with 100% of patients on ADT, with ADT use included as a covariate in the MMRM model (scenario 10).	Excel cost-effectiveness model (redacted and unredacted version)
B17. In the patient_simulations sheet of the Excel model, the formulas in cells AL90 to AL94 include reference to cells AMN90 to AMN94. Is this an error?	The cells AMN90 to AMN94 were removed from the formula (these cells did not contain any data). The corrected formula is now as follows:	Excel cost-effectiveness model (redacted and unredacted version)

	• AL90: AJ90*AG90*E90	
B18. In the patient_simulations sheet of the Excel model the formula in cell T19 is =(O15*(O16*(O18/12)*365.25*O20*O23)+(O17*(O19/12)*365.25*O21*O24)) ,but we think the O15 should be before the first bracket, as O15 is the percentage on ADT so should be applied to both the weighted sertraline and fluoxetine costs. There is a similar issue in O26 (the cost of the cost of zuranolone treatment, as that includes ADT costs for those on baseline ADT). Can you please check?	The suggested edits made by the EAG were implemented in the Excel cost-effectiveness model: • The position of O15 was changed to before the first bracket, Cell T9: =O15*((O16*(O18/12)*365.25*O20*O23)+(O17*(O19/12)*365.25*O21*O24)) • The position of O26 was changed to before the first bracket, Cell O26: =(O8*O11*O14*50)+(O9*O12*O14*40)+(O10*O13*O14*30)+O15*((O16*(O18/12)*365.25*O20*O23)+(O17*(O19/12)*365.25*O21*O24)) As this resulted in small changes to the results, the DSA, PSA, and scenarios have been re-run	Excel cost-effectiveness model (redacted and unredacted version)

Abbreviations: ADT: Antidepressant treatment; CS: Company submission; DSA: Deterministic sensitivity analysis; EAG: External Assessment Group; EPDS: Edinburgh Postnatal Depression Scale; MDD; Major Depressive Disorder; MMRM: Mixed model for repeated measures; NICE: National Institute for Health and Care Excellence; PSA: Probabilistic sensitivity analysis.

The risk of bias assessment for SKYLARK and ROBIN are presented in Table 14 below.

Table 13: Risk of bias assessment for SKYLARK and ROBIN

Study	Deligiannidis, 2021 (Deligiannidis 2021) (Some details in ROBIN CSR (Sage Therapeutics 2019))	SKYLARK trial (Sage Therapeutics 2023)
1. Biases arising from the randomisation process		
1.1 Was the allocation sequence random?	Yes "Initially randomisation codes were generated with a block size of 4 by an independent statistical vendor. Following a protocol amendment, randomisation was performed centrally via an IRT system and randomisation schedules were generated by an independent statistician using SAS v 9.4."	Yes Centrally randomised by interactive voice response system. Patients were stratified by use of antidepressant treatment at baseline
1.2 Was the allocation sequence concealed until participants were enrolled and assigned to interventions?	Yes "Randomisation was performed centrally via an IRT system"	Yes Interactive voice response system. Intervention and placebo were identical in appearance
1.3 Did baseline differences between intervention groups suggest a problem with the randomisation process?	No Demographic characteristics were generally well balanced for both groups and were similar when summarised for the safety set.	No Groups reported to be similar at baseline
Risk of bias judgement	**Low**	**Low**
2. Bias due to deviations from the intended intervention (effect of assignment to intervention)		
2.1 Were participants aware of their assigned intervention during the trial?	No Participants blinded to allocation	Yes Participants were not blind to treatment allocation
2.2 Were carers and people delivering the interventions aware of the participants assigned intervention during the trial?	No "Double-blind. Subjects, clinicians, and the study team were blinded to treatment allocation. Pharmacy staff responsible for dispensing medication reported to be unblinded to randomisation."	No Carers and investigators were blinded to treatment allocation
2.3 If Yes/probably yes/no information to 2.1 or 2.2, were there deviations from the intended intervention that arose because of the trial context?	N/A	N/A
2.4 If Yes/probably yes to 2.3, were these deviations likely to have affected the outcomes?	N/A	N/A
2.5 If yes/possibly yes/no information to 2.4 Were these deviations from intended intervention balanced between groups?	N/A	N/A

Study	Deligiannidis, 2021 (Deligiannidis 2021) (Some details in ROBIN CSR (Sage Therapeutics 2019))	SKYLARK trial (Sage Therapeutics 2023)
2.6 Was an appropriate analysis used to estimate the effect of assignment to intervention?	N/A	Yes Intention to treat analysis were reported
2.7 If No/probably no/no information to 2.6. Was there potential for a substantial impact (on the result) of the failure to analyse the participants in the group to which they had been randomised?	N/A	N/A
Risk of bias judgement	**Low**	**Low**
2. Risk of bias due to deviations from the intended interventions (Effect of adhering to intervention)		
2.1 Were participants aware of their assigned intervention during the trial?	No Participants blinded to allocation	Yes Participants were not blind to treatment allocation
2.2 Were carers and people delivering the interventions aware of the participants assigned intervention during the trial?	No "Double-blind. Subjects, clinicians, and the study team were blinded to treatment allocation. Pharmacy staff responsible for dispensing medication reported to be unblinded to randomisation."	No Carers and investigators were blinded to treatment allocation
2.3 [If applicable] If yes/probably yes/ no information to 2.1 or 2.2 Were important non-protocol interventions balanced across intervention groups?	Yes Baseline antidepressant use similar, no modifications to use allowed during study	N/A
2.4 [If applicable] Were there failures in implementing the intervention that could have affected the outcome?	Probably no 2/76 (2.6%) subjects were randomised to placebo but received at least 1 dose of SAGE-217 in error and are included in the SAGE-217 group for the summary of subjects who received study drug.	N/A
2.5 [If applicable] Was there non-adherence to the assigned intervention regimen that could have affected participant's outcomes?	No Most participants adhered to intervention; overall adherence 99.684%	N/A
2.6. If no/probably no/no information to 2.3, or yes probably yes/no information to 2.4 or 2.5 Was an appropriate analysis used to estimate the effect of adhering to the intervention?	N/A	N/A
Risk of bias judgement	**Low**	**Low**
3. Bias due to missing outcome data		
3.1 Were outcome data available for all, or nearly all participants randomised?	Yes	Yes

Study	Deligiannidis, 2021 (Deligiannidis 2021) (Some details in ROBIN CSR (Sage Therapeutics 2019))	SKYLARK trial (Sage Therapeutics 2023)
	Outcome data available for most (>97%) participants in both arms.	Most outcome analyses (in the CSR) are based on the full Analysis Set which was defined as participants who were randomised, administered blinded IP, had a valid baseline and at least 1 post-baseline efficacy assessment.
3.2 If N/PN/NI to 3.1: Is there evidence that the result was not biased by missing outcome data?	N/A	N/A
3.3 If N/PN/NI to 3.2: Could missingness in the outcome depend on its true value?	N/A	N/A
3.4 It Y/PY/NI to 3.3: Is it likely that missingness in the outcome depended on its true value?	N/A	N/A
Risk of bias judgement	**Low**	**Low**
4. Bias in the measurement of the outcome		
4.1 Was the method of measuring the outcome inappropriate?	No Pre-specified, validated outcome scales	No Validated outcome measures used
4.2 Could measurement or ascertainment of the outcome have differed between intervention groups?	No Same outcomes measured for both in the same way	Probably no The study team/investigators were blinded to treatment allocation
4.3 If N/PN to 4.1 and 4.2: Were outcome assessors aware of the intervention received by the study participant?	No Assessors blinded to treatment allocation	No The study team/investigators were blinded to treatment allocation
4.4 If Y/PY/NI to 4.3: Could assessment of the outcome have been influenced by knowledge of intervention received?	N/A	N/A
4.5 If Y/PY/NI to 4.4: Is it likely that assessment of the outcome was influenced by knowledge of intervention received?	N/A	N/A
Risk of bias judgement	**Low**	**Low**
5. Risk of bias in selection of the reported result		
5.1 Were the data that produced this result analysed in accordance with a pre-specified analysis plan that was finalised before unblinded outcome data were available for analysis?	Yes Pre-specified analysis plan reported	Yes Clinical trial and detailed protocol available
5.2. Is the numerical result being assessed likely to have been selected, on the basis of the results, from multiple eligible outcome measurements (e.g. scales, definitions, time points) within the outcome domain?	No Data reported for all outcomes and timepoints specified	No All outcomes appear to have been reported

Study	Deligiannidis, 2021 (Deligiannidis 2021) (Some details in ROBIN CSR (Sage Therapeutics 2019))	SKYLARK trial (Sage Therapeutics 2023)
5.3 Is the numerical result being assessed likely to have been selected, on the basis of the results, from multiple eligible analyses of the data?	No All pre-specified analyses reported	No All eligible results appear to be reported
Risk of bias judgement	**Low**	**Low**
Overall judgement		
	Low	**Some concerns**

Source: Cited within table

Abbreviations: CSR: Clinical study report; IP: Investigational product; IRT: Interactive response technology; N: No; N/A: Not applicable; NI: No information; PN: Probably no; SAS: Statistical analysis system.

The risk of bias assessment for relevant identified studies from the clinical SLR are presented in Table 15 below.

Table 14: Overall risk of bias results for relevant identified studies from the clinical SLR

Study reference	Randomisation	Deviation from intervention (assignment)	Deviation from intervention (adherence)	Missing outcome data	Outcome measurement	Selection of Reported result	Overall
Neurosteroids							
Kanes 2017 (Kanes 2017)	Low	Low	Low	Low	Low	Low	Low
Meltzer-Brody 2018a (Meltzer-Brody 2018)	Low	Low	Some concerns	Low	Low	Low	Some concerns
Meltzer-Brody 2018b (Meltzer-Brody 2018)	Low	Low	Low	Low	Low	Low	Low
NCT03228394 (Marinus Pharmaceuticals 2017)	Low	Low	Low	High	Low	Some concerns	High
NCT03460756 (Marinus Pharmaceuticals 2018)	Low	High	Low	High	High	Some concerns	High
Deligiannidis, 2021 (Deligiannidis 2021) CSR (Sage Therapeutics 2019)	Low	Low	Low	Low	Low	Low	Low
SKYLARK trial (Sage Therapeutics 2023)	Low	Low	Low	Low	Low	Low	Low
SSRIs							
Appleby, 1997 (Appleby 1997)	Some concerns	Low	High	High	Low	Some concerns	High
Bloch, 2012 (Bloch 2012)	Low	Low	Low	Low	Low	Some concerns	Some concerns
Hantsoo, 2014 (Hantsoo 2014)	High	Low	Low	Low	Low	Some concerns	High
Milgrom, 2015 (Milgrom 2015)	Low	Low	Some concerns	Low	Low	Low	Some concerns
Misri, 2004 (Misri 2004)	Some concerns	Some concerns	Some concerns	Low	Low	Some concerns	Some concerns
O'Hara 2019 (O'Hara 2019)	Low	Low	Low	High	Low	Low	High
Sharp 2010 (Sharp 2010)	Some concerns	High	High	High	High	Some concerns	High
Wisner 2015 (Wisner 2015)	Low	Low	Low	Low	Low	Some concerns	Some concerns
Wisner 2006, (Wisner 2006)	Low	Low	Low	High	Low	Some concerns	High
Yonkers 2008 (Yonkers 2008a, Yonkers 2008b)	Low	Low	Low	High	Low	Some concerns	High
Tricyclics							
Chibanda 2014 (Chibanda 2014)	Low	High	High	High	High	Some concerns	High
NPTs							
Myczkowski 2012 (Myczkowski 2012)	Low	Low	Some concerns	Low	Low	Low	Some concerns
Cooper 2003 (Cooper 2003)	Low	Some concerns	Some concerns	Some concerns	High	Some concerns	High

Source: Cited within table

Abbreviations: AE: Adverse events; BIMF: Barkin Index of Maternal Functioning; CGI: Clinical Global Impression; CGI-I: Clinical Global Impression Improvement; CGI-S: Clinical Global Impression Severity; C-SSRS: Columbia-Suicide Severity Rating Scale; EPDS: Edinburgh Postnatal Depression Score; HAM-A: Hamilton Anxiety Rating Scale; HAM-D: Hamilton Rating Scale for Depression; HAM-D-17: Hamilton Rating Scale for Depression - 17 items; MADRS: Montgomery-Åsberg Depression Rating Scale; NPT: Non-Pharmacological Interventions; PHQ-9: Patient Health Questionnaire - 9 items; SF-36: Short Form Health Survey - 36 items.

NICE National Institute for Health and Care Excellence

Single Technology Appraisal
Zuranolone for treating postnatal depression [ID6431]
Professional organisation submission

Thank you for agreeing to give us your organisation's views on this technology and its possible use in the NHS.

You can provide a unique perspective on the technology in the context of current clinical practice that is not typically available from the published literature.

To help you give your views, please use this questionnaire. You do not have to answer every question – they are prompts to guide you. The text boxes will expand as you type.

Information on completing this submission

Please do not embed documents (such as a PDF) in a submission because this may lead to the information being mislaid or make the submission unreadable

We are committed to meeting the requirements of copyright legislation. If you intend to include **journal articles** in your submission you must have copyright clearance for these articles. We can accept journal articles in NICE Docs.

Your response should not be longer than 13 pages.

About you

1. Your name	
2. Name of organisation	College of Mental Health Pharmacy
3. Job title or position	Specialist Perinatal Pharmacist
4. Are you (please select Yes or No):	An employee or representative of a healthcare professional organisation that represents clinicians? Yes A specialist in the treatment of people with this condition? Yes A specialist in the clinical evidence base for this condition or technology? Yes Other (please specify):
5a. Brief description of the organisation (including who funds it).	The College of Mental Health Pharmacy is a registered charity Which aims to benefit individual care through advancing education and research in the practice of mental health pharmacy. We are funded by our membership and by our corporate partnership scheme. Details of our corporate partner and previous partners can be found here - https://www.cmhp.org.uk/about-us/corporate-partners/
5b. Has the organisation received any funding from the manufacturer(s) of the technology and/or comparator products in the last 12 months? [Relevant manufacturers are listed in the appraisal matrix.] **If so, please state the name of manufacturer, amount, and purpose of funding.**	No
5c. Do you have any direct or indirect links with, or funding from, the tobacco industry?	No

The aim of treatment for this condition

6. What is the main aim of treatment? (For example, to stop progression, to improve mobility, to cure the condition, or prevent progression or disability.)	For the treatment of severe postnatal depression. The main aim of the treatment is to reduce symptoms of depression, restore functioning, and prevent relapse or recurrence**.**
7. What do you consider a clinically significant treatment response? (For example, a reduction in tumour size by x cm, or a reduction in disease activity by a certain amount.)	A reduction of at least 50% in depression scale scores (such as the Edinburgh Postnatal Depression Scale) would be considered clinically significant.
8. In your view, is there an unmet need for patients and healthcare professionals in this condition?	Postnatal depression is a significant mental health concern affecting many women in the UK. Current treatment options, such as selective serotonin reuptake inhibitors (SSRIs) and cognitive-behavioural therapy (CBT), often require several weeks to alleviate symptoms, which can be a considerable burden during the postpartum period. This delay in therapeutic effect highlights an unmet need for treatments that provide more rapid relief. Access to a treatment that provides rapid symptom relief could significantly improve outcomes for mothers experiencing postnatal depression. While SSRIs and other antidepressants are commonly prescribed, they are not specifically licensed for postnatal depression, meaning their use relies on clinical judgment rather than regulatory approval for this indication. Licensed treatments should be used in preference whenever available, as they have been rigorously evaluated for safety and efficacy in the target population, providing clearer guidance for healthcare professionals and greater reassurance for patients.

What is the expected place of the technology in current practice?

9. How is the condition currently treated in the NHS?	Currently NICE guidance (CG192) recommends: *1.8.3 For a woman with moderate or severe depression in pregnancy or the postnatal period, consider the following options:* ● *a high-intensity psychological intervention (for example, CBT)* ● *a TCA, SSRI or (S)NRI if the woman understands the risks associated with the medication and the mental health problem in pregnancy and the postnatal period and:* ○ *she has expressed a preference for medication or* ○ *she declines psychological interventions or* ○ *her symptoms have not responded to psychological interventions* ● *a high-intensity psychological intervention in combination with medication if the woman understands the risks associated with the medication and the mental health problem in pregnancy and the postnatal period and there is no response, or a limited response, to a high-intensity psychological intervention or medication alone.*
9a. Are any clinical guidelines used in the treatment of the condition, and if so, which?	● NICE Clinical Guideline CG192 Antenatal and postnatal mental health: clinical management and service guidance ● NICE Guideline NG222 Depression in adults: treatment and management
9b. Is the pathway of care well defined? Does it vary or are there differences of opinion between professionals across the NHS? (Please state if your experience is from outside England.)	Yes, clear pathway of care of women with postnatal depression. This would normally involve a stepped approach involving GP, Specialist Community Perinatal Mental Health Teams and Specialist inpatient Mother and Baby Units.
9c. What impact would the technology have on the current pathway of care?	If zuranolone were to be licensed in the UK for post-natal depression, it could significantly impact the current pathway of care by offering a faster-acting, licensed treatment option specifically designed for this condition. Unlike traditional antidepressants, which can take several weeks to become effective and are often used off-

	license, zuranolone has shown rapid symptom relief within days. This could reduce the reliance on longer treatment courses with SSRIs and potentially decrease the need for hospital admissions or crisis interventions. Additionally, its short 14-day treatment regimen may improve adherence and recovery rates, allowing mothers to engage in bonding and daily activities more quickly. The introduction of zuranolone could streamline treatment decisions for healthcare professionals by providing a licensed, evidence-based option for postnatal depression, ultimately improving outcomes for both patients and the wider healthcare system. It would be likely that the prescribing and monitoring would be done by a Specialist Perinatal Mental Health Team as per current care pathway with current options.
10. Will the technology be used (or is it already used) in the same way as current care in NHS clinical practice?	Yes, the treatment will be used as an alternative to prescribing an antidepressant in the current care pathway. Treatment response and course is significantly shorter than traditional antidepressant options.
10a. How does healthcare resource use differ between the technology and current care?	Healthcare resource use could differ significantly compared to current treatment options. Traditional antidepressants, such as SSRIs, require long-term treatment and regular follow-ups to monitor effectiveness and side effects, often necessitating ongoing GP or specialist input. Additionally, due to their delayed onset of action, some patients may require crisis interventions, inpatient care, or additional psychological support while waiting for symptom improvement. In contrast, zuranolone's rapid onset of action—demonstrating symptom relief within days—could reduce the need for prolonged treatment, frequent follow-ups, and emergency mental health interventions. Its short 14-day regimen may also improve adherence and decrease long-term medication costs. Furthermore, by providing faster symptom relief, it could lessen the impact of PPD on mother-infant bonding, reducing the need for extended perinatal mental health services. While the upfront cost of zuranolone may be higher than traditional antidepressants, the potential savings from reduced hospital admissions, crisis care, and prolonged treatment make it a potentially more efficient use of healthcare resources. Mostly in cost of Zuranolone compared to standard antidepressant costs. Speed of treatment response is reported to be quicker with Zuranolone therefore use of healthcare services may well be shorter in those receiving this intervention.
10b. In what clinical setting should the technology be	Zuranolone would be prescribed only for severe postnatal depression and managed primarily by specialist perinatal mental health teams in secondary care (community specialist perinatal mental health teams or inpatient

used? (For example, primary or secondary care, specialist clinics.)	Mother and Baby Units). Its rapid onset of action and potential sedative effects require specialist oversight for safe initiation and monitoring. Unlike traditional antidepressants managed in primary care, zuranolone would be used within perinatal psychiatry services, ensuring appropriate assessment and diagnosis, risk management, and follow-up. While primary care may support post-treatment monitoring, prescribing and initial management would remain within specialist services. Specialist Perinatal Mental Health Teams are now well established across the country.
10c. What investment is needed to introduce the technology? (For example, for facilities, equipment, or training.)	Introducing zuranolone in the UK would require investment in specialist perinatal mental health services, including training for clinicians on prescribing, monitoring, and managing side effects. Clear prescribing pathways and patient education materials would also be essential. Importantly, zuranolone's cost is expected to be significantly higher than traditional antidepressants, requiring careful consideration of its financial impact on healthcare budgets.
11. Do you expect the technology to provide clinically meaningful benefits compared with current care?	Yes. Unlike traditional antidepressants, which can take weeks to show effects, zuranolone has demonstrated rapid symptom relief within days, which could be critical in reducing suffering and improving mother-infant bonding. Its short 14-day treatment course may also improve adherence compared to long-term antidepressant use. Additionally, by offering a licensed treatment specifically for PPD, zuranolone could provide a more targeted and evidence-based option, reducing the reliance on off-license prescribing.
11a. Do you expect the technology to increase length of life more than current care?	Zuranolone is unlikely to directly increase length of life more than current care, as it is specifically designed to address severe postnatal depression rather than to have a direct impact on mortality. However, by providing rapid relief of depressive symptoms, it may significantly improve the quality of life for mothers suffering from postnatal depression, potentially reducing the long-term psychological and physical health risks associated with untreated or inadequately treated depression. This could indirectly contribute to better overall wellbeing and, in some cases, improve long-term health outcomes by preventing the chronic effects of severe depression. The focus of zuranolone would be more on enhancing recovery and mental health rather than directly affecting life expectancy. There may be a reduction in completed suicides (which is still a leading cause of death in the perinatal period) due to a more rapid improvement in symptoms.
11b. Do you expect the technology to increase health-related quality of life more than current care?	As above. Its rapid onset of action could provide quick symptom relief, helping to reduce the emotional and psychological burden on mothers. This improvement in mental health can have a positive impact on mother-baby attachment, as mothers are more likely to engage in bonding activities and caregiving when their depression symptoms are alleviated sooner. Enhanced attachment can lead to better long-term developmental and emotional outcomes for the child. Additionally, by addressing postnatal depression more effectively and promptly, zuranolone may prevent the long-term consequences of untreated depression, such as chronic mental health issues or difficulties in parenting, further improving both maternal and child well-being.

12. Are there any groups of people for whom the technology would be more or less effective (or appropriate) than the general population?	More appropriate: • Women with severe postnatal depression Less appropriate: • Women with mild to moderate postnatal depression.

The use of the technology

13. Will the technology be easier or more difficult to use for patients or healthcare professionals than current care? Are there any practical implications for its use (for example, any concomitant treatments needed, additional clinical requirements, factors affecting patient acceptability or ease of use or additional tests or monitoring needed.)	Zuranolone may be easier to use for patients compared to current care, particularly due to its rapid onset of action and short 14-day treatment course. Unlike traditional antidepressants, which can require several weeks of daily medication and frequent follow-ups to monitor effectiveness and side effects, zuranolone's shorter treatment duration could improve patient adherence and reduce the burden of long-term therapy. For healthcare professionals, while the prescribing and monitoring process may be more straightforward, careful management is still required, especially given its sedative effects. Specialists would need to assess patients for suitability, monitor for side effects, and ensure proper follow-up. Training for healthcare professionals in perinatal psychiatry may also be necessary to ensure appropriate prescribing and patient support. A key practical consideration for both patients and healthcare providers is the recommendation to hold breastfeeding while taking zuranolone, as the medication's safety in breastfeeding mothers has not been established. Patients may need guidance on alternative feeding options during treatment.
14. Will any rules (informal or formal) be used to start or stop treatment with the technology? Do these include any additional testing?	No additional test required before starting. No formal or informal rules associated.
15. Do you consider that the use of the technology will result in any	Yes, the use of zuranolone could result in substantial health-related benefits that may not be fully captured in the quality-adjusted life year (QALY) calculation, particularly when considering the improvement in parent-infant bonding and attachment. Rapid symptom relief from severe postnatal depression can facilitate more positive

substantial health-related benefits that are unlikely to be included in the quality-adjusted life year (QALY) calculation?	interactions between mother and child, which is crucial for the child's early emotional and developmental well-being. The improved bonding and attachment can have lasting effects on the child's mental and emotional health, leading to better long-term outcomes in areas such as social development, cognitive abilities, and emotional regulation. The broader, long-term benefits of strengthened parent-child attachment—such as improved family dynamics, reduced risk of child behavioural problems, and better maternal mental health in subsequent years—may not be adequately reflected in the QALY framework. These indirect but meaningful benefits of zuranolone, particularly in promoting healthier family relationships, are significant but may not be fully captured in traditional cost-effectiveness analyses.
16. Do you consider the technology to be innovative in its potential to make a significant and substantial impact on health-related benefits and how might it improve the way that current need is met?	Yes, zuranolone is innovative in its potential to make a significant impact on health-related benefits, particularly due to its distinct mechanism of action compared to traditional treatments for postnatal depression. Unlike conventional antidepressants, such as SSRIs, which can take weeks to show effects, zuranolone acts much more quickly, with symptom improvement often seen within days. This rapid onset addresses an urgent need for mothers experiencing severe postnatal depression, allowing for quicker relief and better functioning.
16a. Is the technology a 'step-change' in the management of the condition?	Yes, zuranolone represents a 'step-change' in the management of post-natal depression due to its rapid onset of action compared to traditional treatments and innovative mechanism of action. While conventional antidepressants like SSRIs can take weeks to show effectiveness, zuranolone provides symptom relief within days, offering faster recovery for mothers suffering from severe postnatal depression. Traditional treatments of depression have focused on targeting neurotransmitters in the brain. These have not been specific to postnatal depression. Zuranolone targets the changes in hormones, specific to some presentations of postnatal depression, unlike any other treatment available. This treatment opens the door to perinatal mental health conditions being recognised as distinct to those at other stages of a women's life.
16b. Does the use of the technology address any particular unmet need of the patient population?	Yes. Current treatments, such as traditional antidepressants, often take weeks to show effectiveness, leaving many women struggling with debilitating symptoms during a critical time in their lives. Zuranolone's rapid onset of action, with symptom relief often seen within days, provides a much-needed solution for women who are experiencing the acute distress of severe PPD. This faster-acting treatment can help improve maternal well-being,

	reduce the emotional burden on mothers, and enhance mother-infant bonding more quickly, ultimately improving long-term outcomes for both the mother and child.
17. How do any side effects or adverse effects of the technology affect the management of the condition and the patient's quality of life?	The side effects of zuranolone, including sedation, dizziness, and somnolence, may impact the management of the condition and the patient's quality of life. These effects could pose challenges for new mothers who need to stay alert for caregiving and infant care. Additionally, there is a recommendation from the World Health Organisation to avoid co-sleeping while taking the medication due to the increased sedative effects, which could raise safety concerns for both mother and baby. Another important consideration is the need to hold breastfeeding while taking zuranolone, as its safety during lactation has not been established. This could be an additional challenge for mothers who wish to breastfeed, requiring alternative feeding options during treatment. Despite these side effects, zuranolone's short 14-day treatment course and rapid symptom relief may ultimately improve maternal well-being by alleviating severe depression more quickly, allowing for better long-term outcomes.

Sources of evidence

18. Do the clinical trials on the technology reflect current UK clinical practice?	Largely. As clinical trials were completed in the US, diagnosis of depression was completed using the DSM 5 requirements, rather than ICD 11.
18a. If not, how could the results be extrapolated to the UK setting?	This would not be expected to have a significant impact as the two broadly overlap.
18b. What, in your view, are the most important outcomes, and were they measured in the trials?	Improvement in depressive symptoms. Yes.
18c. If surrogate outcome measures were used, do	

they adequately predict long-term clinical outcomes?	
18d. Are there any adverse effects that were not apparent in clinical trials but have come to light subsequently?	Not aware.
19. Are you aware of any relevant evidence that might not be found by a systematic review of the trial evidence?	No
20. How do data on real-world experience compare with the trial data?	Unaware of any real-world data being available due to the treatment only recently being available in the US.

Equality

21a. Are there any potential equality issues that should be taken into account when considering this treatment?	
21b. Consider whether these issues are different from issues with current care and why.	

Topic-specific questions

22. How often is treatment for postnatal depression augmented with additional antidepressants, antipsychotics, or electroconvulsive therapy? Which patient or disease characteristics would you typically consider as appropriate for augmentation?	Augmentation strategies are generally used when first-line treatments fail to provide adequate symptom relief. This can occur in approximately 30-50% of patients with moderate to severe depression. However, the exact frequency varies based on the severity of the condition and the patient's response to initial treatments. Augmentation with an additional antidepressant, an antipsychotic or ECT may be considered when there is an inadequate response to a number of antidepressants, commonly at least 2-3. Current local guidelines support the use of augmenting agents after 3 alternative antidepressants have been tried. **Appropriate Patient Characteristics for Augmentation** • **Severity**: Augmentation is more likely to be considered in patients with **severe postpartum depression**, particularly if symptoms are not improving with initial treatments. This may include patients with **psychotic features**, **severe functional impairment**, or **suicidal ideation**. • **Treatment Resistance**: For patients who have not responded to one or more trials of standard antidepressants, **augmentation** may be needed. This is common in **treatment-resistant depression** and is a key factor when considering the addition of antipsychotics or ECT. • **Co-morbid Conditions**: In cases where patients have co-occurring psychiatric conditions, such as **bipolar disorder** or **psychotic disorders**, adjunctive treatments like **antipsychotics** may be required. Patients with **severe anxiety** or **sleep disturbances** may benefit from **mirtazapine** or **quetiapine** for their sedative effects. • **Psychotic Features**: If the patient is experiencing **psychotic features** (delusions, hallucinations), **antipsychotics** or **ECT** may be appropriate to manage these more severe symptoms effectively. • **Urgency**: In situations where there is a **high risk of suicide** or **severe functional impairment**, faster-acting treatments, such as **ECT**, may be required to stabilize the patient more quickly.

Key messages

23. In up to 5 bullet points, please summarise the key messages of your submission.	• First licensed treatment for postnatal depression in the UK • Rapid onset of action – faster than all comparative options • Short length of treatment – likely improving adherence • Improvement in both maternal and infant health – quicker improvement in maternal mental health will support improvement mother-infant bonding • Only likely to be beneficial to those with severe postnatal depression where there is a biological cause.

Thank you for your time.

Please log in to your NICE Docs account to upload your completed submission.

Your privacy

The information that you provide on this form will be used to contact you about the topic above.

Please select YES if you would like to receive information about other NICE topics - YES

For more information about how we process your personal data please see our privacy notice.

NICE National Institute for Health and Care Excellence

Single Technology Appraisal
Zuranolone for treating postnatal depression [ID6431]
NHS organisation submission

Thank you for agreeing to give us your views on the technology and the way it should be used in the NHS.

The Department of Health and Social Care and the Welsh Government provide a unique perspective on the technology, which is not typically available from the published literature. NICE believes it is important to involve NHS organisations that are responsible for commissioning and delivering care in the NHS in the process of making decisions about how technologies should be used in the NHS.

To help you give your views, we have provided a template. The questions are there as prompts to guide you. You do not have to answer every question. Short, focused answers, giving a Department of Health and Social Care and Welsh Government perspective on the issues you think the committee needs to consider, are what we need.

NICE National Institute for Health and Care Excellence

About you

Your name	
Name of your organisation	NHS England
Please indicate your position in the organisation	Department of Health and Social Care or Welsh Government in general? • Commissioning services for the Department of Health and Social Care or Welsh Government specific to the condition for which NICE is considering this technology? • Responsible for quality of service delivery in the ICB (e.g. medical director, public health director, director of nursing)? • A specialist in the treatment of people with the condition for which NICE is considering this technology? • A specialist in the clinical evidence base that is to support the technology (e.g. participation in clinical trials for the technology)? • Other (please specify): • National Specialty Advisor Mental health Pharmacy NHS England on behalf Of The national mental health Perinatal team
Do you have any links with, or funding from, the tobacco industry? Please declare any direct or indirect links to, and receipt of funding from the tobacco industry	No

What is the expected place of the technology in current practice?

How is the condition currently treated in the NHS? Is there significant	In line with Antenatal and postnatal mental health: clinical management and service guidance Clinical guideline CG192 and NICE Depression in adults: treatment and management NICE guideline [NG222]

geographical variation in current practice? Are there differences in opinion between professionals as to what current practice should be? What are the current alternatives (if any) to the technology, and what are their respective advantages and disadvantages?	Pharmacological treatment including antidepressants. Limitations of antidepressants include time taken for a response and tolerability/adverse effects Non pharmacological treatment including evidence based talking treatments.

To what extent and in which population(s) is the technology being used in your local health economy? **Is there variation in how it is being used in your local health economy?** **Is it always used within its licensed indications? If not, under what circumstances does this occur?** **What is the impact of the current use of the technology on resources?** **What is the outcome of any evaluations or audits of the use of the technology?** **What is your opinion on the appropriate use of the technology?**	If the real world experience of use of this treatment demonstrates a substantial benefit in a UK setting , this treatment will increase the speed of recovery and could mean that admission to secondary care mental health services or specialist mother & baby units can be avoided. There remain a number of unknows, such as the safety of repeated courses of this treatment, the risks of dependence, the impact of sedation on the mother and her bonding with the baby.

Potential impact on the NHS if NICE recommends the technology

What impact would the guidance have on the delivery of care for patients with this condition?	If this treatment proves to be effective and without undue adverse effects (eg sedation, dependence) etc this treatment could offer significant benefits to both the mother and child
In what setting should/could the technology be used – for example, primary or	Given the lack of UK based clinical trial data , in the first instance there may be merit in restricting use to specialist services within mother and baby units.

secondary care, specialist clinics? Would there be any requirements for additional resources (for example, staff, support services, facilities or equipment)?	
Can you estimate the likely budget impact? If this is not possible, please comment on what factors should be considered (for example, costs, and epidemiological and clinical assumptions).	To be advised by NHSE Medicines value team
Would implementing this technology have resource implications for other services (for example, the trade-off between using funds to buy more diabetes nurses versus more insulin pumps, or the loss of funds to other programmes)?	If the technology proves to be effective , a reduction in length of stay within mother & baby units can be expected.
Would there be any need for education and training of NHS staff?	

Equality

Please let us know if you think that this appraisal: **Could exclude from full consideration any people protected by the equality legislation who fall within the patient population for which [the treatment(s)] is/are/will be licenced** **Could lead to recommendations that have a different impact on people protected by the equality legislation than on the wider population, e.g. by making it more difficult in practice for a specific group to access the technology** **Could lead to recommendations that have any adverse impact on people with a particular disability or disabilities.**	
Please tell us what evidence should be obtained to enable the committee to identify and consider such impacts.	

NICE is committed to promoting equality of opportunity, eliminating unlawful discrimination and fostering good relations between people with particular protected characteristics and others.

Other issues

Please include here any other issues you would like the appraisal committee to consider when appraising this technology	In summary • Brexanolone and zuranalone have certainly opened up the opportunity to treat this diagnosis in a novel way. A safe, rapid onset treatment, with a sustained response, for post-natal depression would be very welcome and potentially life altering for many women and children. • The trials so far have been for severe depression and have been controlled with placebo arms – even the placebo arms are showing good improving – reinforcing the benefits of good nursing care for women with PND. PMHS in UK offer more than this and would proactively treat with SSRIs and psychological interventions as indicated and so a RCT between zuranolone and TAU would be more helpful. • There are key limitations with zuranolone – it is sedating – which is problematic for new mothers who need to care for their baby and potentially very risky should co-sleeping occur. • Breast feeding is stopped and we know that breast feeding has enormous benefits for mother and baby when it is permitted and possible. More evidence is required re safety in terms of breastfeeding. • Trials show that it is associated with increased risk for drug dependence compared to TAU. Also risk of physical dependence and risk of fetal abnormality (in women who get pregnant again whilst on drug presumably). Therefore risk of dependence and discontinuation effects also need to be quantified

	• In order to show the benefits of zuranolone in the UK further studies are needed. They would need to take into account the robust perinatal mental health services which are available across the UK which supports proactive bio-psycho-social treatments for women affected by severe post natal depression. The studies would have to show a significant increase in speed of efficacy – which in turn would reduce the impact on mother, baby and the family. • It was felt that due to the sedating effects, before it is offered mainstream to all women with severe post natal depression, it could perhaps be offered within a Mother and Baby Unit setting where it could be closely monitored/controlled more easily - the key question would be would it get people out of the MBU quicker?. Then potentially offered more widely within community PMH teams once more data is available and it can be done safely. It was noted that it would be important that it didn't result in a sizeable increase in referrals to MBU's for this purpose.

Your privacy

The information that you provide on this form will be used to contact you about the topic above.

Please select YES if you would like to receive information about other NICE topics - YES

For more information about how we process your personal data please see our privacy notice.

NICE National Institute for Health and Care Excellence

Single Technology Appraisal

Zuranolone for treating postnatal depression [ID6431]

Clinical expert statement

Information on completing this form

In part 1 we are asking for your views on this technology. The text boxes will expand as you type.

In part 2 we are asking you to provide 5 summary sentences on the main points contained in this document.

Please do not embed documents (such as a PDF) in a submission because this may lead to the information being mislaid or make the submission unreadable. Please type information directly into the form.

Do not include medical information about yourself or another person that could identify you or the other person.

We are committed to meeting the requirements of copyright legislation. If you want to include **journal articles** in your submission you must have copyright clearance for these articles. We can accept journal articles in NICE Docs. For copyright reasons, we will have to return forms that have attachments without reading them. You can resubmit your form without attachments, but it must be sent by the deadline.

Combine all comments from your organisation (if applicable) into 1 response. We cannot accept more than 1 set of comments from each organisation.

Please underline all confidential information, and separately highlight information that is submitted as 'confidential [CON]' in turquoise, and all information submitted as 'depersonalised data [DPD]' in pink. If confidential information is submitted, please also

send a second version of your comments with that information redacted. See Health technology evaluations: interim methods and process guide for the proportionate approach to technology appraisals (section 3.2) for more information.

The deadline for your response is **5pm** on **<insert deadline>.** Please log in to your NICE Docs account to upload your completed form, as a Word document (not a PDF).

Thank you for your time.

We reserve the right to summarise and edit comments received, or not to publish them at all, if we consider the comments are too long, or publication would be unlawful or otherwise inappropriate.

Comments received are published in the interests of openness and transparency, and to promote understanding of how recommendations are developed. The comments are published as a record of the comments we received, and are not endorsed by NICE, its officers or advisory committees.

Part 1: Treating postnatal depression and current treatment options

Table 1 About you, aim of treatment, place and use of technology, sources of evidence and equality

1. Your name	Livia Martucci
2. Name of organisation	Royal College of Psychiatrists (RCPsych), South London and Maudsley NHS Foundation Trust (SLaM)
3. Job title or position	Chair of the Perinatal Faculty, RCPsych, consultant in perinatal psychiatry and clinical lead for perinatal services, SLaM
4. Are you (please tick all that apply)	☒ An employee or representative of a healthcare professional organisation that represents clinicians? ☒ A specialist in the treatment of people with postnatal depression? ☒ A specialist in the clinical evidence base for postnatal depression or technology? ☐ Other (please specify):
5. Please disclose any past or current, direct or indirect links to, or funding from, the tobacco industry.	No links, direct or indirect
6. What is the main aim of treatment for postnatal depression? (For example, to stop progression, to improve mobility, to cure the condition, or prevent progression or disability)	To stop progression or alleviate symptoms so that they subside – many antidepressant have a partial benefit which can still have a significant impact
7. What do you consider a clinically significant treatment response? (For example, a reduction in tumour size by x cm, or a reduction in disease activity by a certain amount)	An improvement in mood defined as regaining ability to feel enjoyment, engage meaningfully in the care of the infant and the mother-infant relationship, improvement in sleep and appetite, improved focus and concentration, improved motivation to engage with activities such as work, hobbies, social contact.
8. In your view, is there an unmet need for patients and healthcare professionals in postnatal depression?	Detection and diagnosis can be limited, which means more prolonged illness and in turn greater difficulties in treatment. Available treatments are often only

	partially effective. Therefore there are unmet needs for patients and healthcare professionals.
9. How is postnatal depression currently treated in the NHS? • Are any clinical guidelines used in the treatment of the condition, and if so, which? • Is the pathway of care well defined? Does it vary or are there differences of opinion between professionals across the NHS? (Please state if your experience is from outside England.) • What impact would the technology have on the current pathway of care?	NICE guidelines on the treatment of antenatal and postnatal maternal mental health Pathways of care are well defined by NICE but require updating (guidelines dated 2014) Augmentation of current available treatments with Zuranolone may improve outcomes and overall effectiveness of pathway of care (patients may become more responsive to non pharmacological treatments if their pharmacological treatment is optimised)
10. Will the technology be used (or is it already used) in the same way as current care in NHS clinical practice? • How does healthcare resource use differ between the technology and current care? • In what clinical setting should the technology be used? (for example, primary or secondary care, specialist clinic) • What investment is needed to introduce the technology? (for example, for facilities, equipment, or training)	Pharmacological treatments for postnatal depression have longer courses compared to Zuranolone (months at least compared to 2 weeks) Only in specialist perinatal services, community or Mother and Baby Units, secondary or tertiary care Some training may be required.
11. Do you expect the technology to provide clinically meaningful benefits compared with current care?	A review on available evidence of effectiveness shows that current evidence on effectiveness is limited (Wilson et al, 2025).

• Do you expect the technology to increase length of life more than current care? • Do you expect the technology to increase health-related quality of life more than current care?	Yes, probably, as a contributor to preventing maternal deaths by suicide Yes, probably, by improving maternal mental state and as a result the mother-infant relationship and maternal ability to childcare.
12. Are there any groups of people for whom the technology would be more or less effective (or appropriate) than the general population?	Patients with additional comorbidities e.g. personality disorders, attachment disorders and bonding problems, psychotic illness, anxiety disorders may benefit from this medication to a limited extent and if effective it will be one of the contributors to their treatment plans. It is possible that mothers with postnatal depression and no further complicating factors may respond better to this medication
13. Will the technology be easier or more difficult to use for patients or healthcare professionals than current care? Are there any practical implications for its use? (For example, any concomitant treatments needed, additional clinical requirements, factors affecting patient acceptability or ease of use or additional tests or monitoring needed)	The treatment should not be more difficult to use than current ones as it is an oral medication, but clinicians may feel it will require additional clinical monitoring. Side effects will need to be taken into account
14. Will any rules (informal or formal) be used to start or stop treatment with the technology? Do these include any additional testing?	The treatment should only be prescribed or discontinued following assessment by a consultant specialised in perinatal psychiatry, and with their supervision.
15. Do you consider that the use of the technology will result in any substantial health-related benefits that are unlikely to be included in the quality-adjusted life year (QALY) calculation? • Do the instruments that measure quality of life fully capture all the benefits of the technology or have some been missed? For example, the treatment regimen	This treatment will be administered as an oral tablet – similarly to other antidepressant treatments.

may be more easily administered (such as an oral tablet or home treatment) than current standard of care	
16. Do you consider the technology to be innovative in its potential to make a significant and substantial impact on health-related benefits and how might it improve the way that current need is met? • Is the technology a 'step-change' in the management of the condition? • Does the use of the technology address any particular unmet need of the patient population?	Zuranolone is different as it is a brief course of treatment compared to other antidepressants that require treatment for months. Furthermore, it has a different pharmacological make up. It will add to the possibility of augmenting current treatments with a novel pharmacological agent
17. How do any side effects or adverse effects of the technology affect the management of the condition and the patient's quality of life?	There is limited data on side effects or adverse effect of Zuranolone as there is limited evidence about this treatment
18. Do the clinical trials on the technology reflect current UK clinical practice? • If not, how could the results be extrapolated to the UK setting? • What, in your view, are the most important outcomes, and were they measured in the trials? • If surrogate outcome measures were used, do they adequately predict long-term clinical outcomes? • Are there any adverse effects that were not apparent in clinical trials but have come to light subsequently?	The clinical trials have been limited but carried out in a way similar as in the UK Improvement in mental state is the principal outcome. This was measured in the trials. Improvement in mother infant relationship and in childcare skills can improve self confidence and as a result prolong the positive outcome.
19. Are you aware of any relevant evidence that might not be found by a systematic review of the trial evidence?	no
20. How do data on real-world experience compare with the trial data?	There is very limited real world experience to my knowledge

21. NICE considers whether there are any equalities issues at each stage of an evaluation. Are there any potential equality issues that should be taken into account when considering this condition and this treatment? Please explain if you think any groups of people with this condition are particularly disadvantaged. Equality legislation includes people of a particular age, disability, gender reassignment, marriage and civil partnership, pregnancy and maternity, race, religion or belief, sex, and sexual orientation or people with any other shared characteristics. Please state if you think this evaluation could • exclude any people for which this treatment is or will be licensed but who are protected by the equality legislation • lead to recommendations that have a different impact on people protected by the equality legislation than on the wider population • lead to recommendations that have an adverse impact on disabled people. Please consider whether these issues are different from issues with current care and why. More information on how NICE deals with equalities issues can be found in the NICE equality scheme. Find more general information about the Equality Act and equalities issues here.	No No No
Key issue 1	

22. How is mild, moderate and severe postnatal depression (PND) diagnosed in clinical practice? What are the key differences between each classification? (relevant section: page 23 EAG report)	Severity of depression is usually diagnosed via clinical interview in specialist services. In primary care it is likely that the EPDS will be used in the UK
23. Baseline PND severity was recorded in the trial by multiple depression measurement tools. Which results would you expect to better reflect the UK PND population - HAMD-17, MASRS or EPDS? (relevant section: pages 42-44 EAG report)	In the UK, out of the scales mentioned, it is more likely that EPDS will be used most likely in primary care. In specialist settings usually clinical progress is documented as expert clinician and Key Performance Indicators such as HoNOS, CARE-10.
24. What population would be most likely to receive zuranolone in the NHS – people with mild, moderate or severe PND?	Only patients with moderate to severe PND will receive Zuranolone, as they will have access to a perinatal psychiatrist's care.
Key issue 2	
25. Does the scope accurately represent established clinical management (ECM) for PND in the NHS? (relevant section: pages 24-25 EAG report)	Established clinical management for PND in primary care is as a first line psychological treatment or parent infant psychotherapy/interventions. Antidepressant medication can be added or prescribed in the first instance if patients are not willing to engage in psychological treatment. In moderate to severe PND under the care of secondary perinatal psychiatric services often antidepressant therapy is first line or prescribed alongside psychological treatment/PIP.
26. Do the concomitant treatments permitted in the SKYLARK and ROBIN accurately represent ECM for PND? If not, how do you expect that this will impact the generalisability of the clinical trial results? (relevant section: pages 37-38 EAG report)	ECM for PND does not exclude use of concomitant pharmacological treatments e.g. antipsychotics or other treatments such as ECT in cases of severe PND, therefore the representation of ECM in ROBIN and SKYLARK is limited and not fully representative of current NHS care.
27. Do you anticipate zuranolone being prescribed in the NHS with or without established clinical management? (relevant section: pages 37-38 EAG report)	I would anticipate prescription of Zuranolone in NHS as a concomitant treatment to ECM
Key issue 3	
28. The SKYLARK and ROBIN trials excluded people with mental health comorbidities and who had attempted	A large proportion of patients with PND who access secondary care have comorbidities and a serious risk profile. Excluding these patients would limit

suicide. What proportion of the eligible PND population in the NHS would not have met the trials' inclusion criteria and what impact do you think this has on the generalisability of the clinical trial results? (relevant section: pages 34-36 EAG report)	strongly the generalisability of the results, considering that Zuranolone would be prescribed by perinatal psychiatrists.
29. Are the types of ADTs used and proportion of people having ADTs at baseline reflective of UK NHS practice? (also key issue 5; relevant section: pages 34-36 and 64-65 EAG report)	In primary care it is more usual that first line antidepressant i.e. SSRIs be prescribed. In secondary care, SNRIs are also prescribed, and treatment is often augmented with other agents i.e. Antipsychotics or mood stabilisers.
30. What proportion of people with PND seen by perinatal psychiatrists would be considered non-responders to ADTs prescribed before referral to secondary care? (relevant section: pages 34-36 EAG report)	It is difficult to give a numeric estimate. PND seen in secondary care is often concomitant with other psychiatric diagnoses or longstanding psychological issues possibly stemming from trauma or significant social concerns. ADTs may contribute to the treatment plan but their effectiveness is limited by clinical complexity. In average about one third of patient have a significant response with ADT - significant meaning that their quality of life is no longer affected by their symptoms, a third will have a limited effect with symptoms still impacting their quality of life but with a meaningful and marked improvement, and another third of the patients will not have noticed a significant change to their symptoms.
Key issue 4	
31. How long would you expect the treatment effect from a 14-day treatment course of zuranolone to last? (relevant section: page 66-70 EAG report)	There is limited evidence on long term efficacy of Zuranolone at present and it is currently not possible to estimate it.
32. Do you anticipate the treatment benefit from zuranolone to decrease over time? Several treatment waning scenarios are described on pages 67 to 70 of the EAG report. Which of these seems most plausible? (relevant section: pages 66-70 EAG report)	To my knowledge there isn't enough evidence to support a sustained improvement. In my opinion based on experience, perpetuating factors such as social difficulties, unresolved trauma, childhood adversity, personality disorder, comorbid psychiatric diagnoses will have an impact on long term outcomes.
Key issue 6	
33. Does a change in EPDS score indicate a change in PND clinical diagnosis (e.g., mild, moderate or severe PND) or a change in subsequent risk of depression (e.g.,	A change in the EPDS score is a strong indicator of change in severity of illness.

major depressive disorder [MDD] or chronic MDD)? (relevant section: page 71 EAG report)	
34. How does EPDS score relate to a clinical diagnosis of mild, moderate or severe depression? If unknown, please estimate (relevant section: page 71 EAG report).	Scores above 10 would indicate presence of depression. Scores above 13 indicate a high chance of depression. Severe depression usually falls in EPDS scores above 20. However, with lower scores but positive answers to questions concerning suicidality, depression may be considered severe. NB this is an estimate.
Key issue 7	
35. The EAG's clinical advisors suggest that PND would last for one year after birth. After that, PND and MDD symptoms are similar and PND would be treated as MDD. Is this assumption accurate and would PND be treated as MDD 1 year after birth? (relevant section: page 74 EAG report)	PND is not always diagnosed right away and may persist beyond the first postnatal year and indeed further. Hence the extension of perinatal secondary care to 24 months. PND may maintain perinatal specific characteristics i.e. notions of incompetence as a mother, estrangement from the infant, morbid anxieties about the health of the infant, difficulties in bonding with the infant, thoughts of suicide by violent means. As such specialist care may be needed beyond the first postnatal year
36. Would a patient with unresolved symptoms of severe postnatal depression after 18 months be treated in primary or secondary care? (relevant section: page 74 EAG report)	Most perinatal services are expanding to 24 months as part of the NHS Long Term plan. Some services will accept referrals beyond 12 months postnatal, and most services will keep treating patients at 18 months.
Resource use and costs applied in the model	
37. Would the population taking zuranolone be on an antidepressant treatment 18 months post birth? (relevant section: pages 76-77 EAG report)	It is not unusual that patients may need chronic treatment with ADs. Some patients with PND have MDD and PND was one episode of illness rather than a new onset disorder. these patients will most often be seen in perinatal psychiatric services, and may continue their treatment beyond the perinatal period.
38. Is Byford 2011 reflective of NHS postnatal depression care and population 18 months after birth? (relevant section: pages 76-77 EAG report)	I do not believe the Byford 2011 study can be generalised to the perinatal population, as some patients will go into full remission before 12 months postnatal, but others will continue to experience symptoms and require both antidepressants and psychological treatment/psychotherapy beyond 18 months. The perinatal population looked after in secondary care is complex and will not likely follow a recovery model that is linear.
Key issue 8	

39. Are there any benefits of zuranolone not accounted for in the model? (relevant section: pages 94-95 EAG report)	I am not aware of any further benefits not accounted for in the EAG report.

Part 2: Key messages

In up to 5 sentences, please summarise the key messages of your statement:

Limited evidence on long term effectiveness of Zuranolone

More evidence is required to establish long term effectiveness

The population with PND looked after in secondary care is often complex and will require a multifactorial care plan, not limited to medication

Zuranolone should only be prescribed by a specialist perinatal psychiatrist

Prescription of Zuranolone will often be concomitant with other treatments e.g. other antidepressants, antipsychotics, mood stabilisers, ECT, psychological treatment, psychotherapy, social interventions

Thank you for your time.

Your privacy

The information that you provide on this form will be used to contact you about the topic above.

☐ **Please tick this box** if you would like to receive information about other NICE topics.

For more information about how we process your personal data please see our privacy notice.

NICE National Institute for Health and Care Excellence

Single Technology Appraisal

Zuranolone for treating postnatal depression [ID6431]

Clinical expert statement

Information on completing this form

In part 1 we are asking for your views on this technology. The text boxes will expand as you type.

In part 2 we are asking you to provide 5 summary sentences on the main points contained in this document.

Please do not embed documents (such as a PDF) in a submission because this may lead to the information being mislaid or make the submission unreadable. Please type information directly into the form.

Do not include medical information about yourself or another person that could identify you or the other person.

We are committed to meeting the requirements of copyright legislation. If you want to include **journal articles** in your submission you must have copyright clearance for these articles. We can accept journal articles in NICE Docs. For copyright reasons, we will have to return forms that have attachments without reading them. You can resubmit your form without attachments, but it must be sent by the deadline.

Combine all comments from your organisation (if applicable) into 1 response. We cannot accept more than 1 set of comments from each organisation.

Please underline all confidential information, and separately highlight information that is submitted as 'confidential [CON]' in turquoise, and all information submitted as 'depersonalised data [DPD]' in pink. If confidential information is submitted, please also

send a second version of your comments with that information redacted. See Health technology evaluations: interim methods and process guide for the proportionate approach to technology appraisals (section 3.2) for more information.

The deadline for your response is **5pm** on **<insert deadline>.** Please log in to your NICE Docs account to upload your completed form, as a Word document (not a PDF).

Thank you for your time.

We reserve the right to summarise and edit comments received, or not to publish them at all, if we consider the comments are too long, or publication would be unlawful or otherwise inappropriate.

Comments received are published in the interests of openness and transparency, and to promote understanding of how recommendations are developed. The comments are published as a record of the comments we received, and are not endorsed by NICE, its officers or advisory committees.

NICE National Institute for Health and Care Excellence

Part 1: Treating postnatal depression and current treatment options

Table 1 About you, aim of treatment, place and use of technology, sources of evidence and equality

1. Your name	Heather O'Mahen
2. Name of organisation	NHS England/University of Exeter
3. Job title or position	
4. Are you (please tick all that apply)	☒ An employee or representative of a healthcare professional organisation that represents clinicians? ☒ A specialist in the treatment of people with postnatal depression? ☒ A specialist in the clinical evidence base for postnatal depression or technology? ☐ Other (please specify):
5. Please disclose any past or current, direct or indirect links to, or funding from, the tobacco industry.	No links to the tobacco industry
6. What is the main aim of treatment for postnatal depression? (For example, to stop progression, to improve mobility, to cure the condition, or prevent progression or disability)	1. To reduce distress/suffering associated with persistent, very low mood and/or anhedonia 2. To improve functioning.
7. What do you consider a clinically significant treatment response? (For example, a reduction in tumour size by x cm, or a reduction in disease activity by a certain amount)	Minimally Clinically Important Difference for Patient Health Questionnaire-9 – 1.7-5 "Clinically meaningful" change: Moderately severe: a MCID of 20% (e.g., 1.7 on the Patient Health Questionnaire-9)

	Chronic, “treatment resistant” depression: 30-32%. (Löwe, B., Unützer, J., Callahan, C. M., Perkins, A. J., & Kroenke, K. (2004). Monitoring depression treatment outcomes with the patient health questionnaire-9. *Medical care*, *42*(12), 1194-1201. Button, K. S., Kounali, D., Thomas, L., Wiles, N. J., Peters, T. J., Welton, N. J., ... & Lewis, G. (2015). Minimal clinically important difference on the Beck Depression Inventory-II according to the patient's perspective. *Psychological medicine*, *45*(15), 3269-3279. Kounali, D., Button, K. S., Lewis, G., Gilbody, S., Kessler, D., Araya, R., ... & Lewis, G. (2022). How much change is enough? Evidence from a longitudinal study on depression in UK primary care. *Psychological Medicine*, *52*(10), 1875-1882.)
8. In your view, is there an unmet need for patients and healthcare professionals in postnatal depression?	Yes. Current NICE approved treatments successfully treat about 50% of individuals. “Treatment resistance” is as yet unresolved, and more chronic, relapsing forms of postnatal depression have a stronger relationship with prospective negative child outcomes. Stein, A., Pearson, R. M., Goodman, S. H., Rapa, E., Rahman, A., McCallum, M., ... & Pariante, C. M. (2014). Effects of perinatal mental disorders on the fetus and child. *The Lancet*, *384*(9956), 1800-1819.
9. How is postnatal depression currently treated in the NHS? • Are any clinical guidelines used in the treatment of the condition, and if so, which?	• 2014 NICE Antenatal and Postnatal guidelines recommend either antidepressants or Cognitive Behavioural Therapy/Interpersonal Psychotherapy.

• Is the pathway of care well defined? Does it vary or are there differences of opinion between professionals across the NHS? (Please state if your experience is from outside England.) • What impact would the technology have on the current pathway of care?	• Yes, the pathway is reasonably well defined (though now that perinatal mental health services have expanded would benefit from updating). Yes, there are differences of opinion on care, related in part to focus on the causes of the problem. At the moderately severe to severe end psychiatry may focus on depressed mood and functioning and biological /medical treatment; psychological providers may focus on psychosocial causes, particularly history of trauma (of which in chronic/mod severe to severe there are high rates of trauma). Agreement – combined approaches may be particularly helpful. Fu, C., Li, C., Wan, X., Yang, Y., Zhang, S., & Hu, J. (2024). The relationship between adverse childhood experiences and postpartum depression: a systematic review and meta-analysis. *Trauma, Violence, & Abuse*, *25*(4), 3066-3081. Oh, W., Muzik, M., McGinnis, E. W., Hamilton, L., Menke, R. A., & Rosenblum, K. L. (2016). Comorbid trajectories of postpartum depression and PTSD among mothers with childhood trauma history: Course, predictors, processes and child adjustment. *Journal of affective disorders*, *200*, 133-141. Wajid A, van Zanten SV, Mughal MK, Biringer A, Austin MP, Vermeyden L, Kingston D (2019) Adversity in childhood and depression in pregnancy Archives of women's mental health doi: 10.1007/s00737-019-00966-4 • Zuralonone requires initial administration under supervision, as there are sedating effects. This is a significant difference from other forms of treatment for PND, and could be particularly problematic for mothers of new babies – especially when there is little other family support and when there are other children living at home. These individuals may require more intensive supervision/observation (i.e., in hospital setting) and this would significantly impact costs/feasibility.

	• Evidence so far is for women with severe PND. The technology would therefore require that a more careful and thorough diagnostic assessment of women and birthing persons be conducted, and that this assessment would specifically and robustly assess for the severity, chronicity and treatment response history of the depression. Currently, diagnoses are rarely recorded in care, and it is unclear how specifically severity, chronicity and treatment response history are assessed. This would need to be standardised in a new care pathway.
10. Will the technology be used (or is it already used) in the same way as current care in NHS clinical practice? • How does healthcare resource use differ between the technology and current care? • In what clinical setting should the technology be used? (for example, primary or secondary care, specialist clinic) • What investment is needed to introduce the technology? (for example, for facilities, equipment, or training)	As above. • Need to standaridise assessment of severe depression (and this would be a much smaller proportion of women than the broader PND population). • Need to understand the numbers of women with severe PND who would require more intensive professional supervision (e.g., MBU admission for start of treatment).
11. Do you expect the technology to provide clinically meaningful benefits compared with current care? • Do you expect the technology to increase length of life more than current care? • Do you expect the technology to increase health-related quality of life more than current care?	• Possibly. Perhaps the most promising aspect of Zuralolone is its speed of action. Suicide is the leading cause of maternal deaths from 6 weeks to the first year postnatally and is frequently driven by hopelessness and entrapment. A fast-acting treatment that can make a measurable change to mood provides the greatest opportunity for improving hopelessness and entrapment and saving lives that might otherwise be lost to suicide. It may be particularly appropriate for these individuals in terms of safety of administration as acutely suicidal individuals often require hospitalisation. • Possibly. The existing trial data shows promise, but it is critical to have more robust data, particularly in comparison to treatment-as-usual as offered in the UK. It should be noted that although the initial smaller trial showed very

	positive results, larger subsequent trials had mixed/equivocal results and that there are significant side sedating side effects associated with the technology which could pose risks to the baby and other children living in the household. Further, the medication may be inappropriate for women who are committed to breastfeeding, as they are required to discontinue breastfeeding to take the medicine. The fast acting nature of the technology, however, could promote quicker recovery of functioning and have positive effects for the larger family. Longer-term data is strongly needed.
12. Are there any groups of people for whom the technology would be more or less effective (or appropriate) than the general population?	Yes. More effective: The evidence is for people with severe depression and those who are acutely suicidal. It may also be appropriate for people with chornic/treatment resistant depression. Less effective: Mothers with severe/chronic depression who are committed to breastfeeding, those without family/close support who may find it challenges to manage parenting tasks under the sedating effects of the medication, mothers with less than severe PND.
13. Will the technology be easier or more difficult to use for patients or healthcare professionals than current care? Are there any practical implications for its use? (For example, any concomitant treatments needed, additional clinical requirements, factors affecting patient acceptability or ease of use or additional tests or monitoring needed)	Both easier and more difficult. Easier – fast-acting nature More difficult – requires close oversight. Sedating. More costly. Requires discontinuation of breastfeeding.

14. Will any rules (informal or formal) be used to start or stop treatment with the technology? Do these include any additional testing?	As above: severity/chronicity/suicidality/history of response to other treatments (to start). Stopping: response to treatment (clinically significant change), side effects
15. Do you consider that the use of the technology will result in any substantial health-related benefits that are unlikely to be included in the quality-adjusted life year (QALY) calculation? • Do the instruments that measure quality of life fully capture all the benefits of the technology or have some been missed? For example, the treatment regimen may be more easily administered (such as an oral tablet or home treatment) than current standard of care	Benefits on social functioning and impact on family unlikely to be adequately captured.
16. Do you consider the technology to be innovative in its potential to make a significant and substantial impact on health-related benefits and how might it improve the way that current need is met? • Is the technology a 'step-change' in the management of the condition? • Does the use of the technology address any particular unmet need of the patient population?	It is a step-change in terms of speed of effect. Later and larger scale trials do not show that it is definitively superior and there is evidence lacking on the longer-term benefits of the technology (beyond 4 weeks). A number of mothers will not be able to take it. It is costly.
17. How do any side effects or adverse effects of the technology affect the management of the condition and the patient's quality of life?	Sedation – potentially dangerous for parents of infants and other children in terms of risk of neglect, accidental smothering of infants if parent falls asleep with them. Can't breastfeed Cannot get pregnant again while on the medication (this may be particularly difficult for mothers who have experienced stillbirth/infant death and would like to get pregnant again)
18. Do the clinical trials on the technology reflect current UK clinical practice?	No.

• If not, how could the results be extrapolated to the UK setting? • What, in your view, are the most important outcomes, and were they measured in the trials? • If surrogate outcome measures were used, do they adequately predict long-term clinical outcomes? • Are there any adverse effects that were not apparent in clinical trials but have come to light subsequently?	• They should be compared to usual care in the UK which is now of an internationally exemplary standard. • Speed of effect on low mood/anhedonia and functioning most important outcomes. They were measured, but with equivocal results in the larger/later trials. • No. It is critical to do the work to get long-term clinical outcomes. Depression is a problem that has a fluctuating course, even with severe depression. • Not that I am aware of.
19. Are you aware of any relevant evidence that might not be found by a systematic review of the trial evidence?	No
20. How do data on real-world experience compare with the trial data?	I cannot comment on this.
21. NICE considers whether there are any equalities issues at each stage of an evaluation. Are there any potential equality issues that should be taken into account when considering this condition and this treatment? Please explain if you think any groups of people with this condition are particularly disadvantaged. Equality legislation includes people of a particular age, disability, gender reassignment, marriage and civil partnership, pregnancy and maternity, race, religion or belief, sex, and sexual orientation or people with any other shared characteristics. Please state if you think this evaluation could • exclude any people for which this treatment is or will be licensed but who are protected by the equality legislation	Yes. There are (a) significant inequalities in rates of negative maternity outcomes for women from Black/Afro-Caribbean and S. Asian ethnicities, and they are more likely to have traumatic maternity experiences. These place them at increased risk for mental health problems, but (b) they are less likely to be identified and receive care. It is not clear if the data from the original study reflects outcomes for these women. Women who have low incomes and those with significant social complexities are considerably more likely to have severe and recurrent forms of depression, and therefore might be more appropriate for the technology – but they also are the least likely to have the kinds of social support that could arrange to be consistently present and supervising them while taking the technology. They may be especially likely to need external supervision.

• lead to recommendations that have a different impact on people protected by the equality legislation than on the wider population • lead to recommendations that have an adverse impact on disabled people. Please consider whether these issues are different from issues with current care and why. More information on how NICE deals with equalities issues can be found in the NICE equality scheme. Find more general information about the Equality Act and equalities issues here.	
Key issue 1	
22. How is mild, moderate and severe postnatal depression (PND) diagnosed in clinical practice? What are the key differences between each classification? (relevant section: page 23 EAG report)	This is a significant issue. The best way to do this is through a standardised semi-structured diagnostic clinician administered clinical interview, using the interview cut-offs for mild, moderate, severe depression. At present point, the most likely professionals trained to deliver these interviews are psychiatrists, though in some cases clinical psychologists and other professionals may be trained to administer them. The second means is through standardised assessments using pre-determined cut-offs, but these are typically not intended to take the place of semi-structured diagnostic interviews.
23. Baseline PND severity was recorded in the trial by multiple depression measurement tools. Which results would you expect to better reflect the UK PND population - HAMD-17, MASRS or EPDS? (relevant section: pages 42-44 EAG report)	The HAMD-17 and EPDS have both been used widely in research studies in the UK and studies of cut-offs with these measures have been comparable in UK perinatal and other populations in the U.S. to those in N. America and Europe.
24. What population would be most likely to receive zuranolone in the NHS – people with mild, moderate or severe PND?	The evidence base is for people with severe PND.

	Note on page 40 of the document prepared in evidence for zuranolone – it states (And is highlighted) that "changes were in favour of zuranolone treatment irrespective of PND severity." (in the Skylark study). Given that the study inclusion criteria was that individuals had to have *severe PND* I can only assume that they meant that differences in severity were within those with PND, *not within the full symptom range in PND.*
Key issue 2	
25. Does the scope accurately represent established clinical management (ECM) for PND in the NHS? (relevant section: pages 24-25 EAG report)	Clinical management in most of the trials was not comparable to what is offered in the UK.
26. Do the concomitant treatments permitted in the SKYLARK and ROBIN accurately represent ECM for PND? If not, how do you expect that this will impact the generalisability of the clinical trial results? (relevant section: pages 37-38 EAG report)	As above – I would expect that the results for the UK are largely not generalisable.
27. Do you anticipate zuranolone being prescribed in the NHS with or without established clinical management? (relevant section: pages 37-38 EAG report)	Zuranolone should always be prescribed with clinical management – and it is notable to see that a possible recommendation is that it should be administered in the context of admission to an MBU or other suitable inpatient facility – would significantly change the impact on the mother and family and increase costs considerably.
Key issue 3	
28. The SKYLARK and ROBIN trials excluded people with mental health comorbidities and who had attempted suicide. What proportion of the eligible PND population in the NHS would not have met the trials' inclusion criteria and what impact do you think this has on the generalisability of the clinical trial results? (relevant section: pages 34-36 EAG report)	At the point of severe depression it would be highly unusual to not have mental health comorbidities. It therefore strongly affects the generalisability of the clinical trial results.
29. Are the types of ADTs used and proportion of people having ADTs at baseline reflective of UK NHS practice?	I am not able to comment on this specific question.

(also key issue 5; relevant section: pages 34-36 and 64-65 EAG report)	
30. What proportion of people with PND seen by perinatal psychiatrists would be considered non-responders to ADTs prescribed before referral to secondary care? (relevant section: pages 34-36 EAG report)	This is challenging to answer for PND specifically as most individuals with PND will be managed by psychiatrists at least at some point and this would require a referral to secondary care (because in the UK this is where psychiatrists sit). ADTs outside the perinatal period may be managed by GPs without psychiatric input, but many GPs remain wary of prescribing ADTs within the perinatal period and therefore request specialist perinatal psychiatric input. The question of how many individuals would be considered non-responders should likely to be extrapolated from non-postnatal populations.
Key issue 4	
31. How long would you expect the treatment effect from a 14-day treatment course of zuranolone to last? (relevant section: page 66-70 EAG report)	So far it appears that 20-30 day follow-up is the only data we've got. It "should" last at least 6 months, and to be equivalent to psychological treatments should last 12 months.
32. Do you anticipate the treatment benefit from zuranolone to decrease over time? Several treatment waning scenarios are described on pages 67 to 70 of the EAG report. Which of these seems most plausible? (relevant section: pages 66-70 EAG report)	This is concerning and without appropriate long-term data this is exactly the risk. This is also consistent with other ADTs where the recommendation is to take the medication for at least 6 months.
Key issue 6	
33. Does a change in EPDS score indicate a change in PND clinical diagnosis (e.g., mild, moderate or severe PND) or a change in subsequent risk of depression (e.g., major depressive disorder [MDD] or chronic MDD)? (relevant section: page 71 EAG report)	The EPDS assesses symptom severity and is a dimensional measure. There have been a number of studies that have tried to determine the best cut-off point for the EPDS that accurately detects risk of having diagnosable Major Depressive Disorder. The best cut-off numbers are 12 or 13. Numbers below this are likely to detect low levels of sad mood. 8-11 are likely persons that may be at risk of MDD. The EPDS is not typically used to detect severity within individuals who have MDD (e.g., are above the cut-off of 12/13). The PHQ-9, however, does have cut-offs representing severity levels for those who have MDD.

34. How does EPDS score relate to a clinical diagnosis of mild, moderate or severe depression? If unknown, please estimate (relevant section: page 71 EAG report).	As above. It does not. The PHQ-9 does. It should be noted that *neither* measure is intended to be diagnostic. Clinical diagnostic semi-structured interviews are the best methods to determine diagnosis.
Key issue 7	
35. The EAG's clinical advisors suggest that PND would last for one year after birth. After that, PND and MDD symptoms are similar and PND would be treated as MDD. Is this assumption accurate and would PND be treated as MDD 1 year after birth? (relevant section: page 74 EAG report)	Internationally there is debate about this. In some diagnostic manuals – e.g., the DSM-5, PND is defined as persisting within 6 weeks following childbirth. In the research literature, PND is typically defined as Major Depression persisting up to 1 year postnatal. Although some (including the documents submitted here) suggest that PND is a unique "disorder" there is scant evidence of this. To diagnose PND, there is no distinction between MDD symptoms and PND. They are the same. Rates of diagnosed depression do not vary within and outside the perinatal period in similarly aged women women. There is scant evidence that hormonal changes following pregnancy are a **direct** causal precipitant of PND (it may be a moderating factor). Most cases of PND are not *de novo.* Many women were depressed in pregnancy, or prior to pregnancy. It is the case that those who have bipolar I, (there is evidence from Ian Jones' and other's genetics research that postpartum psychosis is strongly genetically related to bipolar I) and who relapse postnatally are most likely to have a *depressive relapse* rather than a psychotic relapse. This type of depression is different from PND (and much rarer – bipolar I is prevalent in about 1% of the perinatal population). To answer the question: Yes, PND would be treated as MDD 1 year after birth. PND would be also be treated as MDD in the first postnatal year (with perinatal adaptations).

36. Would a patient with unresolved symptoms of severe postnatal depression after 18 months be treated in primary or secondary care? (relevant section: page 74 EAG report)	If there are no other complicating factors (unlikely) then it is possible that such an individual could be treated in primary care (NHS Talking Therapies does treat individuals with severe depression, and GPs do prescribe antidepressants). However, if the individual had received treatment and not responded it would most likely be the case that they would be referred to secondary care – and could still be seen in perinatal specialist services in most regions of the country.
Resource use and costs applied in the model	
37. Would the population taking zuranolone be on an antidepressant treatment 18 months post birth? (relevant section: pages 76-77 EAG report)	Potentially.
38. Is Byford 2011 reflective of NHS postnatal depression care and population 18 months after birth? (relevant section: pages 76-77 EAG report)	Postnatal depression care in the UK has changed considerably since 2015, including expanding perinatal specialist services to 24 months postnatally.
Key issue 8	
39. Are there any benefits of zuranolone not accounted for in the model? (relevant section: pages 94-95 EAG report)	I think they accurately cover them.

NICE National Institute for Health and Care Excellence

Part 2: Key messages

In up to 5 sentences, please summarise the key messages of your statement:

Click or tap here to enter text.

Click or tap here to enter text.

Click or tap here to enter text.

Click or tap here to enter text.

Click or tap here to enter text.

Thank you for your time.

Your privacy

The information that you provide on this form will be used to contact you about the topic above.

☒ **Please tick this box** if you would like to receive information about other NICE topics.

For more information about how we process your personal data please see our privacy notice.

Zuranolone for treating postnatal depression (ID6431): A single Technology Appraisal

Produced by: Bristol Technology Assessment Group, University of Bristol

Authors: Mary Ward, Senior Research Associate, Bristol TAG, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol

Eve Tomlinson, Senior Research Associate in Evidence Synthesis, Bristol TAG, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol

Catalina Lopez-Manzano, Senior Research Associate in Evidence Synthesis, Bristol TAG, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol

Elsa MR Marques, Associate Professor in Health Economics, Bristol TAG, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol

Jelena Savović, Associate Professor in Epidemiology and Evidence Synthesis, Bristol TAG, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol

Deborah M Caldwell, Professor in Epidemiology and Public Health, Bristol TAG, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol

Chris Cooper, Senior Research Fellow in Health Technology Assessment and Information Science, Bristol TAG, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol

Dr Karl Scheeres, Consultant Perinatal Psychiatrist, New Horizons Mother and Baby Unit, Southmead Hospital, Avon and Wiltshire Partnership NHS Trust

Nicky Pedwell, Mental Health Midwife, University Hospitals Bristol and Weston NHS Foundation Trust

Nicky J. Welton, Professor in Statistical and Health Economic Modelling, Bristol TAG, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol

Correspondence to: Nicky J. Welton, Professor in Statistical and Health Economic Modelling, Bristol TAG, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol

Date completed: Date completed 22/05/2025

Source of funding: This report was commissioned by the NIHR Evidence Synthesis Programme as project number NIHR175326.

Declared competing interests of the authors:
None of the authors has any conflicts of interest to declare. We note that the company worked with researchers at the University of Bristol on an analysis of Avon Longitudinal Study of Parents and Children (ALSPAC) data used in their submission. However, the University of Bristol did not receive payment for this work, and the authors of this report were unaware of this work and not involved in it in any way.

Acknowledgements
We would like to thank Dr Jonathan Evans (Senior Lecturer, Centre for Academic Mental Health, University of Bristol) for clinical advice relating to this project. We would also like to thank Nicola Horler, Bristol TAG, for providing administrative support.

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

This report should be referenced as follows:
Ward M, Tomlinson E, Lopez-Manzano C, Marques EMR, Savović J, Caldwell DM, Cooper C, Scheeres K, Pedwell N, Welton NJ. Zuranolone for treating severe postnatal depression (ID6431): A single Technology Appraisal. Bristol Technology Assessment Group, 2025.

Contributions of authors
Tomlinson, Lopez-Manzano, Caldwell and Savović, summarised and critiqued the clinical effectiveness data reported within the company's submission. Lopez-Manzano, Caldwell and Welton critiqued the statistical aspects of the submission. Cooper critiqued the company's search strategy. Welton, Marques, and Ward critiqued the health economic analysis submitted by the company and conducted additional scenario analyses. Scheeres and Pedwell provided clinical advice. All authors were involved in drafting and commenting on the final report.

[redacted]

Abbreviations

Abbreviation	Definition
ADT	Anti-depressant therapy
AEs	Adverse Events
ALSPAC	Avon Longitudinal Study of Parents and Children
BMI	Body Mass Index
CGI-I	Clinical Global Impression - Improvement
CGI-S	Clinical Global Impression - Severity
CI	Confidence Interval
CS	Company Submission
CSR	Clinical Study Report
C-SSRS	Colombia Suicide Severity Rating Scale
CYP	Cytochrome P450
DSA	Deterministic Sensitivity Analysis
DSM-5	Diagnostic and Statistical Manual of Mental Disorders 5
EAG	Evidence Assessment Group
ECG	Echocardiogram
ECM	Established Clinical Management
EMA	European Medicines Agency
eMIT	electronic Market Information Tool
EPDS	Edinburgh Postnatal Depression Scale
EQ-5D-3L	EuroQol 5 dimensions 3 level questionnaire
FDA	Food and Drug Administration
GABAA	Gamma Amino Butyric Acid Type A
GBP	Great British Pound
HAM-A	Hamilton Rating Scale for Anxiety
HAM-D	Hamilton Rating Scale for Depression
HAMD-17	17-item Hamilton Rating Scale for Depression
HRQoL	Health-Related Quality of Life
HRU	Healthcare Resource Utilisation
ICER	Incremental Cost Effectiveness Ratio
ITC	Indirect Treatment Comparison
LOCF	Last Observation Carried Forward
LSM	Least-Square Means
MADRS	Montgomery-Åsberg Depression Rating Scale
MAIC	Matched Adjusted Indirect Comparison
MDD	Major Depressive Disorder
MMRM	Mixed Model for Repeated Measures
N/A	not applicable
NHS	National Health Service
NICE	National Institute for Health and Care Excellence

PHQ-9	Patient Health Questionnaire 9
PND	post-natal depression
PSA	Probabilistic Sensitivity Analysis
PSS	Personal Social Services
QALY	Quality-Adjusted Life Year
RCT	Randomised Controlled Trial
RoB2	Risk of Bias 2 tool
ROBIS	Risk Of Bias In Systematic reviews
SAE	Serious Adverse Event
SD	standard deviation
SE	Standard error
SF-36	36-item Short Form Health Survey
SLR	Systematic Literature Review
SMR	Standardised Mortality Rate
SSRI	Selective Serotonin Reuptake Inhibitor
STA	Single Technology Appraisal
TEAE	Treatment-emergent Adverse Event
TECH-VER	TECH-VER Model Verification Checklist

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 EXECUTIVE SUMMARY

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

- Section 1.1 provides an overview of the key issues.
- Section 1.2 provides an overview of key model outcomes and the modelling assumptions that have the greatest effect on the ICER.
- Sections 1.3 to 1.6 explain the key issues in more detail.

Background information on the condition, technology and evidence and information on non-key issues are in the main EAG report.

1.1 Overview of the EAG's key issues

All issues identified represent the EAG's view, not the opinion of the National Institute of Health and Care Excellence (NICE). Table 1 provides an overview of the EAG's key issues:

Table 1 Summary of key issues

ID 6431	Summary of issue	Report sections
Key Issue 1	The most appropriate population for treatment with zuranolone. The scope does not restrict by severity of post-natal depression (PND), however the trial evidence is mostly from a population of severe PND, and the EAG heard zuranolone is likely to be used in severe patients in practice.	Sections 2.2, 3.2.1.2 and 4.2.3
Key Issue 2	Definition of Established Clinical Management (ECM) in the company's two studies and NHS practice. In the SKYLARK trial only 15% of participants were on anti-depressant therapy (ADT), whereas the EAG heard that the majority of severe PND patients would be on ADT at the time of referral. Aditionally, patients were not permitted to change treatment during the trial, which also may not be reflective of NHS ECM.	Section 3.2.1.3.1
Key Issue 3	Applicability of the SKYLARK trial to NHS practice – differences between the trial	Section 3.2.1.2

	and the target population. There are differences between the trial population and the population where zuranolone would be used in NHS practice in mental health comorbidities and baseline antidepressant use.	
Key Issue 4	Extrapolation of treatment effect. The company assume that the treatment effect predicted at 8 months post-birth (based on 1.5 months of follow-up data) continue until 29 years post-birth	Section 4.2.6
Key Issue 5	Concomitant anti-depressant therapy (ADT) use. The company assume 15.3% concomitant ADT use, whereas the EAG heard this would be much higher in NHS practice for severe PND patients.	Section 4.2.4 and 3.2.1.2
Key Issue 6	Short-term evidence sources inform long-term costs and effects in the model.	Sections 4.2.6.3, 4.2.7.5, and 4.2.8.2
Key Issue 7	Choice of tool and validity of utility estimates.	Section 4.2.7
Key Issue 8	Uncaptured costs and benefits. The impact of reductions in Edinburgh Postnatal Depression Scale (EPDS) score on children and partner are not included in model. Zuranolone cannot be used in patients who are breast-feeding, and may have other uncaptured longer-term side effects.	Section 7

The key differences between the company's preferred assumptions and the EAG's preferred assumptions are:

- the company assumes the treatment effect for zuranolone persists for 29 years, whereas the EAG assumes that the effect begins to wane at 1-year post-treatment initiation.
- the company assumes 15.3% ADT use, whereas the EAG assumes 90% ADT use.
- the company models a mixed severity PND population, whereas the EAG assumes all patients have severe PND.
- the company use utilities for PND based on SF6D, and assume these are relevant over a 29 year time period, whereas the EAG prefer to use EuroQol 5 dimensions (EQ-5D) utilities, and use PND utilities in the short-term following birth and Major Depressive Disorder (MDD) utilities in the longer term

1.2 Overview of key model outcomes

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

Overall, the technology is modelled to affect QALYs by:

- Reducing severity of depression in the short-term
- Maintaining a reduced severity of depression long-term

Overall, the technology is modelled to affect costs by:

- Increased treatment acquisition cost to add zuranolone to Established Clinical Management (ECM)
- Reduced healthcare resource use due to a higher proportion of patients being in remission from post-natal depression (PND)

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

- The extrapolation of treatment effect over the 29 year time-horizon
- The way that health-state utilities are modelled in the long-term
- Proportion of patients with baseline ADT use (when treatment effect waning is applied)
- Severity of patient population (when treatment effect waning is applied)

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

Key Issue 1 The most appropriate population for treatment with zuranolone

Report section	Sections 2.2, 3.2.1.2 and 4.2.3
Description of issue and why the EAG has identified it as important	The scope does not restrict by severity of PND, however the randomised controlled trial (RCT) evidence is from a population that was mostly severe PND with some upper-range moderate PND. Furthermore the EAG heard from its clinical advisers that zuranolone would likely be used in patients with severe PND referred to a perinatal psychiatrist (secondary care). The company's model is described as for a PND population, without restricting by severity in line with the scope. However, the treatment effect estimates are based on results from the SKYLARK trial, which was mostly a severe PND population with some upper-range moderate PND, with the long-term model based on Avon Longitudinal Study of Parents and Children (ALSPAC) data from a moderate to severe population. It is unknown how

	the clinical or cost-effectiveness results would change if the model were based on a population of patients with mild or moderate PND.
What alternative approach has the EAG suggested?	The EAG prefers to restrict to a severe PND population to match the expected use in the NHS, and the majority of patients in the SKYLARK trial. The EAG requested a subgroup analysis by severity, but the company argued that this was not possible due to small numbers, and lack of long-term data by severity. The EAG has conducted scenario analyses which restricts the results to be based on those simulated patients classified as moderate at baseline (according to the EPDS state ≤19 in company's model) and for those severe at baseline (according to the EPDS state >19 in the company's model). However, note that this is not equivalent to a subgroup analysis by severity, as the EPDS short-term model does not include an interaction with baseline severity, and the long-term model is still based on a combined moderate to severe population.
What is the expected effect on the cost-effectiveness estimates?	Restricting to a more severe population increases the ICER slightly from £12,704 to £12,839, whereas restricting to a moderate population reduces the ICER to £12,289. Note however, that these results do not include any interaction between severity and treatment effect, and so the ICERs may differ in practice.
What additional evidence or analyses might help to resolve this key issue?	For the guidance to be relevant to the full population RCT evidence including mild and moderate PND patients would be required. An analysis of the ALSPAC dataset restricting to patients with severe PND would provide a better estimate of the long-term model. However, the results are unlikely to be sensitive to the long-term model.

Key Issue 2 Definition of Established Clinical Management (ECM) in the company's two studies and NHS practice.

Report section	3.2.1.3.1
Description of issue and why the EAG has identified it as important	The NICE scope specifies Established Clinical Management (ECM) as a comparator for the appraisal. The Company Submission (CS) reports evidence of clinical effectiveness for zuranolone plus ECM versus ECM, from two trials of zuranolone vs placebo (in which participants on stable anti-depressant therapy (ADT) or psychological therapies at baseline were eligible). Only baseline use of ADT is documented in the CS, with a low proportion of patients receiving ADT at baseline (15% in SKYLARK). It is not clear what ECM represents for the remaining 85% of participants. The EAGs clinical advisors noted that while referral practice varies across the UK, it is likely that a severe population would already be receiving ADT at the point of referral. Participants in SKYLARK were prevented from initiating new therapies (or changing the doses or regimens of their existing therapies) during the 45 day follow-up of the trial, which also may not reflect ECM in NHS practice. For the results of the company's trial to apply to ECM reflective of NHS practice, there would need to be no interaction between the zuranolone effect and components of ECM. This can only be explored in subgroup analyses for ADT, since details of other baseline therapies are not provided. The CS reports subgroup analyses for baseline ADT use. However, company studies were not powered to detect subgroup effects and results are imprecise (CS Figure 22, SKYLARK and CS Figure 23, ROBIN). No subgroup analyses are available for participants receiving psychological therapies, other ECM, or no additional treatments, at baseline from either trial. It is therefore uncertain whether the results for the ECM used in the company's trials generalise to that used in NHS practice.
What alternative approach has the EAG suggested?	In clarification question A3, the EAG requested subgroup information on psychological therapies received at baseline, but this was not captured as a

	standalone variable in either ROBIN or SKYLARK and so could not be provided by the company. In the economic model, scenario analyses were run for the proportion of baseline ADT use (see Key Issue 5), but this did not include interactions between zuranolone and baseline ADT use due to insufficient data.
What is the expected effect on the cost-effectiveness estimates?	It is unclear what the effect of using ECM that matched NHS practice would have on the cost-effectiveness estimates. However, there is some indication of an attenuation of effect for those with baseline ADT use and so may expect the ICER to increase.
What additional evidence or analyses might help to resolve this key issue?	The key additional data would be RCT data with a higher proportion of patients on ADT reflective of NHS practice. In the absence of this, additional information on the interaction between Selective Serotonin Reuptake Inhibitors (SSRIs) and zuranolone could be obtained from trials in MDD,[1] although the EAG acknowledges that MDD and PND are not the same condition.

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

Key Issue 3 Applicability of the SKYLARK trial to NHS practise – differences between the trial and the target population

Report section	3.2.1.2
Description of issue and why the EAG has identified it as important	EAG's clinical advisors noted that zuranolone will initially only be prescribed by perinatal psychiatrists, after being referred to secondary care. Referral practice varies across the UK, however it is likely that the severe population would already be receiving ADT at the point of referral - our clinical advisors reported that >90% patients they see in perinatal mental health services are already on ADT. As mentioned in Key Issue 2, only 15% of participants in SKYLARK were receiving ADT at baseline. Participants with other mental health comorbidities and higher risk patients were excluded from the trials, whereas these patients will be treated by the perinatal mental health services. The trial sample may therefore not fully represent those potentially considered for treatment with zuranolone in the UK.

What alternative approach has the EAG suggested?	Given the evidence available the EAG does not have any alternative suggestions. See Key Issue 5 for discussion of the proportion of patients taking ADT.
What is the expected effect on the cost-effectiveness estimates?	It is unclear what the impact of this would be on the cost-effectiveness results.
What additional evidence or analyses might help to resolve this key issue?	Additional evidence that could provide useful insight for this issue would be a pragmatic trial or a cohort study based in the UK that reflects the target population in the UK. However, we are not aware of such studies of zuranolone.

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

Key Issue 4 Extrapolation of treatment effect

Report section	Section 4.2.6
Description of issue and why the EAG has identified it as important	The company use a prediction model to extrapolate their 1.5 month follow-up data from SKYLARK a further 1.5 months (to 8 months post-birth), and then assume that the treatment difference predicted at 8 months post-birth continues until 29 years post-birth. The EAGs clinical advisors found this assumption to be implausible, especially since zuranolone treatment is a single course treatment for 14 days. The cost-effectiveness results are sensitive to assumptions about the extrapolation of the treatment effect.
What alternative approach has the EAG suggested?	The EAG explores various treatment waning assumptions in scenario analyses.
What is the expected effect on the cost-effectiveness estimates?	Including a waning of treatment effect has a large impact on the ICER. The ICER increases from £12,704 with no waning to £49,103, £89,131, £188,542, or £303,851 respectively if the EPDS curves start to wane at 3months post-treatment initiation and converge after 10, 5, 2, or 1 years respectively. If the EPDS curve starts to wane at 6 months post-treatment initiation and converge after a further 6 months the ICER is £299,728, and if the EPDS curve starts to wane at 1 year post-treatment initiation and converge after a further 1 year the ICER is £156,898.
What additional evidence or analyses might help to resolve this key issue?	This uncertainty would require longer-term follow-up data, ideally from the SKYLARK trial. Given that this data is unlikely to be available, future

	longitudinal data may provide insight on the long-term impact of zuranolone on long-term EPDS trajectories.

Key Issue 5 Concomitant anti-depressant therapy (ADT) use

Report section	Section 4.2.4 and 3.2.1.2
Description of issue and why the EAG has identified it as important	The company assumes that 15.3% of patients who would receive zuranolone would take concomitant ADT, based on the data from the SKYLARK trial. However, the EAG heard that most (>90%) patients would be already taking ADT before being referred to perinatal psychiatrists where zuranolone would be prescribed, and so the proportion of patients using ADT is likely to be much higher than 15.3%. The subgroup analyses for HAMD-17 at 15 days from the SKYLARK and ROBIN trials for baseline ADT use differed, with ROBIN indicating a benefit in both subgroups, whereas SKYLARK only showing a benefit in those without baseline ADT use. The EAG acknowledge that the studies were not powered to detect subgroup effects. However, the company said that they were unable to run a scenario analysis using the ADT subgroup from SKYLARK due to small numbers of patients. The impact of this on cost-effectiveness results is therefore unclear.
What alternative approach has the EAG suggested?	In response to clarification questions, the company ran a scenario setting 100% of patients on baseline ADT in the short-term prediction model for EPDS scores, which slightly increased the ICER. The EAG have further explored this with scenarios with 50%, 80%, and 90% on baseline ADT. Note however, that the impact of ADT use on treatment effects was not included. The company did not provide the subgroup results for EPDS score by baseline ADT use, and so the EAG were not able to run a subgroup analysis to explore the impact of this.
What is the expected effect on the cost-effectiveness estimates?	Increasing baseline ADT use from 15.3% to 50%, 80%, 90%, and 100% increases the ICER from £12,704 to £12,839, £13,337, £13,205, and £13,400 respectively. Note however that these results do not account for an interaction between ADT use and treatment effect.

What additional evidence or analyses might help to resolve this key issue?	Subgroup analyses for EPDS scores at 1.5 months follow-up by baseline ADT use. Results from SKYLARK and ROBIN could be pooled in a meta-analysis for the baseline ADT use subgroup to increase power.

Key Issue 6 Short-term evidence sources inform long-term costs and effects in the model

Report section	Sections 4.2.6.3, 4.2.7.5, and 4.2.8.2
Description of issue and why the EAG has identified it as important	The CS base case models long-term costs and effects over 5 health states defined by EPDS categories of 0, 1-6, 7-13, 14-18, 18-30. The EAG argues this classification is arbitrary and does not necessarily link to a clinical definition of severity of depression. It is not possible to determine from the company's submitted evidence by how many categories in the EPDS scale would PND score need to change in the acute period to avoid a chronic diagnosis of MDD in the longer-term. In the same way, utilities and long-term costs are informed by short-term studies (ROBIN SF36 questionnaires, Byford's 12 months resource use) and applied equally across the model for 29 years, which the EAG does not find plausible.
What alternative approach has the EAG suggested?	The issue of immaturity of data could only be resolved with longer follow-up of patients in the SKYLARK and ROBIN trials. The company could have provided scenarios based on long-term evidence from MDD studies for extrapolation of costs and utilities.
What is the expected effect on the cost-effectiveness estimates?	It is unclear what the impact on the ICER would be. If the differences in utilities and costs reduces over time, failing to account for this would overestimate the ICER in favour of zuranolone.
What additional evidence or analyses might help to resolve this key issue?	Providing additional scenarios based on MDD would have required a large systematic review of MDD evidence which was not feasible to undertake within the timeframe of this report.

Key Issue 7 Choice of tool and validity of utility estimates

Report section	Section 4.2.7
Description of issue and why the EAG has identified it as important	The CS informs utilities for the health states of the economic model using the SF6D, based on

	responses to the SF36. This is not the NICE recommended tool. The utility values between categories report a very large drop in utilities between categories, particularly for patients with no diagnoseable disease. The company pools estimates by categories for both groups and does not adjust for baseline imbalances, potential confounders, or missing data.
What alternative approach has the EAG suggested?	In clarification questions, the EAG suggested using EQ-5D utilities estimated by mapping PHQ-9 questionnaires to EQ-5D-3L utilities, by EPDS category, which the company provided in their response based on data from the ROBIN trial. In scenario analyses, the EAG suggested applying the same utility for EPDS scale of 0 as scale of 1-6, as they portray health states without disease and closer approximate to the SF6D population norms for patients 25-34 years of age. The EAG also present a scenario as above but additionally using utilities for MDD in the period beyond 12 months post-treatment initiation.
What is the expected effect on the cost-effectiveness estimates?	In sensitivity analyses provided by the company at clarification questions, using mapped EQ-5D utilities increased the ICER from £12,704 to £14,600. Using the EAG's preferred assumptions, including the EQ-5D mapped utilities for the PND period and EQ-5D utilities for MDD beyond the PND period, increased the ICER from to £12,704 to £22,939).
What additional evidence or analyses might help to resolve this key issue?	In the absence of EQ-5D data from the ROBIN and SKYLARK trials, the company could have also mapped utility values from responses to the SKYLARK PHQ-9 questionnaires and adjust these by baseline utility score, group allocation, and any other potential confounding factors, such as ADT therapy for example, and missing data.

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

Key Issue 8 Uncaptured costs and benefits

Report section	Section 7
Description of issue and why the EAG has identified it as important	The impact of reductions in EPDS score on children and partner are not included in model. Zuranolone may have other rare but serious longer-term side effects, such as decreased consciousness, and

	potential for abuse. However, the EAG notes these were not observed in the zuranolone studies. Zuranolone cannot be used in patients who are breast-feeding. Support for patients to either continue breastfeeding following treatment or stop breastfeeding, and dealing with possible side effects would be required in practice.
What alternative approach has the EAG suggested?	None.
What is the expected effect on the cost-effectiveness estimates?	It is expected that including the impact of treatment with zuranolone on the patients' children and partner would reduce the ICER. However, support for decisions around breast-feeding and dealing with uncaptured side effects would be expected to increase the costs of zuranolone treatment and the ICER. In response to clarification questions the company states that it is willing to work with stakeholders to "deliver effective support for patients without additional costs to the healthcare system."
What additional evidence or analyses might help to resolve this key issue?	The EAG is not aware of any evidence that could resolve this uncertainty. Longer term follow-up of ROBIN and SKYLARK trial patients may have picked up other potential harms, if they do exist.

1.7 Summary of EAG's preferred assumptions and resulting ICER

The EAG's preferred assumptions differ from the company's base case as follows:

- Reparameterisation of the long-term Mixed Model for Repeated Measures (MMRM) model so that the ALSPAC MMRM model is used to calculate long-term Edinburgh Postnatal Depression Scale (EPDS) scores for the ECM arm rather than the zuranolone with ECM arm. The EPDS scores for the zuranolone with ECM arm are then calculated by adding the zuranolone treatment effect to the ECM arm scores. EAG scenario 1.
- The zuranolone treatment effect wanes linearly to zero, with waning starting at 1 year post-treatment initiation and lasting 1 year (i.e. the EPDS scores for the two arms converge at 2 years post-treatment initiation). EAG scenario 2g. **(Key Issue 4)**
- 90% of patients are on ADT at baseline. EAG scenario 3c. **(Key Issues 2 and 5)**
- Standardised mortality rate (SMR) of 2 for patients with EPDS ≥ 13. Company scenario S1.
- Using a cohort of patients with severe EPDS scores (>19) at baseline. EAG scenario 5a. **(Key Issue 1)**

- Using EPDS utility values mapped from ROBIN PHQ-9 to EQ-5D for the PND period (until 1 year post-treatment initiation, i.e. until 17 months post-birth) and using MDD-specific utilities for the period beyond PND (17 months post-birth onwards). EAG scenario 6b. **(Key Issue 7)**

Table 2 provides the company's base-case results and the impact of each of the EAGs preferred assumptions alone and all together (the EAG base case). Probabilistic results are shown for the company and EAG base cases. For further details of the exploratory and sensitivity analyses done by the EAG, see sections 6.1 and 6.2. Modelling errors identified and corrected by the EAG are described in section 5.3.

Table 2 Summary of EAG's preferred assumptions and ICER

Scenario	Incremental costs	Incremental QALYs	ICER (change from company base case)
Deterministic results (zuranolone with ECM vs ECM alone)			
0. Company's base case (after clarifications, and including EAG's corrections)	[redacted]	[redacted]	£12,704
1. Reparameterisation of long-term MMRM model	[redacted]	[redacted]	£12,725 (+£21)
2. Treatment effect waning starts at 1 year post-treatment initiation and lasts 1 year **(Key Issue 4)**	[redacted]	[redacted]	£156,898 (+£144,194)
3. 90% of patients are on ADT at baseline **(Key Issues 2 and 5)**	[redacted]	[redacted]	£13,205 (+£501)
4. SMR of 2 for patients with EPDS ≥ 13	[redacted]	[redacted]	£12,415 (-£289)
5. Patients with severe EPDS scores (>19) at baseline **(Key Issue 1)**	[redacted]	[redacted]	£12,839 (+£135)
6. Using EPDS utility values mapped from ROBIN PHQ-9 to EQ-5D for the PND period, and using MDD utilities for the MDD period **(Key Issue 7)**	[redacted]	[redacted]	£22,939 (+£10,235)
EAG's preferred base case (0+1+2+3+4+5+6)	[redacted]	[redacted]	£194,554 (+£181,850)
Probabilistic results (zuranolone with ECM vs ECM alone)			
0. Company's base case (after clarifications, and including EAG's corrections)	[redacted]	[redacted]	£12,706
EAG's preferred base case (0+1+2+3+4+5+6)	[redacted]	[redacted]	£188,102 (+£175,396)

2 INTRODUCTION AND BACKGROUND

This report provides a critique of the evidence submitted by the company (Biogen) in support of zuranolone for treating PND in adults. It considers the company evidence submission and the company's executable model received on 12/03/25.[2] It also considers the company's response to clarification questions from the EAG received on 17/04/25 and 19/04/25.

2.1 Critique of the company's proposed place of the technology in the treatment pathway and intended positioning of the intervention.

Section 1.3 of the Company Submission (CS) provides an overview of PND and position of the technology in the treatment pathway. Full details of the technology, its mechanisms of action and the intended positioning of the intervention are described in sections 1.2 of the company submission (CS).[2]

Postnatal depression (PND) affects over 10% of individuals within the first year after birth (postpartum). However, onset is possible antenatally typically within the last trimester of pregnancy. UK data show rising prevalence, from 10.3% in 2014 to 23.9% in 2020.[3] The aetiology of PND is not fully understood and, although onset can occur without identifiable triggers, key risk factors include a personal history of mental illness, limited social support, interpersonal difficulties, recent adverse life events, and psychosocial stress. Symptoms include persistent low mood, anxiety, irritability, fatigue, sleep disturbance, impaired concentration, appetite changes, intrusive thoughts, guilt, bonding difficulties, and reduced enjoyment in caregiving. PND is also associated with adverse outcomes for relationships with partners and child development, including cognitive, emotional, and behavioural challenges.

There are currently no pharmacological treatments with marketing approval specifically for PND in the UK. However, some of the treatments licensed for major depressive disorder (MDD) are used off-license for PND. NICE guideline CG192 recommends facilitated self-help in mild to moderate cases of PND and high-intensity psychological therapy (e.g. CBT), with antidepressants, for moderate to severe cases.[4] CG192 recommends that antidepressants should be considered for patients with a history of severe depression. Combination therapy is advised if there is limited response to single treatments, and specialist referral is recommended for severe or treatment-resistant cases. It has been reported that in the UK up to 60% of women with clinically relevant symptoms either do not seek or cannot access treatment.[5-7]

The CS states that zuranolone will initially be prescribed by psychiatrists within perinatal mental health services (secondary care) and EAG clinical advisors agreed with this position. The EAG heard from clinical experts that patients referred to perinatal psychiatrists in NHS secondary care settings will be those with severe PND, most of whom (>90%) will have already been prescribed antidepressants as part of established clinical management (ECM).

The NICE scope and CS states that zuranolone is an adjunct therapy to ECM, and the EAG's clinical advisors agreed that patients would continue their existing therapies whilst taking zuranolone (including antidepressants). However, the company states in the clarification to EAG questions A4 and A6 that "*zuranolone is expected to fit into the existing clinical pathway as a treatment option to be used alone or as an adjunct to therapeutic alternatives used as part of the ECM, regardless of the history of prior ADT* [antidepressant therapy] *use*". However, the EAG notes that zuranolone monotherapy is not under consideration in the current appraisal and was not included in the NICE scope. Please see sections 2.2 and 3.2.1.3 for further critique.

At the time of writing, zuranolone has not received a market authorisation in the UK or in the EU. It is currently only approved by the US Food and Drug Administration (FDA), without restriction by disease severity.

2.2 Critique of company's definition of decision problem

Table 3 summarises the decision problem as outlined in the NICE scope and provides a summary of how this was addressed in the company submission, and the EAGs comments on this.

The population for whom the CS seeks approval for zuranolone is adults with PND unrestricted by disease severity. However, the evidence for clinical effectiveness comes from two trials (SKYLARK and ROBIN) which recruited participants with a HAMD-17 score ≥26 at baseline, which represents a severe population. A post-hoc analysis using MADRS scores is reported as evidence of mixed baseline PND severity (SKYLARK: [redacted] moderate and [redacted] severe; ROBIN: [redacted] moderate and [redacted] severe PND; CS, section 2.8.2.2). However, SKYLARK's mean baseline MADRS scores suggest that the "moderate" subgroup only includes the upper range of moderate, and therefore cannot be considered representative of a "moderate" subgroup in the general population. Additionally, as noted in section 1.3.4.3, the CS states that zuranolone will initially be prescribed by psychiatrists within secondary care services. In the UK, EAG clinical experts advised that patients seen in secondary care perinatal mental health services are those with severe PND, complex PND or those with co-occurring psychiatric conditions. As clinical evidence for zuranolone comes mostly from trial participants with severe PND, and EAG clinical experts indicated this matches the expected use in the NHS, the EAG considers that the appropriate population for consideration in this appraisal should be those with severe PND.

The comparison of interest in the NICE scope is zuranolone plus established clinical management (ECM) compared with ECM. The NICE scope provides examples of treatments that may be considered a part of ECM, such as psychological therapies, antidepressants, combination therapies, antipsychotics and electroconvulsive therapy. EAG clinical experts advised that the majority of patients currently referred to secondary care settings in the NHS are very likely to be taking antidepressants, as part of ECM. In the SKYLARK study, only 15% (30/196) participants were taking ADT at baseline (See sections 3.2.1.2, 3.2.1.3 and

Table 5), and it is not clear what other background treatments (if any) the remaining 85% (166/196) were receiving. Additionally, the EAG note that the NICE scope specifies that the technology under appraisal is zuranolone as an adjunct to ECM and not zuranolone monotherapy (see Key Issue 2 in section 1.3, section 2.1 and Table 3). If the NICE scope had specified that the technology under appraisal was zuranolone monotherapy, then the EAG considers the relevant comparators would have been the components of ECM (including ADT). In that case, comparisons between zuranolone and ADT would need to have been presented. However, the EAG considers that in the context of the NHS, zuranolone as an adjunct to ECM is the most appropriate intervention for this appraisal.

Table 3 Summary of decision problem

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	EAG comment
Population	Adults with PND	As per scope	N/A	The clinical evidence submitted in the CS is narrower than that specified in the NICE scope and is most applicable to patients with severe PND. The EAG believes a narrower scope, restricting to severe patients to be appropriate. The aim of company studies submitted in the CS (SKYLARK and ROBIN) was to determine the efficacy of zuranolone in patients with *severe* disease[8-10]. All patients were classified as having severe PND at baseline, based on the HAMD-17. A post-hoc analysis using MADRS scores is reported as evidence of mixed baseline PND severity (SKYLARK: 89 [46%] moderate and 105 [54%] severe; ROBIN: 61 [41%] moderate and 89 [59%] severe PND; CS, section 2.8.2.2). However, the EAG do not consider that the efficacy of zuranolone in patients with mild or moderate PND can be determined from evidence reported in the CS. Additionally, as patients with high risk of suicide and psychotic symptoms were not eligible for participation in the company's trials, the EAG considers that the applicability of the evidence to an unrestricted population is uncertain. As zuranalone will be available via secondary care services, the EAG also notes it is likely to be offered only to patients with severe PND.
Intervention	Zuranolone with ECM	As per scope	N/A	The EAG agrees that zuranolone with ECM is the appropriate intervention for appraisal. However, EAG clinicians noted that

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	EAG comment
				background ADT use in the company trials (SKYLARK: 15%) was considerably less than expected for patients with severe PND who would be considered for zuranolone in the UK. The EAG therefore has concerns that the results in the CS do not represent the same ECM as used in clinical practice in the NHS. The EAG consider that the subgroup with "baseline ADT" use might be more representative of zuranolone with ECM in patients with severe PND in NHS secondary perinatal mental health settings.
Comparator(s)	ECM without zuranolone which may include:	As per scope	N/A	EAG clinical advisors explained that most, if not all patients with the level of disease severity of those included in the trial, would be offered an intervention (pharmacological, psychological, or both), or to have augmentation with additional drugs and/or psychological therapies. Patients were not allowed to start any new interventions during the trial, which differs from ECM in clinical practice, and in particular may adversely affect outcomes on the placebo with ECM arm more than the zuranolone plus ECM arm.

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	EAG comment

	• Psychological therapies (e.g. CBT; facilitated self-help)
	• Antidepressant treatments:
	• TCAs
	  ○ SSRIs
	  ○ SNRIs
	  ○ Atypical antidepressants (e.g. mirtazapine)
	• High-intensity psychological intervention combined with antidepressant treatments
	• Augmentation with additional antidepressants, antipsychotics, or electroconvulsive therapy
	• Best supportive care			

Outcomes	The outcome measures to be considered include:	As per scope	N/A	Only EPDS and safety outcomes from SKYLARK and SF-6D utilities from ROBIN are used in the economic model. Hospitalisation and child health related outcomes are not included in the trials or economic model. All additional outcomes are reported in the trials and described in the CS, although cognitive function and sleep quality are reported only as sub-items of depression scales.

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	EAG comment
	• depressive symptoms including relapse and remission rates • severity of depression • cognitive function • anxiety • sleep quality • hospitalisation • mortality • child health-related outcomes • adverse effects of treatment (including adverse effects of treatment discontinuation) • 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 differences in costs or	As per scope	N/A	Whilst the EAG has no concerns with the length of the time-horizon used by the company (29 years), the treatment effect observed from the company's trial after 45 days follow-up is extrapolated for the duration of the time-horizon, which the EAG finds implausible.

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	EAG comment
	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.			
Subgroups	If the evidence allows, the following subgroups may be considered: • Previous history of depression • Severity of PND	• Subgroup analysis for severity of PND is provided • Subgroup analysis for previous history of depression is not provided	In SKYLARK, [redacted] of patients were reported to have a history of depression. The subgroup sample size was deemed too small to produce stable assessment on efficacy;	Subgroup analysis for severity is done using MADRS scale, which has somewhat different cut-off scores for moderate vs severe disease compared with HAM-D (all participants were classified as severe based on HAM-D). No subgroup analyses were conducted for the cost-effectiveness analyses, due to small samples. This means it is not possible to fully assess the cost-effectiveness for mild or moderate PND patients, those with baseline ADT use, or those with a previous history of depression.

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	EAG comment
			therefore, the analyses were not conducted.	
Special considerations including issues related to equity or equality	Guidance will only be issued in accordance with the marketing authorisation. Where the wording of the therapeutic indication does not include specific treatment combinations, guidance will be issued only in the context of the evidence that has underpinned the marketing authorisation granted by the regulator.	N/A	N/A	The EAG note that UK marketing authorisation is not yet available for zuranolone.

3 CLINICAL EFFECTIVENESS

3.1 Overview of evidence reported in company submission

The clinical evidence in the CS focuses on two company studies: SKYLARK and ROBIN. SKYLARK is the main trial for this submission and is a randomised, double-blinded, study comparing zuranolone 50mg (the intended licensed dose) to placebo. The EAG's critique of SKYLARK is presented in section 3.2.1. ROBIN is also a randomised, double-blinded, placebo-controlled study, but it compares zuranolone 30mg to placebo. As this is not the intended licensed dose, the company presents the ROBIN trial as supportive evidence only. The EAG's critique of the ROBIN trial is presented in section 3.2.2.

Table 4 provides an overview of the outcomes reported in the company submission, the data sources used for each treatment comparison, if and how the outcome informed the economic model, whether the outcome is within the NICE scope and whether the outcome is recommended by the European Medicines Agency (EMA).[11]

Table 4 Overview of clinical evidence included in the company submission

NICE scope category	Key outcomes	Recommended by EMA	Source of data	Outcome	Treatment comparison	Results reported in CS	Included in company model
Depressive symptoms including relapse and remission rates	HAMD-17	Yes	SKYLARK	Primary	Direct, randomised (vs. placebo)	Yes	No
			ROBIN	Primary	Direct, randomised (vs. placebo)	Yes	No
Severity of depression	HAMD-17	Yes	SKYLARK	Primary	Direct, randomised (vs. placebo)	Yes	No
			ROBIN	Primary	Direct, randomised (vs. placebo)	Yes	No
	EPDS	Yes	SKYLARK	Secondary	Direct, randomised (vs. placebo)	Yes	Yes
			ROBIN	Secondary	Direct, randomised (vs. placebo)	Yes	No
	MADRS	Yes	SKYLARK	Secondary	Direct, randomised (vs. placebo)	Yes	No
			ROBIN	Secondary	Direct, randomised (vs. placebo)	Yes	No
	PHQ-9	Yes*	SKYLARK	Secondary	Direct, randomised (vs. placebo)	Yes	No**
	CGI-S	Yes	SKYLARK	Secondary	Direct, randomised (vs. placebo)	Yes	No
	CGI-I	Yes	SKYLARK	Secondary	Direct, randomised (vs. placebo)	Yes	No
	CGI-I	Yes	ROBIN	Secondary	Direct, randomised (vs. placebo)	Yes	No
Cognitive function	HAMD-17 (subitem - retardation)	No***	SKYLARK	Secondary	Direct, randomised (vs. placebo)	Yes	No
	MADRS (subitem – concentration difficulties	No***	SKYLARK	Secondary	Direct, randomised (vs. placebo)	Yes	No
			ROBIN	Secondary	Direct, randomised (vs. placebo)		No
Anxiety	HAM-A	Yes	SKYLARK	Secondary	Direct, randomised (vs. placebo)	Yes	No
			ROBIN	Secondary	Direct, randomised (vs. placebo)	Yes	No
Sleep quality	HAM-D subitem	No	SKYLARK	Secondary	Direct, randomised (vs. placebo)	Yes	No
			ROBIN	Secondary	Direct, randomised (vs. placebo)	Yes	No
	MADRS subitem	No	SKYLARK	Secondary	Direct, randomised (vs. placebo)	Yes	No
			ROBIN	Secondary	Direct, randomised (vs. placebo)	Yes	No
	EPDS subitem		ROBIN	Secondary	Direct, randomised (vs. placebo)		No

NICE scope category	Key outcomes	Recommended by EMA	Source of data	Outcome	Treatment comparison	Results reported in CS	Included in company model
Hospitalisation	Not reported						
Mortality	Death (AEs)		SKYLARK	Secondary	Direct, randomised (vs. placebo)	Yes	No
Child health-related outcomes	Not reported						
adverse effects of treatment (including adverse effects of treatment discontinuation)	TEAEs, maximum severity of TEAE, TEAE leading to dose reduction, TEAE leading to treatment discontinuation, TEAE leading to withdrawal from the study, SAE, death, TEAEs with ⩾5%		SKYLARK	Secondary	Direct, randomised (vs. placebo)	Yes	Yes
			ROBIN	Secondary	Direct, randomised (vs. placebo)	Yes	No
HRQoL	SF-36		ROBIN	Secondary	Direct, randomised (vs. placebo)	Yes	Yes

*Not specifically mentioned but EMA says “other validated scales might be acceptable as well”. ** “Responses from PHQ-9 from the ROBIN trial were mapped onto EQ-5D-3L utilities using Mukuria’s mapping model and applied in a sensitivity analysis in the model. *** EMA states there is no consensus on gold standard to evaluate this outcome.

3.2 Critique of the company trials

3.2.1 SKYLARK

3.2.1.1 Study design

SKYLARK (NCT04442503) was a phase 3, randomised, double-blind, placebo-controlled, outpatient trial. A full description can be found on Section 2.3.1.1 of the CS. Skylark was conducted between June 2020 and April 2022 in 82 sites in the US, Spain and United Kingdom. This study enrolled a total of 200 women aged 18-45 years, ≤12 months postnatal, with a PND diagnosis ((baseline HAMD-17 total score of ≥26). Further inclusion criteria are outlined in section 3.2.2.2 (population). Participants were randomised 1:1 to self-administer zuranolone 50 mg (n=98) or matching placebo (n=98) once daily. The study consisted of an up to 28 days screening period, a 14-day treatment course, and follow-up through day 45. The EAG considers this to be an appropriate study design for the aim of the trial, although there are concerns in sections 3.2.1.2 (population) regarding a lack of alignment between the trial and the decision problem specified by NICE and the marketing authorisation for zuranolone.

3.2.1.2 Population

Eligibility criteria are presented in table 7 in the CS and in the CS appendices, section B.6.1.1. Participants were eligible for inclusion if they were aged 18 to 45 years, ≤12 months postnatal, had a major depressive episode that began during the third trimester of pregnancy or ⩽4 weeks postnatal, had a baseline HAMD-17 total score of ≥26, and had stopped lactation or agreed to stop breastfeeding from day 1 of the trial until 7 days after the last dose of the study medication. Participants were not eligible if they had active psychosis or history of bipolar disorder, psychotic disorders, nonfebrile seizures, schizophrenia and/or schizoaffective disorder, a history of a substance use disorder (including benzodiazepines), if they had attempted suicide or had risk of suicide associated with the current episode of PND, or had an index pregnancy resulting in miscarriage, still birth, or neonatal/infant death, or terminated parental rights. Additional exclusion criteria included a recent history or active clinically significant manifestations of metabolic, hepatic, renal, haematological, pulmonary, cardiovascular, gastrointestinal, musculoskeletal, dermatological, urogenital, or neurological disorders; eyes, ears, nose, and throat disorders, a body mass index (BMI) ⩽18 or ⩾45 kg/m2, and a clinically significant abnormal 12-lead electrocardiogram at Screening or Day 1, or a history of sleep apnoea, gastric bypass surgery, gastric sleeve or lap band, or any procedures that interfere with gastrointestinal transit.

The EAG notes that the SKYLARK trial does not align with the decision problem because it enrolled women with severe PND (HAMD-17 total score of ≥26), whereas the decision problem is broader, specifying "adults with PND". The company suggests that the "*unique mode of action of zuranolone supports its efficacy in all PND patients, irrespective of disease severity*" (p52 of the CS), however the EAG's clinical advisors consider there to be no available evidence to support this. The EAG's clinical advisors also noted that the HAMD-17

scale (and other scales) are not commonly used in UK clinical practice to diagnose and manage patients or to measure recovery. Instead, clinical judgement of severity of symptoms is undertaken, taking into account functional impact and risk. However, the EAG acknowledges that HAMD-17 is a validated scale widely used in research and appropriate for outcome measurement in trial settings. EPDS is a screening tool for identifying people at risk for PND and was not developed for the measurement of depression severity in research settings,[12] although the use of EPDS and PHQ-9 is suggested in a NICE Scenario for monitoring PND in clinical practice.[13] Additionally, the EAG experts noted that the exclusion of patients with psychosis and attempted suicide reduces the generalisability to UK practice, as patients with these characteristics would be referred to and treated within their practice (i.e. within perinatal mental health services), although they acknowledge that such exclusions are common in clinical trials.

Baseline characteristics were balanced across treatment groups (Table 5) and the EAG does not note any concerns regarding their comparability. However, the EAG's clinical advisors reported that the US-majority trial sample differs from patients seen in UK clinical practice in certain characteristics. They noted that the UK has a lower proportion of Hispanic/ Latina patients and higher proportion of South Asian patients, and it is unclear whether this difference would have an impact on clinical effectiveness. Additionally, as zuranolone will initially only be prescribed by perinatal psychiatrists (after being referred to secondary care), most patients considered for zuranolone for PND in the UK will likely already be taking antidepressants for PND - likely >90%, according to EAG clinical advisors. Our advisors also noted that a large proportion patients seen by perinatal psychiatrists would be considered non-responders to antidepressants prescribed before referral to secondary care. In comparison, only 15/98 (15.3%) in the zuranolone 50 mg group and 15/98 (15.3%) in the placebo group of the SKYLARK trial. The trial sample may therefore not fully represent those considered for treatment with zuranolone in the UK. The types of antidepressants used by patients at baseline are also reported in Table 6. The most commonly used baseline antidepressant in SKYLARK was sertraline hydrochloride (zuranolone: 7.1% and placebo: 11.3%) followed by escitalopram oxalate (zuranolone: 3.1% and placebo: 2.1%) and bupropion hydrochloride (zuranolone: 4.1% and placebo: 0%). Our clinical advisors have indicated that sertraline is similarly likely to be the most prescribed antidepressant in the UK population, with escitalopram and citalopram also in use. Bupropion and vilazodone are not licenced in the UK. Paroxetine is very rarely used, while fluoxetine is used less frequently now compared to its past use. However, EAG clinical advisors did not think the differences in the types of baseline ADT used in SKYLARK and the UK NHS setting would introduce significant issues in terms of the applicability of the trial evidence.

The EAG asked the company to outline how many patients who were on baseline antidepressants would be considered non-responders to antidepressants and whether non-response to antidepressants was an inclusion criterion for ROBIN (clarification question A4). The company did not provide the number of patients who would be considered non-responders but noted that non-response was not an inclusion criterion for the trial.

The company stated, in response to EAG clarification question A2 and A3, that psychological therapies were not captured as standalone variables at baseline, and data on parity and gravidity were also not collected, so this information is not available for the SKYLARK trial.

Table 5 SKYLARK (NCT04442503): Patient Demographics and Baseline Characteristics [amended from the CS Table 9, CSR Table 11 and table 14.1.4.5.1, and response to clarification question A2]

Characteristics	Zuranolone 50 mg (n=98)	Placebo (n=98)
Age, mean (SD), years	30.0 (5.9)	31.0 (6.0)
Age group, years, n (%)		
18-24	[redacted]	[redacted]
25-45	[redacted]	[redacted]
Race, n (%)		
White	68 (69.4)	69 (70.4)
Black/African American	25 (25.5)	18 (18.4)
Other[a]	5 (5.1)	11 (11.2)
Hispanic/Latina ethnicity, n (%)	33 (33.7)	42 (42.9)
BMI, mean (SD), kg/m2	30.9 (6.3)	29.6 (6.3)
HAMD-17 total score, mean (SD)	28.6 (2.5)	28.8 (2.3)
CGI-S score, mean (SD)	5.0 (0.7)	4.9 (0.6)
Country, n (%)		
USA	95 (96.9)	96 (98.0)
Rest of the world	3 (3.1)	2 (2.0)
Baseline ADT use, n (%)	15 (15.3)	15 (15.3)
Sertraline / sertraline hydrochloride	[redacted]	[redacted]
Escitalopram oxalate	[redacted]	[redacted]
Bupropion hydrochloride	[redacted]	[redacted]
Fluoxetine hydrochloride	[redacted]	[redacted]
Paroxetine	[redacted]	[redacted]
Vilazodone hydrochloride	[redacted]	[redacted]
Baseline HAMD-17 total score, n (%)		
<28 HAMD-17 total score	[redacted]	[redacted]
≥28 HAMD-17 total score	[redacted]	[redacted]
History of PND, n (%)		
First episode	81 (82.7)	87 (88.8)
Recurrent episode	17 (17.3)	11 (11.2)
Onset of PND, n (%)		
Third trimester	34 (34.7)	31 (31.6)
≤4 weeks postnatal	64 (65.3)	67 (68.4)

Source: Company Submission.[2]

The safety set was composed of patients who received ≥1 dose of the assigned blinded treatment.

[a] Other included Asian, American Indian, Alaska Native, Native Hawaiian/Pacific Islander, multiple, other race, and/or not reported.

Abbreviations: ADT: Antidepressant Therapy; BMI: Body Mass Index; CGI-S: Clinical Global Impressions – Severity; HAMD-17: 17-item Hamilton Rating Scale for Depression; PND: Postnatal Depression; SD: Standard Deviation; USA: United States of America.

3.2.1.3 Interventions

The interventions assessed in the SKYLARK trial are described in the CS in table 7 and in section 2.3.2.1. Participants were randomised to receive zuranolone 50 mg (n=98) taken orally as two 25 mg capsules once daily for 14 days, or matching placebo (n=98).

Participants who were unable to tolerate their assigned intervention were allowed to decrease their dosage to 40 mg for the remainder of the treatment course. 16 participants (16.3) in the zuranolone arm and 1 (1%) in the placebo arm reduced their dose to 40 mg, and 4 (4.1%) and 2 (2%) discontinued the treatment course respectively (CS, table 25).

3.2.1.3.1 Concomitant medications and interventions

Participants who were randomised into SKYLARK were allowed to continue concomitant use of ADTs if they were on a stable dose for ≥30 days prior to first study treatment dose, or to continue psychotherapies if they had been receiving them on a regular schedule for at least 30 days prior. Participants were prohibited from starting any other antidepressant or antianxiety medications until after completion of the Day 45 visit. Patients taking benzodiazepines, barbiturates, Gamma Amino Butyric Acid Type A ($GABA_A$) modulators, antipsychotics and psychostimulants were not eligible for enrolment, as well as patients receiving strong inhibitors of cytochrome p450 or strong CYP3A inducers.

Concomitant interventions received by SKYLARK participants are listed in Clinical Study Report (CSR) table 14.1.4.2 (concomitant non-psychotropic medications), 14.1.4.3 (on-treatment concomitant non-psychotropic medications), 14.1.4.4 (post-treatment concomitant non-psychotropic medications), 14.1.4.6.1 (concomitant psychotropic medications), 14.1.4.7 (on-treatment psychotropic medications), 14.1.4.8 (post-treatment psychotropic medications), and 14.1.4.9 (concomitant antidepressant medications). [redacted] patients modified their antidepressant use during the treatment period (including stopping existing ADTs, increasing or decreasing ADT dosages, or starting new ADTs). [redacted] participant receiving placebo and [redacted] patients receiving zuranolone started a new ADT during follow-up.

The company notes that allowing concomitant use of previously established ADTs and psychological therapies is aligned with ECM, and therefore the trial matches the interventions and comparators stated in the decision problem (zuranolone with ECM vs. ECM) (CS, page 49). The EAG does not fully agree with this statement. First, it is repeatedly stated in the trial registration,[14] trial protocol,[15] and main journal publication,[10] that the SKYLARK trial was designed to compare zuranolone against placebo, and does not mention a comparison of zuranolone + ECM vs placebo + ECM. While the trial permitted the background use of ADT, the EAG clinical advisers have expressed concerns regarding the comparability of the observed use of ADT and psychotherapies in SKYLARK with that in the

UK population, as previously discussed in the Population section of this report (3.2.1.2). EAG's clinical advisors also noted that most of their patients would likely continue their existing therapies whilst taking zuranolone (including antidepressants). Moreover, the trial does not align with the comparators in the decision problem, as the decision problem included augmentation with additional antidepressants, antipsychotics, or electroconvulsive therapy and these were prohibited in the SKYLARK trial (antipsychotics and electroconvulsive therapy were prohibited through completion of the Day 45 visit). This is particularly concerning for the comparator (placebo) arm. Patients whose symptoms were severe enough to require augmentation of pharmacological and/or psychological interventions were not allowed to receive them until the end of follow-up. This may have led to worse outcomes in the placebo arm because a higher proportion may may have required augmentation during the trial since they were not taking zuranolone. Therefore, the EAG considers that the ECM used in the SKYLARK trial does not match the ECM used in NHS clinical practice.

3.2.1.4 Efficacy results

Results of the SKYLARK trial are presented in section 2.6.1 of the CS. Efficacy analyses were completed using the "Full analysis set", which included all patients who received at least one dose of study drug with both a valid baseline and at least one postbaseline score on the HAMD-17, HAM-A, MADRS, CGI-S, EPDS, and PHQ-9, or at least one postbaseline CGI-I score. 98 participants allocated to zuranolone and 97 assigned to placebo were included in the efficacy analyses.

HAMD-17: The primary endpoint of the SKYLARK trial was change from baseline HAMD-17 score at Day 15 as pre-specified in the trial protocol, and it is reported in section 2.6.1.1 of the CS. This outcome was not included in the company's economic model. The least-square means (LSM) change from baseline was −15.6 for the zuranolone 50 mg group compared with −11.6 for the placebo, showing strong evidence of improvement in depressive symptoms when comparing zuranolone and placebo, with a LSM treatment difference of −4.0 (95% CI: −6.3, −1.7; p=0.001).

Furthermore, strong evidence of improvement in depressive symptoms measured by this scale was found at additional timepoints, starting from Day 3 and extending into Day 45 (Table 6). Additionally, the company reports results for change from baseline in HAMD-17 subscales and individual item scores at day 15, HAMD-17 response and remission rate at Day 15 and Day 45 in the CS section 2.6.1.

Table 6 SKYLARK: LSM change from baseline in HAMD-17 score (Full Analysis Set) (Reproduced from the CS, table 14).

Study visit	Zuranolone 50 mg N=98		Placebo N=97		LSM difference			p-value
	Mean	SE	Mean	SE	Mean	95% CI lower	95% CI upper	
Baseline[a]	28.6	2.49	28.8	2.34	-	-	-	-
Day 3[b]	-9.5	0.70	-6.1	0.71	-3.4	-5.4	-1.4	0.001
Day 8	-13.2	0.76	-9.5	0.76	-3.7	-5.8	-1.6	0.001
Day 15[c]	**-15.6**	**0.82**	**-11.6**	**0.82**	**-4.0**	**-6.3**	**-1.7**	**0.001**
Day 21	-15.7	0.87	-13.3	0.87	-2.5	-4.9	-0.1	0.045
Day 28[b]	-16.3	0.88	-13.4	0.88	-2.9	-5.4	-0.5	0.020
Day 45[b]	-17.9	0.90	-14.4	0.90	-3.5	-6.0	-1.0	0.007

Source: Company submission.[2]
Results are presented as LSM and SE CFB unless otherwise noted.
[a] Baseline values are reported as mean ± SD.
[b] Key secondary endpoint.
[c] Primary endpoint.
Primary and key secondary endpoints were adjusted for multiplicity. All other secondary endpoints were not adjusted for multiplicity and are to be interpreted with nominal p-values.
Abbreviations: CFB: Change form Baseline; CI: Confidence Interval; HAMD-17: 17-item Hamilton Depression Rating Scale; LSM: Least Squares Mean; SE: Standard Error

EPDS: Change from baseline in EPDS total score is reported for several timepoints in section 2.6.1.3 of the CS. This was a pre-specified, secondary outcome in the trial protocol,[15] and it was included in the company's economic model. The trial found strong evidence of greater improvement in EPDS scores among participants receiving zuranolone compared with those receiving placebo at most timepoints, including Day 3 (LSM difference [95% CI]: −1.5 [−2.9, −0.1]; P=0.03), Day 8 (LSM difference [95% CI]: −2.2 [−3.8, −0.5]; P=0.01), Day 15 (LSM difference [95% CI]: −2.0 [−3.8, −0.1]; P=0.04), and Day 45 (LSM difference [95% CI]: −2.4 [−4.5, −0.3]; P=0.03). There was some evidence of improvement at other timepoints, but 95% CI included the null value.

Table 7 SKYLARK: LSM change from baseline in EPDS total score (Full Analysis Set) (Reproduced from the CS, table 21)

Study visit	Zuranolone 50 mg N=98		Placebo N=97		LSM difference			Nominal p-value
	Mean	SE	Mean	SE	Mean	95% CI lower	95% CI upper	
Day 3	-3.8	0.49	-2.3	0.49	-1.5	-2.9	-0.1	0.032
Day 8	-8.4	0.60	-6.2	0.59	-2.2	-3.8	-0.5	0.010
Day 15	**-10.3**	**0.66**	**-8.4**	**0.66**	**-2.0**	**-3.8**	**-0.1**	**0.038**
Day 21	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Day 28	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Day 45	-12.2	0.76	-9.8	0.76	-2.4	-4.5	-0.3	0.028

Source: Company submission.[2]
Abbreviations: CI: Confidence Interval; EPDS: Edinburgh Postnatal Depression Scale; LSM: Least Squares Mean; SE: Standard Error.

MADRS: Change from baseline in MADRS total score is reported for several timepoints in section 2.6.1.3 of the CS. This was a pre-specified, secondary outcome in the trial protocol, but it was not included in the company's economic model. Measurements for MADRS were scheduled at Days 1, 8, 15, 28 and 45.[15] All timepoints show strong evidence of an improvement in MADRS scores when comparing patients receiving zuranolone against participants receiving placebo (Table 8). Change from baseline in MADRS individual items at Day 15 is also reported in the CS, table 20.

Table 8 SKYLARK: LSM change from baseline in MADRS total score (Full Analysis Set) (Reproduced from the CS, Table 19)

Study visit	Zuranolone 50 mg N=98		Placebo N=97		LSM difference			Nominal p-value
	Mean	SE	Mean	SE	Mean	95% CI lower	95% CI upper	
Baseline[a]	35.5	5.37	35.0	4.81	-	-	-	-
Day 3	-16.4	1.11	-11.8	1.10	-4.6	-7.7	-1.5	0.004
Day 15	**-19.7**	**1.20**	**-14.6**	**1.21**	**-5.1**	**-8.4**	**-1.7**	**0.003**
Day 28	-20.2	1.23	-16.8	1.22	-3.4	-6.8	0	0.051
Day 45	-22.5	1.27	-17.8	1.28	-4.7	-8.3	-1.1	0.010

Source: Company submission.[2]
Results are presented as LSM and SE CFB unless otherwise noted.
[a]Baseline values are reported as mean ± SD.
Primary and key secondary endpoints were adjusted for multiplicity. All other secondary endpoints were not adjusted for multiplicity and are to be interpreted with nominal p-values.
Abbreviations: CFB: Change form Baseline; CI: Confidence Interval; LSM: Least Squares Mean; MADRS: Montgomery-Åsberg Depression Rating Scale; SE: Standard Error.

CGI: Change from baseline in CGI-S score is reported for several timepoints in section 2.6.1.3 of the CS. This was a pre-specified, secondary outcome in the trial protocol (key secondary

outcome at Day 15), but it was not included in the company's economic model. Strong evidence of greater improvement in CGI-S scores was observed when comparing participants receiving zuranolone against participants receiving placebo at most timepoints. CGI-I response rate at different timepoints was also reported in the CS, section 2.6.1.3, showing strong evidence of improved outcomes of zuranolone when compared with placebo at most timepoints.

Study visit	Zuranolone 50 mg N=98		Placebo N=97		LSM difference			p-value
	Mean	SE	Mean	SE	Mean	95% CI lower	95% CI upper	
Baseline[a]	5.0	0.66	4.9	0.58	-	-	-	-
Day 3	-1.0	0.10	-0.7	0.10	-0.2	-0.5	0	0.093
Day 8	-1.7	0.12	-1.1	0.12	-0.6	-0.9	-0.3	0.0004
Day 15[b]	**-2.2**	**0.14**	**-1.6**	**0.14**	**-0.6**	**-0.9**	**-0.2**	**0.005**
Day 21	-2.3	0.14	-1.9	0.14	-0.5	-0.9	-0.1	0.022
Day 28	-2.3	0.14	-1.9	0.14	-0.4	-0.8	0	0.044
Day 45	-2.6	0.14	-2.1	0.14	-0.5	-0.9	-0.1	0.012

Source: Company submission.[2]

[a] Baseline values are reported as mean ± SD.

[b] Key secondary endpoint.

Multiplicity was accounted for when analysing primary and key secondary endpoints. All other secondary endpoints were not adjusted for multiplicity and are to be interpreted with nominal p values.

Abbreviations: CI: Confidence Interval; CGI-S: Clinician Global Impressions – Severity; LSM: Least Squares Mean; SE: Standard Error.

HAM-A: Change from baseline in HAM-A score is reported for several timepoints in section 2.6.1.3 of the CS. This was a secondary outcome in SKYLARK, and it was not used in the economic model.

3.2.1.4.1 Subgroup analyses

The efficacy of zuranolone was assessed in several pre-specified subgroups (Figure 22, section 2.8.1 of the CS): race, age, baseline antidepressant use, baseline BMI, baseline HAMD-17 total score, onset of PPD, and HAM-A score. The company notes that the results of subgroup analyses were similar to the overall HAMD-17 total score analyses, with zuranolone being favoured in all subgroups except for the ADT use at baseline. The company also highlights that several groups included small number of patients, and the study was not powered to detect efficacy differences in these subgroups. The EAG consider these sub analysis were adequate.

Section 2.8.2 of the CS discusses additional, post-hoc subgroup analysis, covering the subgroups indicated by the decision problem: previous history of depression and severity of PND.

In section 2.8.2.1, the CS noted that although previous psychiatric history had been collected during screening, the subgroup sample size [redacted] was considered too small to produce a stable assessment of efficacy. Therefore, an analysis was initially not provided. However, it was later added in response to the EAG's request to present the results of this sub-analysis as part of the clarification question A5, in line with the NICE scope. [redacted], although the company noted that this subgroup was small, and that the study was not powered to detect a treatment difference. The EAG note this sub analysis and the company's notes are adequate, and reflect the subgroups indicated in the scope.

In section 2.8.2.2, the company discusses results of HAMD-17 scores by PND severity. The CS noted that they used the MADRS score to categorise depression severity because it focuses more on psychological symptoms. The company also noted that they used literature-established thresholds for the MADRS to categorise participants baseline severity as moderate (total score ≤34) or severe (total score ≥35), with [redacted] participants categorised as having moderate depression at baseline, and [redacted] considered as having severe PND. The company reports that following this categorisation there was strong evidence of greater HAMD-17 score improvements at Day 15 and Day 45 in the moderate subgroup and at Day 15 in the severe subgroup. Results were similar for the severe subgroup at Day 45, but confidence intervals included the null value (Figure 1).

Figure 1 SKYLARK: Forest plot of LSM (95% CI) treatment difference in change from baseline in HAMD-17 total score by MADRS severity subgroups at Day 15 and Day 45 (Full Analysis Set) (Reproduced from the CS, Figure 24)

Source: Company submission.[2]
[1] Post-hoc subgroup analysis
Results are from a MMRM with treatment, baseline HAMD-17 total score, antidepressant use at baseline, assessment time point, and time point-by-treatment interaction as fixed effects.
Abbreviations: CI: Confidence Interval; HAMD-17: 17-item Hamilton Depression Rating Scale; LS: Least Squares Mean; MADRS: Montgomery-Åsberg Depression Rating Scale; MMRM: Mixed Model for Repeated Measures.

The EAG note that following literature-established thresholds for HAMD-17,[16] the scale used to determine eligibility and chosen as primary outcome, all patients included in the SKYLARK trial are considered to have severe PND at baseline (a HAMD-17 score ≥24 is classified as severe depression). Additionally, and to have a better idea of the severity degree of the subgroups included in this sub analysis, the EAG requested mean values and SDs of the MADRS, EPDS and HAMD-17 baseline scores for the moderate and severe post-hoc subgroups. The company's response is illustrated in Table 10 SKYLARK: Mean MADRS, EPDS and HAMD-17 score at Baseline by MADRS Severity – Full Analysis Set (Reproduced from clarification question A9 response). According to MADRS thresholds,[17] patients with scores between 20 and 34 are considered to have moderate depression. SKYLARK's mean baseline MADRS scores suggest that the "moderate" subgroup only includes the upper range of moderate, and therefore cannot be considered representative of a "moderate" subgroup in the general population. A similar picture is observed when considering EPDS severity thresholds: patients with scores between 14 -19 are considered to have moderate PND, while patients with scores ≥19 are classified as severe PND.[18] To summarise, the EAG note that the "moderate" subgroup in SKYLARK is composed of patients with upper range moderate and severe depression, or purely composed of patients with severe depression depending on the scale, so this sub analysis should not be considered representative of a subgroup of patients with moderate PND.

The EAG also note that there appears to be a disproportionate number of participants from the "severe" subgroup at baseline with missing data for MADRS score analyses on Day 15 and Day 45, when compared to participants from the "moderate" subgroup. By Day 15, [redacted] of baseline participants from the "moderate" subgroup were not included in the analysis, and by Day 45 there were [redacted] of participants missing, while for the "severe" group, missing participants were [redacted] and [redacted] respectively (Table 9). This could potentially impact the validity of the results in the "severe" subgroup, however it is not possible to assess its potential extent without additional data on distribution between interventions groups available.

Table 9 [redacted]

[redacted]	[redacted]		[redacted]		[redacted]	
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Additionally, the EAG note that the numbers per subgroup in Figure 1 column "Count" do not add up [redacted]. The

EAG acknowledge that this might be due to missing data or a typo, and this is not expected to have a significant impact on the validity of the results.

Table 10 SKYLARK: Mean MADRS, EPDS and HAMD-17 score at Baseline by MADRS Severity – Full Analysis Set (Reproduced from clarification question A9 response)

Baseline score (SD)	Zuranolone 50 mg Moderate PND			Zuranolone 50 mg Severe PND		
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
HAMD	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
MADR	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
EPDS	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

3.2.1.5 Safety results

Safety results from the SKYLARK trial are reported in section 2.11.1 of the CS for the "Safety set", which included all patients who received at least one dose of study drug (zuranolone 50mg n=98, placebo n=98). Adverse events (AE)s from SKYLARK input to the company's economic model.

On-treatment treatment emergent adverse events (TEAEs) were reported by 60.2% (59/98) and 41.8% (41/98) of patients receiving zuranolone 50 mg and placebo, respectively. In the zuranolone group, this was most frequently somnolence (26.5%), dizziness (13.3%), sedation (11.2%), headache (9.2%), diarrhoea (6.1%), nausea (5.1%), urinary tract infection (5.1%), and COVID-19 (5.1%). In the placebo group, the most frequently reported TEAEs were headache (13.3%), dizziness (10.2%), nausea (6.1%), and somnolence (5.1%). Most TEAES in both groups were mild or moderate.

Six patients reported TEAES leading to discontinuation of the study drug: 4/98 (4.1%) in the zuranolone 50 mg group and 2/98 (2%) in the placebo group. Two patients (2%) presented serious adverse events (SAEs), both were receiving zuranolone 50 mg and were considered unrelated to study drug.

3.2.1.6 Health-related quality of life

There were no outcomes pertaining to health-related quality of life in SKYLARK.

3.2.1.7 Protocol deviations

Major protocol deviations from SKYLARK were reported in section 10.2 and table 14.1.2.1 of the CSR. In the zuranolone group, [redacted] participants [redacted] had at least one major protocol deviation [redacted] Three patients were excluded from the per-protocol set due to a major

protocol deviation in this group. In the placebo group, [redacted] participants ([redacted] had at least one major protocol deviation [redacted] patients were excluded from the per-protocol set in the placebo group due to having significant deviations that could affect efficacy outcomes. Minor protocol deviations were not reported.

3.2.1.8 Risk of bias assessment (RoB 2)

The company assessed risk of bias for SKYLARK at the trial-level, using the Downs and Black Checklist in the CS section 2.5 and Appendix B.5.[19] The company considered the study to be at low risk of bias, but highlighted that a potential risk of functional unblinding was raised by the FDA. This was because of an imbalance between the groups in treatment discontinuation due to AEs (zuranolone n=4, placebo n=2) and in patients needing dose reductions due to AEs (14% of zuranolone group, 1% of placebo group). The company highlight in the risk of bias assessment that the study does not provide estimates of the random variability in the data for the main outcomes, and that the characteristics of patients lost to follow-up were not described. The company also noted that it was not possible to determine if the subjects who were prepared to participate were representative of the entire population from which they were recruited, and if the staff, places, and facilities where the patients were treated representative of the treatment the majority of patients receive. The company also raise concerns regarding whether the randomised intervention assignment was concealed from both patients and health care staff until recruitment was complete and irrevocable, and that it was not possible to determine if losses of patients to follow-up were taken into account.

The EAG consider the Risk of Bias 2 (RoB 2) tool to be a more appropriate tool for risk of bias assessment in randomised controlled trials, as it is more robust and facilitates assessment at the outcome-level.[20] In clarification question C6, the EAG asked the company to repeat risk of bias assessment with RoB 2 for the outcomes EPDS and HAMD-17 at day 45. The company, in return, presented a RoB 2 assessment, however it was at the trial level rather than the outcome level (Table 14 in the clarification questions appendix). Similarly to the company's Downs and Black Checklist assessment, the trial was judged at low risk of bias in all domains, even though they highlight that participants were not blind to treatment allocation (unclear if this is in reference to the potential functional unblinding raised by the FDA, or an error as the study was designed as "double-blind"). However, the overall assessment is not aligned with the domain-level ratings and was judged at "some concerns" without further justification for this decision.

The EAG completed a RoB 2 assessment for the following outcomes: EPDS, HAMD-17, MADRS, PHQ-9 and Adverse Events which have been summarised in Table 11 and further detailed in Appendix 9.2.1. The EAG used the CS, the published trial report and the associated trial protocol to complete the RoB 2 assessment. All outcomes were judged at some concerns of risk of bias overall. The domain "measurement of the outcome" was

judged at some concerns for all outcomes at all timepoints because outcome assessors were likely aware of the intervention received due to functional unblinding and this may have influenced assessment using subjective scales. Moreover, the domain “missing outcome data” was also judged at some concerns for efficacy outcomes at timepoints Day 21 to Day 45, due to data not being available for most participants who were enrolled into the study. All other domains were judged at low risk of bias.

Table 11 Risk of bias in key outcomes in the SKYLARK trial, assessed using RoB 2

EAG assessment							
Outcome	**Domain judgement**						**EAG Rationale for overall judgment**
	1	**2**	**3**	**4**	**5**	**Overall**	
EPDS (*Day 3– 5 **Day 21-45	Low	Low	Low*/Some concerns**	Some concerns	Low	Some concerns	Some concerns of risk of bias from the domain "measurement of the outcome" because outcome assessors were likely aware of the intervention received, due to functional unblinding, and this may have influenced assessment using subjective scales. Some concerns of risk of bias from the domain "missing outcome data" for timepoints Day 21 to Day 45, due to data not being available for most participants who were enrolled into the study. All other domains judged at low risk of bias.
HAMD-17 *Day 3– 5 **Day 21-45	Low	Low	Low*/Some concerns**	Some concerns	Low	Some concerns	
MADRS *Day 3– 5 **Day 21-45	Low	Low	Low*/Some concerns**	Some concerns	Low	Some concerns	
PHQ-9 *Day 3– 5 **Day 21-45	Low	Low	Low*/Some concerns**	Some concerns	Low	Some concerns	
Adverse effects (AEs) of treatment	Low	Low	Low	Some concerns	Low	Some concerns	

Domains: 1: Randomisation process; 2: deviation from intended intervention; 3: missing outcome data; 4: measurement of the outcome; 5: selective outcome reporting. AEs = adverse events.

3.2.2 ROBIN

3.2.2.1 Study design

Section 2.3.1.2 of the CS outlines the design of the ROBIN trial (NCT02978326). ROBIN was a phase 3, randomised, double-blind, outpatient trial, conducted between January 2017 and December 2018 in 27 US sites. A total of 153 women aged 18-45 years, ≤6 months postnatal, with PND (baseline HAMD-17 total score of ≥26) were enrolled in the study. Further inclusion criteria are outlined in section 3.2.2.2 (population). Participants were randomised using a 1:1 ratio to receive once-daily zuranolone 30mg (n=77) or placebo (n=76). The study was conducted in two parts: Part A and Part B. Part A used an oral zuranolone solution and was closed to enrolment when an oral capsule formulation became available, which was used in Part B (protocol amendment 2). Only one participant was screened for Part A before it was closed to enrolment, so data collected in Part A are not included in analyses and not focused on in the CS. Part B consisted of a 28-day screening period, a 14-day treatment period, and follow-up until day 45. The EAG considers the ROBIN study design to be appropriate for the aim of the trial, although raises concerns in sections 3.2.2.2 (population) and 3.2.2.3 (interventions) about a lack of alignment between the trial and the decision problem specified by NICE and the marketing authorisation for zuranolone.

3.2.2.2 Population

Participants were eligible for inclusion if they were aged 18 to 45 years, ≤6 months postnatal, had a major depressive episode without psychosis (as per the Diagnostic and Statistical Manual of Mental Disorders 5 (DSM-5)) that began during the third trimester of pregnancy or ≤4 weeks postnatal, had a baseline HAMD-17 total score of ≥26, and had stopped lactation or agreed stop breastfeeding from day 1 of the trial until 7 days after the last dose (CS Table 7). Participants were not eligible if they had a known allergy to zuranolone, active psychosis (per investigator assessment), had attempted suicide with the current PND episode, or had a history of seizures, bipolar disorder, schizophrenia, schizoaffective disorder, alcoholism or drug addiction in the 12 months before screening. Other exclusion criteria included recent history or active clinically significant renal, hepatic, metabolic, pulmonary, cardiovascular, haematological, gastrointestinal, dermatological, musculoskeletal, urogenital, or neurological disorders; eyes, ears, nose, and throat disorders. Allowed concomitant medications are discussed in section 3.2.2.3.1 of this report. Similarly to the SKYLARK trial (section 3.2.1.2), the EAG notes concerns regarding the alignment of the ROBIN trial population with the NICE scope due to it focusing on women with severe PND (not “adults with PND”) and exclusion of patients with psychosis and attempted suicide who would be seen in UK practice (although it is acknowledged that the exclusions of higher risk patients are common in clinical trials). This introduces uncertainty in the generalisability of the trial results to the population outlined in the NICE scope.

Baseline demographic characteristics, presented in Table 12, were balanced across treatment groups and the EAG has no concerns about the comparability of the groups within the study. Concerns discussed regarding demographic characteristics in the SKYLARK trial (section 3.2.1.2) also apply to the ROBIN trial, reducing its generalisability to UK clinical

practice. These include that the EAGs clinical advisors noted that the UK has a higher proportion of South Asian and lower proportion of Hispanic/ Latina patients than the trial, although they said this is not likely to affect estimated clinical effectiveness of zuranolone, and PND patients in the UK have higher antidepressant use than the trial sample. Regarding the type of antidepressant, the EAGs clinical advisors noted that sertraline is also likely to be the most prescribed antidepressant within UK practice, with escitalopram and citalopram also frequently used. They highlighted that bupropion and vilazodone are not licensed in the UK and that fluoxetine and paroxetine are now rarely used. Data on baseline psychological therapy use, parity, and gravidity were not available for the ROBIN trial (clarification question A2 and A3).

Table 12 ROBIN (NCT02978326): Patient Demographics and Baseline Characteristics [amended from the CS Table 10, CSR Table 14.1.3.3, and response to clarification question A2 and A3]

Characteristics	Zuranolone 30 mg (n=76)	Placebo (n=74)
Age, mean (SD), years	29.3 (5.4)	27.4 (5.3)
Race, n (%)		
White	44 (58)	40 (54)
African American	31 (41)	31 (42)
Other[a]	1 (1)	3 (4)
Hispanic/Latino ethnicity, n (%)	16 (21)	18 (24)
Weight, mean (SD), kg	85.1 (19)	80.2 (24)
HAMD-17 total score, mean (SD)	28.4 (2)	28.8 (2)
Family history of PND, n (%)	10 (13)	10 (14)
Onset of PND, n (%)		
Third trimester	32 (42)	31 (42)
≤4 weeks after delivery	44 (58)	43 (58)
History of MDD, n (%)	[redacted]	[redacted]
Baseline ADT use, n (%)	[redacted]	[redacted]
Sertraline/ Sertraline hydrochloride	[redacted]	[redacted]
Bupropion/ Bupropion hydrochloride	[redacted]	[redacted]
Citalopram hydrochloride	[redacted]	[redacted]
Fluoxetine hydrochloride	[redacted]	[redacted]
Paroxetine/ Paroxetine hydrochloride	[redacted]	[redacted]
Duloxetine hydrochloride	[redacted]	[redacted]
Escitalopram oxalate	[redacted]	[redacted]

[a]Other included Asian, Native Hawaiian/Pacific Islander, and more than 1 race. Abbreviations: ADT: Antidepressant Therapy; HAMD-17: 17-item Hamilton Rating Scale for Depression; PND: Postnatal Depression; SD: Standard Deviation; MDD: Major Depressive Disorder.

3.2.2.3 Interventions

The interventions used in the ROBIN trial are outlined in the CS in Table 7. Participants in the intervention group (n=76) received zuranolone 30mg, administered orally once daily in the evenings with food for 14 days. The company state that zuranolone 30mg used in ROBIN was in a capsule formulation with higher relative bioavailability – the EAG is unable to confirm this, so this remains an uncertainty, and ultimately this is a different dose to the

intended licensed dose of 50mg. The placebo group (n=74) received matching placebo capsules, administered orally once daily in the evenings with food for 14 days.

A dose reduction from zuranolone 30mg to 20mg was permitted for patients who were unable to tolerate the 30mg dose. Tolerability was assessed by occurrence of a severe or moderate AE of special interest (sedation, somnolence, dizziness, euphoric mood, confusion, drowsiness, inebriation (feeling drunk), or fatigue) judged by the investigator to be related to study drug.[21] Three participants reduced dose to 20mg (CSR Table 14.1.5). Additionally, one participant discontinued the study as they were unable to tolerate 30mg and did not take 20mg.

Four patients were administered the study drug by a health professional in clinic and all others self-administered at home. The reason for this is not stated. However, the EAG notes that there was a protocol amendment (no4, v5.0, 31 August 2017) which changed study drug administration from in the clinic or at home under supervision of a healthcare professional to self-administration by subjects (with follow-up call from the site every evening on days 1-14) (CSR section 9.8.1).

3.2.2.3.1 Concomitant medications

To be eligible for the ROBIN trial, patients who were taking psychotropic medications, such as ADTs, were required to have been on a stable dose for more than 30 days before day 1 of the trial and to delay the start or alteration of their treatment regimen until after the treatment period and day 15 assessments were completed. The EAGs clinical advisors noted that the requirement for patients to be on a stable dose is reasonable and would likely be the case in UK practice.

Atypical antipsychotics for the treatment of psychotic symptoms and anticonvulsants were not permitted. Benzodiazepines were to be avoided, but participants on a stable dose for at least 30 days before day 1 were allowed on a case-by-case basis. Likewise, hypnotics for sleep/ insomnia were to be avoided and their use was allowed on a case-by-case basis. Known strong inhibitors of CYP3A4 and any CYP inducers (e.g. rifampin, ritonavir) were prohibited within 14 days or 5 half-lives (whichever is longer) prior to receiving the first dose of study drug and throughout the trial. The EAGs clinical advisors raised no concerns with these criteria, although they noted that in the UK, PND patients may be on hypnotics and benzodiazepines.

A summary of all concomitant medications used by patients in the trial is provided in CSR Table 14.1.3.2. In the participants who used an ADT at baseline, [redacted] in the zuranolone group and [redacted] in the placebo group started a new ADT during the follow-up period of the trial (CSR Table 14.1.3.4). [redacted] had an increase in the dose of their baseline ADT during the treatment period or follow-up. In those who did not use an ADT at baseline, [redacted] in the zuranolone group and [redacted] in the placebo group

started an ADT during follow-up. In the total sample, [redacted] started a new ADT during the treatment period (this was disallowed by the trial).

3.2.2.4 Efficacy results

Results of the ROBIN trial are presented in section 2.6.2 of the CS. The EAG notes that no efficacy results from ROBIN input into the company's economic model. The only outcome from ROBIN that inputs into the model is health-related quality of life, which is outlined in section 3.2.2.6.

Efficacy analyses were conducted using the "efficacy set", which included all randomised patients in Part B who completed at least one dose of zuranolone 30mg and had a valid baseline and at least one post-baseline efficacy assessment (explained in section 2.4.2 of the CS). This included 76 patients for zuranolone 30mg and 74 placebo. The company notes that the efficacy set in ROBIN corresponds to the "full analysis set" in the SKYLARK trial.

HAMD-17: Change in baseline in HAMD-17 total score was the primary outcome of the ROBIN trial, as pre-specified in the trial protocol.[21] Reported in section 2.6.2.1 of the CS, at day 15, the least squares mean (LSM) change from baseline was −17.8 for the zuranolone 30 mg group and −13.6 for the placebo group (mean difference −4.2, 95% CI −6.9 to −1.5, *P* = .003). Numerically greater reductions from baseline HAMD-17 score were shown at all measured timepoints from day 3 to day 45 in the zuranolone 30mg group (Day 3: LSM difference (95% CI) −2.7 (−5.1, −0.3), nominal P=0.025; Day 45: LSM difference (95% CI) −4.1(−6.7, −1.4); nominal P=0.003).

EPDS: Change from baseline in EPDS total score is also reported in the CS section 2.6.2.3. This was a pre-specified outcome in the trial protocol. The company reports that, compared to the placebo group, patients treated with zuranolone 30mg had [redacted] (LSM difference [redacted] (Table 24 in the CS).

MADRS: MADRS total score data are presented in the CS Appendix B.7.2.1 and Table 21 of the CSR. MADRS was pre-specified in the trial protocol. The company reports that, at day 15, numerically greater improvements in depressive symptoms were shown in the zuranolone 30mg group compared to placebo (LSM difference (95% CI) −4.6 (−8.3, −0.8); nominal P=[redacted]). Apart from day 3, this was maintained until day 45 (LSM difference (95% CI) −5.8 (−9.4, −2.2); nominal P=0.002) (Table 22 in the CS Appendix B.7.2.1).

The following outcomes are also reported in the appendix of the CS (none of these into the company's economic model): HAMD-17 response rate and remission rate, change from baseline in HAMD-17 subscales and individual item scores at day 15, CGI-I response rate and HAM-A (anxiety) total score.

3.2.2.4.1 Subgroup analyses

The efficacy of zuranolone 30mg in the ROBIN trial was assessed in several pre-specified subgroups (Figure 23, section 2.8.1 of the CS): race, age, baseline BMI, baseline anti-depressant use, onset of PND, family history of PND. The company notes that zuranolone was favoured in all subgroups. However, many of these included small numbers of patients and ROBIN was not powered to detect differences in efficacy within subgroups.

Figure 25 in section 2.8.2 of the CS presents a post-hoc subgroup analysis of treatment difference in change from baseline of HAMD-17 total score by MADRS severity subgroups at day 15 and day 45. MADRS is an additional clinician-rated scale used in the trial which the company says focuses more on psychological symptoms and less on somatic or anxiety-related symptoms than the HAMD-17. There were [redacted] in the moderate group (threshold used: baseline MADRS total score ≤34) and [redacted]) in the severe group (threshold used: baseline MADRS total score ≥35). The company report that [redacted]

The company provided a subgroup analysis for history of MDD or not, at day 15 and 45, in response to EAG clarification question A5. [redacted] in the ROBIN trial had history of MDD. The LSM change from baseline of the HAMD-17 total score was reported to [redacted].

3.2.2.5 Safety results

AEs from the ROBIN trial are reported in section 2.11.2 of the CS for the "safety set", which included all participants who were administered study intervention (zuranolone 30mg n=78, placebo n=73). AEs do not input into the company's economic model from this trial and pertain to zuranolone dose of 30mg, which is not the intended licensed dose of 50mg.

At least one treatment-emergent adverse event (TEAE) was reported by 47/78 (60.3%) participants in the zuranolone 30mg group and 38/73 (52.1%) in the placebo group. In the zuranolone group, this was most frequently (≥5%) somnolence (15.4%), headache (9%), dizziness (7.7%), upper respiratory tract infection (7.7%), diarrhoea (6.4%), and sedation (5.1%). In the placebo group, TEAEs were most frequently somnolence (11%), headache (12.3%), nausea (8.2%), dizziness (5.5%), vomiting (5.5%), abnormal dreams (5.5%), and hyperhidrosis (5.5%) (CS Table 26). [redacted]. Severe TEAEs were reported by three patients in the zuranolone group (sedation n=1, confusional state n=1, migraine n=1) and three in the placebo group (back pain/muscle spasms n=1, headache/oropharyngeal pain n=1, menorrhagia n=1). One patient in the zuranolone group had a TEAE (severe intermittent sedation) that led to treatment discontinuation.

One SAE was also reported in each group. One patient experienced a confusional state and sedation starting on day 3 in the zuranolone group, leading to hospitalisation. This resolved within seven hours and the patient was reduced to 20mg zuranolone and completed the treatment period. In the placebo group, one patient had pancreatitis during follow-up, which resolved following cholecystectomy.

The company note that no clinically significant changes were reported in echocardiogram (ECG), vital signs, or clinical laboratory parameters, and there was no increase in suicidal ideation or behaviour (measured by the Colombia Suicide Severity Rating Scale (C-SSRS)).

3.2.2.6 Health-related quality of life

The EAG emphasises that health-related quality of life data from ROBIN pertain to zuranolone 30mg and not the intended licensed dose (50mg).

Short Form Health Survey (SF-36): SF-36 data from ROBIN input into the company's economic model. SF-36 change from baseline to day 45 is presented in section 2.6.2.3 of the CS. At day 45, SF-36 outcome data were available for [redacted] participants in the placebo group and [redacted] in the zuranolone 30mg group (CSR Table 14.2.8.1). By day 45, the zuranolone group demonstrated improvements compared with placebo across the five SF-36 domains: social functioning LS mean [redacted] vs [redacted], mental health [redacted] vs [redacted], physical functioning [redacted] vs [redacted], role physical [redacted] vs [redacted], , and bodily pain [redacted] vs [redacted], and in the mental component summary score [redacted] vs [redacted]. Change from baseline in other SF-36 domains was noted to be numerically greater with zuranolone vs. placebo (Figure 21 in the CS).

SF-36 is also reported at day 15 in the CSR (Table 14.2.8.1 and 14.2.8.2). At this timepoint, SF-36 data were available for 73/74 (99%) participants in the placebo group and 74/76 (97%) in the zuranolone 30mg group (CSR Table 14.2.8.1). The company report that improvements from baseline at day 15 were numerically greater with zuranolone than placebo, but the difference was not statistically significant.

Patient Health Questionnaire (PHQ-9): Responses to the PHQ-9 were also collected in the ROBIN trial, as pre-specified in the protocol. These data are not reported in the CS, and do not input into the company's economic model, but are outlined in the CSR (Tables 14.2.6.1 and 14.2.6.2). At day 45, data were available for [redacted] participants in the placebo group and [redacted] in the zuranolone 30mg group. The LS mean change (SE) was [redacted] for the placebo group and [redacted] for the zuranolone 30mg group. The treatment difference and corresponding 95% CI was [redacted] The company notes that

improvement in PHQ-9 was numerically greater at all measured timepoints from day 3 to day 45 for the zuranolone group versus placebo group.

3.2.2.7 Protocol deviations

Section 10.2 and Table 14.1.6 of the CSR reports protocol deviations. In the zuranolone group, [redacted] participants [redacted] had at least one protocol deviation. Of these, [redacted] participants [redacted] had a minor only protocol deviation and [redacted] had at least one major protocol deviation [redacted]. No patients were excluded from the per-protocol set due to a major efficacy deviation in the zuranolone group. In the placebo group, [redacted] participants [redacted] had at least one protocol deviation. Of these, [redacted] participants [redacted] had a minor only protocol deviation and [redacted] had at least one major protocol deviation [redacted]. [redacted] patients were excluded from the per-protocol set in the placebo group due to having significant deviations that could affect efficacy outcomes.

The company reported that for [redacted] of patients in total who had a major protocol deviation that may have affected efficacy, the reason was due to the day 15 visit being conducted "out of window". [redacted] participants had their day 15 visit ±1 day out of window, [redacted] had it ±2 days out of window, [redacted] had it ±3 days out of window, and [redacted] did not complete it. All day 15 visits conducted out of window were used in the efficacy analyses. No further information on protocol deviations was provided to the EAG.

3.2.2.8 Risk of bias assessment (RoB 2)

In the CS section 2.5 and Appendix B.5, the company assessed risk of bias for ROBIN at the trial-level, using the Downs and Black Checklist.[19] The company considered the study to be at a low risk of bias but noted that the potential risk of functional unblinding was raised by the FDA. This was because of an imbalance between the groups in treatment discontinuation due to AEs (zuranolone n=1, placebo n=0) and in patients needing dose reductions due to AEs (4% of zuranolone group, 0% of placebo group). The company highlight in the risk of bias assessment that the study does not provide estimates of the random variability in the data for the main outcomes and intervention assignment was not concealed from both patients and health care staff until recruitment was complete and irrevocable. The company noted they could not determine whether losses of patients to follow-up were taken into account, whether participants were representative of the entire population from which they were recruited, or whether the staff, places, and facilities where the patients were treated representative of the treatment the majority of patients receive.

The EAG consider the Risk of Bias 2 (RoB 2) tool to be a more appropriate tool for risk of bias assessment in randomised controlled trials, as it is more robust and facilitates assessment at the outcome-level.[20] In clarification question C6, the EAG asked the company to repeat risk of bias assessment with RoB 2 for the outcomes HAMD-17 and SF-36 at day 45. The

company responded with a RoB 2 assessment, however it was at the trial level rather than the outcome level (Table 14 in the clarification questions appendix). Similarly to the company's Downs and Black Checklist assessment, the trial was judged at low risk of bias in all domains and overall using RoB 2. The company did not raise issues with allocation concealment or functional unblinding, which had been highlighted in the previous assessment (CS Appendix B.5).

The EAG completed a RoB 2 assessment for the following outcomes: EPDS, HAMD-17, SF-36, MADRS, and PHQ-9. An outline of this is presented in Table 13 and shown in full in Appendix 9.2.2. The EAG used the CS, the published trial report and the associated trial protocol to complete the RoB 2 assessment. All outcomes were judged at some concerns of risk of bias overall. This was due to a judgement of some concerns for the domain "measurement of the outcome", because outcome assessors were likely aware of the intervention received due to functional unblinding and this may have influenced assessment using subjective scales. All other domains were judged at low risk of bias.

Table 13 EAG RoB 2 assessment for key outcomes in the ROBIN trial

EAG assessment at outcome level							
Outcome	**Domain judgement**						**EAG Rationale for overall judgment**
	1	**2**	**3**	**4**	**5**	**Overall**	
EPDS	Low	Low	Low	Some concerns	Low	Some concerns	Some concerns of risk of bias from the domain "measurement of the outcome" because outcome assessors were likely aware of the intervention received, due to functional unblinding, and this may have influenced assessment using subjective scales. All other domains judged at low risk of bias.
HAMD-17	Low	Low	Low	Some concerns	Low	Some concerns	
SF-36	Low	Low	Low	Some concerns	Low	Some concerns	
MADRS	Low	Low	Low	Some concerns	Low	Some concerns	
PHQ-9	Low	Low	Low	Some concerns	Low	Some concerns	

Domains: 1: Randomisation process; 2: deviation from intended intervention; 3: missing outcome data; 4: measurement of the outcome; 5: selective outcome reporting. AEs = adverse events.

3.3 Critique of the systematic review

The company submitted a systematic literature review (SLR) of clinical effectiveness. The SLR identified two studies that the company state are relevant to the decision problem: SKYLARK and ROBIN. All other studies were excluded. The EAG critiqued the SLR of clinical effectiveness using the Risk of Bias in Systematic reviews (ROBIS) tool (see section 9.1 Appendix).[22] In clarification question C1, the EAG requested the protocol for the clinical effectiveness SLR.[23] The EAG consider that the review broadly adheres to established conduct guidance.

The eligibility criteria for the SLR mostly reflected the NICE scope. The EAG asked the company to clarify whether best supportive care was included in the systematic review of clinical effectiveness, as per the NICE scope (clarification question C3). The company responded that best supportive care was not included in the review because NICE guidelines for the management of antenatal and postnatal mental health do not list it as a treatment option and do not provide a definition for it.

Regarding study identification and selection, the EAG queried with the company why a single reviewer (rather than two reviewers, or one reviewer checked by a second) assessed the search results and removed "obviously irrelevant records such as those about ineligible diseases" (clarification question C5). The company responded that single reviewer screening was only used to exclude publications at the initial review stage with the Cochrane randomised controlled trial (RCT) classifier.

Regarding study appraisal, as discussed in section 3.2.2.8 of this report, the EAG asked for more information on the company's risk of bias assessment (clarification question C7), due to the use of an old tool conducted at the trial level and a lack of detail around how many individuals conducted risk of bias assessment. The EAG also asked the company to repeat assessment with RoB 2 (clarification question C6). The company responded with a repeated risk of bias assessment using RoB 2, at the trial level only, and clarified that one reviewer conducted the assessment and one independently checked it.

3.4 Critique of the Indirect Treatment Comparisons (ITC)

The CS presents evidence for zuranolone as an adjunct to ECM compared to ECM alone, which is the comparison listed in the NICE scope. However, because ECM can comprise a single therapy, the company explored potential indirect treatment comparisons (ITC) for zuranolone vs ADTs. The company identified an ITC comparing zuranolone versus SSRIs and combination therapies (SSRIs + psychological therapies)[24]. The ITC was conducted by the company, but analyses were not submitted as part of the CS and the EAG is not able to assess the company's conclusions. The CS notes that the ITC analysis was limited by high levels of heterogeneity between placebo responses across trials. To overcome this, study authors conducted a Matched Adjusted Indirect Comparison (MAIC) to compare zuranolone with SSRIs, but the CS notes that this analysis also had limitations, and there was a large

reduction in effective sample size due to poor overlap of study populations. In section 2.10.2 of the CS the company outline a variety of reasons why they did not consider it appropriate to conduct an ITC for this appraisal.

The NICE scope sets the comparison of interest for the appraisal as zuranolone as an adjunct to ECM compared to ECM alone. In NHS secondary care settings, the EAG heard clinical advice that ECM is SSRIs and the EAG's view is therefore that the appropriate comparison is zuranolone plus SSRI versus SSRI (in those for whom ECM is SSRI alone). The EAG note that direct evidence on this comparison can be obtained from the baseline ADT subgroup from SKYLARK, albeit with a very small sample size (n=30/183). Any indirect comparison would also have to use the same baseline ADT subgroup from SKYLARK, and the further assumptions and uncertainty introduced by the ITC or MAIC would not add to the already uncertain evidence for this comparison. Therefore, the EAG does not consider there to be any value in conducting an ITC for zuranolone plus SSRI versus SSRI.

If zuranolone were to be considered as a monotherapy, as the company suggests in their response to the EAG clarification question A6, then the components of ECM would become eligible comparators and an ITC would then become essential for this appraisal, with the limitations mitigated as much as possible. However, the EAG heard from their clinical advisors that zuranolone would be used in addition to ECM in the majority of severe PND patients and not as monotherapy, so the EAG does not consider this ITC to be necessary. This may not be the case for patients with milder PND.

3.5 Conclusions of the clinical effectiveness section

The company's submitted evidence is broadly, but not fully, in line with the final NICE scope. The population specified in the scope is broader than the company's submitted evidence based on the SKYLARK and ROBIN trials, which focus on a sub-population of patients who have severe PND at baseline (as measured by the HAMD-17 scale, or upper-range moderate to severe as measured by MDRS). The EAG also heard from clinical advisors that zuranolone is likely to be used for severe patients, in secondary care NHS perinatal services.

The majority of participants in SKYLARK were from the USA (>97%) and only five participants were from other countries (UK and Spain, numbers not available by individual country). With respect to ethnic background, Black and Hispanic/Latina participants formed 60% of those recruited to SKYLARK. The EAGs clinical advisors noted that the UK has a higher proportion of South Asian and lower proportion of Hispanic/ Latina patients than the trial, although they did not expect this to affect estimated clinical effectiveness of zuranolone.

The comparison of interest in the NICE scope is zuranolone plus ECM compared with ECM alone. The CS reports evidence for this comparison based on two placebo-controlled randomised clinical trials (SKYLARK and ROBIN), in which zuranolone was evaluated against placebo. Eligibility criteria for these trials permitted inclusion of participants receiving stable antidepressant therapy and/or stable psychological interventions at baseline. However, the

CS provides limited detail regarding the composition of ECM. Specifically, baseline use of ADT is the only component of ECM that is explicitly documented, with reported use being relatively low: only 15% of participants in the SKYLARK trial were receiving ADT at baseline. The nature of ECM for the remaining 85% of trial participants is undefined and the EAG consider it is possible that most of these participants were not receiving any background therapy at baseline. The company were not able to provide data on use of psychological therapies at baseline. This raises uncertainty regarding the generalisability and representativeness of the comparator arm for NHS practice.

3.5.1 Is there evidence of clinical effectiveness?

The estimates of clinical effectiveness in the CS come from the SKYLARK trial, with additional supporting evidence from the ROBIN trial (with the latter evaluating a different formulation of zuranolone at a lower dose of 30mg). In SKYLARK, depression severity was measured at multiple timepoints, with change from baseline in HAMD-17 score at Day 15 as the primary effectiveness endpoint. There was strong evidence of improvement in depressive symptoms for zuranolone compared with placebo in the trial population. This outcome was not included in the company's economic model. Strong evidence of improvement in depressive symptoms (measured by HAMD-17) was found at all other timepoints, starting from Day 3 and extending into Day 45, the latter providing the longest follow-up. This outcome provides the most reliable and robust evidence of effectiveness as HAMD-17 is a validated scale developed for measuring depression symptom scores in research settings. The HAMD-17 subgroup analyses by severity (based on the MADRS scale) did not show differences in the treatment effect sizes between the patients with severe vs moderate baseline MADRS scores. In the subgroup analysis by baseline ADT use, placebo was favoured in the subgroup of participants on baseline ADT, but these estimates were very imprecise due to small numbers of patients on ADT.

Change from baseline in EPDS total score at 6 different timepoints was a pre-specified, secondary outcome in the trial protocol,[15] with estimates from all 6 time points contributing to the estimation of the short-term EPDS model included in the company's economic model. There was evidence of greater improvement in EPDS scores among participants receiving zuranolone compared with those receiving placebo at most timepoints, including Day 3, Day 15 and Day 45. [redacted]

Safety results from the SKYLARK trial also feed into the company's economic model. Treatment emergent adverse events (TEAEs) were reported by 60.2% and 41.8% of patients receiving zuranolone 50 mg and placebo, respectively. In the zuranolone group the most frequent TEAEs were somnolence, dizziness, sedation, headache, diarrhoea, nausea, urinary tract infection, and COVID-19. In the placebo group, the most frequently reported TEAEs were headache, dizziness, nausea, and somnolence. Most TEAES in both groups were mild or moderate. Six patients reported TEAES leading to discontinuation of the study drug, four in the zuranolone 50 mg group and two in the placebo group. Two patients receiving

zuranolone 50 mg developed serious adverse events considered to be unrelated to study drug.

There was no evidence pertaining to health-related quality of life (HRQoL) outcomes reported in the SKYLARK trial. Instead, the company used data from Short Form Health Survey (SF-36) from the ROBIN trial as a source of HRQoL inputs for the economic model. Change from baseline in SF-36 to Day 45 showed greater improvements in the zuranolone 30mg group compared with the patients receiving placebo across the five SF-36 domains (social functioning, mental health, physical functioning, role physical and bodily pain) and in the mental component summary score. Change from baseline in other SF-36 domains was noted to be numerically greater with zuranolone compared with placebo. SF-36 is also reported at day 15 in the CSR; the improvements from baseline at day 15 were numerically greater with zuranolone than placebo, but the difference was not statistically significant.

3.5.2 Uncertainties regarding the reliability of the clinical effectiveness data

The EAG has some concerns that the effectiveness outcome chosen to input into the economic model was EPDS, rather than the primary outcome of the SKYLARK trial, HAMD-17. EAG notes that EPDS is a screening tool for identifying people at risk for PND and was not developed for the measurement of depression severity in research settings, although the use of EPDS is suggested in a NICE Scenario for monitoring PND in clinical practice.[13] The EPDS is also a secondary outcome and the SKYLARK trial may not have been powered to estimate effectiveness based on data from EPDS. The EAG considers that HAMD-17 would have been preferable for feeding into the economic model as HAMD-17 is a validated scale developed for measuring depression symptoms in research settings. HAMD-17 at Day 15 provides the most reliable and robust evidence of effectiveness as this was also the prespecified primary outcome for SKYLARK; as such the trial was powered to estimate this endpoint. HAMD-17 at Day 45 provides evidence of sustained effectiveness at the longest measured follow-up. The company provided further justification for the choice of EPDS during the factual accuracy check. There is some uncertainty about the effectiveness of zuranolone in a population with a high proportion of patients already on ADT.

The EAG has some concerns with regards to the risk of bias in the SKYLARK trial in the measurement of the outcomes due to a likely risk of functional unblinding, and due to a higher proportion of missing outcome data at days 21 and 45 for HAMD-17, EPDS, MADRS and PHQ-9.

Since HRQoL data were not collected in the SYKLARK trial data from the ROBIN trial were used for the economic model. The EAG emphasises that HRQoL data from the ROBIN trial pertain to zuranolone in a different formulation at a lower dose of 30mg and not the intended licensed dose (50mg).

4 COST EFFECTIVENESS

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

The company conducted a systematic literature review to identify previous cost-effectiveness analyses. The EAG appraised the SLR using a modified ROBIS tool. In clarification question C1, the EAG requested the protocol for the cost effectiveness SLR.[23] The EAG consider that the review broadly adheres to established conduct guidance, with some minor concerns relating to study eligibility criteria (See Appendix 1, section 9.1.3).

The company identified three previous cost-effectiveness studies (Table 27, CS). Bauer et al (2016)[25] presented a cost-benefit analysis of non-pharmacological interventions which the company and EAG agree is not relevant for this appraisal. O'Callahan (2024)[26] and Eldar-Lissai (2020)[27] were cost-utility analyses for zuranolone vs SSRIs and bexanolone vs SSRIs respectively, both set in the US and with similar model structure (6-state Markov model with 5 states based on EPDS scores and death) and assumptions. For both these models the company argue that there is a lack of granularity from using discrete health-states, and that the way patients move from the short-term to the long-term part of the model lacked plausibility. They conclude that none of the previous models were appropriate and that a de novo model is required. The EAG agrees that a discrete state Markov model is not the most appropriate structure to model continuous outcomes, such as EPDS scores, and that an alternative approach is warranted.

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

4.2.1 NICE reference case checklist

Table 14 below summarises the EAG's comments on the CS, in relation to the NICE reference case.

Table 14 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	Perspective appropriate for patient. Impact on children and partner not included in model.
Perspective on costs	NHS and PSS	No concerns
Type of economic evaluation	Cost–utility analysis with fully incremental analysis	No concerns
Time horizon	Long enough to reflect all important differences in costs or outcomes between the technologies being compared	Time horizon appropriate to capture long-term impact of PND. However, treatment effects are extrapolated for a long time (29 years) which is

		beyond evidence available (1.5months follow-up), and stopping treatment (after 14 days). Furthermore, no evidence is provided on how categories of PND would translate to longer term MDD or chronic MDD conditions.
Synthesis of evidence on health effects	Based on systematic review	Not applicable
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.	The company measured health effects using the QALY, as per NICE guidelines. However, QALYs were estimated based on SF6D utilities (derived from ROBIN's patients' responses to the SF36), rather than EQ-5D utilities. The company argued the SF-36 was best placed to capture the full impact on quality of life for this patient group. The EAG disagreed because the EQ-5D questionnaire is the NICE preferred tool to capture health effects and includes a specific domain on depression and anxiety and thus applied EQ-5D utilities instead.
Source of data for measurement of health-related quality of life	Reported directly by patients and/or carers	The company's utilities were sourced from responses to the SF36 by patients in the company's ROBIN trial, with 45 days follow-up, applied over 29 years. The EAG used utilities derived from mapped disease-specific questionnaires (PHQ-9) to the EQ-5D in company's ROBIN trial for the PND period (provided by the company at clarification questions), and used on utilities for health states of "remission" and "non-remission" after that, based on a previous TA for MDD.

Source of preference data for valuation of changes in health-related quality of life	Representative sample of the UK population	The company applied Brazier's UK preference weights to SF36 responses to estimate SF6D utility scores. The EAG has no concerns. Preference sources used by EAG's utilities were also UK tariffs.
Equity considerations	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	Zuranolone cannot be used in patients who are breast-feeding. Support for patients to either continue breastfeeding following treatment or stop breastfeeding would be required in practice.
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	The company used an NHS and Personal Social Services (PSS) perspective as per NICE guidance and sourced resource use from older studies in the literature with updated recent unit costs. The company also amended units of resource used in the first 18 months based on their clinical advisors' opinion. For the study informing the following years, the EAG broadly agrees with the costing methods, but disagrees that a study of 1-year follow-up costs in MDD patients could be applied for 27.5 years in the PND population.
Discounting	The same annual rate for both costs and health effects (currently 3.5%)	No concerns
PSS, personal social services; QALYs, quality-adjusted life years; EQ-5D, standardised instrument for use as a measure of health outcome.		

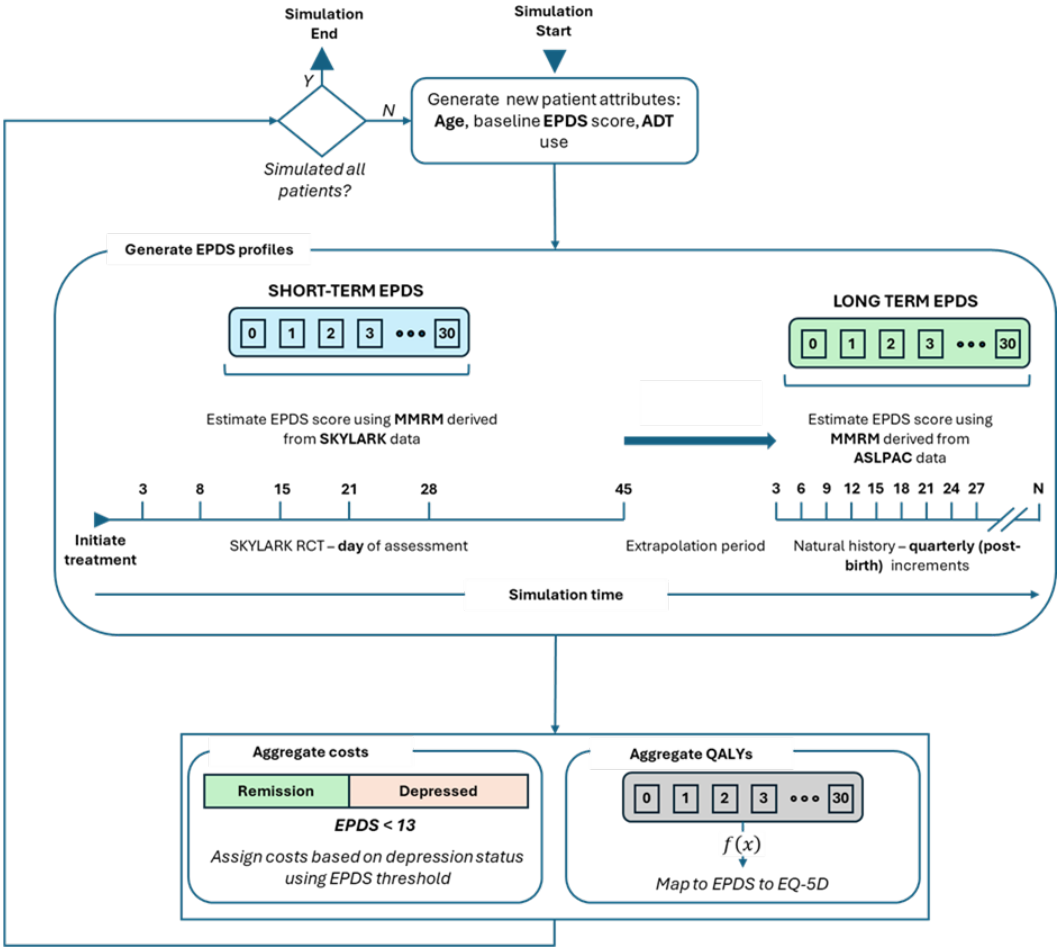
4.2.2 Model structure

To overcome the limitations of previous Markov models (see section 4.1) the company used a patient-level simulation model that simulates individual patients' characteristics (age, baseline EPDS score, and ADT use), EPDS scores in the short and long-term, and the associated cost and health outcomes. The model assumes patients start treatment at 5 months following birth, as this was the average from the SKYLARK trial which recruited

patients whose PND began in the third trimester or post-birth. The EAG considers a patient simulation model to be appropriate to capture changes in EPDS score over time that avoids discretisation, and can account for non-linear effects of baseline characteristics on patient trajectories.

Treatment effectiveness is captured in the model via EPDS trajectories over time (Figure 2) with a short-term model up to 3-months post-treatment initiation, and a long-term model thereafter (up to 29 years post-birth). The company assume that the long-term EPDS scores on zuranolone + ECM are the same as those observed in the Avon Longitudinal Study of Parents and Children (ALSPAC) cohort, and apply a treatment effect to this for patients on ECM. The ALSPAC cohort were not treated with zuranolone and so the EAG considers the results from the ALSPAC cohort to be most appropriate for the ECM arm, rather than the zuranolone + ECM arm, and prefer to parameterise the model this way so that the treatment effect for zuranolone + ECM is applied to the long-term ECM EPDS score. This is mathematically equivalent to the company's approach, with only a small difference in results due to the truncation of the EPDS score to lie in the range 0 – 30, as seen in the EAG scenario analysis (sections 6.1-6.2). However, the EAG prefer this approach as a more intuitive basis from which to explore the impact of extrapolation of treatment effects.

Figure 2 Detailed model overview for Zuranolone + ECM arm (CS Fig. 27)



Abbreviations: ADT: Antidepressant Treatment; ALSPAC: Avon Longitudinal Study of Parents and Children; EPDS: Edinburgh Postnatal Depression Scale; EQ-5D: EuroQoL 5-Dimension; MMRM: Mixed Model for Repeated Measures; RCT: Randomised Controlled Trial.

4.2.3 Population

The company states that the population modelled is adults with PND. However, the company do note in their submission that zuranolone is "… expected to be initially prescribed mainly by perinatal psychiatrists within the community perinatal mental health services, and therefore, most likely be used for the treatment of patients with moderate to severe PND". The company simulate patient characteristics based on the distributions of age, baseline EPDS, and ADT use from SKYLARK (CS Table 30). No subgroup analyses were performed. In their response to clarification questions the company explained that the population includes both those whose PND begins before childbirth and those whose PND begins after childbirth.

The company effectively models a mostly severe PND population, as all key model inputs on population characteristics are sourced from the SKYLARK trial, which recruited predominately severe PND patients (see section 3.2.1.4.1). The EAG heard from their clinical advisors that zuranolone would be prescribed by perinatal psychiatrists for severe patients many of whom are likely to already be taking ADT, and would continue ADT whilst taking zuranolone. The EAG requested a subgroup analysis for the moderate and severe subgroups separately in clarification questions, but the company argued that they could not provide this because they did not have long-term EPDS trajectories from the ALSPAC cohort by severity. The company do however have trajectories from ALSPAC with 8-month EPDS score as a covariate, and they have predictions of 8-month EPDS from SKYLARK for given baseline EPDS score, and so could construct a model for the severe subgroup from SKYLARK. The sub-group analyses for HAMD-17 from the company submission (CS Figures 24-25) suggest that the treatment effect of zuranolone + ECM does not depend on severity, and so a subgroup analysis for the economic model can be constructed by using the results from simulated patients with higher baseline EPDS scores only. To explore the impact of this the EAG presents scenarios which only include the results from simulated patients with a baseline EPDS score >19 and which only include the results from simulated patients with a baseline EPDS score ≤19.

4.2.4 Interventions and comparators

The intervention is zuranolone 50 mg (two 25 mg capsules), taken once daily for 14 days, in addition to ECM. Dose reductions from 50mg to 40mg were included in the model, based on the proportion of dose reductions in SKYLARK. ECM is modelled to include 15.3% of patients taking concomitant ADT, based on the proportion with baseline ADT used in SKYLARK.

The EAG heard that zuranolone would be used in addition to ECM and so is happy with the way the company has modelled this. However, the EAG heard that most patients (>90%) would be already taking ADT before being referred to perinatal psychiatrists where zuranolone would be prescribed, and so the proportion of patients using ADT is likely to be

much higher than the 15.3% seen in SKYLARK. In clarification question B3 the EAG asked for a cost-effectiveness analysis for the subgroup of patients taking ADT. The company explained that a proper scenario analysis using the ADT subgroup from SKYLARK was not possible due to small numbers of patients. However, because they forced baseline ADT use as a covariate in their Mixed Model for Repeated Measures (MMRM) model that predicts EPDS scores at 3 months after initiating treatment, the company were able to provide results for a scenario where 100% of patients had baseline ADT use, which slightly increases the ICER (Table 10 of company's response too clarification questions). The EAG appreciates that the sample size from SKYLARK is too small to estimate a subgroup effect for ADT use, but notes that this does also reflect concerns with the generalisability of the results from the SKYLARK trial, as discussed in section 3.2.1.2. The EAG explores the impact of the level of baseline ADT used in the MMRM model that forces the inclusion of baseline ADT, with scenarios for % baseline ADT use, and uses 90% in its base-case. Note however, that these scenarios only capture the impact of ADT on EPDS, and not an interaction with zuranolone.

4.2.5 Perspective, time horizon and discounting

The economic evaluation took the perspective of the NHS and social care payer, as per NICE reference case.

The company applied a 3.5% discount rate to costs and health outcomes, in line with the NICE reference case. The company did not include a half-cycle correction in the model, as they reasoned that since there were minimal changes to EPDS scores over time, the gradient of change was not large enough to require half-cycle correction. The EAG agrees that this is unlikely to make a big difference to the results.

The time horizon for the model was 29 years based on ALSPAC data availability. Whilst the EAG agrees that the availability of ALSPAC data means the natural history EPDS scores for ECM can be estimated over the time-horizon, it has concerns about the assumed duration of the treatment effect of zuranolone + ECM. The treatment difference between zuranolone + ECM and ECM is extrapolated a long way from the available data from SKYLARK (45 days follow-up), making an implausibly strong assumption of continued treatment effect for over 28 years after treatment was given. The EAG explores various treatment waning assumptions in scenario analyses (see section 4.2.6).

4.2.6 Treatment effectiveness and extrapolation

Treatment effectiveness is captured in the model via EPDS trajectories over time (Figure 2) with a short-term model up to 3-months post-treatment initiation, and a long-term model thereafter (up to 29 years post-birth).

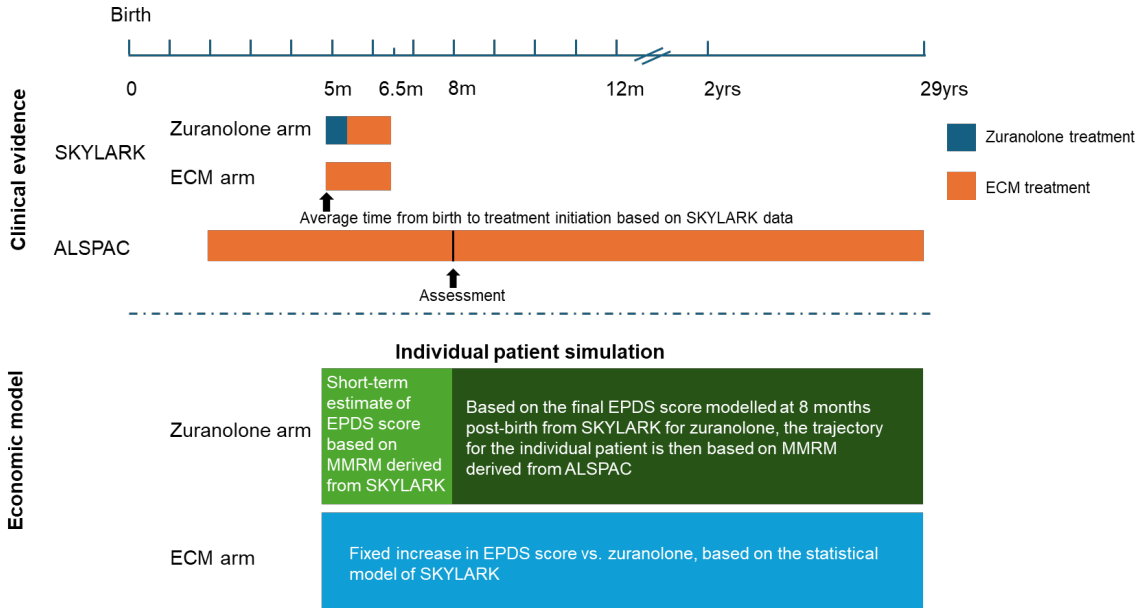
4.2.6.1 Short-term EPDS model and short-term extrapolation

The company built a prediction model based on the 1.5 month follow-up data from SKYLARK and then used it to predict EPDS trajectories in the short term up to 3 months post-treatment initiation. They fitted MMRM models to account for the repeated EPDS

observations, followed a structured variable selection process, and due to the small sample size used a 4-fold cross validation approach to validate the prediction model. Baseline EPDS, time, and baseline ADT use were "forced" into the model. A scenario which does not force baseline ADT use into the model was conducted, but only made a very small change to the ICER (CS Table 57). Non-linearity in EPDS score over time was captured by fitting a "broken-stick" shaped curve.

The EAG do not have any concerns with the approach to model fitting, variable selection and validation of the prediction model. However the prediction model extrapolates the covariate and treatment effects observed in the 1.5 month follow-up period of SKYLARK for a further 1.5 months, and this introduces some uncertainty as whether the treatment differences observed at 1.5 months would continue with the same trend in the subsequent 1.5 months. The company provides functionality in their Excel model to run a scenario where they instead use "last observation carried forward" (LOCF) which fixes the EPDS score at 1.5 months over the following 1.5 months, although they do not include results from this scenario in the CS. The EAG provides cost-effectiveness results for this (EAG scenario 4) in sections 6.1-6.2, which shows that it slightly increases the ICER.

Figure 3 Schematic of data used in the economic model (CS Fig. 26)



Abbreviations: ALSPAC: Avon Longitudinal Study of Parents and Children; ECM: Established Clinical Management; EPDS: Edinburgh Postnatal Depression Scale; m: months; MMRM: Mixed Model for Repeated Measures; yrs: years.

4.2.6.2 Long-term EPDS model and extrapolation

The company models EPDS scores beyond 3 months post treatment initiation (8 months post-birth) for zuranolone plus ECM using a long-term EPDS model. The long-term EPDS model is based on a model fitted to the ALSPAC longitudinal cohort of pregnant women with expected dates of delivery in 1991 to 1992 in Bristol, UK. A model was fitted to a sample of [redacted] women who had an EPDS score ≥13 at 2 months post-birth, an EPDS measure at 8 months post-birth as a baseline measure for the long-term model, and at least one further

EPDS measure. A similar model-fitting approach to the short-term prediction model was taken, using a MMRM model considering EPDS (8 months post-birth) and age as covariates. The statistical analysis plan indicates that no imputation or adjustments were made for missing data. The selected model included baseline EPDS but not age as covariates. The company makes mention of a scenario analysis including age, but the EAG could not find any results from this in the company's report and does not think it would impact the results (as it would capture the same effect as time for a given starting EPDS).

The EAG considers the lack of imputation or adjustment for missing data to be a limitation in this analysis, as there was substantial missing data (Table 2 of the ALSPAC statistical analysis plan). Lack of adjustment for missing EPDS data would introduce bias in the predicted EPDS trajectories. The EAG also has concerns about the relevance of data from pregnancies in 1991 – 1992, when ECM may have been different. However, the EAG acknowledges that this is the best long-term evidence available for moderate to severe PND patients, and that the cost-effectiveness results are unlikely to be sensitive to the actual trajectories other than through differences between the trajectories across treatments.

The company assumes that the difference in EPDS between treatment arms at 3 months post treatment initiation (which is 8 months post-birth), continues into the long-term for the full time horizon of the model (29 years post-birth). The EAG's clinical advisors found this assumption to be implausible, especially since zuranolone treatment is a single course treatment given for 14 days and there is no data on treatment efficacy beyond 45 days. In clarification question B1 the EAG asked the company to provide analyses to explore different treatment waning scenarios. In response, the company suggested that there were 2 ways that this could be applied in their model:

1. The EPDS scores for patients treated with zuranolone + ECM scores worsen with time and eventually converges with the ECM curve. The company argue that this lacks clinical plausibility and there is no reason why it would worsen in the zuranolone + ECM arm and not for ECM.

2. The EPDS scores for patients treated with ECM improve with time and eventually converges with the zuranolone + ECM curve. The company argue that this lacks clinical plausibility too and there is no reason why EPDS would improve for ECM but not also zuranolone + ECM arm.

In their response to clarification question B1 the company proposed an alternative way to model waning of treatment effect in their model by assuming patients would eventually reach remission. They modelled this by assuming that patients with an EPDS $\geqslant$13 at some fixed point in time would have their EPDS scores gradually decreased to 12.99 (classified as 'remission') over a period of 24 months, and this was applied to both arms. The resulting EPDS trajectories for a variety of fixed time points are show in Figure 4. The company state that this "results in a convergence of EPDS scores between zuranolone + ECM and ECM for

patients who have persistent PND". Whilst the curves for the two arms move a little closer together they still remain substantially separated for the duration of the model time-horizon. The ALSPAC data shows the EPDS trajectory for ECM at the time the ALSPAC cohort had PND. It does not show the trajectory for patients on zuranolone, and does not provide evidence on the comparative difference in EPDS trajectories over time for zuranolone vs ECM. Therefore the EAG believes the company's model makes a strong assumption that is not substantiated by the available evidence.

Figure 4 Average EPDS scores over time for the company's treatment waning scenarios in response to clarification questions

N.B. The y-axis is cropped for clarity. Mean baseline EPDS score is 21.1.

The EAG disagrees with the company's interpretation of the waning scenarios. It is easier to conceptualise this if the long-term model from the ALSPAC analysis is used to model ECM. The resulting EPDS trajectory is what the model predicts if zuranolone is not added to ECM, and this is left unchanged. The EPDS trajectory for zuranolone + ECM at 3 months post-treatment initiation is reduced by the EPDS difference from the short-term prediction model. But the treatment benefit from zuranolone does not last forever and will eventually diminish to no difference from the trajectory the patient would be on without zuranolone. At this point the two curves will come together. The EAG has implemented this where the long-term model is used for ECM, and the treatment effect of zuranolone + ECM is applied to the ECM curve with waning beginning at 3 months post-treatment initiation and set so the lines converge after a further 6 months, 1 year, 2 years, 5 years, or 10 years (EAG Scenarios 2a-e). These resulting average trajectories are shown in Figure 5. The EAG also

implemented further scenarios where waning starts at 6 months or 1 year post-treatment initiation, and lasts for 6 months or 1 year respectively (EAG Scenarios 2f-g). These resulting average trajectories are shown in Figure 6, and also in Figure 7 for easier comparison over a 3 year time-horizon. It can be seen that without waning (the company's base-case), the benefits of reduced EPDS for zuranolone continue throughout the 29 year follow-up, whereas for the various waning scenarios, the curves converge after a period of time to follow the trajectories from the long-term ALSPAC model. The EAG provides results for these scenarios in sections 6.1 and 6.2.

The extrapolation of the treatment effect is highly uncertain, however, the EAG prefers the scenario where waning begins 1 year (post treatment initiation), and lasts for 1 year, as this acknowledges that there may be treatment benefit beyond the period typically considered PND (1 year following birth), but that this would not continue for an implausibly long time.

Figure 5 Average EPDS scores over time for the EAG's treatment waning scenarios where waning starts at 3 months and lasts for between 6 months and 10 years

N.B. The y-axis is cropped for clarity. Mean baseline EPDS score is 21.1.

Figure 6 Average EPDS scores over time for the EAG's treatment waning scenarios where waning starts at 6 months and lasts 6 months, or starts at 1 year and lasts 1 year

N.B. The y-axis is cropped for clarity. Mean baseline EPDS score is 21.1.

Figure 7 Average EPDS scores over time for the EAG's treatment waning scenarios where waning starts at 6 months and lasts 6 months, or starts at 1 year and lasts 1 year (plotted over a 3 year period)

4.2.6.3 Immaturity of trials data informing selection of EPDS-based health states in the model

The ROBIN and SKYLARK trials informing the EPDS-based categorisation of health states in the model only have short-term 45day follow-up data. It is therefore not clear how a change in patients EPDS score relates to a change in clinical diagnosis of PND, or even subsequent MDD and chronic MDD. The company's trials are too short to inform the proportion of patients per group who would develop MDD and chronic MDD after PND over the 29 years duration of the model.

The company categorizes health states based on EPDS scores into 5 categories: EPDS=0, 1-6, 7-13, 14-18, 19-30. The company does not provide clinical justification for this categorisation or how it may relate to a clinical diagnosis of mild, moderate, or severe disease. For example, scores of 1-6 indicate no presence of disease, and therefore patients in this category may be no different from those in the EPDS=0 category but are treated differently in the model.

At clarification questions, the EAG queried how many patients in the company's trials provided data, in particular quality-of-life data, to inform parameters of the model in the five chosen EPDS categories and the company advised that those data were not available. The EAG notes this categorization could be considered arbitrary, as there is no evidence that a move between these categories would mean that patients would be clinically different, and that these categorisations would influence patients' treatment pathway or care.

In their clarification response, the company provided a scenario using 10 different EPDS categories, based on scores 0, 1-3, 4-6, 7-9, 10-12, 13-15, 16-18, 19-21, 22-24, 25-30. These were chosen after a review of the utility scores distribution so that no fewer than 10 observation informed each category. The EAG notes again that these categories are arbitrary and a movement between them would not necessarily mean a meaningful change in clinical diagnoses or treatment. The EAG chose to use the 10 categories for the EPDS in its base case because it provides a finer granularity for EPDS to correspond with the way EPDS is treated as a continuous variable in the companys model.

4.2.6.4 Adverse events

The company include adverse events that occur in more than 5% of patients from the SKYLARK, and use the proportions of patients that experience each for zuranolone + ECM and ECM from SKYLARK in the model. The company do not include adverse events for SSRIs since ADT use is assumed to be the same across both arms. The EAG considers the company's approach to modelling the frequency of adverse events to be reasonable.

4.2.6.5 Mortality

In their base-case the company applies general population mortality rates equally to both arms, as there is no evidence that treatment with zuranolone affects mortality. In a scenario analysis the company apply a standardised mortality ratio (SMR) to reflect an increased suicide/mortality risk for patients with an EPDS score of ⩾13. The SMR is sourced from a literature review for patients with major depressive disorders (MDD) compared to the general population, which ranged from 1.5 to 2. The company use an SMR of 2 in their scenario, which very slightly reduces the ICER.

The EAG agrees that there is no evidence of a direct effect of zuranolone on mortality, but considers it reasonable that there may be an indirect effect through changes in EPDS score. The SMR used by the company in their scenario is based on an old study and on patients with MDD, rather than PND, and so it is unclear whether this would apply to PND patients under current standard of care. However, the EAG feels it is appropriate to include the impact of EPDS on mortality, and include the SMR of 2 in their base-case.

4.2.7 Health related quality of life

The model required UK preference-based utility tariffs over the 29 year period where it captures the health benefits of the intervention. The company's base case modelled utilities by five EPDS categories with scores of 0, 1-6, 7-13, 14-18, and 19-30 based on ROBIN's patients' SF-6D utility scores published in O'Callaghan 2024.[26] After clarification questions, the company provided a scenario with EQ-5D utilities mapped from the PHQ-9 questionnaires, using Mukuria's 2025 mapping model.[28]

4.2.7.1 Face-validity of utility estimates in the PND period

The EAG noted a considerable difference in SF-6D utility scores between EPDS categories used in the model. The UK population norm for the EQ-5D utility scores in women 25-34 years is 0.93 and for the SF-6D is 0.803 for women aged 30-34 years.[29] The trial population reported a utility score of 0.935 for patients in the category of EPDS=0, which aligns with the population norm for EQ-5D scores but not SF-6D scores. The EAG queried at clarification questions how many patients informed the utility estimates derived for each EPDS category, but the company replied that they were not available. The drop in utilities from 0 to 1-6 in EPDS, where there is still no diagnosable mental health condition is notably large, from 0.935 to 0.795, and a further large drop to 0.659 for category score of 7-13 of very mild disease. The company did not provide evidence to support the face-validity of these estimates. There is no rationale for patients in category EPDS=0 to be clinically different from those in category EPDS 1-6 and this suggests that the value of 0.795, which is closer to the SF-6D population norm, could be applied to all patients with scores from 0 to 6, when measuring utility using the SF-36 instrument.

4.2.7.2 Choice of tool to estimate utilities

The NICE reference case recommends using utilities estimated from the EQ-5D tool. At clarification questions, the company explained that although both the EQ-5D and the SF6D

tools capture the impact of mental health, [30, 31] the SF36 may be better suited to capture depressive symptoms and is superior to the EQ-5D-3L in terms of validity and responsiveness. The EAG recognizes that the SF36 is a widely used tool to measure HRQoL in patients with mental health conditions, but would argue that this is less the case for depression and anxiety, for which the EQ-5D has a dedicated domain, and EQ-5D is an appropriate measure. The EAG also notes that Stolz[31] is comparing the 3 level EQ-5D tool against the SF36 tool, not the 5L tool, a more recent version of the EQ-5D and widely recognised as more sensitive than the 3L version to measure HRQoL.

The EAG requested sensitivity analyses to the utility estimates at clarification questions and suggested using mapping models in the literature to map patients' responses to disease-specific questionnaires in the SKYLARK and ROBIN studies to EQ-5D-3L utilities. The company provided additional estimates in response by mapping responses to the PHQ-9 in the ROBIN study to EQ-5D-3L utilities. Although data were available for the SKYLARK study, the company restricted to responses in the ROBIN study to match with CS base case which uses SF-6D utilities from the ROBIN study. The EAG feels that it would have been beneficial to obtain mapping scores from both studies where available. The company provided EQ-5D scores by EPDS category using 10 different categories of the EPDS scale rather than the initial 5 (see section 4.2.6.3). The new categories were defined using visual inspection of histograms and allowing a minimum of 10 observations per category. The EAG notes this categorization of the EPDS scores is arbitrary, and not based on clinical evidence or validated in the literature (see section 4.2.6.3) and it is not clear how changes between categories can impact PND or MDD diagnoses and follow-on treatment.

The company argued that their preferred base case utility values are those derived from the SF36, because it best captures broader aspects of health when compared to the PHQ-9, including physical functioning, pain, and general health perception, as well as social functioning, energy/fatigue, and role limitations, and therefore more sensitive and best placed to inform the model utilities than the PHQ-9. The EAG argues that, unlike the EQ-5D, the SF36 is not designed to estimate utilities for economic evaluation. Although the SF36 may capture other aspects of quality of life beyond health-related quality of life, it produces different utility estimates compared with the EQ-5D tools, it is not NICE's preferred tool to estimate utilities, and reduces the comparability of QALY gains across a multitude conditions when others have been based on EQ-5D utilities. The PHQ-9 is a questionnaire designed and validated to measure depressive symptoms and responsiveness. It is widely used to capture depressive symptoms in PND, and is recommended by NICE guidelines, along with the EPDS, to diagnose and monitor PND. [32] The EAG therefore prefers to use the EQ-5D mapped utilities from the PHQ-9 in the ROBIN trial in its base case to capture the effects on quality of life of the intervention for the duration of their PND period.

4.2.7.3 Analysis of PND utility estimates

The CS estimates SF6D utility scores by EPDS and by pooling estimates from patients in both groups of the trial for the same EPDS category. This assumes that zuranolone would not

affect the quality of life of patients within a category of the EPDS scale; rather it only affects the proportion of patients in these categories. The EAG has not found evidence to suggest this and requested at clarification questions that SF6D scores were adjusted by treatment group. The CS also does not adjust estimates by baseline utility scores as recommended in the literature,[33] which is particularly relevant if there were baseline imbalances between the groups. Finally, the EAG notes that ROBIN is a trial with a relatively small sample size, and some of the utility estimates by category may have been estimated based on very few patients in that category. The EAG requested information on the frequency of questionnaires administered, completion rates, and number of patient questionnaires informing the utility estimates in each question. In response to clarification questions, the company advised that utility values were obtained from the O'Callahan study and information on sample size of respondents per category was not available.

4.2.7.4 Choice of utility tariffs used

The CS used health-related quality of life data from the ROBIN trial, measured by the SF 36 questionnaire to estimate the SF6D score. This score was derived by applying Brazier's UK preference-based weights to estimate utilities.[29] The choice of SF36 questionnaire over the more recent SF36v2 questionnaire was not discussed in the CS. The EAG notes there are differences between the tariffs provided by the SF6D and the SF6Dv2, particularly for patients with mental health conditions,[34] and the CS may have benefitted from justifying the choice of tool in the trials.

4.2.7.5 Utility estimates beyond the PND period

The company's model applies the same utility estimates based on the five EPDS categories for the entire 29 years duration of the model. The differences in utility scores across EPDS categories were based on the ROBIN trial over a very short 45-day follow-up and unlikely to be sustained over 29 years.

The EAG broadly agrees that the results from ROBIN and SKYLARK could be extrapolated to inform quality of life changes in the PND period. Clinical advisors to the EAG suggested that, by convention, PND would last for one year after birth. After that, PND and MDD symptoms are similar and PND would be treated as MDD. The EAG therefore considers it appropriate to use utilities from an MDD population to inform the modelled utilities beyond the PND period. Although the company's SLR did not identify sources of long-term PND utilities, there are estimates of utilities in MDD and chronic MDD in the literature, which could be used in the model to inform quality of life gains beyond the PND period. Health-state utilities for MDD are typically reported separately for "remission" and "non-remission", which aligns with the way the companys model costs where patients are classed as in "remission" if their EPDS score <13, and in "non-remission" if their EPDS≥13. The EAG searched for evidence informing Clinical/NICE guidelines for PND (CG192)[4] and MDD (NG222)[35] and used the NICE committee's preferred utility estimates from TA367 to inform utilities in the "remission"(0.85) and "non-remission" (0.67) health states after PND.[36] The EAG provides a scenario where the MDD utilities are applied immediately after the PND

period (1 year after birth), and a scenario where the MDD utilities are applied 1 year following treatment initiation to coincide with the point at which treatment effect waning is applied, and uses the latter in the EAG base-case.

The EAG acknowledges there are caveats in using this approach. TA367 is an appraisal of a recent antidepressant drug for MDD, not PND. It is largely based on EQ-5D utilities from industry trials (the REVIVE study of vortioxetine) and Sapin et al study,[37] a French study reporting EQ-5D utility values using UK tariffs in MDD patients. Ideally, to inform these longer-term health states of this model, there would have been a systematic review of all MDD and chronic MDD studies collecting utilities data, and a synthesis of UK utility estimates found using meta-analysis. The EAG also acknowledges that they may not be a clear linear link between what clinicians classify as "remission" and "non-remission" and the 13 point EPDS score cut-off. This would mean that all patients with an EPDS above or below 14 would be treated the same, which is a common over-simplification in economic models of depression. Nevertheless, the EAG prefers to use TA367 utilities to model health-related quality of life gains at one year post-intervention because they are derived using EQ-5D utilities, rather than SF6D utilities, as per NICE reference case, it dichotomises health states in "remission" and "non-remission" as traditionally used in the MDD literature, clinical experts have advised that PND at one year post-birth is similar to MDD, and the utility values picked were the NICE committee's preferred values in a recent MDD appraisal.

4.2.8 Resources and costs

The economic evaluation captured costs for 29 years after birth using an NHS and social care payer perspective. The timeframe for the economic analysis was critiqued in section 4.2.5. The model includes the resource use and costs for:

- Treatment with zuranolone and ECM, including zuranolone acquisition costs and costs with concomitant ADTs like selective serotonin reuptake inhibitors (SSRIs)
- Other health and social care resource use needs informed by estimates in the literature

The company performed a SLR of studies reporting resource use and costs that would inform the health states of the model. The EAG notes that this SLR failed to pick up important studies, such as Petrou et al 2002, that the company later uses in the model. The SLR could have also been expanded to include studies in MDD and chronic MDD which could have informed health care resource use and costs after the PND period.

4.2.8.1 Resource use and costs of treatment with zuranolone and concomitant SSRIs

The company presented treatment acquisition costs with zuranolone, with a discount applied. Given that this discount has not yet been agreed, the EAG was requested to use the full list price in this report. The zuranolone 50mg treatment lasts two weeks (two daily pills of 25mg) but it is possible that some patients may require a dose reduction to 40mg. [redacted][38] The company does not include costs associated with delivery of zuranolone treatment. The EAG's clinical advisors agreed that patients eligible for treatment would have been referred

to specialised secondary care services and prescribing zuranolone treatment would not impose an additional outpatient care visit. The company assumes that [redacted] of patients would require a dose adjustment and therefore incur an additional outpatient visit and an additional pack of zuranolone. The EAG accepts the company's assumption in its base case.

SSRIs included were sertraline ([redacted] of patients based on CPRD data analysis) and fluoxetine ([redacted]). Costs for SSRIs were sourced from the electronic Market Information Tool (eMIT) national database and similar to the costs provided in the BNF. Clinical advisors to the EAG suggested other ADTs may be used and the assumed proportion of patients on each SSRI from historical CPRD data may different in current practice. These costs, however, are very small and so alternative costing assumptions would not have a noticeable impact on the ICER. The EAG therefore uses the company's estimates of SSRIs costs in its base case. The EAG disagrees that only 15.3% of patients eligible for zuranolone treatment would be on ADTs at baseline (section 4.2.4), and assumes 90% in its base case.

The EAG also notes that patients may be on other treatments at baseline, like talking therapies, which would be considerably more expensive than SSRIs and increase the costs with ECM. The proportion of patients on non-pharmacological therapies at baseline was not available and given that these costs would be applied equally to both groups compared, including these costs would have no impact on the ICER.

4.2.8.2 Other resource use and costs applied in the model

Additional health and social care resource use throughout the 29 years duration of the model was dichotomised based on EPDS score of <13 as "PND" and ≥13 as "remission". Resource use in the first 18 months post birth was informed by annualized cost estimates reported in Petrou et al (2002)'s study, picked up in an ad hoc search of the literature.[38] The EAG agrees this is a relevant and comprehensive study of PND costs in a UK population, albeit old and no longer reflecting current practice. The company recognised this and adjusted some of the resource use assumptions based on the company's clinical advisors' opinion. The EAG is broadly in agreement with the resource use provided and adjustments made. The company further updates the unit costs for health and social care using recent 2023/24 estimates, which are a year old but still adequate.

To inform resource use and costs beyond 18 months post-birth, the company uses Byford et al 2011 study.[39] The company acknowledges the limitations of using these estimates, namely that this is a study in MDD and not PND, and that it is old study which may not represent current practice, and severity of depression is not measured using the EPDS score, whereas the company uses the cutoff of EPDS=13 to classify those in "remission" and "non-remission". Byford's 12 months resource use estimates is broken down into 3 month costs and applied equally for the remaining 27.5 years of the model duration.

Byford 2011 is a large cohort study of 88,935 adults with MDD and at least 3 ADT prescriptions followed up for 12 months in a large UK primary care database between 2001 and 2006. The EAG disagrees that these estimates of differences in costs observed in Byford's 12 months follow-up could be extrapolated for the remainder or the model

duration. For example, Byford's study reports a difference of nearly 5 GP visits and 3 GP phone calls between "remitters" and "non-remitters", which drives most of the cost difference. The EAG does not agree this difference would be perpetuated over 27.5 years without a waning of effect. The EAG further notes that patients in Byford's study are no longer patients treated in secondary care, as our PND study population. It is unclear whether severe PND patients whose symptoms had not yet resolved by 18 months post-birth, would have continued their treatment in secondary care rather than in primary care, as in Byford's study population. The patients in Byfords study population are all using ADTs for at least some of the time, and it is unclear whether the zuranolone study population would be on ADTs 18 months after birth. As the company recognises, these costs are based on resource use from the UK practice in 2001-2006, and practice may be different now. The company could have adjusted resource use to current clinical practice based on clinical expert opinion, as it did with Petrou's 2002 study. There is also no clinical justification for applying "no remission" costs to patients with EPDS>13 in the model, a limitation the company acknowledges.

Despite these significant limitations, the EAG notes that this relatively large annual difference in costs perpetuated over 27.5 years in the company's model, will have a very small impact on the ICER in the EAG's base case. In its base-case the EAG applies a waning of the treatment effects from 18 months post-birth (section 4.2.6.2), which affects utilities and costs, thus drastically reducing the impact of Byford's cost estimates on the ICER. For this reason, the EAG is happy to use the company's cost estimates in the 18 months post birth.

4.2.8.3 Costs with adverse reactions

Costs with adverse reactions were included in the model. They included somnolence, dizziness, diarrhoea, fatigue, and urinary tract infections. These are applied as a one-off basis for each incidence and assumed a GP visit or antibiotic use for the UTI. Clinical advisors to the EAG suggested that patients on zuranolone could be at risk of decreased consciousness and there may be potential for abuse. These are rare but serious adverse events currently observed in patients on brexanolone, a similar drug for PND in current use. However, the EAG notes these were not observed in the zuranolone studies. These adverse events, although rare, could carry substantial long-term costs, although their impact on the ICER are likely to be small.

5 COST EFFECTIVENESS RESULTS

All results in this section use the patient access scheme price for zuranolone with a simple discount of [redacted] and cost per a course of zuranolone of [redacted] after net discount.

5.1 Cost effectiveness results for company's base case

The deterministic results for the company's base case are shown in Table 15. The model was run with 1,000 patients to ensure model convergence (CS Figure 31). The incremental costs

for zuranolone with ECM compared to ECM alone were [redacted], and the incremental QALYs were [redacted], resulting in an ICER of £12,704 per QALY.

Table 15 Deterministic results for company's base case

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALY gain	ICER (£ per QALY)
Deterministic (1,000 patients)							
ECM alone	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	12,704
Zuranolone with ECM	[redacted]	[redacted]	[redacted]				

Costs expressed in 2024 GBP (£).
Abbreviations: ECM: Established Clinical Management; GBP: Great British Pounds; ICER: Incremental Cost-Effectiveness Ratio; LYG: Life Years Gained; QALY: Quality-Adjusted Life Year.

The probabilistic results for the company's base case are shown in Table 16. The model was run for 1,000 iterations of 1,000 to ensure model convergence (CS Figure 34). The probabilistic results are very similar to the deterministic results. There is a high level of variation in the incremental QALYs, resulting in a wide 95% CI for the ICER.

Table 16 Probabilistic sensitivity analysis results for company's base case

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALY gain	ICER (£ per QALY)
PSA (1,000 patients, 1,000 iterations)							
ECM alone	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	12,706
Zuranolone with ECM	[redacted]	[redacted]	[redacted]				
95% CI (lower, upper)							
ECM alone	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	(5,927, 63,723)
Zuranolone with ECM	[redacted]	[redacted]	[redacted]				

Costs expressed in 2024 GBP (£).
Abbreviations: CI: Confidence Interval; ECM: Established Clinical Management; ICER: Incremental Cost-Effectiveness Ratio; LYG: Life Years Gained; PSA: Probabilistic Sensitivity Analysis; QALY: Quality-Adjusted Life Year.

5.2 Company's sensitivity analyses

5.2.1 Deterministic sensitivity analyses

The company conducted a range of one-way deterministic sensitivity analyses (DSAs) using the 2.5% and 97.5% percentiles of the underlying distributions for each parameter included

in the DSAs. The DSA results for the parameters with greatest impact on the ICER are shown in Figure 8 and Table 17.

Figure 8 Top 10 parameters varied in base case DSA

Abbreviations: ADT: Antidepressant; ALSPAC: Avon Longitudinal Study of Parents and Children; DSA: Deterministic Sensitivity Analysis; EPDS: Edinburgh Postnatal Depression Scale; GP: General Practitioner; ICER: Incremental Cost-effectiveness Ratio; MMRM: Mixed Model for Repeated Measures; QALY: Quality Adjusted Life-Year; SF-6D: Short Form Health Survey – 6 Dimensions.

Table 17 Top 10 DSA results

#	Parameter	Variation	Inc. costs (£)	Inc. QALYs	ICER (£/QALY)
Base case			[redacted]	[redacted]	12,704
1	SKYLARK MMRM: On zuranolone ([redacted], [redacted])	Lower	[redacted]	[redacted]	6,544
		Upper	[redacted]	[redacted]	77,729
2	SF-6D EPDS 1 to 6 ([redacted], [redacted])	Lower	[redacted]	[redacted]	13,732
		Upper	[redacted]	[redacted]	11,854
3	ALSPAC MMRM: Time [redacted], [redacted])	Lower	[redacted]	[redacted]	11,904
		Upper	[redacted]	[redacted]	13,693
4	SKYLARK MMRM: Age ([redacted], [redacted])	Lower	[redacted]	[redacted]	12,069
		Upper	[redacted]	[redacted]	13,345
5	ALSPAC MMRM: Time^2 ([redacted], [redacted])	Lower	[redacted]	[redacted]	12,063
		Upper	[redacted]	[redacted]	13,106
6	SF-6D EPDS 14 to 18 ([redacted], [redacted])	Lower	[redacted]	[redacted]	12,215
		Upper	[redacted]	[redacted]	13,232
7	Unit costs: GP visits (£34, 74)	Lower	[redacted]	[redacted]	12,942
		Upper	[redacted]	[redacted]	12,415
8	SF-6D EPDS 7 to 13 ([redacted], [redacted])	Lower	[redacted]	[redacted]	12,460

		Upper	[redacted]	[redacted]	12,956
9	Baseline EPDS score (20.58, 21.62)	Lower	[redacted]	[redacted]	12,634
		Upper	[redacted]	[redacted]	13,115
10	Age (29.67, 31.33)	Lower	[redacted]	[redacted]	13,026
		Upper	[redacted]	[redacted]	12,741

Abbreviations: ALSPAC: Avon Longitudinal Study of Parents and Children; DSA: Deterministic Sensitivity Analysis; EPDS: Edinburgh Postnatal Depression Scale; GP: General Practitioner; ICER: Incremental Cost-effectiveness Ratio; INMB: Incremental Net Monetary Benefit; MMRM: Mixed Model for Repeated Measures; QALY: Quality-Adjusted Life-Year; SF-6D: 36-item Short Form Health Survey.

5.2.2 Deterministic scenario analyses

The company performed deterministic scenario analyses to explore the impact of the model assumptions and input parameters on the results. These scenarios are described in Table 18, and the results are presented in Table 19.

Table 18 Description of the company's scenario analyses (CS Table 56 plus additional scenarios from CQ response)

#	Parameter	Base case	Scenario
S1	**Mortality**	No additional mortality for patients with PND i.e., general population mortality applied for all patients	Additional mortality is applied for patients who have PND (CS section 3.3.6)
S2	**ADT use at baseline**	The short-term EPDS model from SKYLARK forces the inclusion of ADT use at baseline	The short-term EPDS model with no forced ADT at baseline is applied, and the costs associated with ADT are removed (CS section 3.3.2).
S3	**HRU inputs**	HRU resource use is sourced from Petrou et al. (2002)	The data from Petrou et al. (2002) has been augmented with feedback from clinicians to reflect current clinical practice (CS section 3.5.3)
S4	**PND remission threshold (EPDS)**	An EPDS of <13 indicates PND remission	The remission threshold is changed so that an EPDS of <10 indicates PND remission (CS section 3.2.2).
S5	**Model time horizon**	29 years to align with the long-term data from ALSPAC	Lifetime time horizon of 60 years (CS section 3.2.2). The EPDS score at year 29 is carried forward for the duration of the time horizon.
S6	**Dose reductions**	Additional pack of zuranolone is applied to [redacted]% of patients who require a dose reduction	No dose reduction is applied (CS section 3.5.2)
S7	**EPDS after 10 years**	EPDS scores follow base case MMRM long-term model	EPDS is decreased to 12.99 over 24 months for patients who have not achieved remission after 10 years (CQ response B1)
S8	**EPDS after 5 years**	EPDS scores follow base case MMRM long-term model	EPDS is decreased to 12.99 over 24 months for patients who have not

			achieved remission after 5 years (CQ response B1)
S9	**EPDS after 2 years**	EPDS scores follow base case MMRM long-term model	EPDS is decreased to 12.99 over 24 months for patients who have not achieved remission after 2 years (CQ response B1)
S10	**Baseline ADT use**	15.3% of patients on ADT, with ADT implemented as a covariate in the MMRM	100% of patients on ADT, with ADT implemented as a covariate in the MMRM (CQ response B3)
S11	**EPDS utility mapping approach**	ROBIN SF-6D	EPDS utility values mapped from ROBIN PHQ-9 to EQ-5D (CQ response B10)

Abbreviations: ADT: Antidepressant Treatment; ALSPAC: Avon Longitudinal Study of Parents and Children; EPDS: Edinburgh Postnatal Depression Scale; HRU: Healthcare Resource Utilisation; ICER: Incremental Cost-Effectiveness Ratio; PND: Postnatal Depression;

Table 19 Company's deterministic scenario analysis results, comparing Zuranolone with ECM vs ECM alone

#	Scenario	Inc. costs (£)	Inc. QALYs	ICER (£/QALY)
Base case		[redacted]	[redacted]	**12,704**
S1	Excess mortality for patients with PND	[redacted]	[redacted]	12,415
S2	0% ADT at baseline (alternative short-term model)	[redacted]	[redacted]	12,736
S3	HRU from Petrou + expert opinion	[redacted]	[redacted]	12,566
S4	EPDS < 10 for classifying depression	[redacted]	[redacted]	12,382
S5	Lifetime time horizon (60 years)	[redacted]	[redacted]	10,282
S6	No dose reduction	[redacted]	[redacted]	12,272
S7	EPDS is decreased to 12.99 over 24 months for patients who have not achieved remission after 10 years	[redacted]	[redacted]	15,227
S8	EPDS is decreased to 12.99 over 24 months for patients who have not achieved remission after 5 years	[redacted]	[redacted]	16,526
S9	EPDS is decreased to 12.99 over 24 months for patients who have not achieved remission after 2 years	[redacted]	[redacted]	17,568
S10	100% of patients on ADT with ADT implemented as a covariate in the MMRM	[redacted]	[redacted]	13,400
S11	EPDS utility values mapped from ROBIN PHQ-9 to EQ-5D	[redacted]	[redacted]	14,600

Abbreviations: ADT: Antidepressant Treatment; ECM: Established Clinical Management; EPDS: Edinburgh Postnatal Depression Scale; HRU: Healthcare Resource Utilisation; ICER: Incremental Cost-effectiveness Ratio; INMB: Incremental Net Monetary Benefit; PND: Postnatal Depression; QALY: Quality-Adjusted Life-Year.

5.3 Model validation and face validity check

The EAG applied a validation check based on the TECH-VER Model Verification Checklist framework developed by Büyükkaramikli *et al.* (2019).[40] This included checking the model results adhered to 'Black-box tests' and where potential errors were identified in the black-box tests the EAG underwent the 'White-box tests' of the TECH-VER framework. This involves checking the detailed model calculations, running through the related code or by scrutinizing the formulae in the relevant ranges in a spreadsheet, cell by cell. In addition to these formal checks, further white-box tests were completed by the EAG where anomalies were identified when navigating the model to understand the model calculations. Failed black box tests are described in Table 20.

The EAG also identified some errors in the Excel model code, and these are described in Table 21. The impacts of these errors on the results were minor, except for the incorrect method used to calculate the 95% CIs for the probabilistic sensitivity analysis (PSA) results, where the company's method gave much narrower intervals than the correct method.

Table 20 Failed black box tests from the TECH-VER checklist

TECH-VER black box test	Expected result	Result	Reason for discrepancy after white box testing
Change around the effectiveness-, utility- and safety-related model inputs between two treatment options	Accumulated life-years and QALYs in the model at any time should also be reversed	The new QALY values are very close to the reversed QALYs values, but not identical.	The slight discrepancy is likely to be due to the order in which the model constrains the EPDS to be between 0 and 30. The EAG does not consider this be an error.
Patient-level model specific: Check if correlation in patient characteristics is taken into account when determining starting population	Yes	The only patient-level baseline characteristics in the model are age, baseline EPDS score, and ADT use. These are not correlated in the model, and this is consistent with the Company Submission.	The company's model makes a simplifying assumption that patient characteristics are independent, but this is unlikely to be the case. The EAG does not expect this to have a big impact on the cost-effectiveness results.
Do the explored scenario analyses provide a balanced view on the structural uncertainty (i.e. not always looking at more optimistic scenarios)?	Yes	Company did not include some subgroup analyses that were requested by the EAG (see section 4.2.3).	N/A

Check if the one-way sensitivity analysis (OWSA) includes any parameters associated with joint uncertainty (e.g. parts of a utility regression equation, survival curves with multiple parameters)	No	Regression parameters from the MMRM model for short-term EPDS are estimated jointly, but are included separately in the OWSA	N/A

Table 21 Errors identified in Company's original Excel model

Description	Corrections made to Excel model
Minor error in patient_simulations sheet where cells were subtracted from QALY values. This did not affect results since the incorrectly referenced cells were empty.	Cells AL90 to AL94 of the patient_simulations sheet were updated to remove reference to cells AMN90 to AMN94. This was addressed in Clarification Question B17.
Wrongly positioned brackets when calculating weighted ADT costs in cells T19 and O26 of the patient_simulations sheet. This error had minimal impact on the results due to the low cost of ADTs.	The bracket positions in T19 and O26 of the patient_simulations sheet were corrected. These changes were addressed in Clarification Question B18.
Minor error in patient_simulations sheet where a "/12" was missed in cells M90 and AI90 of the patient_simulations sheet when converting time to months. This error did not affect results since this part of the calculation is multiplied by the treatment indicator which is set to FALSE at that timepoint in cell H90.	M90 and AI90 of the patient_simulations sheet were updated to include the missing "/12".
Errors in mean EPDS score calculations on ECM in EPDS_simulations sheet: the mean scores in column E were calculated over columns ALW to BHY rather than ALW to BYH, and hence only the first 575 simulated ECM patients were included in the plots of mean EPDS scores rather than the full 1000 simulated ECM patients.	The calculations in column E of the EPDS simulations sheet were corrected to reference columns ALW to BYH.
The method used by the company in the psa_plot sheet of their model to calculate the 95% CIs for the PSA results is incorrect. This resulted in much narrower 95% CIs than the correct method.	The EAG calculated the 95% CIs for the PSA results by taking the 2.5% and 97.5% percentiles over the 1000 PSA samples.

6 EVIDENCE ASSESSMENT GROUP'S ADDITIONAL ANALYSES

6.1 Exploratory and sensitivity analyses undertaken by the EAG

The EAG explored the following scenarios:

1. re-parametrisation of the long-term EPDS model to represent ECM on which treatment effects for zuranolone with ECM were applied
2. different treatment effect waning scenarios,
3. the proportions of patients on ADT at baseline
4. using "last observation carried forward" to extrapolate short-term EPDS
5. subgroup analyses by PND severity
6. assumed different utility values for all health states of the model

These are described in more detail in the sections below. All analyses were deterministic.

6.1.1 Re-parameterisation of long-term EPDS model

The company's model uses the ALSPAC MMRM model to calculate the long-term EPDS scores for the zuranolone with ECM arm, and then calculates the EPDS scores for the ECM arm by subtracting the zuranolone treatment effect from those scores. As discussed in section 4.2.2, the EAG feels that the ALSPAC cohort better represents the ECM arm rather than the zuranolone with ECM arm since zuranolone would not have been used in ALSPAC. The EAG therefore re-parameterises the model so that the ALSPAC MMRM model is used to calculate the EPDS scores for the ECM arm, and the EPDS scores for the zuranolone with ECM arm are calculated by adding the zuranolone treatment effect to the ECM arm scores. In both the company's model and the EAG's reparameterisation, the treatment effect for zuranolone with ECM vs ECM alone in the long-term MMRM model is applied as a constant, which is calculated as the difference between the two arms' EPDS scores at 8 months.

- Scenario 1: ALSPAC MMRM model used to calculate long-term EPDS scores for ECM arm rather than Zuranolone with ECM arm

These changes were made in the Excel model by editing cells L91:L205 and AH91:AH205 of the patient_simulations sheet.

This scenario is mathematically equivalent to the company's parameterisation, however results differ slightly due to the order in which the EPDS scores are constrained to be between 0 and 30, but this has minimal impact on the results (Table 22).

6.1.2 Zuranolone treatment effect waning

The company's model holds the zuranolone treatment effect size constant across the 29 year time horizon of the model, however as discussed in section 4.2.6, the EAG and clinical advisors consider this assumption implausible. To explore the impact of this, the EAG has adapted the model to allow for waning of the treatment effect. To achieve this, the EAG applied the treatment waning scenarios to the re-parameterised long-term EPDS model

described in section 6.1.1, which is mathematically equivalent to the company's model and only differs slightly due to the way the EPDS constraints are applied.

The EAG explores scenarios in which the treatment effect wanes linearly to zero, with waning starting 3 months post-treatment initiation (i.e. waning starts at 8 months post birth). This is the point at which the model moves from the short-term MMRM model based on SKYLARK to the long-term MMRM model based on ALSPAC. The EAG ran scenarios where the waning lasts various lengths, from 3 months to 10 years. The EAG also ran scenarios where waning started at 6 months or 1 year post-treatment initiation, and lasted for 6 months or 1 year respectively.

- Scenario 2a: Waning starts 3 months post-treatment initiation, for 6 months
- Scenario 2b: Waning starts 3 months post-treatment initiation, for 1 year
- Scenario 2c: Waning starts 3 months post-treatment initiation, for 2 years
- Scenario 2d: Waning starts 3 months post-treatment initiation, for 5 years
- Scenario 2e: Waning starts 3 months post-treatment initiation, for 10 years
- Scenario 2f: Waning starts 6 months post-treatment initiation, for 6 months
- Scenario 2g: Waning starts 1 year post-treatment initiation, for 1 year

These scenarios were implemented by storing the treatment effect waning multipliers in column V of the patient_simulations sheet, and applying them to the long-term EPDS score calculations for the zuranolone with ECM arm in column L.

6.1.3 Baseline ADT use

The company's base case assumes that 15.3% of patients are on ADT at baseline, but as discussed in section 4.2.4 the EAG heard that a higher proportion (>90%) would be taking ADT in practice. The company also ran scenarios where they set 0% and 100% of patients to be on ADT at baseline (company scenarios S2 and S10 respectively). ADT use at baseline is included as a parameter in the short-term MMRM model.

The EAG ran scenarios with different percentages of patients on ADT at baseline:

- Scenario 3a: 50% of patients on ADT at baseline
- Scenario 3b: 80% of patients on ADT at baseline
- Scenario 3c: 90% of patients on ADT at baseline

These scenarios were implemented by changing cell C12 of the clinical sheet.

6.1.4 Short-term extrapolation of EPDS

In the company's base case, they assume that the slopes of the treatment differences observed at 5.5 months post birth would continue with the same trend until 8 months (see section 4.2.6.1). The company had provided functionality in their model to explore the

impact of using last observation carried forwards (LOCF), but had not provided results for this. The EAG ran the LOCF scenario, where the month 6.5 scores are reused at month 8.

- Scenario 4: use month 6.5 EPDS scores at month 8

This scenario was implemented by selecting LOCF in cell F14 of the patient_simulations sheet.

6.1.5 PND severity

As discussed in section 4.2.3, the EAG heard that zuranolone would be prescribed by perinatal psychiatrists to patients with severe depression. The EAG conducted scenario analyses where the modelled patients are split into two groups according to their baseline EPDS scores.

- Scenario 5a: Population with severe EPDS scores (> 19) at baseline
- Scenario 5b: Population with moderate EPDS scores (≤ 19) at baseline

These scenarios were implemented by calculating the cost-effectiveness results for these two patient subgroups separately. This required sampling EPDS scores from a Normal distribution truncated at EPDS = 19.

The more straightforward way to model this would be to increase the overall sample size, and only use results from patients with severe (or moderate) EPDS scores. However, the structure of the Excel model only allows a maximum of 1000 patients to be simulated, which gives too few simulations for each subgroup. Instead, since the sampling of the EPDS scores relies on saved samples from the Uniform distribution, the EAG transformed these values so that they would map to a truncated Normal distribution. These calculations are saved in cells H17:J17 of the patient simulations sheet in the EAG model.

6.1.6 Utilities

- Scenario 6a: Using a utility estimate of 0.795 for patients with EPDS category of 0, equalizing to category 1-6 and better align with the population norm for the SF-6D
- Scenario 6b: Using EPDS utility values mapped from ROBIN PHQ-9 to EQ-5D (as per company's sensitivity analysis after clarification questions) for the PND period and a further 7 months (up to 12 months post treatment initiation, i.e. 17 months post birth), and using MDD-specific EQ-5D utilities based on TA367 for the period thereafter.
- Scenario 6c: Using EPDS utility values mapped from ROBIN PHQ-9 to EQ-5D (as per company's sensitivity analysis after clarification questions) for the PND period (up to 12 months post birth) and using MDD-specific EQ-5D utilities based on TA367 for the period thereafter.

Scenario 6a was implemented by setting cell C24 of the utilities sheet to be 0.795.

Scenarios 6b and 6c were implemented by duplicating the utility_matrix sheet and calling the duplicate sheet "mdd_utility_matrix". The new mdd_utility_matrix sheet was edited to use MDD utilities (0.85 for "remission" and 0.67 for "non-remission", see section 4.2.7.5) instead of PND utilities. In scenario 6b, cells N93:N328 and AJ93:AJ328 (i.e. 1 year post-treatment initiation onwards) of the patient_simulations sheet were then updated to draw utilities from the mdd_utility_matrix sheet rather than the utility_matrix sheet. In scenario 6c, cells N92:N328 and AJ92:AJ328 of the patient_simulations sheet were updated to draw utilities from the mdd_utility_matrix sheet rather than the utility_matrix sheet, and cells N91 and AJ91 used a weighted average from both sheets. The use of the EQ-5D utilities (instead of SF-6D) for the PND period was selected using the dropdown menu in cell C20 of the utilities sheet.

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

Table 22 reports the results of the EAG's scenario analyses described in section 6.1. The ICER is most sensitive to assumptions about the waning of the treatment effect (Scenario 2). Applying waning effects is likely to increase the ICER, increasing from £12,704 with no waning to £49,103 with waning over a 10-year period (Scenario 2e) to £434,254 with waning over a 6-month period (Scenario 2a). The ICER is also sensitive to the use of MDD-specific utilities in the long-term following the PND period (Scenario 6b) which gave an ICER of £22,939 per QALY gained.

We note that changes in severity of PND at baseline (scenarios 5) and ADT use at baseline (scenarios 3) did not change the ICERs in a meaningful way, however, note these analyses did not allow for any subgroup-specific treatment effect. Also, there is a bigger impact of these scenarios when added to the EAG treatment waning assumption (see section , 6.3, Table 23).

Given that the company's deterministic and probabilistic base-case gave very similar results, we would expect the same overall findings of the EAGs deterministic scenario analyses to also hold in probabilistic analyses.

Table 22: Results of the EAG's scenario analysis (deterministic results)

Scenario	ECM		Zuranolone with ECM		Incremental costs (£)	Incremental QALYs	ICER (£/QALY)	Change from base-case ICER (%)
	Total costs (£)	Total QALYs	Total costs (£)	Total QALYs				
Company base-case	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	12,704	-
Scenario 1 (re-parameterisation of long-term MMRM model)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	12,725	0.17%
Scenario 2a (treatment effect waning starts 3 months post-treatment initiation, for 6 months)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	434,254	3318.25%
Scenario 2b (treatment effect waning starts 3 months post-treatment initiation, for 1 year)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	303,851	2291.77%
Scenario 2c (treatment effect waning starts 3 months post-treatment initiation, for 2 years)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	188,542	1384.12%
Scenario 2d (treatment effect waning starts 3 months post-treatment initiation, for 5 years)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	89,131	601.60%
Scenario 2e (treatment effect waning starts 3 months post-treatment initiation, for 10 years)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	49,103	286.52%
Scenario 2f (treatment effect waning starts 6 months post-treatment initiation, for 6 months)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	299,728	2259.32%

Scenario	ECM		Zuranolone with ECM		Incremental costs (£)	Incremental QALYs	ICER (£/QALY)	Change from base-case ICER (%)
	Total costs (£)	Total QALYs	Total costs (£)	Total QALYs				
Scenario 2g (treatment effect waning starts 1 year post-treatment initiation, for 1 year)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	156,898	1135.03%
Scenario 3a (50% on ADT at baseline)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	12,839	1.06%
Scenario 3b (80% on ADT at baseline)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	13,337	4.98%
Scenario 3c (90% on ADT at baseline)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	13,205	3.94%
Scenario 4 (LOCF to predict EPDS at 8 months)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	13,016	2.46%
Scenario 5a (patients with EPDS > 19 at baseline)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	12,839	1.06%
Scenario 5b (patients with EPDS ≤ 19 at baseline)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	12,289	-3.27%
Scenario 6a (utility estimate of 0.795 for EPDS = 0)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	14,754	16.14%
Scenario 6b (EQ-5D utilities for PND period and different utilities for MDD period, where MDD period starts at 17 months post-birth)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	22,939	80.57%
Scenario 6c (EQ-5D utilities for PND period and different utilities for MDD period, where MDD period starts at 12 months post-birth)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	23,246	82.98%

6.3 EAG's preferred assumptions

The EAG's preferred assumptions differ from the company's base case as follows:

- Re-parameterisation of the long-term MMRM model so that the ALSPAC MMRM model is used to calculate long-term EPDS scores for the ECM arm rather than the zuranolone with ECM arm. The EPDS scores for the zuranolone with ECM arm are then calculated by adding the zuranolone treatment effect to the ECM arm scores. **EAG Scenario 1.**
- The zuranolone treatment effect wanes linearly to zero, with waning starting at 1 year post-treatment initiation and lasting 1 year (i.e. the EPDS scores for the two arms converge at 2 years post-treatment initiation). **EAG Scenario 2g.**
- 90% of patients are on ADT at baseline. **EAG Scenario 3c.**
- SMR of 2 for patients with EPDS ≥ 13. **Company Scenario S1.**
- Using a cohort of patients with severe EPDS scores (>19) at baseline. **EAG Scenario 5a.**
- Using EPDS utility values mapped from ROBIN PHQ-9 to EQ-5D for the PND period (until 1 year post-treatment initiation) and using MDD-specific utilities for the period after. **EAG Scenario 6b.**

In Table 23 the EAG's preferred assumptions are incrementally added to show the overall effect on the deterministic ICER. The probabilistic results for the EAG's base case are presented in Table 25. The cost-effectiveness plane for the EAG's base case is presented in Figure 9, and a plot showing the convergence of the ICER in the probabilistic sensitivity analysis is presented in Figure 10.

The EAG's preferred assumptions increases the ICER from £12,704 per QALY gained in the company's base case to £194,554 in deterministic analysis and from £12,706 to £188,102 in probabilistic analysis. The key drivers of the ICER are the introduction of the waning of the treatment effect and the application of MDD utilities in the long-term following the PND period.

Table 23: EAG preferred assumptions added incrementally and EAG base-case (deterministic results)

EAG Assumption Number	ECM		Zuranolone with ECM		Incremental costs (£)	Incremental QALYs	ICER (£/QALY)	Change from base-case ICER (%)
	Total costs (£)	Total QALYs	Total costs (£)	Total QALYs				
0	Company base case							
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	12,704	-
0+1	+ re-parameterisation of long-term MMRM model (EAG Scenario 1)							
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	12,725	0.17%
0+1+2	+ treatment effect waning starts 1 year post-treatment initiation for 1 year (EAG Scenario 2g)							
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	156,898	1135.03%
0+1+2+3	+ 90% of patients on ADT at baseline (EAG Scenario 3c)							
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	169,289	1232.56%
0+1+2+3+4	+ SMR of 2 for patients with EPDS ≥ 13 (Company Scenario S1)							
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	169,289	1232.56%
0+1+2+3+4+5	+ cohort of patients with EPDS scores >19 at baseline (EAG Scenario 5a)							
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	157,848	1142.51%
0+1+2+3+4+5+6 = EAG Base Case	+ EQ-5D utilities for PND period and different utilities for MDD period (EAG Scenario 6b)							
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	194,554	1431.44%

Table 24: Scenario analysis with EAG base-case but without restricting to a severe population, requested by NICE (deterministic results)

EAG Assumption Number (as defined in Table 23)	ECM		Zuranolone with ECM		Incremental costs (£)	Incremental QALYs	ICER (£/QALY)	Change from base-case ICER (%)
	Total costs (£)	Total QALYs	Total costs (£)	Total QALYs				
0+1+2+3+4+6	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	209,838	1551.75%

Table 25: EAG base-case probabilistic results after 1000 iterations run

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER incremental (£/QALY)	Change from company probabilistic base-case ICER (%)
ECM	[redacted]	[redacted]	[redacted]	-	-	-	-	-
Zuranolone with ECM	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	188,102	1380.42%

Figure 9 Cost-effectiveness plane for EAG base-case

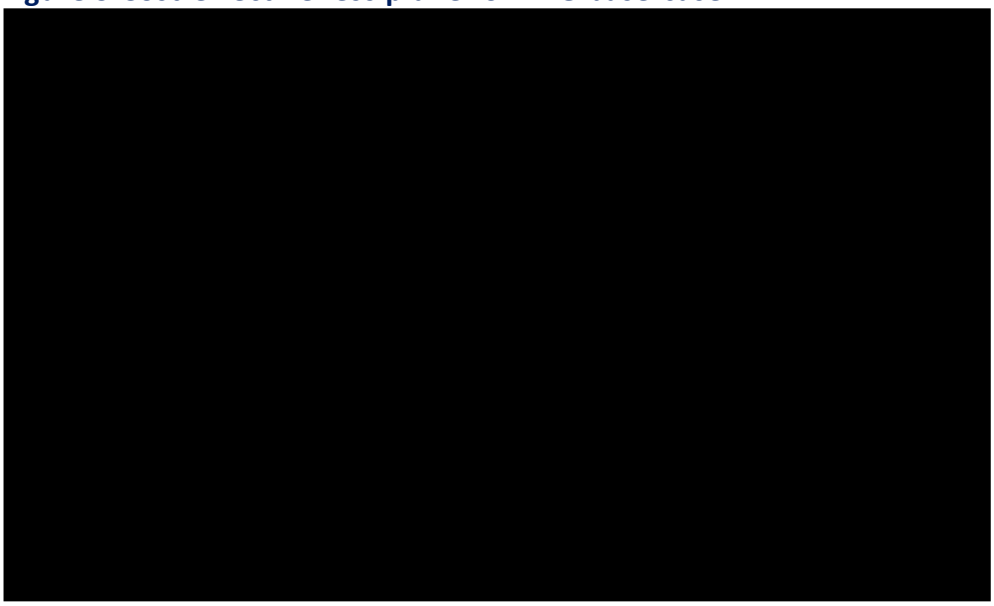


Figure 10 ICER convergence plot for EAG base-case

6.4 Conclusions of the cost effectiveness section

The company have submitted a cost-effectiveness model that aims to address the decision problem defined in the final scope. Whilst the model and scope were for a general PND population, the EAG heard that zuranolone would be used in a severe population, which was also the majority of the population in the company's trials. The EAG therefore consider a severe PND population to be more appropriate. The company also assume 15.3% of patients would be taking ADT at baseline, whereas the EAG heard that the majority of patients would be taking ADT.

The company use a patient-level simulation model that tracks patients trajectories through EPDS scores in the short (from 5 – 8 months post-birth) and long-term (from 8 months – 29 years post-birth), capturing health benefits and costs associated with EPDS. The EAG consider the model structure appropriate. However, the model extrapolates a treatment effect from the SKYLARK trial into the long-term for the full time horizon of the model (29 years post-birth). The EAG believes the trials provide insufficient evidence for a sustained effect over a 29 time-horizon, especially since zuranolone treatment is a single course treatment given for 14 days and there is no data on treatment efficacy beyond 45 days. The EAG's clinical experts also advised they would expect a much shorter effect. There is therefore a high level of uncertainty as to the most appropriate extrapolation of the treatment effect.

The EAG also disagrees with the use of the SF6D index to estimate utilities in the model, and prefers to use EQ-5D utilities, both mapped from disease specific questionnaires in the company's trials (provided by the company after clarification questions) in the PND period, and utilities from previous TAs in MDD in the longer term beyond the PND period. This is because the EQ-5D is NICE's preferred tool to measure utilities and the EAG's clinical experts advised that after 1-year post birth, PND and MDD were similar and treatments would be the same.

The EAG's preferred assumptions increases the ICER from £12,704 per QALY gained in the company's base case to £194,554 in deterministic analysis and from £12,706 to £188,102 in probabilistic analysis. The key drivers of the ICER are the introduction of the waning of the treatment effect and the application of MDD utilities in the long-term following the PND period.

7 Severity and uncaptured costs and benefits

The company states that the submission does not meet the criteria for a severity weight, and the EAG agrees with this.

It is likely that the positive health benefits of zuranolone on mothers' PND symptoms would also positively impact the health and wellbeing and resource use needs of the baby and of the father, partner, or support network of the mother. These would have been very difficult to quantify and have not been captured in the model, but they would decrease the ICER in favour of the intervention.

On the other hand, zuranolone cannot be used in patients who are breast-feeding. Support for patients to either continue breastfeeding following treatment or the stop breastfeeding was not captured in the model. Moreover, the consequences of stopping breastfeeding could have long lasting effects in the health of the baby and are very difficult to measure.

Costs to support decision around breast-feeding would be expected to increase the costs of zuranolone treatment and the ICER. In response to clarification questions, the company states that it is willing to work with stakeholders to "deliver effective support for patients without additional costs to the healthcare system."

Clinical advisors to the EAG also highlighted that patients on zuranolone could be at risk of additional rare but serious side effects such as decreased consciousness, not captured in the ROBIN or SKYLARK trials. These are rare but currently observed in patients treated with brexanolone, a similar drug offered to severe PND patients in the US. Zuranolone, like brexanolone, also has potential for abuse. These additional side effects could bring potential harms and costs for the mother and child which were not captured in the model and would also increase the ICER, although since rare the effect would be small.

8 REFERENCES

1. Parikh SV, Aaronson ST, Mathew SJ, Alva G, DeBattista C, Kanes S*, et al.* Efficacy and safety of zuranolone co-initiated with an antidepressant in adults with major depressive disorder: results from the phase 3 CORAL study. *Neuropsychopharmacology* 2024;**49**(2): 467-475
2. Biogen. 2025. [Company submission] Zuranolone for treating postnatal depression (Accessed 12 March 2025).
3. National Institute for Health and Care Excellence. *[FINAL SCOPE] Zuranolone for treating postnatal depression*.2025. URL: https://www.nice.org.uk/guidance/gid-ta11356/documents/final-scope (Accessed 12 March 2025).
4. National Institute for Health and Care Excellence. *CG192: Antenatal and postnatal mental health: clinical management and service guidance*.2024. URL: https://www.nice.org.uk/guidance/cg192 (Accessed 12 March 2025).
5. Husain N, Lunat F, Lovell K, Miah J, Chew-Graham CA, Bee P*, et al.* Efficacy of a culturally adapted, cognitive behavioural therapy-based intervention for postnatal depression in British south Asian women (ROSHNI-2): a multicentre, randomised controlled trial. *The Lancet* 2024;**404**(10461): 1430-1443
6. Bauer A, Knapp M, Parsonage M. Lifetime costs of perinatal anxiety and depression. *Journal of Affective Disorders* 2016;**192**83-90
7. Annette Bauer, Michela Tinelli, Martin Knapp. *The economic case for increasing access to treatment for women with common mental health problems during the perinatal period*.2022. URL: https://www.lse.ac.uk/cpec/assets/documents/CPEC-Perinatal-Economics-2022.pdf (Accessed 12 March 2025).
8. Biogen. *NCT04442503: A Study to Evaluate the Efficacy and Safety of SAGE-217 in Participants With Severe Postpartum Depression (PPD)*.2023. URL: https://clinicaltrials.gov/study/NCT04442503 (Accessed 12 March 2025).

9. Biogen. *NCT02978326: A Study to Evaluate SAGE-217 in Participants With Severe Postpartum Depression*.2023. URL: https://clinicaltrials.gov/study/NCT02978326 (Accessed 12 March 2025).
10. Deligiannidis KM, Meltzer-Brody S, Maximos B, Peeper EQ, Freeman M, Lasser R*, et al.* Zuranolone for the Treatment of Postpartum Depression. *American Journal of Psychiatry* 2023;**180**(9): 668-675
11. European Medicines Agency. *Clinical investigation of medicinal products in the treatment of depression - Scientific guideline*.2025. URL: https://www.ema.europa.eu/en/clinical-investigation-medicinal-products-treatment-depression-scientific-guideline (Accessed 12 March 2025).
12. Cox JL, Holden JM, Sagovsky R. Detection of Postnatal Depression: Development of the 10-item Edinburgh Postnatal Depression Scale. *British Journal of Psychiatry* 1987;**150**(6): 782-786
13. National Institute for Health and Care Excellence. *Depression - antenatal and postnatal: Scenario: Depression in the postnatal period*.2025. URL: https://cks.nice.org.uk/topics/depression-antenatal-postnatal/management/depression-in-the-postnatal-period/ (Accessed 15 May 2025).
14. Biogen. *[SKYLARK] A Study to Evaluate the Efficacy and Safety of SAGE-217 in Participants With Severe Postpartum Depression (PPD)*.2023. URL: https://clinicaltrials.gov/study/NCT04442503 (Accessed 12 March 2025).
15. [redacted]
16. Zimmerman M, Martinez JH, Young D, Chelminski I, Dalrymple K. Severity classification on the Hamilton depression rating scale. *Journal of Affective Disorders* 2013;**150**(2): 384-388
17. Snaith RP, Harrop FM, Newby DA, Teale C. Grade Scores of the Montgomery—Åsberg Depression and the Clinical Anxiety Scales. *British Journal of Psychiatry* 1986;**148**(5): 599-601
18. McCabe-Beane JE, S. SL, Yelena P, Scott S, and O'Hara MW. The identification of severity ranges for the Edinburgh Postnatal Depression Scale. *Journal of Reproductive and Infant Psychology* 2016;**34**(3): 293-303
19. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *Journal of Epidemiology and Community Health* 1998;**52**(6): 377-384
20. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I*, et al.* RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* 2019;**366**l4898
21. [redacted]
22. Whiting P, Savović J, Higgins JP, Caldwell DM, Reeves BC, Shea B*, et al.* ROBIS: A new tool to assess risk of bias in systematic reviews was developed. *Journal of Clinical Epidemiology* 2016;**69**225-234
23. Biogen. *[Response to Clarification] Zuranolone for treating severe postnatal depression* (Accessed 23 April 2025).
24. Meltzer-Brody S, Gerbasi ME, Mak C, Toubouti Y, Smith S, Roskell N*, et al.* Indirect comparisons of relative efficacy estimates of zuranolone and selective serotonin reuptake inhibitors for postpartum depression. *Journal of Medical Economics* 2024;**27**(1): 582-595

25. Annette Bauer, Martin Knapp, Bayo Adelaja. *Best practice for perinatal mental health care: the economic case*.2016. URL: https://www.pssru.ac.uk/pub/5226.pdf (Accessed 23 April 2025).
26. O'Callaghan L, Elizabeth C, J. JS, Erin F, and Deligiannidis KM. The cost-effectiveness of zuranolone versus selective serotonin reuptake inhibitors for the treatment of postpartum depression in the United States. *Journal of Medical Economics* 2024;**27**(1): 492-505
27. Eldar-Lissai A, Cohen JT, Meltzer-Brody S, Gerbasi ME, Chertavian E, Hodgkins P*, et al.* Cost-Effectiveness of Brexanolone Versus Selective Serotonin Reuptake Inhibitors for the Treatment of Postpartum Depression in the United States. *Journal of managed care & specialty pharmacy* 2020;**26**(5): 627-638
28. Mukuria C, Franklin M, Hinde S. Mapping functions for the PHQ-9 and GAD-7 to generate EQ-5D-3L for economic evaluation. *The European Journal of Health Economics* 2025;**26**(1): 63-70
29. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *Journal of Health Economics* 2002;**21**(2): 271-292
30. LoMartire R, Äng BO, Gerdle B, Vixner L. Psychometric properties of Short Form-36 Health Survey, EuroQol 5-dimensions, and Hospital Anxiety and Depression Scale in patients with chronic pain. *PAIN* 2020;**161**(1):
31. Stolz M, Albus C, Beutel ME, Deter H-C, Fritzsche K, Herrmann-Lingen C*, et al.* Assessment of health-related quality of life in individuals with depressive symptoms: validity and responsiveness of the EQ-5D-3L and the SF-6D. *The European Journal of Health Economics* 2023;**24**(8): 1297-1307
32. Vaidya A, Kleinegris MC, Severens JL, Ramaekers BL, Ten Cate-Hoek AJ, Ten Cate H*, et al.* Comparison of EQ-5D and SF-36 in untreated patients with symptoms of intermittent claudication. *Journal of comparative effectiveness research* 2018;**7**(6): 535-548
33. Manca A, Hawkins N, Sculpher MJ. Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. *Health Economics* 2005;**14**(5): 487-496
34. McDool E, Mukuria C, Brazier J. A Comparison of the SF-6Dv2 and SF-6D UK Utility Values in a Mixed Patient and Healthy Population. *PharmacoEconomics* 2021;**39**(8): 929-940
35. National Institute for Health and Care Excellence. *NG222: Depression in adults: treatment and management*.2022. URL: https://www.nice.org.uk/guidance/ng222 (Accessed 12 March 2025).
36. National Institute for Health and Care Excellence. *TA367: Vortioxetine for treating major depressive episodes*.2015. URL: https://www.nice.org.uk/guidance/ta367 (Accessed 12 March 2025).
37. Sapin C, Fantino B, Nowicki ML, Kind P. Usefulness of EQ-5D in assessing health status in primary care patients with major depressive disorder. *Health Qual Life Outcomes* 2004;**2**20
38. Petrou S, Cooper P, Murray L, Davidson LL. Economic costs of post-natal depression in a high-risk British cohort. *British Journal of Psychiatry* 2002;**181**(6): 505-512
39. Byford S, Barrett B, Despiégel N, Wade A. Impact of Treatment Success on Health Service Use and Cost in Depression. *PharmacoEconomics* 2011;**29**(2): 157-170
40. Büyükkaramikli NC, Rutten-van Mölken MPMH, Severens JL, Al M. TECH-VER: A Verification Checklist to Reduce Errors in Models and Improve Their Credibility. *PharmacoEconomics* 2019;**37**(11): 1391-1408

9 APPENDICES

9.1 Appendix 1: ROBIS assessments

9.1.1 ROBIS: systematic review of clinical effectiveness

This appraisal of the systematic review of clinical effectiveness was conducted with an amended version of ROBIS tool to remove the synthesis domain as this is not relevant to the submission, and a change in the wording of the signalling question 1.2 from "Were the eligibility criteria appropriate for the review question?" to "Were the eligibility criteria appropriate for the scope?" to make it specific to the Single Technology Assessment (STA) context. We also added the answer option "some concerns", to align with more recent risk of bias tools. The EAG's ROBIS assessment was conducted by one reviewer and checked by a second. Any disagreements were resolved by discussion.

DOMAIN 1: STUDY ELIGIBILITY CRITERIA	
1. Did the review adhere to pre-defined objectives and eligibility criteria?	Y
2. Were the eligibility criteria appropriate for the scope?	PY
3. Were eligibility criteria unambiguous?	Y
4. Were any restrictions in eligibility criteria based on study characteristics appropriate (e.g. date, sample size, study quality, outcomes measured)?	PY
5. Were any restrictions in eligibility criteria based on sources of information appropriate (e.g. publication status or format, language, availability of data)?	PY
Risk of bias judgement:	**LOW**

Rationale:

- The protocol for the review was not initially provided by the company. The EAG asked the company for this in clarification question A1 and it was sent by the company. The review adhered to pre-defined eligibility criteria.
- The EAG asked the company, in clarification question C2, whether "atypical antidepressants" (other than mirtazapine) were included in the review as per NICE scope. The company confirmed they were.
- The EAG asked the company, in clarification question C3, whether best supportive care (as specified in the NICE scope) was included by the company. The company excluded studies that compared against no treatment, including "wait list", but do not mention best supportive care. The company responded to question C3 that best supportive care was not included in the review because NICE guidelines for the management of antenatal and postnatal mental health do not list it as a treatment option and do not provide a definition for it.
- Only publications in English language included and limits on publication type (news items, editorials, preprints, opinion pieces, comments, letters, and conference abstracts published before January 2020 were excluded), however the EAG is not concerned about this.

Y=YES, PY=PROBABLY YES, PN=PROBABLY NO, N=NO, NI=NO INFORMATION

DOMAIN 2: IDENTIFICATION AND SELECTION OF STUDIES	
1. Did the search include an appropriate range of databases/electronic sources for published and unpublished reports?	Y
2. Were methods additional to database searching used to identify relevant reports?	Y
3. Were the terms and structure of the search strategy likely to retrieve as many eligible studies as possible?	Y

4.	Were restrictions based on date, publication format, or language appropriate?	PY
5.	Were efforts made to minimize error in selection of studies?	PN
	Risk of bias judgement:	**LOW**

Rationale:

- The searches were conducted in-line with the scope. Whilst not listed in the resources searched, a search strategy for the PsycINFO database was included. The EAG asked the company to clarify in question C4 and the company confirmed that APA PsycINFO was searched via OvidSP interface.
- The company note that a single reviewer assessed the search results and removed "obviously irrelevant records such as those about ineligible diseases". The EAG asked for clarification as to why this was not checked by a second reviewer, or conducted in duplicate, in clarification question C5. The company responded that single reviewer screening was only used to exclude publications at the initial review stage with the Cochrane randomised controlled trial (RCT) classifier.

Y=YES, PY=PROBABLY YES, PN=PROBABLY NO, N=NO, NI=NO INFORMATION

DOMAIN 3: DATA COLLECTION AND STUDY APPRAISAL		
1.	Were efforts made to minimize error in data collection?	Y
2.	Were sufficient study characteristics available for both review authors and readers to be able to interpret the results?	PY
3.	Were all relevant study results collected for use in the synthesis?	PY
4.	Was risk of bias (or methodological quality) formally assessed using appropriate criteria?	Y
5.	Were efforts made to minimise error in risk of bias assessment?	PY
	Risk of bias judgement:	**LOW**

Rationale:

- Data collection processes were appropriate.
- Risk of bias was assessed using the Black and Downs quality checklist. The EAG prefer the use of the Risk of Bias 2 tool which is more robust, designed specifically for the assessment of risk of bias in randomised controlled trials, and facilitates assessment at the outcome level. The company completed an assessment with Risk of Bias 2 tool (RoB2) upon request from the EAG in clarification question C6 – the company did this at the trial-level only.
- The company do not state how many reviewers undertook risk of bias assessment so it was initially unclear if efforts were made to minimise error in risk of bias assessment. The EAG asked the company to clarify this is clarification question C7. The company responded that it was completed by one reviewer and checked by a second independent reviewer.

Y=YES, PY=PROBABLY YES, PN=PROBABLY NO, N=NO, NI=NO INFORMATION

9.1.2 ROBIS: systematic review of HRQoL

This appraisal of the systematic review of clinical effectiveness was conducted with an amended version of ROBIS tool to remove the synthesis domain as this is not relevant to the submission, and a change in the wording of the signalling question 1.2 from "Were the eligibility criteria appropriate for the review question?" to "Were the eligibility criteria appropriate for the scope?" to make it specific to the STA context. We also added the answer option "some concerns", to align with more recent risk of bias tools. The EAG's ROBIS assessment was conducted by one reviewer and checked by a second. Any disagreements were resolved by discussion.

DOMAIN 1: STUDY ELIGIBILITY CRITERIA	
1. Did the review adhere to pre-defined objectives and eligibility criteria?	Y
2. Were the eligibility criteria appropriate for the scope?	Y
3. Were eligibility criteria unambiguous?	Y
4. Were any restrictions in eligibility criteria based on study characteristics appropriate (e.g. date, sample size, study quality, outcomes measured)?	PY
5. Were any restrictions in eligibility criteria based on sources of information appropriate (e.g. publication status or format, language, availability of data)?	PY
Risk of bias judgement:	**LOW**

Rationale:

- The protocol for the review was not initially provided by the company. The EAG asked the company for this in clarification question A1 and it was sent by the company. The review adhered to pre-defined eligibility criteria.
- Only included studies published in the English language and excluded and primary research before 2023 (Table 1 of economic and HRQoL SLR protocol). Also excludes narrative reviews, systematic reviews, non-systematic reviews and case reports. Protocol contains inconsistent inclusion/ exclusion criteria in that it states that conference abstracts published after 2020 are included, but those published before Nov 2023 are excluded.

Y=YES, PY=PROBABLY YES, PN=PROBABLY NO, N=NO, NI=NO INFORMATION

DOMAIN 2: IDENTIFICATION AND SELECTION OF STUDIES	
1. Did the search include an appropriate range of databases/electronic sources for published and unpublished reports?	Y
2. Were methods additional to database searching used to identify relevant reports?	Y
3. Were the terms and structure of the search strategy likely to retrieve as many eligible studies as possible?	Y
4. Were restrictions based on date, publication format, or language appropriate?	PY
5. Were efforts made to minimize error in selection of studies?	PN
Risk of bias judgement:	**LOW**

Rationale:

- The searches aligned with scope and were conducted appropriately.
- The company note that after conducting the search, a single reviewer assessed the search results and removed "obviously irrelevant records such as those about ineligible diseases". Two reviewers then independently assessed title and abstracts and full texts. The EAG asked for clarification as to why this was not checked by a second reviewer, or conducted in duplicate, in clarification question C5. The company responded that single reviewer screening was only used

to exclude publications at the initial review stage with the Cochrane randomised controlled trial (RCT) classifier.

Y=YES, PY=PROBABLY YES, PN=PROBABLY NO, N=NO, NI=NO INFORMATION

DOMAIN 3: DATA COLLECTION AND STUDY APPRAISAL	
1. Were efforts made to minimize error in data collection?	Y
2. Were sufficient study characteristics available for both review authors and readers to be able to interpret the results?	PY
3. Were all relevant study results collected for use in the synthesis?	PY
4. Was risk of bias (or methodological quality) formally assessed using appropriate criteria?	PN
5. Were efforts made to minimise error in risk of bias assessment?	PY
Risk of bias judgement:	**LOW**

Rationale:

- Data collection processes were appropriate.
- The company note that one reviewer assessed risk of bias of each study by considering a list of six issues, rather than using a formal risk of bias tool. The EAG asked the company, in clarification question C8, why a formal tool was not used and how many individuals undertook quality appraisal (i.e. whether risk of bias assessments were checked by a second reviewer). The company replied that NICE do not require validated tools for appraisal of cost or utility studies and the company is not aware of any. Therefore, formal quality assessments were not conducted and instead an informal assessment was completed by one individual verified by a second independent reviewer.

Y=YES, PY=PROBABLY YES, PN=PROBABLY NO, N=NO, NI=NO INFORMATION

9.1.3 ROBIS: systematic review of economic evaluations

This appraisal of the systematic review of clinical effectiveness was conducted with an amended version of ROBIS tool to remove the synthesis domain as this is not relevant to the submission, and a change in the wording of the signalling question 1.2 from "Were the eligibility criteria appropriate for the review question?" to "Were the eligibility criteria appropriate for the scope?" to make it specific to the STA context. We also added the answer option "some concerns", to align with more recent risk of bias tools. The EAG's ROBIS assessment was conducted by one reviewer and checked by a second. Any disagreements were resolved by discussion.

DOMAIN 1: STUDY ELIGIBILITY CRITERIA	
1. Did the review adhere to pre-defined objectives and eligibility criteria?	Y
2. Were the eligibility criteria appropriate for the scope?	Y
3. Were eligibility criteria unambiguous?	PY
4. Were any restrictions in eligibility criteria based on study characteristics appropriate (e.g. date, sample size, study quality, outcomes measured)?	PN
5. Were any restrictions in eligibility criteria based on sources of information appropriate (e.g. publication status or format, language, availability of data)?	PN
Risk of bias judgement:	**SOME CONCERNS**

Rationale:

- The protocol for the review was provided by the company after a clarification request.
- No explanation is provided for limits on publication date (studies published from 2013 to 2023 were eligible).
- The EAG has some concerns that studies on populations with major depressive disorder were excluded. These studies may have been useful to inform the extrapolation beyond the company's trial period, in terms of informing costs and utilities over the 29 year time horizon.
- No explanation is provided for limits to publication type (narrative reviews, systematic literature reviews, non-systematic reviews, case reports, and conference abstracts published before January 2020 were excluded). It is not clear why pre-prints were excluded.
- Protocol contains inconsistent inclusion/ exclusion criteria in that it states that conference abstracts published after 2020 are included, but those published before Nov 2023 are excluded
- A minor discrepancy was noted between the population eligibility criteria in the protocol and in the CS Appendix E (Table 41). In the CS the eligible population is described as 'PPD, Postnatal depression or PND'. In the protocol, it is 'Postpartum depression, Postnatal depression or Perinatal depression'. The EAG do not consider this to impact on the robustness of the SLR but that the abbreviations used in the Appendix are potentially misleading (a reporting concern).

Y=YES, PY=PROBABLY YES, PN=PROBABLY NO, N=NO, NI=NO INFORMATION

DOMAIN 2: IDENTIFICATION AND SELECTION OF STUDIES	
1. Did the search include an appropriate range of databases/electronic sources for published and unpublished reports?	Y
2. Were methods additional to database searching used to identify relevant reports?	Y
3. Were the terms and structure of the search strategy likely to retrieve as many eligible studies as possible?	Y
4. Were restrictions based on date, publication format, or language appropriate?	PY
5. Were efforts made to minimize error in selection of studies?	PN
Risk of bias judgement:	**LOW**

Rationale:

- The searches aligned with scope and were conducted appropriately.
- The company note that after conducting the search, a single reviewer assessed the search results and removed "obviously irrelevant records such as those about ineligible diseases". Two reviewers then independently assessed title and abstracts and full texts. The EAG asked for clarification as to why this was not checked by a second reviewer, or conducted in duplicate, in clarification question C5. The company responded that single reviewer screening was only used to exclude publications at the initial review stage with the Cochrane randomised controlled trial (RCT) classifier.

Y=YES, PY=PROBABLY YES, PN=PROBABLY NO, N=NO, NI=NO INFORMATION

DOMAIN 3: DATA COLLECTION AND STUDY APPRAISAL	
1. Were efforts made to minimize error in data collection?	PY
2. Were sufficient study characteristics available for both review authors and readers to be able to interpret the results?	PY
3. Were all relevant study results collected for use in the synthesis?	NA
4. Was risk of bias (or methodological quality) formally assessed using appropriate criteria?	PN
5. Were efforts made to minimise error in risk of bias assessment?	PY
Risk of bias judgement:	**LOW**

Rationale:

- One reviewer extracted data and completed the critical appraisal. A second 'checked the data points". Disagreement resolved by a third reviewer.
- Extraction form was piloted (no further information given)
- Although trial based economic evaluations were eligible, and non-model based approaches such as cost-consequence, the critical appraisal tool used was for economic model based studies. As only 1 non-model based evaluation was included (of a total of 3 studies), the impact is likely to be inconsequential to final results.

Y=YES, PY=PROBABLY YES, PN=PROBABLY NO, N=NO, NI=NO INFORMATION

9.2 Appendix 2: Risk of Bias 2 assessments

9.2.1 SKYLARK

Domain	Signalling question	EPDS (*Day 3– 5 **Day 21-45	HAMD-17 *Day 3– 5 **Day 21-45	MADRS *Day 3– 5 **Day 21-45	PHQ-9 *Day 3– 5 **Day 21-45	Adverse effects (AEs) of treatment	Comments
Bias arising from the randomization process	1.1 Was the allocation sequence random?	Yes	Yes	Yes	Yes	Yes	"Randomization will be performed centrally via an interactive response technology (IRT) system. Randomization schedules will be generated by an independent statistician. The allocation to treatment group will be based on the randomization schedule. The randomization schedules will be kept strictly confidential, accessible only to authorized personnel until the time of unblinding. The blinding of the study will be broken after the database has been locked".[14]
	1.2 Was the allocation sequence concealed until participants were enrolled and assigned to interventions?	Yes	Yes	Yes	Yes	Yes	
	1.3 Did baseline differences between intervention groups suggest a problem with the randomization process?	No	No	No	No	No	Demographic and baseline clinical characteristics were generally balanced between treatment groups (Table 1).
	Risk of bias judgement	**Low**	**Low**	**Low**	**Low**	**Low**	**No concerns regarding the randomisation process**
Bias due to deviations from intended interventions	2.1.Were participants aware of their assigned intervention during the trial?	PY	PY	PY	PY	PY	Protocol reports: "Participants, clinicians, and the study team will be blinded to treatment allocation", however, the study reports concerns of a higher incidence of treatment -emergent adverse events in the zuranolone group, which might have affected blinding in the study, and as stated in page 55 of the CS, the FDA also raised concerns about a potential risk of functional unblinding.
	2.2.Were carers and people delivering the interventions aware of participants' assigned intervention during the trial?	PY	PY	PY	PY	PY	
	2.3. If Y/PY/NI to 2.1/2.2: Were there deviations from the intended interventions that arose because of the trial context?	PN	PN	PN	PN	PN	Reasons for major protocol deviations were probably not related to the trial context
	2.4. If Y/PY to 2.3: Were these deviations likely to have affected the outcome?	NA	NA	NA	NA	NA	

Domain	Signalling question	EPDS (*Day 3– 5 **Day 21-45	HAMD-17 *Day 3– 5 **Day 21-45	MADRS *Day 3– 5 **Day 21-45	PHQ-9 *Day 3– 5 **Day 21-45	Adverse effects (AEs) of treatment	Comments
	2.5. If Y/PY/NI to 2.4: Were these deviations from intended intervention balanced between groups?	NA	NA	NA	NA	NA	
	2.6. Was an appropriate analysis used to estimate the effect of assignment to intervention?	Y	Y	Y	Y	Y	Efficacy outcomes and adverse events were assessed on the full-analysis set (all randomized participants who received any amount of investigational product and have a valid baseline and at least one valid postbaseline HAMD-17 total score - zuranolone n=98 and placebo n=97).
	2.7. If N/PN/NI to 2.6: Was there potential for a substantial impact (on the result) of the failure to analyse participants in the group to which they were randomised?	NA	NA	NA	NA	NA	NA
	Risk of bias judgement	**Low**	**Low**	**Low**	**Low**	**Low**	**No concerns regarding bias due to deviations from intended interventions**
Bias due to missing outcome data	3.1 Were data for this outcome available for all, or nearly all, participants randomized?	Y*/N**	Y*/N**	Y*/N**	Y*/N**	Y	Fourteen patients from the zuranolone arm and twelve from the placebo arm withdrew from the study prematurely, therefore data was not available for all or nearly all participants for all efficacy outcomes at all timepoints, particularly at Days 21, 38 and 45. • EPDS(Clarification question A7 Table 6) – **Baseline** 97/97 placebo and 98/98 zuranolone; **Day 3**: 94/97 and 97/98; **Day 8**: 95/97 and 93/98; **Day 15**: 89/97 and 91/98; **Day 21**: 81/97 and 84/98; **Day 28**: 85/97 and 76/98; **Day 45**: 85/97 and 84/98 • HAMD-17 (CSR Table 21) – **Baseline** 97/97 placebo and 98/98 zuranolone; **Day 3**: 96/97 and 98/98; **Day 8**: 95/97 and 93/98; **Day 15**: 90/97 and 93/98; **Day 21**: 83/97 and 84/98; **Day 28**: 85/97 and 77/98; **Day 45**: 85/97 and 84/98 • MADRS (CSR Table 14.2.5.1.1) – **Baseline** 96/97 placebo and 98/98 zuranolone; **Day 8**: 95/97 and 93/98; **Day 15**: 90/97 and 92/98; **Day 28**: 82/97 and 76/98; **Day 45**: 85/97 and 84/98

Domain	Signalling question	EPDS (*Day 3– 5 **Day 21-45	HAMD-17 *Day 3– 5 **Day 21-45	MADRS *Day 3– 5 **Day 21-45	PHQ-9 *Day 3– 5 **Day 21-45	Adverse effects (AEs) of treatment	Comments
							• PHQ-9 (CSR Table 14.2.6.1.1) - **Baseline** 97/97 placebo and 98/98 zuranolone; **Day 3**: 94/97 and 97/98; **Day 8**: 95/97 and 93/98; **Day 15**: 90/97 and 91/98; **Day 21**: 81/97 and 84/98; **Day 28**: 85/97 and 77/98; **Day 45**: 85/97 and 84/98
	3.2 If N/PN/NI to 3.1: Is there evidence that result was not biased by missing outcome data?	NA*/N**	NA*/N**	NA*/N**	NA*/N**	Y	There are no sensitivity analyses to assess the robustness of the results for Day 21 to Say 45
	3.3 If N/PN to 3.2: Could missingness in the outcome depend on its true value?	NA*/PY**	NA*/PY**	NA*/PY**	NA*/PY**	NA	Missingness in the outcomes could potentially be associated with symptoms severity
	3.4. If Y/PY/Ni to 3.3: Is it likely that missingness in the outcome depended on its true value?	NA*PN**	NA*PN**	NA*PN**	NA*PN**	NA	It is not likely that missingness in the outcome depended on its true value
	Risk of bias judgement	**Low*/Some concerns****	**Low*/Some concerns****	**Low*/Some concerns****	**Low*/Some concerns****	**Low**	No concerns regarding missing outcome data for efficacy outcomes at timepoints at Days 3-15. There are some concerns of risk of bias due to data not being available for nearly all participants at Days 21-45, however it is not likely that missingness in the outcome depended on its true value.
Bias in measurement of the outcome	4.1 Was the method of measuring the outcome inappropriate?	N	N	N	N	N	EPDS, MADRS and HAMD-17 are standardised scales. All assessments were conducted by raters that have been trained and certified to conduct assessments in this study. Adverse events were a mixture of treatment emergent adverse events, vital signs, clinical laboratory measurements, and ECG. and were measured adequately.
	4.2 Could measurement or ascertainment of the outcome have differed between intervention groups?	N	N	N	N	N	Unlikely that measurement would have differed between groups.
	4.3 If N/PN/NI to 4.1 and 4.2: Were outcome assessors aware of the intervention received by study participants?	PY	PY	PY	PY	PY	Outcome assessors were blind to the interventions by study design, however there are concerns of potential functional unblinding (see domain 2).
	4.4. If Y/PY/NI to 4.3: Could assessment of the outcome have been influenced by knowledge of intervention received?	PN	PN	PN	PN	PN	Efficacy outcomes were measured using standardised scales, based on participants interviews, however the scale's items are subjective, and therefore subject to potential bias.

Domain	Signalling question	EPDS (*Day 3– 5 **Day 21-45	HAMD-17 *Day 3– 5 **Day 21-45	MADRS *Day 3– 5 **Day 21-45	PHQ-9 *Day 3– 5 **Day 21-45	Adverse effects (AEs) of treatment	Comments
	4.5.: If Y/PY/NI to 4.4: Is it likely that assessment of the outcome was influenced by knowledge of the intervention received?	PN					Safety outcomes are a mixture of objective and subjective measures, but even though some assessments could be influenced by knowledge of intervention received, it's not likely this was the case.
	Risk of bias judgement	**Some concerns**	**Some concerns**	**Some concerns**	**Some concerns**	**Some concerns**	**Some concerns of risk of bias in the measurement of the outcome because it is likely outcome assessors were aware of the intervention received, due to functional unblinding, and this may have influenced assessment using subjective scales.**
Bias in selection of the reported result	5.1 Were the data that produced this result analysed in accordance with a pre-specified analysis plan that was finalized before unblinded outcome data were available for analysis?	Y	Y	Y	Y	Y	Protocol not initially provided by company, but sent in response to a request from the EAG, and all outcomes pre-specified.
	5.2 Is the numerical result being assessed likely to have been selected on the basis of the results from: multiple eligible outcome measurements (e.g. scales, definitions, time points) within the outcome domain?	PN	PN	PN	PN	PN	All intended scales and endpoints were reported in the CS, EPDS was the only scale used in the economic model, but it was chosen because it was the only scale available in ALSPAC.
	5.3 Is the numerical result being assessed likely to have been selected on the basis of the results from: multiple eligible analyses of the data?	PN	PN	PN	PN	PN	
	Risk of bias judgement	**Low**	**Low**	**Low**	**Low**	**Low**	**No concerns regarding selection of the reported result**
Overall bias	**Risk of bias judgement**	**Some concerns**	**Some concerns**	**Some concerns**	**Some concerns**	**Some concerns**	**Overall some concerns of risk of bias from the domain "measurement of the outcome" because outcome assessors were likely aware of the intervention received and this may have influenced assessment using subjective scales. Additionally, some timepoints (Day 21 to 45) were considered to have some concerns due to missing outcome data. All other domains judged at low risk of bias.**

9.2.2 ROBIN

Domain	Signalling question	EPDS	HAMD-17	SF-36	MADRS	PHQ-9	Comments
Bias arising from the randomization process	1.1 Was the allocation sequence random?	Yes	Yes	Yes	Yes	Yes	Randomisation codes generated with a block size of 4 by an independent statistical vendor not affiliated with Sage Therapeutics Inc. Randomisation scheme initially performed manually using SAS statistical software version 9.3 (80 patients), then it was contracted to an interactive response technology system vendor (73 patients). Study site-designated pharmacy staff, who dispensed the study drug, were the only study personnel unblinded to the randomisation scheme.
	1.2 Was the allocation sequence concealed until participants were enrolled and assigned to interventions?	Yes	Yes	Yes	Yes	Yes	
	1.3 Did baseline differences between intervention groups suggest a problem with the randomization process?	No	No	No	No	No	Demographic and baseline clinical characteristics were generally balanced between treatment groups (Table 1).
	Risk of bias judgement	**Low**	**Low**	**Low**	**Low**	**Low**	**No concerns regarding the randomisation process**
Bias due to deviations from intended interventions	2.1.Were participants aware of their assigned intervention during the trial?	PY	PY	PY	PY	PY	Patients, clinicians, and study team blinded to treatment allocation. However, the CS notes on p55 that a potential risk of functional unblinding was raised by the FDA, as there was an imbalance in adverse events/ no. patients requiring dose reductions due to AEs in the two groups (more in zuranolone group).
	2.2.Were carers and people delivering the interventions aware of participants' assigned intervention during the trial?	PY	PY	PY	PY	PY	
	2.3. If Y/PY/NI to 2.1/2.2: Were there deviations from the intended interventions that arose because of the trial context?	PN	PN	PN	PN	PN	Reasons for protocol deviations not provided but unlikely to be due to trial context.
	2.4. If Y/PY to 2.3: Were these deviations likely to have affected the outcome?	N/A	N/A	N/A	N/A	N/A	
	2.5. If Y/PY/NI to 2.4: Were these deviations from intended intervention balanced between groups?	N/A	N/A	N/A	N/A	N/A	

Domain	Signalling question	EPDS	HAMD-17	SF-36	MADRS	PHQ-9	Comments
	2.6. Was an appropriate analysis used to estimate the effect of assignment to intervention?	Y	Y	Y	Y	Y	The efficacy set, based on intention to treat, included all patients in the safety set who had a valid baseline and at least 1 post-baseline efficacy assessment (zuranolone 76/77 randomised; placebo 74/76 randomised).
	2.7. If N/PN/NI to 2.6: Was there potential for a substantial impact (on the result) of the failure to analyse participants in the group to which they were randomised?	NA	NA	NA	NA	NA	
	Risk of bias judgement	**Low**	**Low**	**Low**	**Low**	**Low**	**No concerns regarding bias due to deviations from intended interventions**
Bias due to missing outcome data	3.1 Were data for this outcome available for all, or nearly all, participants randomized?	Y	Y	Y	Y	Y	At Day 45, five patients in the placebo group (6.8%) and three in the zuranolone group (3.9%) did not have outcome data. Other timepoints reported below. Overall, three patients in the zuranolone group and 0 in the placebo group were lost to follow-up. • EPDS(CSR Table 14.2.5.1) – **Baseline** 74/74 placebo and 76/76 zuranolone; **Day 3**: 72/74 and 72/76; **Day 8**: 74/74 and 75/76; **Day 15**: 73/74 and 74/76; **Day 21**: 73/74 and 74/76; **Day 45**: 69/74 and 73/76 • HAMD-17 (CSR Table 14.2.1.1.1) – **Baseline** 74/74 placebo and 76/76 zuranolone; **Day 3**: 74/74 and 74/76; **Day 8**: 74/74 and 75/76; **Day 15**: 73/74 and 74/76; **Day 21**: 73/74 and 74/76; **Day 45**: 69/74 and 73/76 • SF-36 (CSR Table 14.2.8.1) – **Baseline** 74/74 placebo and 76/76 zuranolone; **Day 15**: 73/74 and 74/76; **Day 45**: 69/74 and 73/76 • MADRS (CSR Table 14.2.2.1.1) – **Baseline** 74/74 placebo and 76/76 zuranolone; **Day 3**: 74/74 and 74/76; **Day 8**: 74/74 and 75/76; **Day 15**: 73/74 and 74/76; **Day 21**: 73/74 and 74/76; **Day 45**: 69/74 and 73/76 • PHQ-9 (CSR Table 14.2.6.1) - **Baseline** 74/74 placebo and 76/76 zuranolone; **Day 3**: 72/74 and 72/76; **Day 8**: 74/74 and 75/76; **Day 15**: 73/74 and 74/76; **Day 21**: 73/74 and 74/76; **Day 45**: 69/74 and 73/76

Domain	Signalling question	EPDS	HAMD-17	SF-36	MADRS	PHQ-9	Comments
	3.2 If N/PN/NI to 3.1: Is there evidence that result was not biased by missing outcome data?	NA	NA	NA	NA	NA	
	3.3 If N/PN to 3.2: Could missingness in the outcome depend on its true value?	NA	NA	NA	NA	NA	
	3.4. If Y/PY/Ni to 3.3: Is it likely that missingness in the outcome depended on its true value?	NA	NA	NA	NA	NA	
	Risk of bias judgement	**Low**	**Low**	**Low**	**Low**	**Low**	**No concerns regarding missing data**
Bias in measurement of the outcome	4.1 Was the method of measuring the outcome inappropriate?	N	N	N	N	N	All standardised scales.
	4.2 Could measurement or ascertainment of the outcome have differed between intervention groups?	PN	PN	PN	PN	PN	Unlikely that measurement would have differed between groups.
	4.3 If N/PN/NI to 4.1 and 4.2: Were outcome assessors aware of the intervention received by study participants?	PY	PY	PY	PY	PY	Outcome assessors were blind to the interventions by study design, however the CS notes on p55 that potential risk of functional unblinding was raised by the FDA, as there was an imbalance in adverse events/ no. patients requiring dose reductions due to AEs in the two groups (more in zuranolone group).
	4.4. If Y/PY/NI to 4.3: Could assessment of the outcome have been influenced by knowledge of intervention received?	PY	PY	PY	PY	PY	Measurements using standardised scales, however the scale's items are subjective, and therefore subject to potential bias.
	4.5.: If Y/PY/NI to 4.4: Is it likely that assessment of the outcome was influenced by knowledge of the intervention received?	PY	PY	PY	PY	PY	
	Risk of bias judgement	**Some concerns**	**Some concerns**	**Some concerns**	**Some concerns**	**Some concerns**	**Some concerns of risk of bias in the measurement of the outcome because it is likely outcome assessors were aware of the intervention received, due to functional unblinding, and this may have influenced assessment using subjective scales**

Domain	Signalling question	EPDS	HAMD-17	SF-36	MADRS	PHQ-9	Comments
Bias in selection of the reported result	5.1 Were the data that produced this result analysed in accordance with a pre-specified analysis plan that was finalized before unblinded outcome data were available for analysis?	Y	Y	Y	Y	Y	Protocol not initially provided by company, but sent in response to a request from the EAG, and all outcomes pre-specified.
	5.2 Is the numerical result being assessed likely to have been selected on the basis of the results from: multiple eligible outcome measurements (e.g. scales, definitions, time points) within the outcome domain?	PN	PN	PN	PN	PN	The protocol states SF-36 will be assessed at baseline, day 15, and day 45. These data are provided in the CSR but the CS only provides change from baseline to day 45. The EAG notes that PHQ-9 is a pre-specified outcome in the protocol and data are reported in the CSR, but not in the CS. All other outcomes not likely to have been selected based on multiple outcome measurements.
	5.3 Is the numerical result being assessed likely to have been selected on the basis of the results from: multiple eligible analyses of the data?	PN	PN	PN	PN	PN	
	Risk of bias judgement	**Low**	**Low**	**Low**	**Low**	**Low**	**No concerns regarding selection of the reported result**
Overall bias	**Risk of bias judgement**	**Some concerns**	**Some concerns**	**Some concerns**	**Some concerns**	**Some concerns**	**Overall some concerns of risk of bias from the domain "measurement of the outcome" because outcome assessors were likely aware of the intervention received and this may have influenced assessment using subjective scales. All other domains judged at low risk of bias.**

Single Technology Appraisal

Zuranolone for treating postnatal depression [ID6431]

EAG response to factual accuracy check and confidential information check

All confidential information, and information that is submitted as [redacted] is highlighted in turquoise.

Issue 1 The most appropriate population for treatment with zuranolone

Description of problem	Description of proposed amendment	Justification for amendment	EAG Response
Page 13: "The EAG prefers to restrict to a severe PND population to match the expected use in the NHS, and the majority of	The company propose to remove the statement that 'a severe PND population matches the expected use in the NHS', and 'the majority	The company disagree that the appropriate population for this appraisal should be those with severe PND only: • Clinical evidence supporting zuranolone is formed by the SKYLARK and ROBIN studies, which form the basis for the marketing authorisation for zuranolone in the UK. The company believe that the appropriate population for the appraisal is the licensed population. As communicated at the clarification stage, following the company's latest discussions with the Medicines and Healthcare products Regulatory Agency (MHRA), it is likely that the target population for the marketing authorisation of zuranolone in the UK will be limited to [redacted]. • Further, the company would like to highlight that while the focus on the severity thresholds for scales such as HAMD-17, MADRS, and EPDS is crucial for designing and evaluating clinical trials, it may be less directly applicable to real-world clinical practice. In day-to-day patient care, the assessment of PND severity primarily relies on clinical judgment. This has been outlined in the CS (Section 1.3.1.3) and acknowledged by the clinical experts consulted by EAG (page 35 of the EAG report): *"…that the HAMD-17 scale (and other scales) are not commonly*	This is not a factual inaccuracy, and reflects the views of our clinical advisors on the expected use of zuranolone in the NHS. The appropriateness of the population will be considered by the committee. No changes were made. See response to Issue 3 below (p23) regarding the breakdown of the trial population according to severity based on MADRS scale.

patients in the SKYLARK trial."	of patients in the SKYLARK trial'.	*used in UK clinical practice to diagnose and manage patients or to measure recovery. Instead, clinical judgement of severity of symptoms is undertaken, taking into account functional impact and risk."* Additionally, in line with the draft Summary of Product Characteristics from MHRA, zuranolone is expected to be initiated under the supervision of a specialist prescriber; the company feel that the specialist is best placed to assess whether treatment with zuranolone is appropriate. • Additionally, the company believe that the evidence package for zuranolone is sufficient to demonstrate efficacy of the treatment in moderate to severe PND, as the SKYLARK and ROBIN trials included patients with moderate (45.9% and 40.7%, respectively) and severe (54.19% and 59.3%, respectively) disease (based on literature-established MADRS thresholds) – as presented in the CS (Section 2.8.2.2), results of the post-hoc analyses [redacted] (Figure 26 and Figure 27 of the CS), suggesting efficacy in patients with both moderate and severe PND.	

Issue 2 Definition of Established Clinical Management (ECM) in the company's two studies and NHS practice

Description of problem	Description of proposed amendment	Justification for amendment	EAG Response
Page 14: "The CS reports subgroup analyses for	The company propose the wording is updated to reflect the opinion of clinical experts	The company would like to reflect the opinion of clinical experts engaged by EAG (reported on page	This is not a factual inaccuracy. The sentence taken from page 35

baseline ADT use. However, company studies were not powered to detect subgroup effects and results are imprecise (CS Figure 22, SKYLARK and CS Figure 23, ROBIN). No subgroup analyses are available for participants receiving psychological therapies, other ECM, or no additional treatments, at baseline from either trial. It is therefore uncertain whether the results for the ECM used in the company's trials generalise to that used in NHS practice."	engaged by the EAG (reported on page 35) in the summary section of Key issue 2: "The CS reports subgroup analyses for baseline ADT use. However, company studies were not powered to detect subgroup effects and results are imprecise (CS Figure 22, SKYLARK and CS Figure 23, ROBIN). No subgroup analyses are available for participants receiving psychological therapies, other ECM, or no additional treatments, at baseline from either trial. It is therefore uncertain whether the results for the ECM used in the company's trials generalise to that used in NHS practice. **However, EAG clinical advisors did not think the difference in baseline ADT use between SKYLARK and the UK NHS setting would introduce significant issues in terms**	35) in the summary section of Key issue 2. Clarifying input from the clinical advisor that there are no significant issues with applicability of the trial evidence to the NHS setting would support the committee in their deliberations. Finally, the company would like to highlight that zuranolone was granted an Innovation Passport by the MHRA as part of the innovative licensing access pathway (ILAP), demonstrating zuranolone to be an innovative product that addresses area of high unmet need and offering potential benefits to patients in a vulnerable patient population.	does not refer to the same issue that is discussed in Key issue 2. We have amended the sentence on page 35 to improve its clarity and avoid misinterpretation. No changes are made to Key Issue 2 on page 14. Specifically, the statement from page 35 ("*EAG clinical advisors did not think the difference in baseline ADT use between SKYLARK and the UK NHS setting would introduce significant issues in terms of the applicability of the trial evidence.*") was selectively taken from a larger paragraph which discusses potential implications of the differences in **the types of antidepressants** used in SKYLARK and those commonly used in the NHS practice (e.g. some of the baseline antidepressants used in SKYLARK are not licensed in the UK and are thus not used). We have now edited this sentence to avoid its misinterpretation. The new

	of the applicability of the trial evidence."		wording is: "*However, EAG clinical advisors did not think **the differences in the types of baseline ADT used in SKYLARK and the UK NHS setting** would introduce significant issues in terms of the applicability of the trial evidence*." This statement is not relevant to the issues raised in Key Issue 2 (page 14) and the EAG do not agree that this statement should be added to Key Issue 2. This key issue relates to whether the company's claim to have compared zuranolone to ECM is appropriate, given the absence of information about participants' background treatments. The EAG clinicians did not have significant concerns about the baseline comparability **of the type of ADT** being taken by 15% of the participants and the type of ADT is typically used in the NHS. This is not the same as stating that the "ECM" is similar across the trial and NHS practice. It would also be inaccurate to state that our experts

			did not have concerns about the comparability of trial evidence to NHS practice in terms of ECM - they did.
Page 15: "The key additional data would be RCT data with a higher proportion of patients on ADT reflective of NHS practice. In the absence of this, additional information on the interaction between Selective Serotonin Reuptake Inhibitors (SSRIs) and zuranolone could be obtained from trials in MDD."	The company propose to remove the suggestion that data from MDD trials is leveraged to support the interaction between SSRIs and zuranolone.	The company would like to re-iterate that, as outlined in the CS (Section 1.3.1.1), PND and MDD are distinct disorders. PND is defined as a depressive episode characterised by onset in the peripartum period and is driven by pathophysiological mechanisms that differentiate it from MDD. The company feels it is not appropriate to give the perception to the committee that PND and MDD are interchangeable diseases, and data and outcomes from MDD trials are relevant for PND. Further, zuranolone was granted an Innovation Passport by the MHRA as part of the ILAP, demonstrating zuranolone to be an innovative product (potential to be the first treatment targeting the underlying pathophysiology of PND) that addresses area of high unmet need and offering potential benefits to	This is not a factual inaccuracy. The EAG recognises that PND and MDD are different conditions and ideally evidence from a PND population would be preferred, but in the absence of that data from a MDD may give an indication of the expected effect and interaction between SSRIs and zuranalone. We have however added text to acknowledge that MDD and PND are not the same condition: "*In the absence of this, additional information on the interaction between Selective Serotonin Reuptake Inhibitors (SSRIs) and zuranolone could be obtained from trials in MDD,*[1] ***although the EAG*** **acknowledges** ***that MDD and PND are not the same condition***."

		patients in a vulnerable patient population.	

Issue 3 Applicability of the SKYLARK trial to NHS practice

Description of problem	Description of proposed amendment	Justification for amendment	EAG Response
Page 14: "Participants in SKYLARK were prevented from initiating new therapies (or changing the doses or regimens of their existing therapies) during the 45 day follow-up of the trial, which also may not reflect ECM in NHS practice."	The company propose that the wording is updated: "Participants in SKYLARK were prevented from initiating new therapies (or changing the doses or regimens of their existing therapies) during the 45 day follow-up of the trial. Although this may not reflect ECM in NHS practice, **preventing initiation of new therapies during the trial follow-up is not uncommon, as it could potentially confound the assessment of clinical efficacy of the investigational therapy**	The company would like to provide the committee with additional context on the rationale for such approach and support the decision making: • Preventing the initiation of new pharmacotherapies or increasing the dose of concomitant treatments during the follow-up period of any trial is not uncommon. • In case of ROBIN and SKYLARK, dose adjustments of concomitant ADTs and/or switching to/initiation of a new ADT during the 45-day follow-up period could introduce significant confounding variables,	This is not a factual inaccuracy. No changes were made. The EAG acknowledges that preventing initiation of new therapies during trial follow-up is not uncommon in clinical trials and this is done to strengthen internal validity of the assessment of clinical effectiveness. These approaches can however reduce external validity of the trial evidence, which is the point made by the EAG in this key issue. Therefore the EAG does not think it is appropriate to add the proposed explanation.

	(in this case, zuranolone).”	thereby compromising the ability to delineate the therapeutic effects attributable to zuranolone from those potentially arising from new pharmacotherapies or adjustments to existing pharmacotherapies. • The company believe that such approach was necessary to mitigate concerns around introducing confounders and support the decision-making.	
Page 26: “Patients were not allowed to start any new interventions during the trial, which differs from ECM in clinical practice, and in particular may adversely affect outcomes on the placebo with ECM arm more than the zuranolone plus ECM arm.”	The company propose that the wording is updated: “Patients were not allowed to start any new interventions during the trial, which differs from ECM in clinical practice, and in particular may adversely affect outcomes on the placebo with ECM arm more than the zuranolone plus ECM arm. **However, preventing initiation of new therapies during the trial follow-up is not uncommon, as it could potentially confound the assessment of clinical efficacy of the investigational therapy (in this case, zuranolone).”**		This is not a factual inaccuracy. No changes were made. See the explanation above.
Page 23:	The company propose to remove the statement that	The company disagree that the appropriate population for this	This is not a factual inaccuracy. The appropriateness of the

"The population for whom the CS seeks approval for zuranolone is adults with PND unrestricted by disease severity. However, the evidence for clinical effectiveness comes from two trials (SKYLARK and ROBIN) which recruited participants with a HAMD-17 score ≥26 at baseline, which represents a severe population. Additionally, as noted in section 1.3.4.3, the CS states that zuranolone will initially be prescribed by psychiatrists within secondary care services. In the UK, EAG clinical experts advised that patients seen in secondary care perinatal mental health services are those with severe PND, complex PND or those with co-	the appropriate population for consideration in this appraisal should be those with severe PND only.	appraisal should be those with severe PND only: • The company believe that the appropriate population for the appraisal is the licensed population. As communicated at the clarification stage, following the company's latest discussions with the MHRA, it is likely that the target population for the marketing authorisation of zuranolone in the UK will be limited to [redacted]. • The company believes that the evidence package for zuranolone is sufficient to demonstrate efficacy of the treatment in moderate to severe PND, as the SKYLARK and ROBIN trials included patients with moderate (45.9% and 40.7%, respectively) and severe (54.19% and 59.3%, respectively) disease (based on literature-established MADRS thresholds) – as	population will be considered by the committee. However, the EAG has added additional text on page 23, about the breakdown of the trial population according to severity based on MADRS scale, as follows: "The population for whom the CS seeks approval for zuranolone is adults with PND unrestricted by disease severity. However, the evidence for clinical effectiveness comes from two trials (SKYLARK and ROBIN) which recruited participants with a HAMD-17 score ≥26 at baseline, which represents a severe population. **A post-hoc analysis using MADRS scores is reported as evidence of mixed baseline PND severity (SKYLARK: [redacted] moderate and [redacted] severe; ROBIN: [redacted] moderate and [redacted] severe PND; CS, section 2.8.2.2). However, SKYLARK's mean baseline MADRS scores**

occurring psychiatric conditions. As clinical evidence for zuranolone comes from trial participants with severe PND, and EAG clinical experts indicated this matches the expected use in the NHS, the EAG considers that the appropriate population for consideration in this appraisal should be those with severe PND."		presented in the CS (Section 2.8.2.2), results of the post-hoc analyses [redacted] (Figure 26 and Figure 27 of the CS), suggesting efficacy in patients with both moderate and severe PND. • Additionally, in line with the draft Summary of Product Characteristics from MHRA, zuranolone is expected to be initiated under the supervision of a specialist prescriber; the company feel that the specialist is best placed to assess whether treatment with zuranolone is appropriate.	**suggest that the "moderate" subgroup only includes the upper range of moderate, and therefore cannot be considered representative of a "moderate" subgroup in the general population.**" … "As clinical evidence for zuranolone comes **mostly** from trial participants with severe PND, and EAG clinical experts indicated this matches the expected use in the NHS, the EAG considers that the appropriate population for consideration in this appraisal should be those with severe PND."
Page 34: "The EAG notes that the SKYLARK trial does not align with the decision problem because it enrolled women with severe PND (HAMD-17 total score of ≥26),	The company propose that this sentence is removed.		This is not a factual inaccuracy. No changes were made.

whereas the decision problem is broader, specifying “adults with PND”.			
Page 35: “Additionally, as zuranolone will initially only be prescribed by perinatal psychiatrists (after being referred to secondary care), most patients considered for zuranolone for PND in the UK will likely already be taking antidepressants for PND - likely >90%, according to EAG clinical advisors. Our advisors also noted that a large proportion patients seen by perinatal psychiatrists would be considered non-responders to antidepressants prescribed before referral to secondary	The company propose that the following sentence is removed: “The trial sample may therefore not fully represent those considered for treatment with zuranolone in the UK.”	The company would like to emphasise the following points: • The company disagree with the EAG position on using the ADT baseline use in SKYLARK as the main criterion of whether the trial population is generalizable to the NHS setting. As stated by the clinical advisors consulted by the EAG (page 35 of the EAG report) *“EAG clinical advisors did not think the difference in baseline ADT use between SKYLARK and the UK NHS setting would introduce significant issues in terms of the applicability of the trial evidence.”* • As communicated at the clarification stage, following the company’s latest discussions with the MHRA, it is likely that the target	This is not a factual inaccuracy. No changes were made. As explained under Issue 2 above, the sentence from page 35 of the EAG report was overinterpreted by the company. The expert opinion given in this sentence refers only to the differences in the specific types of antidepressants used in SKYLARK and the NHS practice, not to the proportion of participants on baseline ADT.

care. In comparison, only 15/98 (15.3%) in the zuranolone 50 mg group and 15/98 (15.3%) in the placebo group of the SKYLARK trial. The trial sample may therefore not fully represent those considered for treatment with zuranolone in the UK.”		population for the marketing authorisation of zuranolone in the UK will be limited to [redacted], based on the clinical evidence provided by the SKYLARK and ROBIN trials. • The company believes that the evidence package for zuranolone is sufficient to demonstrate efficacy of the treatment in moderate to severe PND, as the SKYLARK and ROBIN trials included patients with moderate (45.9% and 40.7%, respectively) and severe (54.19% and 59.3%, respectively) disease (based on literature-established MADRS thresholds) – as presented in the CS (Section 2.8.2.2), results of the post-hoc analyses [redacted] (Figure 26	

		and Figure 27 of the CS), suggesting efficacy in patients with both moderate and severe PND.	
Page 45-46: "All outcomes were judged at some concerns of risk of bias overall. The domain "measurement of the outcome" was judged at some concerns for all outcomes at all timepoints because outcome assessors were likely aware of the intervention received due to functional unblinding and this may have influenced assessment using subjective scales."	The company propose that the wording is updated: "All outcomes were judged at some concerns of risk of bias overall. The domain "measurement of the outcome" was judged at some concerns for all outcomes at all timepoints because outcome assessors were likely aware of the intervention received due to functional unblinding and this may have influenced assessment using subjective scales. **As concluded by the FDA, despite the risk of functional unblinding, the robust efficacy findings from SKYLARK overcome confounding from this potential bias**."	The company would like to highlight that as presented in the CS (Section 2.5), the FDA concluded that the robust efficacy findings from SKYLARK and ROBIN overcome confounding from the potential bias introduced by functional unblinding.	This is not a factual inaccuracy. No changes were made.

Page 48: "Similarly to the SKYLARK trial (section 3.2.1.2), the EAG notes concerns regarding the alignment of the ROBIN trial population with the NICE scope due to it focusing on women with severe PND (not "adults with PND") and exclusion of patients with psychosis and attempted suicide who would be seen in UK practice. This introduces uncertainty in the generalisability of the trial results to the population outlined in the NICE scope. "	The company propose that the wording is updated: "Similarly to the SKYLARK trial (section 3.2.1.2), the EAG notes concerns regarding the alignment of the ROBIN trial population with the NICE scope due to it focusing on women with severe PND (not "adults with PND") and exclusion of patients with psychosis and attempted suicide who would be seen in UK practice, **however, it is acknowledged that such exclusions are common in clinical trials."**	The company would like to highlight that the exclusion of patients with high risk of suicide and psychotic symptoms is a common practice in clinical trials.	This is not a factual inaccuracy. Although the exclusion of patients with high risk of suicide and psychotic symptoms is a common practice in clinical trials, this does not mitigate the issues raised. Nevertheless, the EAG has included this edit, given that there was a similar statement on page 35 with regards to the SKYLARK trial. The edited text on page 48 now reads: "Similarly to the SKYLARK trial (section 3.2.1.2), the EAG notes concerns regarding the alignment of the ROBIN trial population with the NICE scope due to it focusing on women with severe PND (not "adults with PND") and exclusion of patients with psychosis and attempted suicide who would be seen in UK practice **(although it is acknowledged that the exclusions of higher risk patients are common in clinical trials)**. This introduces uncertainty in the generalisability of the trial

			results to the population outlined in the NICE scope. “
Page 59: “The EAG has some concerns that the effectiveness outcome chosen to input into the economic model was EPDS, rather than the primary outcome of the SKYLARK trial, HAMD-17. EAG notes that EPDS is a screening tool for identifying people at risk for PND and was not developed nor validated for diagnosis of PND, disease monitoring, or for the measurement of depression severity in either clinical or research settings.”	The company propose that concerns relating to the use of EPDS in the economic model are not applicable and should be removed.	EPDS is the most frequently used depression screening tool in clinical practice and is recommended for monitoring of PND patients by the NICE Guideline CG192: “1.4.9: *Discuss and plan how symptoms will be monitored (for example, by using validated self-report questionnaires, such as the Edinburgh Postnatal Depression Scale [EPDS], Patient Health Questionnaire [PHQ-9] or the 7-item Generalized Anxiety Disorder scale [GAD-7])”* and *“1.5.8: At all contacts after the first contact with primary care or the booking visit, the health visitor, and other healthcare professionals who have regular contact with a woman in pregnancy and the postnatal period (first year after birth), should consider:* • *asking the 2 depression identification questions and*	We have amended the sentence on page 59 to reflect the suggested use of EPDS for disease monitoring in the NICE scenario. The amended sentence now reads: “EAG notes that EPDS is a screening tool for identifying people at risk for PND and was not developed ~~nor validated for diagnosis of PND, disease monitoring, or~~ for the measurement of depression severity in ~~either clinical or~~ research settings,[12] **although the use of EPDS is suggested in a NICE Scenario for monitoring PND in clinical practice.[13]”** A similar sentence on page 35 was also amended accordingly, for consistency. It is not a factual inaccuracy for the EAG to raise the disparity between the primary outcome for the trial and the outcome used in

		the GAD-2 (see recommendation 1.5.4) as part of a general discussion about her mental health and wellbeing and • *using the EPDS or the PHQ-9 as part of monitoring.“* As explained in the CS (Section 3.2.2), the EPDS was chosen as the preferred endpoint to assess PND symptoms because it is a commonly used to measure PND in clinical practice, and was the tool used to collect patient outcomes in ALSPAC, which is the best available source of data on the long-term impact of PND on patients, and further supports use of EPDS in the real-world clinical practice. Therefore, the use of EPDS from SKYLARK allows for the short-term model to be linked with the long-term ALSPAC analysis, allowing for an estimate of the EPDS trajectory of the modelled PND patients over a 29-year time horizon based on the EPDS score and age. Using HAMD-17 as the model outcome (as suggested by the EAG) would	the model. No changes were made to the remaining text in the paragraph. However, we have added a sentence to refer to the company’s justification for the use of EPDS: “The company provided further justification for the choice of EPDS during the factual accuracy check.”

		prevent the company from leveraging the rich data set offered by the ALSPAC study, which allows for long-term estimations natural history outcomes.	

Issue 4 Extrapolation of treatment effect over long-term

Description of problem	Description of proposed amendment	Justification for amendment	EAG Response
Page 67: "The company assumes that the difference in EPDS between treatment arms at 3 months post treatment initiation (which is 8 months post-birth), continues into the long-term for the full time horizon of the model (29 years post-birth). The EAG's clinical advisors found this assumption to be implausible, especially since zuranolone treatment is a single	The company proposes that the longer-term assumptions are described more clearly, the current wording implies that the model applies a long-term zuranalone treatment effect.	The company would like to clarify the following points: • The underling clinical assumption in the model is that zuranalone would improve EPDS outcomes between day 0 and 3-months (8-months post-birth) – from this point onwards, zuranalone no longer has an explicit impact on future outcomes, and the EPDS attained at month 3 are extrapolated into the longer term using observations from ALSPAC. • An alternative way to think about and operationalise	The current wording accurately describes the assumption in the company's model. The zuranolone arm is extrapolated using ALSPAC, but the ECM arm is obtained by assuming a constant difference in EPDS equal to that from the model 8-months post-birth. This means that the treatment difference 8-months post-birth continues into the long-term. The EAG has therefore not amended this wording. The company is making the case for EPDS 8-month post-birth to be a surrogate endpoint for the long-term EPDS trajectory, arguing that

course treatment given for 14 days and there is no data on treatment efficacy beyond 45 days”		the longer-term model would be to apply the ALSPAC MMRM model to both the zuranolone + ECM and ECM arms separately. Now, if one assumed that two different patients (one treated with ECM and one treated with zuranolone) both achieved an EPDS score of 11 at month 3 in the model (8-months post-birth), both patients would have the same EPDS trajectory for the model time horizon. • The model does not assume that there is a longer-term benefit perse for zuranalone, rather the assumption is that there would be no reason why the EPDS gains achieved in the short-term would not then follow the same trajectory as the observed data in ALSPAC. The company are not aware of any clinical reasons why improvements	the treatment effect attained by 8-months post birth would have shifted the EPDS trajectory permanently. This means that the trajectory of a patient whose EPDS has been lowered by zuranolone is now equivalent to the trajectory of someone who had that lower EPDS on ECM. This is a strong assumption, and we do not have any comparative evidence on long-term trajectories for zuranolone vs ECM.

		in EPDS would only diminish in the zuranalone arm (and not in the ECM arm) and follow a different trajectory to that established by the ALSPAC dataset.	
Page 68: "The company state that this "results in a convergence of EPDS scores between zuranolone + ECM and ECM for patients who have persistent PND". Whilst the curves for the two arms move a little closer together they still remain substantially separated for the duration of the model time-horizon, which the EAG believes is a strong assumption that is not substantiated by the available evidence."	The company propose removing the statement "which the EAG believes is a strong assumption that is not substantiated by the available evidence."	As demonstrated by the real-world data from the ALSPAC cohort, the long-term EPDS score trajectory of a patient is dependent on the EPDS score at 8 months, and from 8 months onwards patient's EPDS remains stable. There is no evidence to substantiate the EAG assumption that a patient previously treated with zuranolone would have a marked increase in their EPDS score compared to a patient in the ECM arm. Additionally, there is a lack of justification as to why there is a need for the long-term EPDS curves to converge over time. The approach proposed by the company more accurately reflects real world clinical practice whereby patients who have persistent depression (EPDS ≥13) may be expected to receive further	The ALSPAC data shows the EPDS trajectory for patients on what was ECM at the time the ALSPAC cohort had PND. It does not show the trajectory for patients on zuranolone, and does not provide evidence on the comparative difference in EPDS trajectories over time for zuranolone vs ECM. It is therefore the case that the assumptions are "not substantiated by the available evidence", and so this is not a factual inaccuracy. We have however edited this to clarify, so the paragraph now reads: "Whilst the curves for the two arms move a little closer together they still remain substantially separated for the duration of the model time-horizon. The ALSPAC data shows the EPDS trajectory for ECM at

		intervention to decrease their EPDS scores, regardless of the previous treatments a patient may have received. The company provides further commentary in the justification for amendment to Issue 4, Pg 67.	the time the ALSPAC cohort had PND. It does not show the trajectory for patients on zuranolone, and does not provide evidence on the comparative difference in EPDS trajectories over time for zuranolone vs ECM. Therefore the EAG believes the company's model makes a strong assumption that is not substantiated by the available evidence. "
Page 69: "The extrapolation of the treatment effect is highly uncertain, however, the EAG prefers the scenario where waning begins 1 year (post treatment initiation), and lasts for 1 year, as this acknowledges that there may be treatment benefit	The company would like the report remove the concept of treatment waning in the long-term.	As described in Issue 4, long-term EPDS scores are based on the EPDS achieved at month 3 rather than an ongoing treatment effect. The company provides further commentary in the justification for amendment to Issue 4, Pg 67.	This is not a factual inaccuracy. No change made.

beyond the period typically considered PND (1 year following birth), but that this would not continue for an implausibly long time.”			

Issue 5 Concomitant anti-depressant therapy (ADT) use in the company’s two studies and NHS practice

Description of problem	Description of proposed amendment	Justification for amendment	EAG Response
Page 65: “The EAG explores the impact of the level of baseline ADT used in the MMRM model, with scenarios for % baseline ADT use, and uses 90% in its base-case. Note however, that these scenarios only capture the impact of ADT on EPDS, and not an interaction with zuranolone.”	The company would like to amend the wording to the following: “The EAG explores the impact of the level of baseline ADT used in the MMRM model **that forces the inclusion of baseline ADT,** with scenarios for % baseline ADT use, and uses 90% in its base-case. Note however, that these scenarios only capture the impact of ADT on EPDS, and not an interaction with zuranolone”	ADT use at baseline does not lead to the best fitting model, rather the base case short-term MMRM model forced baseline ADT use.	Edit made, as suggested.

Issue 6 Short-term evidence sources inform long-term costs and effects in the model

Description of problem	Description of proposed amendment	Justification for amendment	EAG Response
Page 76-77: "The EAG does not agree this difference would be perpetuated over 27.5 years without a waning of effect. The EAG further notes that patients in Byford's study are no longer patients treated in secondary care, as our PND study population. It is unclear whether severe PND patients whose symptoms had not yet resolved by 18 months post-birth, would have continued their treatment in secondary care rather than in primary care, as in Byford's study population."	The company propose that the reference to waning effect is removed.	As described in Issue 4, long-term EPDS scores are based on the EPDS achieved at month 3 rather than an ongoing treatment effect. The company provides further commentary in the justification for amendment to Issue 4, Pg 67.	This is not a factual inaccuracy. No change made.

Issue 7 Choice of tool and validity of utility estimates

Description of problem	Description of proposed amendment	Justification for amendment	EAG Response
Page 75: "The EAG acknowledges there are caveats in using this approach. TA367 is an appraisal of a recent antidepressant drug for MDD, not PND. It is largely based on EQ-5D utilities from industry trials (the REVIVE study of vortioxetine) and Sapin et al study,37 a French study reporting EQ-5D utility values using UK tariffs in MDD patients. Ideally, to inform these longer-term health states of this model, there would have been a systematic review of all MDD and chronic MDD studies collecting utilities data, and a synthesis of UK utility estimates found using meta-analysis. The EAG also acknowledges	The company propose that the EAG report is amended to include an additional limitation of using the TA367 "remission" and "non-remission" utility values in the long-term, specifically there would be an underlying assumption that a patient with an EPDS of 14 would have the same utility score as a patient with an EPDS of 30.	Notwithstanding the MDD and PND are different diseases with different pathophysiology, one of the additional issues with using the MDD utility scores from TA367 is a lack of granularity at specific EPDS scores. The mapped EQ-5D values from ROBIN provide the most granular estimate of utility based on EPDS scores, and are therefore more appropriate to use in the long-term compared to TA367, which only consider "remission" or "non-remission". The company believe that it is implausible to assume that a patient with EPDS score of 15 would have the same utilities as a patient with EPDS score of 30, even though both would technically be considered as "non-remission" patients.	The EAG agrees with the company in that there is an assumption that patients in "non-remission" are treated the same regardless of score above 13. It has therefore already stated in Page 75 "EAG also acknowledges that they may not be a clear linear link between what clinicians classify as "remission" and "non-remission" and the 13 point EPDS score cut-off." The EAG has now added: "This would mean that all patients with an EPDS above or below 14 would be treated the same, which is a common oversimplification in economic models of depression."

that they may not be a clear linear link between what clinicians classify as “remission” and “non-remission” and the 13 point EPDS score cut-off. Nevertheless, the EAG prefers to use TA367 utilities to model health-related quality of life gains at one year post-intervention because they are derived using EQ-5D utilities, rather than SF6D utilities, as per NICE reference case, it dichotomises health states in “remission” and “non-remission” as traditionally used in the MDD literature, clinical experts have advised that PND at one year post-birth is similar to MDD, and the utility values picked were the NICE committee’s preferred values in a recent MDD appraisal.”			Further to the above, clinical advice to the EAG have strongly advised that PND is only treated as a different condition in the first 12 months post-birth. After that, EAG clinicians advised that PND and MDD are “virtually indiscernible” and treated the same. Note that the company model costs based on remission, and so the same limitation applies to the way costs have been captured in the model.

Issue 8 Uncaptured costs and benefits

Description of problem	Description of proposed amendment	Justification for amendment	EAG Response
Page 19-20: "Zuranolone may have other rare but serious longer-term side effects, such as decreased consciousness, and potential for abuse."	The company would like to amend the wording to the following: "Zuranolone may have other rare but serious longer-term side effects, such as decreased consciousness, and potential for abuse. **However, the EAG notes that addiction or withdrawal symptoms were not observed in the zuranolone studies.**"	At the clarification stage, in the response to question B11, the company stated that there is no evidence in the clinical trial that zuranolone leads to addiction or suggesting the potential for abuse. As addiction (in the SKYLARK study captured as withdrawal symptoms assessed by 20-item Physician Withdrawal Checklist) was not observed, there were no incidence rates that could be included in the economic model. Due to a one-off, short-course nature of zuranolone treatment, and the fact that – in line with the draft Summary of Product Characteristics from MHRA – zuranolone is expected to be initiated under the supervision of a specialist prescriber, patients are not expected to have access to extra supply of	The EAG is happy to expand on this point to clarify in pages 19-20 and 77: "Zuranolone may have other rare but serious longer-term side effects, such as decreased consciousness, and potential for abuse. **However, the EAG notes these were not observed in the zuranolone studies.**"
Page 77: "Clinical advisors to the EAG suggested that patients on zuranolone could be at risk of decreased consciousness and there may be potential for abuse. These are rare but serious adverse events currently observed in patients on	The company would like to amend the wording to the following: "Clinical advisors to the EAG suggested that patients on zuranolone could be at risk of decreased consciousness and there may be potential for abuse. These are rare but serious adverse events currently observed in patients on brexanolone, a similar drug for PND in current use.		We have added "**However, the EAG notes these were not observed in the zuranolone studies.**"

brexanolone, a similar drug for PND in current use.”	**However, the EAG notes that addiction or withdrawal symptoms were not observed in the zuranolone studies.”**	zuranolone and therefore the long-term risk of addiction is limited. Somnolence has been included as an adverse event in the economic model, and therefore the model accounts for the risk of decreased consciousness. Finally, the company would like to highlight that there are no head-to-head studies of zuranolone and brexanolone and therefore it is not appropriate to make direct comparisons between the two treatments. There are different administration models for zuranolone (a short-course oral treatment) and brexanolone, which required an in-patient treatment model due to the 60-hour infusion time and boxed warning for confusional state. The company would like to highlight that brexanolone is no longer distributed in the US,	

		and is not approved in any other country or region.	

Issue 9 Typographical errors, formatting errors and minor text alterations

Description of problem	Description of proposed amendment	Justification for amendment	EAG Response
Page 36: Table 5: ≥28 HAMD-17 total score: 65 (66.6); 66 (67.3)	The company propose the following change: ≥28 HAMD-17 total score: 65 (**66.3**); 66 (67.3)	Incorrect % of patients receiving zuranolone 50 mg with ≥28 HAMD-17 total score	Edit made as suggested.
Page 42: "The company also noted that they used literature-established thresholds for the MARS to categorise participants baseline severity as moderate (total score ≤34) or severe (total score ≥35), with 89 (45.9%) participants categorised as having moderate depression at baseline, and	The company propose updating the text to: "The company also noted that they used literature-established thresholds for the **MADRS** to categorise participants baseline severity as moderate (total score ≤34) or severe (total score ≥35), with 89 (45.9%) participants categorised as having moderate depression at baseline, and 105 (54.19%) considered as having severe PND."	Typographical error	Edit made as suggested.

105(54.19%) considered as having severe PND.”			
Page 43: “The EAG also note that there appears to be a disproportionate number of participants from the “severe” subgroup at baseline with missing data for HAMD-17 score analyses on Day 15 and Day 45, when compared to participants from the “moderate” subgroup.”	The company propose updating the text to: “The EAG also note that there appears to be a disproportionate number of participants from the “severe” subgroup at baseline with missing data for **MADRS** score analyses on Day 15 and Day 45, when compared to participants from the “moderate” subgroup.”	Typographical error	Edit made as suggested.
Page 44: “On-treatment treatment emergent adverse events (TEAEs) were reported by 60.2% (59/89) and 41.8% (41/89) of patients receiving zuranolone 50 mg and placebo, respectively.”	The company propose updating the text to: “On-treatment treatment emergent adverse events (TEAEs) were reported by 60.2% (59/**98**) and 41.8% (41/**98**) of patients receiving zuranolone 50 mg and placebo, respectively.”	Typographical error	Edit made as suggested.

Issue 10 Incorrect confidentiality marking

Location of incorrect marking	Description of incorrect marking	Amended marking	EAG Response
Page 36, Section 3.2.1.2	The CIC marking currently provided is insufficient for the following item in Table 5: ≥28 HAMD-17 total score: [redacted]	Please update the CIC marking to this text as provided: ≥28 HAMD-17 total score: [redacted]	We apologise for missing these confidentiality markings and have corrected them all. Edit made as suggested.
Page 42, Section 3.2.1.4.1	The CIC marking currently provided is insufficient: "In section 2.8.2.1, the CS noted that although previous psychiatric history had been collected during screening, the subgroup sample size ([redacted]) was considered too small to produce a stable assessment of efficacy."	Please update the CIC marking to this text as provided: "In section 2.8.2.1, the CS noted that although previous psychiatric history had been collected during screening, the subgroup sample size ([redacted]) was considered too small to produce a stable assessment of efficacy."	Edit made as suggested.
Page 42, Section 3.2.1.4.1	The CIC marking currently provided is insufficient: "The company also noted that they used literature-established	Please update the CIC marking to this text as provided:	Edit made as suggested. In the same sentence "MARS" was also corrected to "MADRS".

	thresholds for the MARS to categorise participants baseline severity as moderate (total score ≤34) or severe (total score ≥35), with [redacted] participants categorised as having moderate depression at baseline, and [redacted] considered as having severe PND.”	“The company also noted that they used literature-established thresholds for the MARS to categorise participants baseline severity as moderate (total score ≤34) or severe (total score ≥35), with [redacted] participants categorised as having moderate depression at baseline, and [redacted] considered as having severe PND.”	
Page 43, Section 3.2.1.4.1	The CIC marking currently provided is insufficient: [redacted]	Please update the CIC marking to this text as provided: [redacted]	Edit made as suggested.
Page 44, Section 3.2.1.4.1	The CIC marking currently provided is insufficient: “Additionally, the EAG note that the numbers per subgroup in	Please update the CIC marking to this text as provided:	Edit made as suggested.

	Figure 1 column “Count” do not add up ([redacted]). The EAG acknowledge that this might be due to missing data or a typo, and this is not expected to have a significant impact on the validity of the results.”	“Additionally, the EAG note that the numbers per subgroup in Figure 1 column “Count” do not add up ([redacted]). The EAG acknowledge that this might be due to missing data or a typo, and this is not expected to have a significant impact on the validity of the results.”	
Page 44-45, Section 3.2.1.7	The CIC marking currently provided is insufficient: “Major protocol deviations from SKYLARK were reported in section 10.2 and table 14.1.2.1 of the CSR. In the zuranolone group, [redacted] participants [redacted] had at least one major protocol deviation ([redacted]). [redacted] patients were excluded from the per-protocol set due to a major protocol deviation in this group. In the placebo group, [redacted] participants ([redacted]) had at least one major	Please update the CIC marking to this text as provided: “Major protocol deviations from SKYLARK were reported in section 10.2 and table 14.1.2.1 of the CSR. In the zuranolone group, [redacted] participants [redacted] had at least one major protocol deviation ([redacted]). [redacted] patients were excluded from the per-protocol set due to a major protocol deviation in this	Edit made as suggested.

	protocol deviation ([redacted]). [redacted] patients were excluded from the per-protocol set in the placebo group due to having significant deviations that could affect efficacy outcomes. Minor protocol deviations were not reported."	group. In the placebo group, [redacted] participants ([redacted]) had at least one major protocol deviation ([redacted]). [redacted] patients were excluded from the per-protocol set in the placebo group due to having significant deviations that could affect efficacy outcomes. Minor protocol deviations were not reported."	
Page 50, Section 3.2.2.3.1	The CIC marking currently provided is insufficient: "A summary of all concomitant medications used by patients in the trial is provided in CSR Table 14.1.3.2. In the participants who used an ADT at baseline, [redacted] in the zuranolone group and [redacted] in the placebo group started a new ADT during the follow-up period of the trial (CSR Table 14.1.3.4). [redacted] participants had an increase in the dose of their baseline ADT during the	Please update the CIC marking to this text as provided: "A summary of all concomitant medications used by patients in the trial is provided in CSR Table 14.1.3.2. In the participants who used an ADT at baseline, [redacted] in the zuranolone group and [redacted] in the placebo group started a new ADT during the follow-up period of the trial (CSR Table 14.1.3.4). [redacted] participants had	Edit made as suggested.

	treatment period or follow-up. In those who did not use an ADT at baseline, [redacted] in the zuranolone group and [redacted] in the placebo group started an ADT during follow-up. In the total sample, [redacted] patients started a new ADT during the treatment period (this was disallowed by the trial)."	an increase in the dose of their baseline ADT during the treatment period or follow-up. In those who did not use an ADT at baseline, [redacted] in the zuranolone group and [redacted] in the placebo group started an ADT during follow-up. In the total sample, [redacted] patients started a new ADT during the treatment period (this was disallowed by the trial)."	
Page 53, Section 3.2.2.6	The CIC marking currently provided is insufficient: "***Patient Health Questionnaire (PHQ-9):*** Responses to the PHQ-9 were also collected in the ROBIN trial, as pre-specified in the protocol. These data are not reported in the CS, and do not input into the company's economic model, but are outlined in the CSR (Tables 14.2.6.1 and 14.2.6.2). At day 45, data were available for [redacted] participants in the placebo group and [redacted] in the zuranolone	Please update the CIC marking to this text as provided: "***Patient Health Questionnaire (PHQ-9):*** Responses to the PHQ-9 were also collected in the ROBIN trial, as pre-specified in the protocol. These data are not reported in the CS, and do not input into the company's economic model, but are outlined in the CSR (Tables 14.2.6.1 and 14.2.6.2). At day 45, data were available for	Edit made as suggested.

	30mg group. The LS mean change (SE) was [redacted] for the placebo group and [redacted] for the zuranolone 30mg group. The treatment difference and corresponding 95% CI was [redacted]. The company notes that improvement in PHQ-9 was numerically greater at all measured timepoints from day 3 to day 45 for the zuranolone group versus placebo group."	[redacted] participants in the placebo group and [redacted] in the zuranolone 30mg group. The LS mean change (SE) was [redacted] for the placebo group and [redacted] for the zuranolone 30mg group. The treatment difference and corresponding 95% CI was [redacted]. The company notes that improvement in PHQ-9 was numerically greater at all measured timepoints from day 3 to day 45 for the zuranolone group versus placebo group."	
Page 53-54, Section 3.2.2.6	The CIC marking currently provided is insufficient: "Section 10.2 and Table 14.1.6 of the CSR reports protocol deviations. In the zuranolone group, [redacted] participants ([redacted] had at least one protocol deviation. Of these, [redacted] participants [redacted] had a minor only protocol deviation and [redacted] had at least one major protocol deviation	Please update the CIC marking to this text as provided: "Section 10.2 and Table 14.1.6 of the CSR reports protocol deviations. In the zuranolone group, [redacted] participants ([redacted] had at least one protocol deviation. Of these, [redacted] participants [redacted] had a minor only protocol deviation and	Edit made as suggested.

	( ). patients were excluded from the per-protocol set due to a major efficacy deviation in the zuranolone group. In the placebo group, participants ( ) had at least one protocol deviation. Of these, participants ( ) had a minor only protocol deviation and had at least one major protocol deviation ( ). patients were excluded from the per-protocol set in the placebo group due to having significant deviations that could affect efficacy outcomes. The company reported that for of patients in total who had a major protocol deviation that may have affected efficacy, the reason was due to the day 15 visit being conducted “out of window”. participants had their day 15 visit ±1 day out of window, had it ±2 days out of window, had it ±3 days out of	had at least one major protocol deviation ( ). patients were excluded from the per-protocol set due to a major efficacy deviation in the zuranolone group. In the placebo group, participants ( ) had at least one protocol deviation. Of these, participants ( ) had a minor only protocol deviation and had at least one major protocol deviation ( ). patients were excluded from the per-protocol set in the placebo group due to having significant deviations that could affect efficacy outcomes. The company reported that for of patients in total who had a major protocol deviation that may have affected efficacy, the reason was due to the day 15 visit	

	window, and [redacted] did not complete it. All day 15 visits conducted out of window were used in the efficacy analyses. No further information on protocol deviations was provided to the EAG.”	being conducted “out of window”. [redacted] participants had their day 15 visit ±1 day out of window, [redacted] had it ±2 days out of window, [redacted] had it ±3 days out of window, and [redacted] did not complete it. All day 15 visits conducted out of window were used in the efficacy analyses. No further information on protocol deviations was provided to the EAG.”	