Ganaxolone for treating seizures caused by CDKL5 deficiency disorder in people 2 years and over [ID3988]: A Single Technology Appraisal

Produced by

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

Authors

Darren Burns[1]
Zoe Philips[1]
Justin Matthews[2]
Ash Bullement[1]
Laura Trigg[2]
Simon Briscoe[3]
Joseph Symonds[4]
Sam Amin[5]
Dougal Hargreaves[6]
G.J. Melendez-Torres[2]
Caroline Farmer[2]

[1] Delta Hat, Ltd
[2] Peninsula Technology Assessment Group (PenTAG), University of Exeter Medical School, Exeter
[3] University of Exeter Medical School, Exeter
[4] Royal Hospital for Children, Glasgow
[5] University Hospitals Bristol and Weston NHS
[6] Imperial College London

Correspondence to Caroline Farmer

Ganaxolone for treating seizures caused by CDKL5 deficiency disorder in people 2 years and over [ID3988]: A Single Technology Appraisal

Produced by	**Peninsula Technology Assessment Group (PenTAG)** **University of Exeter Medical School** **South Cloisters** **St Luke's Campus** **Heavitree Road** **Exeter** **EX1 2LU** 3.09 South Cloisters, St Luke's Campus, Heavitree Road, Exeter, EX1 2LU; c.farmer@exeter.ac.uk
Date completed	23/01/2023
Source of funding	This report was commissioned by the NIHR Evidence Synthesis Programme as project number NIHR135705.
Declared competing interests of the authors	Dr Amin was a trial manager for Marigold and provided advice to the company that informed their submission. He has never received money from the company. In his role as national lead for the CDKL5, Dr Amin has sought funding from industry on behalf of the organisation.
Acknowledgments	The authors acknowledge the administrative support provided by Mrs Sue Whiffin and Ms Jenny Lowe (both PenTAG).
Rider on responsibility for document	The views expressed in this report are those of the authors and not necessarily those of the NIHR Evidence Synthesis Programme. Any errors are the responsibility of the authors.
This report should be referenced as follows	Burns, Phillips, Matthews, Bullement, Trigg, Briscoe, Symonds, Amin, Hargreaves, Melendez-Torres, Farmer. Ganaxolone for treating seizures caused by CDKL5 deficiency disorder in people 2 years and over [ID3988]: A Single Technology Appraisal. Peninsula Technology Assessment Group (PenTAG), 2023.
Copyright	© 2023, PenTAG, University of Exeter. Copyright is retained by Marinus Pharmaceuticals Inc. for tables and figures copied and/or adapted from the company submission and other submitted company documents.

Author Contributions:	
Darren Burns	Critical appraisal of the economic evidence and analysis submitted by the company, conducted additional economic analyses and drafted sections of the report
Zoe Phillips	Critical appraisal of the economic evidence and analysis submitted by the company, and drafted sections of the report
Justin Matthews	Critical appraisal of the statistical analyses conducted by the company throughout its submission, and drafted sections of the report
Ash Bullement	Critical appraisal of the economic evidence submitted by the company, conducted additional economic analyses and drafted sections of the report
Laura Trigg	Critical appraisal of the clinical trials of ganaxolone and drafted sections of the report
Simon Briscoe	Critical appraisal of the company's literature search strategies
Joseph Symonds	Expert clinical advice about CDKL5 deficiency disorder and its treatment

Author Contributions:	
Sam Amin	Expert clinical advice about CDKL5 deficiency disorder and its treatment
Dougal Hargreaves	Dr Hargreaves is a general paediatrician with expertise in epileptic disorders. He provided clinical advice about epilepsy amongst children and NHS services.
G.J. Melendez-Torres	Critical appraisal of the company submission, writing and editorial input
Caroline Farmer	Project lead, critical appraisal of the company submission, writing and editorial input

Table of Contents

List of key issues

List of tables

List of Figures

Abbreviations

A&E	Accident and emergency
ADAMS	Anxiety, depression and mood scales
AE	Adverse event
ASMs	Anti-seizure medications
CDD	CDKL5 deficiency disorder
CDKL5	Cyclin-depended Kinase-like 5
CEAC	Cost-effectiveness acceptability curve
CGI	Clinical Global impressions
CGI-CSID	CGI of change in seizure intensity, duration and severity
CI	Confidence interval
CS	Company submission
CSHQ	Childrens sleep habit questionnaire
CSR	Clinical Study Report
CVI	Cortical Visual Impairment
DS	Dravet syndrome
DSU	Decision Support Unit
EAG	External Assessment Group
ECM	Established clinical management
EQ-5D	EuroQol five dimension
FS	Focal Seizures
GNX	Ganaxolone
GP	General practitioner
HL	Hodges-Lehmann
HRQoL	Health-related quality of life
HSUV	Health State utility value
HTA	Health technology assessment
ICER	Incremental cost-effectiveness ratio
IRQ	Inter Quartile Range
ITT	Intention-to-treat
KOL	Key opinion leader
LGS	Lennox-Gastaut Syndrome
MECP2	methly-CpG-binding protein 2
MHRA	Medicines and Healthcare products Regulatory Agency
NA	Not applicable
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NMA	Network meta-analysis
NR	Not reported
OLE	Open label extension
ONS	Office of National Statistics

A&E	Accident and emergency
OR	Odds Ratio
OWSA	One-way sensitivity analysis
PBO	Placebo
PCSF	Percentage change in 28 day seizure frequency
PSA	Probabilistic sensitivity analysis
PSSRU	Personal Social Services Research Unit
QA	Quality assessment
QALY	Quality-adjusted life year
RCT	Randomised controlled trial
SD	Standard deviation
SE	Standard error
SF	Seizure frequency
SFD	Seizure-free days
SLR	Systematic literature review
SmPC	Summary of product characteristics
SMR	Standardised Mortality Ratio
SUDEP	Sudden Death From Epilepsy
TA	Technology Appraisal
TEAE	Treatment emergent adverse events
USD	United States Dollar
VNS	Vagus Nerve simulation
VS	Versus
WTP	Willingness to pay

1. EXECUTIVE SUMMARY

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

Section 1.1 provides an overview of the key issues. Section 1.2 provides an overview of key model outcomes and the modelling assumptions that have the greatest effect on the ICER. Sections 1.3 to 1.6 explain the key issues in more detail. Background information on the condition, technology and evidence and information on other issues identified by the EAG are in the main report.

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

1.1. Overview of the EAG's key issues

A brief overview of the key issues identified by the EAG in their appraisal of the company submission (CS) is provided in Table 1. Further detail of the issues is provided in Sections 1.3, 1.4, 1.5, and 1.6.

Broadly speaking the key clinical issues related to the extent of a long-term treatment effect in the open-label extension of the pivotal trial.

In terms of cost effectiveness issues, the EAG noted several key issues. These have varying impacts on the cost-effectiveness of ganaxolone (GNX), though generally increase the ICER.

Table 1: Summary of key issues

ID	Summary of issues	Report section(s)
Long-term treatment effect	The EAG identified quality concerns with the OLE of Marigold, which increase uncertainty in the trial results beyond the double-blind period (>17 weeks). The concerns include a high rate of attrition that is associated with treatment outcome, and the risk that some reductions in SF may be driven by regression towards the mean.	3.2.2.5 and 4.2.6.1
Model structure	The company used a simple model structure, which limits its ability to represent the condition and likely treatment pathway. The potential impact of this on the results was unclear.	4.2.2
Seizure frequency	The company's model structure imposed many assumptions on the distribution and behaviour of seizure frequency, as	4.2.6.1

ID	Summary of issues	Report section(s)
	well as the effect of GNX. The net effect of these was likely to be an optimistic estimate of the clinical benefit of GNX.	
Consistency of disease proxies throughout the model	The company's base-case model used different diseases to proxy CDD mortality and healthcare resource use compared to patient HRQoL, creating inconsistency. Using the same disease to inform all of these considerably worsened the cost-effectiveness of GNX	4.2.7 and 4.2.8
Modelling errors	Correcting the errors in the company cost effectiveness model had a considerable impact on the ICER.	6.1.1; 6.1.2; 6.1.3; 6.1.4; 6.1.5; 6.1.6; and 6.1.7
Disease severity modifier and caregivers	The company base case included a severity multiplier of 1.7 for both incremental patient and caregiver QALYs. The NICE methods guidance is unclear about whether a severity multiplier should be applied to caregiver QALYs, though the EAG were of the view that this was not appropriate.	6.2.4.2

Abbreviations: CDD, CDKL5 deficiency disorder; EAG, external assessment group; GNX, ganaxolone; HRQoL, health-related quality of life; ICER, incremental cost effectiveness ratio; OLE, open-label extension; QALY, quality-adjusted life-year; SF, seizure frequency

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

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

	Company's preferred assumption	EAG preferred assumption	Report Sections
Discontinuation rate	This was calculated based on the number of participants and the number of discontinuations at the end of the Marigold OLE	This was calculated based on the exposure time in Marigold (i.e., time at risk of discontinuation) and the number of discontinuations	4.2.6.2 and 6.2.1
Health-related quality of life	Lo *et al.* vignette based on people with TSC	Auvin *et al.* based on people with LGS, which was consistent with inputs for HCRU and mortality	4.2.7; 4.2.8; 4.2.6.3; and 6.2.4
Dynamics of the treatment effect	The treatment effect from the end of the double-blind period of Marigold (17 weeks) applied from baseline with no transition or accumulation over time	The treatment effect was linearly interpolated based on half-cycle corrected data from the double-blind period of Marigold week 0-4 (titration period) and week 4-17 (maintenance period)	4.2.6.1 and 6.2.5
Cost of hospitalisation	Long-stay cost used based on Mangatt *et al.*	Short-stay cost based on the short average length of stay reported in Chin *et al.*	4.2.8

	Company's preferred assumption	EAG preferred assumption	Report Sections
Wastage	No wastage	10% wastage based on clinical expert advice	6.2.6
Severity modification for caregivers	Severity modifier applied to caregiver utilities, based on the QALY shortfall in patients (i.e., not based on caregiver QALY shortfall)	The EAG interpreted the NICE methods guide to exclude caregivers from disease severity modification. However, as this was unclear, this report presents the EAG preferred base case both with and without the severity modifier applied to caregivers	6.2.4.2

Abbreviations: EAG, external assessment group; HCRU, health care resource utilisation; LGS, Lennox-Gastaut syndrome; OLE, open-label extension; QALY, quality-adjusted life-year; TSC, Tuberous Sclerosis Complex

1.2. Overview of key model outcomes

NICE technology appraisals compare how much a new technology improves length of life (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 the frequency of seizures experienced by patients. Given improvements in seizure frequency are associated with improved health-related quality of life, GNX is modelled to generate more QALYs compared to established clinical management.

Overall, the technology is modelled to increase costs due to the cost of GNX for as long as patients are assumed to remain on treatment (in addition to the costs of established clinical management), and leads to a reduction in costs associated with hospitalisation and the use of rescue medications.

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

- Assumptions affecting seizure frequency, and the ability of GNX to affect it
- Selection of an appropriate source for utility data, and the implementation of the data
- The baseline age of the cohort at initiation of GNX
- Assumptions relating to the average length of stay for epilepsy-related hospitalisations

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

The EAG did not identify any key issues with regard to the decision problem for this appraisal.

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

Key Issue 1: Uncertainty surrounding clinical effects in the Marigold OLE

Report sections	**3.2.2.5 and 4.2.6.1**
Description of issue and why the EAG has identified it as important	The company argued that there was evidence of a sustained treatment effect of GNX in the Marigold OLE, however the EAG had concerns about the interpretation of these data. **1. Regression to the mean** Clinical experts to the EAG advised that people with CDD may initiate treatment for seizures following an exacerbation in seizure frequency. One expert described this as applicable to clinical trials also and would be like starting treatment at "the crest of a wave" of seizures. If this was the case, then a natural decline in seizure frequency would occur during trial follow-up, known as a 'regression towards the mean'. During the double-blind phase of Marigold, a significant minority of people in both treatment arms experienced reductions in seizure frequency, and it was unclear how many of these would have occurred naturally. However, relative effect sizes are able to generate an estimate of whether GNX delivered a benefit over and above ECM. In the OLE, however, there was no comparator arm, and it was therefore unclear to what extent reductions in seizure frequency were related to treatment. **2. Missingness due to treatment outcome** Participants receiving GNX in the double-blind phase of Marigold were permitted to discontinue treatment and not enter the OLE, and all participants in the Marigold OLE were permitted to discontinue at any time. Approximately 40% of participants receiving GNX withdrew from the trial before the latest data cut of the OLE, some of whom withdrew due to a lack of efficacy and some who withdrew for ambiguous reasons that the EAG considered could have been influenced by treatment efficacy (e.g. 'clinician decision'). The withdrawal of participants with a poor treatment response could cause an artificial drop in seizure frequency at follow-up timepoints.
What alternative approach has the EAG suggested?	It was not possible for the EAG to resolve this issue within its appraisal using the available data. Overall, the EAG considered the data from the double-blind phase of Marigold to be the highest quality data for decision-making, and that data from the OLE should be interpreted with extreme caution.
What is the expected effect on the cost-effectiveness estimates?	There was uncertainty surrounding the effect of GNX beyond the 17-week treatment period of the double-blind phase of Marigold, which had implications for modelling the long-term treatment effect within the lifetime horizon of the company model.
What additional evidence or analyses	There was limited information in the CS on the way in which participants in Marigold were recruited, though it is known that inclusion criteria included >16 major motor seizures per 28 days in a historical period. To assess the

Report sections	3.2.2.5 and 4.2.6.1
might help to resolve this key issue?	plausibility of a regression to the mean phenomenon: further trial details that indicate whether or not trial participants were more likely to be recruited when SF was intense; and longer term (>17 weeks) evidence (e.g. RWE or related disease) about stability/constancy of SF rates. Further data cuts from the Marigold OLE are expected (latest data cut to inform the CS was [redacted]). It would also be preferable to correct bias in the submitted SF analysis in the OLE phase using a missing data analysis which estimates SF for the full trial cohorts (i.e., analyses SF for all patients, including withdrawals).

Abbreviations: CDD, CDKL5 deficiency disorder; CS, company submission; EAG, External Assessment Group; GNX, ganaxolone; SF, seizure frequency; OLE, open-label extension; RWE, Real World Evidence

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

Key Issue 2: Model structure

Report sections	4.2.2
Description of issue and why the EAG has identified it as important	The company model was a simple Markov state- transition model with two primary health states (alive and dead) which may not have captured the full impact of the disease or treatment pathway, and may be considered atypical for NICE technology appraisals of genetic epileptic syndromes.
What alternative approach has the EAG suggested?	In its appraisal, the EAG suggested some alternative model structures which could (theoretically) be considered, though it is beyond the remit of the EAG to develop these further (and not possible with data the EAG was able to access).
What is the expected effect on the cost-effectiveness estimates?	The impact on the cost-effectiveness estimates was unclear.
What additional evidence or analyses might help to resolve this key issue?	Beyond re-developing the cost-effectiveness model using alternative structures, no additional analyses would help resolve this issue. However, provision of further justification for the choice of model structure (and dismissal of alternatives) may increase confidence in the structure chosen.

Abbreviations: EAG, External Assessment Group; NICE, National Institute for Health and Care Excellence.

Key Issue 3: Application of seizure frequency

Report sections	4.2.6, 6.1, and 6.2.5
Description of issue and why the EAG has identified it as important	The company's overall approach to capturing SF for both treatment arms incorporated a large number of assumptions which had a considerable impact on cost-effectiveness results. For example, only primary seizures were considered in the base case model, while secondary and tertiary seizures were omitted. The company also assumed that the distribution of SF observed in the Marigold trial was representative of UK clinical practice, could best be represented with a lognormal distribution and would not change over time. The company assumed that treatment effects were instantaneous and maintained provided the patient remains on treatment, reverting to baseline immediately after discontinuation of treatment. They also assumed it was appropriate to apply a HL shift directly to distributional

Report sections	4.2.6, 6.1, and 6.2.5
	parameters to model the treatment effect, and that treatment did not impact seizure type or severity. Further, the EAG identified an error in the application of the treatment effect in that the treatment effect of GNX was applied as a percentage reduction directly to the mean [redacted] distribution fit, which was a mathematical error due to it violating the product rule of logarithms.
What alternative approach has the EAG suggested?	The EAG disagreed that only primary seizures were relevant to the decision problem. Some data suggested that the treatment effect of GNX may differ by seizure type and incorporating all seizure types may have therefore better reflected the scope of the appraisal. However, given that different types of seizure may be associated with different costs and utilities, the scenario analysis considering 'all seizures' may be considered conservative. The EAG implemented a ''fix' for the application error within its base-case analysis and explored a number of other scenarios related to the application of treatment effect, including interpolation of the effect to account for time-varying treatment effects within the observed period (per Marigold evidence at 4 and 17 weeks).
What is the expected effect on the cost-effectiveness estimates?	The ICER increased substantially when addressing this error in application, and again when interpolating the treatment effect. The ICER fell slightly when using the maintenance period efficacy for interpolation between weeks 4 and 17.
What additional evidence or analyses might help to resolve this key issue?	No further evidence needed for implementation errors. However, statistical analysis of the GNX/GNX cohort in the Marigold OLE could provide more up to date data with longer follow-up on GNX treated patients (acknowledging the need to address Key Issue 1). Clinical opinion may also help to resolve uncertainty relating to the generalisability of SF observed in the Marigold trial to UK clinical practice.

Abbreviations: EAG, External Assessment Group; GNX, ganaxolone; HL, Hodges-Lehmann; ICER, incremental cost-effectiveness ratio; SF, seizure frequency

Key Issue 4: Utility values

Report sections	4.2.7
Description of issue and why the EAG has identified it as important	The utility values used to populate the model were taken from published vignette studies and were subject to limitations. As there was no survival benefit associated with GNX, the utility values were important drivers of the cost-effectiveness results, applying to both patients and caregivers.
What alternative approach has the EAG suggested?	The EAG preferred the utility values reported by Auvin *et al.* as these were more granular with respect to SF and were based on the same proxy condition used for both medical resource use frequencies and mortality (LGS). [redacted]

Report sections	4.2.7
What is the expected effect on the cost-effectiveness estimates?	Depending on the choices made to populate the model, the cost-effectiveness results may improve or worsen. Implications are presented in Section 6.2 of this report. In the EAG's base-case analysis (Auvin *et al.,* correcting the implementation of these utilities to absolute values), the ICER is increased substantially – see Section 6.3.
What additional evidence or analyses might help to resolve this key issue?	Clinical expert opinion may be sought on the applicability of different proxy conditions, and whether the source condition should be consistent for resource use, mortality, and HRQoL

Abbreviations: CDD, CDKL5 deficiency disorder; EAG, External Assessment Group; GNX, ganaxolone; HL, Hodges-Lehmann; HRQoL, health-related quality of life; ICER, incremental cost-effectiveness ratio; LGS, Lennox-Gastaut syndrome; SF, seizure frequency.

Key Issue 5: Miscellaneous model errors and unsubstantiated assumptions

Report sections	4.2.6, 4.2.8, 6.1, 6.2.6
Description of issue and why the EAG has identified it as important	The company model contained numerous errors. The errors with the largest impact on the ICER were: the incorrect application of the HL shift estimate to model treatment effect on SF distribution; not implementing age adjustment for caregivers (assuming them to be ageless); truncation of the SF distribution at 400 seizures; correction to incorrect age adjustment of patients; and correction of rescue medication cost estimates. In addition, the company's implementation of one-way sensitivity analyses was incorrect and the calculation of probabilistic ICERs, leading to an underestimation of the impact of individual parameter uncertainty on modelled outcomes. Key unsubstantiated assumptions included the instantaneous and infinitely durable nature of the treatment effect, a lack of any wastage of GNX.
What alternative approach has the EAG suggested?	The EAG corrected the objective errors in the modelling, and presented a base-case without the unsubstantiated assumptions made by the company
What is the expected effect on the cost-effectiveness estimates?	The EAG corrected company base case ICER was substantially higher than the company's base-case ICER. The EAG preferred base-case ICER was substantially higher than the willingness to pay threshold.
What additional evidence or analyses might help to resolve this key issue?	The EAG resolved a number of errors in the company model. To validate assumptions in the model with a large impact on the ICER, longer-term follow up data on the efficacy of GNX would be required.

Abbreviations: EAG, External Assessment Group; GNX, ganaxolone; ICER, incremental cost-effectiveness ratio; LGS, Lennox-Gastaut Syndrome; SF, seizure frequency.

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

Key Issue 6: Application of severity modifier

Report sections	6.2.4.2
Description of issue and why the EAG has identified it as important	The company applied a severity multiplier of 1.7 for both incremental caregiver and patient QALYs. The NICE methods guidance describes the severity modification applying to those "living with the disease", and the EAG was uncertain if this was also intended to applicable to caregivers.
What alternative approach has the EAG suggested?	The EAG explored scenarios with and without the severity modifier applied to caregiver QALYs.
What is the expected effect on the cost-effectiveness estimates?	The choice of severity modifier has a meaningful impact on the ICER.
What additional evidence or analyses might help to resolve this key issue?	Not applicable.

Abbreviations: EAG, External Assessment Group; QALY, quality-adjusted life year(s).

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

The EAG generated a base-case ICER of [redacted] with the implementation of the severity modifier for caregivers, and [redacted] without.

Table 3: Summary of EAG's preferred assumptions and ICER

Scenario	Incremental cost	Incremental QALYs	ICER (change from company base case)
Company's base case	[redacted]	[redacted]	£22,200
Correction 1: Incorrectly implemented treatment effect	[redacted]	[redacted]	[redacted]
Correction 2: Implementation of Lo *et al.* utilities	[redacted]	[redacted]	[redacted]
Correction 3: Age adjustment for caregivers	[redacted]	[redacted]	[redacted]
Correction 4: SMR based on wrong values from Chin et al	[redacted]	[redacted]	[redacted]
Correction 5: Using EAG AUC function and increasing SF upper limit to 1000	[redacted]	[redacted]	[redacted]
Correction 6: Age adjust patients	[redacted]	[redacted]	[redacted]
Correction 7: Rescue medication	[redacted]	[redacted]	[redacted]
EAG corrected company base case	[redacted]	[redacted]	[redacted]
EAG 1: Discontinuation rate based on exposure time in Marigold study	[redacted]	[redacted]	[redacted]

Scenario	Incremental cost	Incremental QALYs	ICER (change from company base case)
EAG 2: Use of the Marigold maintenance HL	[redacted]	[redacted]	[redacted]
EAG 3: Use of Auvin *et al.* (with absolute values and caregiver utilities) (Key issue 4)	[redacted]	[redacted]	[redacted]
EAG 4: Interpolation of the treatment effect (Key issues 2 & 3)	[redacted]	[redacted]	[redacted]
EAG 5: Including 10% wastage	[redacted]	[redacted]	[redacted]
EAG 6: Hospitalisation short stay based on Chin *et al.*	[redacted]	[redacted]	[redacted]
EAG 7: Severity modifier applied to patients only (Key issue 6)	[redacted]	[redacted]	[redacted]
EAG's preferred base case (Caregiver severity 1.7x)	[redacted]	[redacted]	[redacted]
EAG's preferred base case (Caregiver severity 1x)	[redacted]	[redacted]	£868,980 (+£846,780)

Abbreviations: EAG, External Assessment Group; ICER, incremental cost-effectiveness ratio; QALY, quality adjusted life year

2. INTRODUCTION AND BACKGROUND

2.1. Introduction

In this report, the External Assessment Group (EAG) provides a review of the evidence submitted by Marinus Pharmaceuticals for an appraisal of ganaxolone (GNX) for the treatment of seizures in people with Cyclin-depended Kinase-like 5 (CDKL5) Deficiency Disorder (CDD).

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

The company provided an overview of the burden of CDD in the target population in section B.1.2 and B.1.3 of the CS.

The CDKL5 gene, found on the X chromosome, encodes a protein responsible for normal brain function. [1] Estimated incidence is one in 40,000-60,000 live births, with a ratio of 1:4 males to females. [2] Though occurrence is more common in females, males commonly experience higher seizure frequency and increased brain atrophy. [3] A deficiency in the CDKL5 genes causes early onset seizures and developmental arrest. [4] Other symptoms include hypotonia, cortical visual impairments (CVI), sleep and gastrointestinal disturbance and autonomic dysfunction. Until recently (2005), CDD was considered to be a variant of Rett Syndrome, a neurological disorder resulting in similar symptoms. [5] However, those subsequently identified as having CDD were more severely affected and had a younger onset of seizures.

People with CDD experience a 90% onset of disease by the age of three months, and after a brief 'honeymoon' period where seizures temporarily remit, most people with CDD experience frequent seizures throughout their lives. Fehr et al (2016) reported that fewer than half of CDD patients experience a seizure free period of more than two months. [6] The most common seizure types experienced by people with CDD are epileptic spasms and tonic seizures, which are often clustered together. Many people with CDD are prescribed multiple anti-seizure medications (ASMs). However, polypharmacy has been identified as a risk of patients' wellbeing and is associated with an increased risk of adverse events.

People with CDD experience severe impairments to everyday functioning, and fewer than a quarter of people are able to walk independently or verbally communicate. [7] Clinical advice to the EAG was that it was difficult to determine if impairments experienced by people with CDD are caused by their development disorder, epilepsy, or other mechanisms of the condition.

However, seizures may cause harm to the brain and impair functioning ability and increase risk of sudden death from epilepsy (SUDEP).

Due to the severity of the condition, caregiver burden is very high. Mean mental health scores on the SF-12 were lower for CDD caregivers than the general population.Among CDD caregivers, those with children with gastrostomy feeding had better mental health scores but lower physical health scores. [8] Additionally, emotional wellbeing was significantly worse than for caregivers for children with Rett or Down's Syndrome. [9]

The EAG noted that the company provided an accurate summary of evidence on CDD and disease burden. The EAG considered the level of functional impairment to be a major driver of health-related quality of life (HRQoL). Clinical advice to the EAG highlighted that seizures vary in severity, meaning seizure frequency alone may not be a reliable marker or HRQoL. The company's description of CDD stages were reflective of the high level of uncertainty of CDD. Notably, due to the recent disease classification, there is no long-term natural history data showing the course of the disease and typical life expectancy of those with CDD. The EAG considered the company's description of the comorbidities well researched and to incorporate relevant evidence. There was less evidence presented on the impact of CDD on the mental health of people and their caregivers, which are likely to be significant.

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

The company provided an overview of the current treatment options for people with CDD and the proposed treatment pathway with ganaxolone (GNX) in Section B.1.3.3 of the CS (Document B).

While NICE guidelines exist for epilepsies [NG217], including genetic epilepsies in children, there are no existing guidelines specifically for CDD. Currently, there is no curative treatment for CDD, relying on broad ASMs. CDD is classed as a drug-resistant epilepsy, which is defined by not achieving seizure control after two or more anti-seizure medications. [10]

Prior to a CDD diagnosis, children exhibiting seizures are treated with steroid medication. Diagnosis may take some months, after which treatment would switch to more specific ASMs. The median number of ASMs prescribed was six (0-33) across a person's lifetime, [11] more frequently levetiracetam, topiramate, clobazam and phenobarbital. NICE currently recommends the use of sodium valproate as a first line therapy for tonic and tonic-clonic seizures in those unlikely to have children in the future, followed by lamotrigine or levetiracetam, but the

prescription of lamotrigine in children under 13 was off-label. Due to the relatively new distinction of CDD from Rett syndrome, there is a lack of evidence on the impact and efficacy of ASMs.

A recent longitudinal study showed that around a quarter (82/312, 26%) of people with CDD reported cannabinoid use to aid seizure control, with around two-thirds reporting improvements in seizure control. [12] Caregivers also reported benefits of cannabinoid for cognition, sleep and mood, with most patients reporting no adverse effects, although the evidence from cannabinoid use for epileptic syndromes is uncertain. Currently, the NHS prescribes epidiolex, a highly purified CBD, for Lennox-Gastaut syndrome (LGS) (TA615) and Dravet Syndrome (DS) (TA614), both rare and severe forms of epilepsy. As some people with CDD are also diagnosed with LGS, this means that they would be eligible to receive epidiolex. Some people with CDD follow a ketogenic diet to aid seizure control, though evidence for the efficacy of this is also uncertain. Vagus nerve simulation (VNS) delivers electrical pulses to the vagus nerve and is an accepted form of treatment for refractory epilepsy. In a CDD specific study, two thirds of patients experienced an improvement in seizure activity. [13] Alternatively, surgical treatments for seizure control may also be used, with a significant, but short-lasting impact. Other symptoms of CDD are managed using treatments such as serotonin for sleep disturbances, or for patients with feeding difficulties, a gastrostomy tube may be used.

The EAG generally agreed with the company's description of current service provision for CDD. However, the EAG were unclear about whether GNX would be used as a first line treatment, or whether clinicians may only prescribe GNX if people had not responded to other treatments. The EAG were also unclear about the anticipated duration of treatment with GNX, for example whether a minimum treatment period is needed to determine if there will be a response, and whether those showing a response would be expected to receive the treatment for life. Clinical experts advised that any clinical response should be evident by 6 months, at which point, non-responders should be withdrawn. The EAG were concerned that this would not be the case if other treatment options were also not considered effective, increasing the risks associated with polypharmacy, but considered that due to safety and impact of HRQoL, withdrawal would typically occur for most non-responders.

2.4. Critique of company's definition of decision problem

The company statement regarding the decision problem was presented in Section 1 of the CS (Document B). The company position and the EAG response is provided in Table 4 below.

Table 4: Summary of decision problem

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	EAG comment
Population	People 2 years of age or older with seizures caused by CDD	As per the scope	NA	The EAG agreed that the evidence submitted by the company was consistent with the NICE decision problem.
Intervention	Ganaxolone (ZTALMY®)	As per the scope	NA	The EAG agreed that the evidence submitted by the company was consistent with the NICE decision problem.
Comparator(s)	Established clinical management (ECM) without ganaxolone	Established clinical management, although restrictions were placed on use of cannabidiol.	NA	The EAG considered the decision problem submitted by the company was consistent with the NICE scope. ECM was considered to consist of ASMs and steroids as well as non-pharmacological treatments such as a ketogenic and vagus nerve stimulation. The EAG agreed with the company's descriptions of established clinical management, but highlighted the exclusion of cannabidiol, with the exception of epidiolex during the trial, which may not reflect real world use. However, did not consider this would have a major impact on trial findings.
Outcomes	The outcome measures to be considered include: • Seizure frequency (overall and by seizure type)	The clinical evidence was consistent with the NICE scope, though the company's economic model did not consider seizure severity or differences in adverse events	The company stated that there are no reliable methods for estimating the severity of seizures, and therefore this was not considered in the model.	The EAG agreed that the evidence submitted by the company was consistent with the NICE decision problem. However, the EAG noted that the use of seizure frequency as a primary outcome measure may not be entirely representative of disease severity, as advice from clinical experts suggested that impacts from seizures

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	EAG comment
	• Proportion of people seizure-free (overall and by seizure type) • Seizure severity • Adverse effects of treatment • Health-related quality of life	between GNX and ECM.		are heterogeneous. The EAG agreed with the company that there are no reliable measures of the severity of seizures, though noted that this limits consideration of the potential effect of GNX. The company reported comparable rates of treatment-emergent adverse events between GNX and ECM in Marigold, and therefore assumed that the impact of AEs was equivalent in the model. However, the EAG noted that rates of drug-related AEs were higher in the GNX arm. There was no clear evidence that treatment with GNX increases the risk if AEs with significant resource implications, and so the EAG did not consider that differences in this assumption would have a major effect on the ICER.
Economic analysis	The reference case stipulates that the cost effectiveness of treatments should be expressed in terms of incremental cost per quality-adjusted life year. The reference case stipulates that the time horizon for estimating clinical and cost effectiveness should be sufficiently long to reflect any differences in costs or outcomes between the	No costs for genetic testing were included	The company analysis was consistent with the NICE reference case. The company stated that all people with CDD would receive a genetic test prior to starting ganaxolone, and therefore the availability of ganaxolone would not lead to a change in testing costs. However, the company	The EAG agreed with the company's rationale with respect to the testing costs, as CDD diagnosis was only able to be confirmed after genetic testing. The EAG understood that genetic testing for CDD is likely to have already occurred before ganaxolone is a administered. The EAG noted that the time horizon in the model was updated to 100 years at clarification from the original 75 years. This implied that people with CDD were able to exceed a life expectancy of 100 years, considering the mean starting age in the model is [redacted]. Despite

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	EAG comment
	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 of any managed access arrangement for the intervention will be taken into account. The economic modelling should include the costs associated with diagnostic testing for CDKL5 gene mutations in people with CDD who would not otherwise have been tested. A sensitivity analysis should be provided without the cost of the diagnostic test.		also acknowledged that there are adults with CDD who have not received a genetic test and would not be likely to receive one in current practice.	the lack of long-term survival data in CDD, clinical advice to the EAG was that this was highly unlikely. Additionally, when considering a life-time horizon, the assumptions around the baseline age of caregivers became highly uncertain.
Subgroups	NA	NA	NA	NA
Special considerations including issues related to equity or equality	NA	NA	NA	NA

Abbreviations EAG, Evidence Assessment Group; NA, not applicable; NICE, National Institute for Health and Care Excellence

3. CLINICAL EFFECTIVENESS

3.1. Critique of the methods of review(s)

The company undertook a systematic literature review (SLR) to identify evidence for the clinical effectiveness of GNX. A single search was conducted to identify relevant evidence, along with all evidence required to inform the company's economic model (see Section 4.1). The EAG assessment of the company's SLR for clinical effectiveness is presented in Table 5.

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

Systematic review step	Section of CS in which methods are reported	EAG assessment of robustness of methods
Searches	B.2.1. Appendix D	Acceptable. The company searched a combination of bibliographic databases, conference websites, clinical trials registries, websites of relevant organisations, google scholar, and review of reference lists of relevant studies. The strategy used appeared appropriate, although the terms used to conduct supplementary searches were not reported in the CS. At clarification (question A1), the company submitted the terms used to search one such resource, which were appropriate and provided reassurance that other sources were appropriately searched.
Inclusion criteria	B.2.1	Excellent. A comprehensive SLR was conducted to identify evidence for the CS.
Screening	Appendix D	Excellent. Double screening with involvement of a third reviewer was used to select relevant publications at all screening levels.
Data extraction	Appendix D	Acceptable. A single reviewer conducted data extraction with review by a senior reviewer and involvement of a third reviewer where required.
Tool for quality assessment of included study or studies	Appendix D	Poor. The NICE checklist for comparative trials was used for the Marigold double-blind phase, which was acceptable. However, only the minimum criteria were evaluated, and no account was made of variation in bias across outcomes (for example, where outcomes showed differences at baseline or were susceptible to measurement issues). The same checklist was used for the Marigold OLE and Phase IIa trial, which was not appropriate. This approach does not consider the risks relevant to trials without a control group and where group allocation is not random.

Systematic review step	Section of CS in which methods are reported	EAG assessment of robustness of methods
Evidence synthesis	NA	No evidence synthesis was conducted by the company, which was considered appropriate.

Abbreviations: CS, Company submission; EAG, External Assessment Group; OLE, open-label extension; SLR, systematic literature review.

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

3.2.1. Studies included in the clinical effectiveness review

The CS described two studies (shown in Table 6), including one double-blind randomised-controlled trial (RCT; Marigold) with an open label single arm extension (Marigold OLE), and a small phase IIa single-arm study with an extension for those who showed a response to treatment (Study 1042-0900). The latter study was small (n=7 and n=4 in the extension period) and was used by the company as supporting evidence for the RCT only.

The EAG identified a further double-blind, placebo-controlled trial to evaluate GNX for treating seizures in infants with CDD (aged 6-months to 2 years), though this trial had yet to begin recruiting (final data cut estimated December 2024; NCT05249556) and was not considered further in the appraisal.

Table 6: Clinical evidence included in the CS

Study name and acronym	Study design	Population	Intervention	Comparator	Study type
Marigold (1042-CDD-3001) [14,15] NCT03572933	Double-blind RCT with 17-weeks follow-up	People aged 2 – 19 years with CDD and ≥16 major motor seizures per 28 days (N=101)	GNX + ECM. Titration period = 4 weeks, full dose = 13 weeks	ECM	Clinical efficacy and safety
Marigold OLE NCT03572933	Single-arm extension to Marigold with further follow-up available in the CS (February 2021) and in the company's clarification response (June 2021). Study ongoing and expected to complete data collection in December 2022, with data available in Q1/2023	All those completing Marigold and still meeting eligibility criteria	GNX + ECM. Titration period for people receiving placebo during Marigold = 4 weeks	NA	Long-term clinical effectiveness and safety
Phase IIa study (1042-0900) NCT02358538	Open-label, single arm proof-of-concept study with 26-weeks follow-up	People with rare genetic epilepsies, including PCDH19 (n=11), LGS (n=7), continuous spikes in slow wave (n=2), and CDD (n=7)	GNX + ECM.	NA	Clinical effectiveness and safety
Phase IIa study extension (1042-0900) NCT02358538	Extension period with 52-weeks follow-up	Participants in the initial Phase IIa follow-up who attended all study visits and showed a ≥35% improvement in mean seizure	GNX + ECM.	NA	Clinical effectiveness and safety

Study name and acronym	Study design	Population	Intervention	Comparator	Study type
		frequency. Participants with CDD n=4.			

Abbreviations: CDD, CDKL5 deficiency disorder; ECM, established clinical management; GNX, ganaxolone; LGS, lennox-gastaut syndrome; NA, not applicable; OLE, open-label extension; PCDH19, Protocadherin 19; RCT, randomised controlled trial

3.2.2. Description and critique of the design of the studies

3.2.2.1. Design of the studies

The pivotal study for GNX in this indication was the Marigold RCT and its OLE. The availability of a high-quality RCT in such a rare disease area was notable, and the EAG considered that the follow-up (17-weeks plus data of at least 1 year in the latest data cut of the OLE) would be sufficient to determine whether treatment with GNX was effective for reducing seizures as compared to existing treatments, which typically lose their effect after 3-months. The latest available data cut for the Marigold OLE provided in the CS was 24th February 2021, though at clarification the company provided evidence for a subsequent data cut-off of [redacted]. Only a subset of outcomes were presented at this later data cut-off and, given the timeline, the EAG was unclear why these were not provided with the original CS. At clarification (QA10), the company stated that data for a cut-off in [redacted] would be available by [redacted].

An overview of the Marigold trial design was shown in Figure 6 in the CS (p.38). An 8-week period was used to collect data on seizure frequency used to determine participant eligibility for the trial (see Section 3.2.2.2), after which the double-blind phase began with a 6-week 'baseline' period for the collection of seizure frequency data to be used as baseline measures. Following the baseline phase, participants allocated to GNX entered a 4-week period in which GNX was titrated to reach the target dose that they received for a further 13-weeks. Primary analyses for Marigold were based on the full 17-week period incorporating both the titration and target dose trial periods, though sensitivity analyses were also conducted restricted to the period when participants were receiving the full target dose. In the OLE, participants allocated to placebo in Marigold were unblinded and switched to GNX. As with the original Marigold trial, GNX was titrated to the full target dose over a 4-week period.

The EAG considered the initial phase of the Phase IIa as supporting evidence for Marigold, though due to the small sample size, it agreed with the company that the data was limited for the purposes of decision-making. The EAG considered that the extension period of the Phase IIa study was not suitable for evaluating the clinical effectiveness and safety of GNX, due to the risk that the eligibility criteria excluded those with poor efficacy or safety data, and that this would have a notable effect amongst a small sample.

3.2.2.2. Population

Study eligibility criteria

Eligibility criteria for the Marigold trial were provided in Table 9 of the CS (page 42).

Inclusion criteria for the trial included those aged 2 – 21 years. The lower age limit was considered appropriate and was in line with the NICE scope, though the EAG considered the upper age limit to be restrictive given that the NICE scope and company decision problem included people with CDD with no upper age limit. There is a great deal of uncertainty about the typical survival of people with CDD, owing to a lack of long-term data, though in the company's survival estimations, 65% of patients may survive to reach ~53 years old. The EAG considered it plausible that the effects of CDD on a person's life may worsen over time, as health may be impacted by the cumulative effect of neurodevelopmental impairment. Overall, despite the uncertainty surrounding survival, the EAG considered that the lack of data in those aged over 21 years presents uncertainty for the long-term outcomes of treatment.

The trial was restricted to people with CDD for whom ≥2 previous anti-seizure medications (ASMs) had failed to control their seizures, and thus GNX was evaluated as a third-plus line therapy. The anticipated marketing authorisation for GNX [redacted]. Clinical advice to the EAG was that those in the trial may have received ASMs both prior to and following a diagnosis of CDD. In practice, broad-spectrum ASMs are commonly used to treat seizures while awaiting a diagnosis of CDD, which may take several months. Following diagnosis, alternative ASMs would be used. If GNX became available in practice, the EAG was uncertain whether this would be used first line following a diagnosis of CDD, or whether people would only receive GNX following a failure to respond to other ASMs (as in the trial). The EAG understood that few people with CDD may achieve a satisfactory response to other ASMs, and therefore the trial population may nevertheless be comparable with a first line population in practice.

The inclusion criteria permitted participants to be receiving a stable regimen of up to four ASMs at baseline, not including non-pharmacological treatments. Polypharmacy for seizures in practice was common, and evidence suggests that people will CDD receive a lifetime average of six ASMs (range 0-18).[11] Clinical advice to the EAG was that people with CDD often receive between 2 – 4 ASMs concurrently, which was consistent with the trial participants. With regard

to seizures, participants were required to be experiencing ≥ 16 major motor seizures per 28-day period, as assessed over an 8-week period prior to the trial. The EAG were uncertain how representative this was of seizure frequency in the target population, though noted evidence that some people may experience fewer seizures that this.[6] In Key Issue 1, the EAG considered the possibility that the trials included people who were experiencing a temporary exacerbation in seizures, necessitating consideration of ASM. This issue is discussed further in Section 3.2.2.5.

The inclusion criteria for the phase IIa trial were not reported in the CS but were available to the EAG from the CSR[14] provided by the company. Compared to Marigold, the criteria required a [redacted] seizure frequency [redacted] trial CSR, p. 18). At clarification, the EAG requested a rationale for the change, but the company stated that they did not have access to this information. The company suggested that more restrictive criteria may have been used for a smaller trial that was to be used as exploratory and a proof-of-concept evaluation, which the EAG agreed was plausible despite the broader uncertainty.

Baseline characteristics

Participants in Marigold were most frequently from the United States (US; 41.6%), followed by Italy (14.9%) and Russia (13.9). Seven participants (6.9%) were from the UK. Overall, trial arms appeared comparable. The EAG identified baseline quality of life scores for Marigold from the trial CSR appendices (ref) provided by the company at clarification. These were comparable between arms and were also comparable to total scores reported in a published study using the same scale with a sample of people with CDD.[16] The EAG noted there to be a difference in the median percentage of seizure-free days (SFD) between trial arms, though no further differences in seizure-free outcomes were noted and as quality of life was also comparable, the EAG did not consider this to be a major concern. However, this was noted when considering findings for this outcome.

Baseline characteristics were considered to be representative of the likely population of people with CDD in the UK who would be eligible for GNX, though as discussed above, the EAG noted that no participants were treatment naïve.

Table 7: Baseline characteristics of participants in the included trials

	Marigold		Phase II
	Ganaxolone (N=50)	**Placebo (N=51)**	**Ganaxolone (N=7)**
Demographics			
Age, mean (SD)	6.8 (4.7)	7.7 (4.4)	[redacted]
Female sex, n (%)	39 (78%)	41 (80.4%)	6 (85.7%)
Weight, mean (SD)	[redacted]	[redacted]	-
Age at diagnosis	-	-	-
CDD recorded in participants' medical history at baseline, n (%)	[redacted]	[redacted]	-
Confirmed pathogenic CDKL5 variants identified at baseline, n (%)	[redacted]	[redacted]	-
Age at first seizure, median (range)	[redacted]	[redacted]	-
Measurements during the baseline period			
Total number of seizures per 28 days, median (range)	-	-	[redacted]
Number of bilateral tonic seizures per 28 days, median (range)	[redacted]	[redacted]	-
Number of people who exhibited bilateral tonic seizures, n (%)	[redacted]	[redacted]	-
Number of major motor seizures per 28 days, median (range)	54.0 [redacted]	49.2 [redacted]	-
Number of seizure-free days per 28 days, median (range)	[redacted]	[redacted]	[redacted]
Treatment history			
Use of ASMs at start of trial, n (%)	49 (98.0%)	48 (94.1%)	[redacted]

	Marigold		**Phase II**
	Ganaxolone (N=50)	**Placebo (N=51)**	**Ganaxolone (N=7)**
Use of non-pharmacological treatment for seizures at start of trial, n (%)	29 (58.0%)	26 (51.0%)	-
Number of previous ASMs, median (range)	7 (2 – 16)	7 (1 – 14)	-
Number of concurrent ASMs, mean (SD)	2.6 (1.39)	2.2 (1.14)	-

3.2.2.3. Intervention

The intervention for the included trials was GNX in combination with established clinical management (ECM), including adjunctive treatment with up to four ASMs.

In the Marigold trial, participants were treated with GNX as an oral suspension in accordance with the licensed dose in the US: 50mg/mL taken three times daily. [17] A weight-based method was used to titrate the dose in children weighing under 28kg, and a standard titration schedule was used for other participants. No information was provided in the CS about the tapering of treatment in the event of discontinuation, which was notable given that a steady reduction in ASM is needed to reduce the risk of a rebound in seizures following withdrawal. [17]

The majority of participants achieved the maximum dosage of GNX, though dose reductions due to adverse events (AEs) were needed in 22% (11/50) of those in the GNX arm and 23.5% (12/51) in the placebo arm. Participants in the GNX arm in the double-blind phase continued on their final dose throughout the OLE.

Mean (SD) treatment exposure length was 113.0 (23.32) days in the double-blind trial, and [redacted] days in the OLE (data cut-off February 2021). Adherence to the medication was moderately high: [redacted]% of participants in the GNX arm received treatment on 90% of the days in the double-blind phase.

The company prohibited the use of cannabidiol as an adjunctive treatment in the double-blind phase of Marigold unless participants had a stable, pre-existing prescription of epidiolex. Conversely, use of cannabidiol was permitted as an adjunct to GNX during the OLE. The EAG understood that the use of cannabidiol to control seizures was common for people with CDD, and that the exclusion of this as an option during the double-blind phase of the trial was excluding an established method of managing seizures. The EAG also considered that the variation in approach between the double-blind and OLE phases of Marigold was not substantiated. However, given the unregulated nature of cannabidiol that was not provided on prescription, the EAG did not consider it unreasonable to exclude this from the double-blind phase of the trial.

The CS also described that a small number of participants (10.9%) were following a ketogenic diet during the double-blind phase of Marigold to manage seizures. More than half of participants (58.0%) were also receiving other non-pharmacological therapies, such as physiotherapy, speech rehabilitation and occupational therapy.

3.2.2.4. Comparator

Only the Marigold double-blind phase involved a comparator to GNX, which was a placebo administered in addition to ECM including use of up to four concurrent ASMs. The placebo method used was also an oral suspension administered to the same schedule. A similar number of participants were following a ketogenic diet during the double-blind phase (13.7%) and were receiving non-pharmacological therapies (51.0%).

3.2.2.5. Outcomes

The outcomes reported in the included trials of GNX are summarised in Table 8, and the EAG provides an appraisal of the specific outcomes measured in the sections below. As discussed in Section 2.4, the EAG considered that the outcomes reported were consistent with the scope for this appraisal.

Outcomes measured included consideration of the impact of GNX on seizure outcomes and safety, as well as broader functional and HRQoL outcomes. With some exceptions, overall the EAG considered that detail about some clinical outcomes were limited both within the CS and the main report documents for the trial CSRs, so the EAG requested appendices to the trial CSRs during clarification (QC4), as these contained full data tables for measured outcomes. The company provided these for the Marigold trial but not the Phase IIa trial, and no trial CSR was provided for the Marigold OLE, which at clarification the company confirmed was because no such document exists. The CSR appendices for Marigold were provided later than the clarification response deadline, meaning that the EAG were unable to explore these in full detail, meaning that further relevant outcome data may have been measured.

Outcome reporting was most comprehensive for the Marigold double-blind phase. Very few outcomes were reported in the CS for the Phase IIa study, which the company explained was due to the small sample size of this trial and its lesser importance for informing the CS and economic model. Some outcomes were also not reported for the Marigold OLE. In the CS, data for the Marigold OLE was limited to the February 2021 data cut, though at clarification (QA12), the company provided additional data for a subset of clinical outcomes from the [redacted] data cut.

It was unclear whether the Marigold trial included sufficient follow-up to evaluate the full way in which treatment would be used in practice. The company did not specify the likely duration of treatment with GNX in clinical practice and no stopping rule was considered within the

company's economic model (Section 4.2.6.1). Clinical advice to the EAG was that people with CDD may be treated for a minimum of 6-months, at which point those not exhibiting a response would discontinue treatment. GNX may then be used up to a maximum of 2-years, at which point people may be discontinued to consider whether there was ongoing benefit. While the Marigold OLE provided some longer-term data that may be used to inform the use of a 2-year treatment period, the EAG identified concerns about the quality of these data for decision-making (see Key Issue 1, and Sections 3.2.2.5 and 3.2.2.6).

Table 8: Outcomes reported in the included trials

Outcome	Marigold	Marigold OLE	Phase IIa study	Phase IIa extension
Seizure outcomes				
Number of major motor seizures per 28 days	✓	✓	✓ (CSR)	✓ (CSR)
Number of other/all seizure types	✓	×	✓	✓
% of participants who experienced a response in major motor seizures	✓	✓	×	×
% of participants who experienced a response in all seizure types	✓	×	✓ (CSR)	✓ (CSR)
Number of seizure-free days	✓	✓	✓	✓
Duration of time seizure-free	✓ (CSR)	×	✓ (CSR)	✓ (CSR)
Proportion of people seizure-free	×	×	×	×
CGI-I parent report	✓	✓	✓	✓
CGI-I clinician report	✓	✓	✓	✓
CGI of change in seizure intensity, duration and severity (CGI-CSID)	✓	✓	×	×
Use of rescue medication	✓(CSR)	×	×	×
HRQoL and functioning				
QI-disability scale	✓	×	×	×
CGI of change in attention	✓	✓	×	×
Parenting stress	✓(CSR)	×	×	×
Children's sleep habit questionnaire (CSHQ)	✓(CSR)	×	×	×
Anxiety, depression and mood scales (ADAMS)	✓(CSR)	×	×	×
Safety				
Adverse events	✓	✓	✓ (CSR)	×

Seizure outcomes

Following infancy, people with CDD experience seizures that are both generalised (affecting both sides of the brain) and focal (affecting one side of the brain). People often experience a combination of difference seizure types, including generalised tonic, generalised clonic, absence, and drop seizures, and focal seizures that can cause a broad range of symptoms (depending on where in the brain the seizure occurs). It is typically challenging to measure the frequency and duration of seizures in everyday life as reliable, physiological measures of seizure activity can be invasive and/or are restricted to hospital settings. This would not be appropriate for trials of seizure treatments in CDD, where people typically experience seizures every day.

The EAG noted a number of concerns with the measurement of seizures within the trials of GNX. These issues were common across seizure research and did not represent a failing in the way that the trials were conducted or analysed. However, they nevertheless affected the reliability of the trial findings and their interpretation. A summary of the issues is shown in Table 9, with further discussion below.

Table 9: Measurement issues associated with seizure outcomes in the clinical trials

Measurement issue	EAG comment
Physiological measures that provide a more accurate method for assessing seizures would not have been appropriate for use in trials of GNX, and therefore seizure frequency was assessed using carer and clinician reported outcomes. The frequency of seizure outcomes and participants' use of rescue medication were assessed using daily electronic diary (e-diary) entries completed by caregivers	The EAG considered that these methods were the best available to the company for the trials, however there were several limitations to this approach: • These measures may be less reliable for certain types of seizures, e.g. absence, drop, and focal seizures may be less visible and/or noticeable to caregivers during their day-to-day activities. Measures of generalised clonic seizures may therefore be most reliably assessed using this method. • Self-report measures of count data can be burdensome for caregivers alongside their daily activities, which can sometimes lead to unreliable measurements if caregivers attempt to complete diary entries retrospectively. It is plausible that measurements become less accurate over time if caregivers struggle to manage the burden over the long-term. • Caregivers and clinicians may not be able to determine some changes in the effects of seizures, for example small changes in intensity or the presence of certain after-effects, particularly in context of the broader health issues experienced by people with CDD. • Subjective outcomes are vulnerable to bias within open-label designs, meaning that seizure outcomes during the Marigold OLE and the Phase IIa trial were more uncertain.
There were no definitive measures of seizure severity or duration. Carer and clinician perceptions of seizure intensity and duration were measured using the CGI-CSID	Clinical advice to the EAG was that severity includes consideration of the impact and duration of the seizure, as well as any after-effects (for example, fatigue over several days). The EAG considered that carer and clinician reports would not be able to accurately represent the full impacts of seizures on people with CDD, despite their knowledge and experience of participants. In particular, the EAG considered that small changes in severity may be challenging to detect against a complex condition with many impacts on people's health and function.
Some people may experience a sudden increase in seizures that occur very closely together, which is defined as a cluster. This is challenging to measure	Due to challenges in measuring cluster seizures, the company defined each cluster as one seizure. This was a simplistic approach that inevitably under-estimated seizure count.

Measurement issue	EAG comment
as there may be little break between individual seizures	
Some people with CDD may experience a steady rate of seizures, while in others, seizure frequency can vary naturally over time	There was limited evidence about the rate of change in seizure frequency over time and so the length of trial follow-up that would be needed to account for natural variation over time. The EAG was also aware of evidence that it is rare but possible for people with CDD to experience prolonged periods of time without seizures and that these may not be captured within the timeline of clinical trials.
People with CDD may receive new treatment for seizures following an increase in severity, which may also be true for the decision to enter a clinical trial. This means that seizure frequency in some people would be expected to regress naturally towards the mean over time.	There was limited information about the methods of recruitment used for the clinical trials, and whether longer treatment history was collected in addition to measuring seizure frequency during the 4-week baseline period. It is therefore unclear whether a proportion of the trial sample entered the trial during an exacerbation in seizures. The EAG were also unclear about the typical length of seizure exacerbations, and whether a regression to the mean would be discernible within the 17-week double-blind phase of Marigold. This issue is included in Key Issue 1.
People may experience exacerbations in seizure frequency following the withdrawal of a treatment, particularly if medications are withdrawn too quickly. [17]	There was limited information in the CS about the way GNX and other treatments were discontinued, and no outcome data were included for those who withdrew from treatment. If withdrawal from GNX was associated with an increase in seizure severity, this should be considered in clinical and cost effectiveness analyses.

The company took several steps to account for the subjective measurement of seizure frequency in the trials; for example, the first 17-week phase of Marigold was double-blind and in all studies they conducted quality checks on diary entries and removed data points that appeared erroneous. The company also conducted separate analyses according to different seizure types, including analyses limited to seizures they considered 'countable' (the latter were not reported in the CS, but were identified by the EAG from the trial CSR). While the EAG considered these steps to be appropriate, the EAG noted that these would not account for limitations in the measures. Firstly, blinding halted at the end of the double-blind phase of Marigold, at which point outcomes in the OLE would be subject to an increased risk of overestimating treatment effects (see Section 3.2.2.6). Secondly, quality checks on diary entry data are necessarily conservative (to avoid the deletion of valid data), and therefore do not resolve issues with the reliability of the data. Thirdly, while the exclusion of uncountable seizures increased the reliability of measurement, outcomes did not account for the full spectrum of seizures experienced by people with CDD. Finally, there were no steps open to the company to improve the accuracy and sensitivity of carer and clinician reported measures of seizure severity and duration, and changes in these outcomes may be undetected in the clinical trials.

The EAG identified the risk of a regression to the mean effect in the clinical trials as a key issue in this appraisal (Key Issue 1). Clinical advice to the EAG was that people with CDD may receive a new treatment for seizures following an increase in seizure severity, with one advisor describing seizures as being at a 'crest of a wave' at the start of clinical trials for ASMs in general. Inclusion criteria for the Marigold trial specified a requirement for >16 major motor seizures per 28 days in a historical period, though a longer treatment history for participants was not reported (and plausibly not measured). The EAG considered it to be plausible that some participants in the sample may experience improvements in SF due a regression towards the mean effect. During the double-blind phase of Marigold, any natural decline could be accounted for through relative comparisons between the two treatment arms (though absolute outcomes, including absolute thresholds for response, would incorporate any natural decline that occurred). However, once entering the OLE, there was no comparator arm, and therefore all outcomes may be affected by any regression to the mean effect. For this reason, the EAG was concerned about the validity of seizure frequency outcomes in the OLE and considered that this weakened the company's assertion of a sustained treatment effect for GNX. Finally, with regards the measurement of seizures, trials did not evaluate whether those withdrawn from

GNX experienced an exacerbation in seizure frequency, which is a common effect of withdrawal. The CS did not clearly describe the strategy used to withdraw GNX from participants who discontinued the trial (see Section 3.2.2.3), and therefore the EAG was unable to discuss with clinical advisors whether the approach was able to reduce the risk of exacerbation. The EAG considered it plausible that some participants may have experienced an exacerbation in SF following withdrawal, however without clinical data the EAG was unable to consider how this would affect the clinical and cost effectiveness of GNX.

The EAG identified two issues regarding the analysis of seizure outcomes in the Marigold OLE that affected the interpretation of the results. Firstly, the EAG were concerned with outcome data based on a pooled population of the two group (i.e. those on GNX throughout the double-blind and OLE phases [GNX/GNX] and those switched from placebo to GNX during the OLE [PBO/GNX], company to clarification question A12). The EAG viewed these data to be more uncertain that data presented separately for each group, given that variations in outcomes might be expected depending on whether GNX was received during the double-blind phase or the OLE. For example, changes in blinding, longer experience with outcome measures, different rules about permitted background care, and different rules concerning discontinuation from treatment may all influence treatment outcomes.

Secondly, a number of participants discontinued from Marigold either prior to or during the OLE. Of 101 patients randomised, 88 proceeded to the OLE, and 31 had discontinued at the data cut reported in the CS (doc B, Table 11). Of these 31, 12 (38.7%) withdrawals were reported as being due to ‘lack of efficacy', and the EAG considered it plausible that more ambiguous reasons for discontinuation (e.g. clinician judgement) may also have been informed by efficacy outcomes. The EAG therefore considered it plausible that participants who discontinued from the trial were experiencing higher SF, which lends further uncertainty to claims of a sustained treatment effect for GNX. The EAG considered it a major concern that the company did not conduct any analyses exploring the impact of missingness from OLE data. The EAG identified as a key issue for this appraisal (Key Issue 1).The EAG reviewed the analysis of seizure frequency conducted by the company, which calculated the percentage change in 28-day seizure frequency (PCSF) for an individual as (described in clarification question B2):

$$(f(t_1)_i - f(t_2)_i)/f(t_1)_i \times 100$$

where $f(t_1)_i$ was the 28-day seizure frequency for individual 'i' at baseline, and $f(t_2)_i$ was the same at the end of the 17 week double-blind period. The EAG noted that this approach can be seen as adjusting for baseline SF, which the EAG agreed was logical.

The company utilised the Hodges-Lehmann estimator of location shift in PCSF between the trial arms. This was a nonparametric estimator of the median difference robust to outliers, which the EAG believed was a judicious method considering the very wide variations in measured SF (see responses to clarification queries A8 and B8). However, as described in Section 4, issues arise when applying this estimate to model cost-effectiveness.

HRQoL and functioning

No HRQoL or functional outcomes were reported for the OLE of Marigold or the Phase IIa trial.

Participant quality of life was measured in Marigold using the QI-disability scale, which is a parent-completed measure of quality of life in children and adolescents with intellectual disability. The scale authors describe it as appropriate for use in both children and adults with CDD[18] and a published study has used QI-disability in a CDD population (ref). [16] The scale includes 32 questions across 6 domains: social interaction, physical health, independence, positive emotions, leisure and outdoors, and negative emotions. To the knowledge of the EAG, there was no validated threshold for a clinically meaningful change in QI-disability for people with CDD.

The company assessed several other measures in Marigold to explore whether treatment with GNX affected other outcomes important to the lives of people with CDD and their caregivers, including attention, sleep habits, mood and anxiety, and parenting stress. The EAG did not identify any additional outcomes that would have been relevant for inclusion.

Safety

The company assessed both drug-related and treatment-emergent adverse events in clinical trials of GNX. In response to a query from the EAG at clarification (QB27), the company stated that they considered the assessment for identifying drug-related AEs to be unreliable, due to the subjective assessment needed to determine if AEs were caused by the drug. To some extent the EAG agreed that there may not always be definitive evidence that AEs have been caused by the drug under evaluation, but noted that these judgements are made by experienced clinicians, and that this method is frequently used across clinical trials. The EAG considered that

inspection of both drug-related and treatment-emergent rates of AEs may be informative for evaluating the safety of treatments.

Overall, the EAG considered that measurement of AEs in clinical trials of GNX may be challenging due to the heterogenous nature and severity of the effects of CDD on the health and functioning of people with CDD. Moreover, GNX was delivered as an adjunctive to ECM, which included a range of permitted medications for seizures and other health concerns. In these circumstances, relative comparisons of AEs are the most reliable method for determining treatment safety. However, the EAG considered that the small sample size of the trials would increase uncertainty about these outcomes, particularly for AEs with low event rates.

3.2.2.6. Critical appraisal of the design of the studies

Critical appraisal checklists for Marigold, its OLE, and the Phase IIa trial were reported in the CS appendices (appendix D). The minimum criteria were evaluated within the checklists, and the company used only the checklist for randomised trials for all three assessments, rather than using a checklist for non-randomised/uncontrolled trials.

The EAG considered that the company's assessment of Marigold was acceptable, though it did not account for potential variation in bias across outcomes. Notably, the EAG considered that a difference in baseline in SFD between treatment arms would at minimum increase the risk of bias for this outcome. The company also did not comment on the potential risks of bias due to issues with outcome measurement (discussed in Section 3.2.2.5).

A similar number of participants in both arms opted to continue from the double-blind phase of Marigold into the OLE, though ≥10% of participants discontinued. Reasons for discontinuation were reported by the company in the CS and included reasons related to trial outcomes (i.e. safety and efficacy of treatment). This issue was not thoroughly assessed in the company's appraisal. Discontinuation during the OLE was assessed by the company as being non-problematic, even though further discontinuations were due to treatment outcomes, and declining sample size over time would have affected the robustness of data at follow-up. The EAG agreed with the company assessment that the lack of blinding in the OLE was a potential source of bias. All outcomes for this appraisal were subjective outcome measures, and therefore susceptible to bias in open-label designs. The EAG considered that pooling of data in the OLE of the GNX/GNX and PBO/GNX arms was particularly problematic, due to changes in the trial protocols between phases (e.g. on blinding, background treatment, and

discontinuation). The company stated that all outcomes measured in the trials were reported, though the EAG were not presented with some outcomes for the OLE that were measured, including quality of life and functional outcomes. Finally, the same measurement issues related to assessing seizures as apply to the double-blind phase of Marigold also applied to the OLE.

The Phase IIa trial was a very small, uncontrolled, open-label trial, which the EAG considered to be at a high risk of bias.

Overall, the EAG considered the double-blind phase of the Marigold trial to be the best quality evidence available for GNX in this indication. Risk of bias was generally considered to be low but the EAG considered that issues relevant to measuring seizure outcomes should be considered when interpreting outcomes. The EAG further considered there to be a number of quality issues with the Marigold OLE that should be considered when interpreting the results.

3.2.3. Description and critique of the results of the studies

3.2.3.1. Clinical effectiveness results

Seizure outcomes

Double-blind phase (up to 17-weeks treatment)

During the double-blind phase of Marigold, there was a greater reduction in median major motor seizure frequency and all seizure frequency in the GNX arm compared to placebo (CS Doc B p. 56-57, 64). There were participants in both arms exhibiting reductions and increases in seizures over the course of the 17-weeks, though participants in the GNX arm were less likely to experience an increase and more likely to experience a decrease in seizures. Using the threshold of 50% reduction in seizures (a common threshold used to determine a meaningful change in seizures), 24.5% of people in the GNX arm experienced a reduction in major motor seizures compared to 9.8% in the placebo arm, and [redacted]% in the GNX arm experienced an increase in major motor seizures compared to [redacted]% in the placebo arm. Rates of response were generally similar for all seizure types (CS Doc B p.63), [redacted] (CSR appendices Table 14.2.5.6.1 and 14.2.5.6.2). The cumulative proportion of people in each arm showing reductions and increases in major motor seizure frequency is shown in Table 10; please note that these figures were estimated from graphs provided by the company (CS Doc B Fig 9, p.59, and clarification response QA5, Fig 1 p.4) and so may lack some accuracy. These data were not available for analyses of all seizure types.

Table 10: Response rate in Marigold DB phase

	Cumulative % change in major motor seizure frequency							
	-80%	-60%	-40%	-20%	+20%	+40%	+60%	+80%
Ganaxolone	7%	22%	32%	60%	[redacted]	[redacted]	[redacted]	[redacted]
Placebo	5%	6%	16%	33%	[redacted]	[redacted]	[redacted]	[redacted]

Source: figures estimated from graphs provided by the company: CS Doc B Fig 9, p.60, and clarification response QA5, Fig A.

Results using the CGI-I showed that caregivers and clinicians were more likely to say that participants in the GNX arm had improved, though differences were marginal and not statistically significant (CD Doc B p.60). However, there was a greater difference in carer reported CGI-CSID, where caregivers were statistically more likely to say that those in the GNX arm showed improvements in seizure intensity/duration/severity (CS Doc B p.61). From the data

presented by the company, it was not possible to determine if carer responses were comparable for both severity and duration.

[redacted] CSR p.57).

There was a small increase in the median percentage of SFD reported by participants in the GNX arm (CS Doc B p.62), though there was no clear difference between arms. [redacted] (trial CSR appendices, Table 14.2.5.3.1 and 14.2.5.3.2).

OLE

Data from the latest data cut of the Marigold OLE ([redacted]) were presented by the company at clarification (QA12).

Reductions in median major motor seizure frequency were reported based on a combined population of those who started and were switched to GNX in Fig A (clarification response p.13) and separately between groups in Fig B (clarification response p.14). The company suggested that the data showed reductions in major motor seizure frequency shown in the double-blind phase [redacted] Response rates using a 50% threshold were higher in the OLE than in the DB phase ([redacted]% vs. [redacted]%, [redacted]%, [redacted]%, and [redacted]% at 1-3, 4-6, 7-9, and 10-12 months, respectively). [redacted]median reductions continued to increase, though the number of participants available for follow-up reduced due to the staggered entry of participants into the OLE.

A higher rate of SFD was shown in the OLE in the PBO/GNX arm than in the GNX arm (no statistical test performed; clarification response p.12).

Carer perceptions of severity and duration of seizure as assessed by the CGI-CSID were comparable between arms, with the PBO/GNX arm showing a rate of improvement comparable with the GNX arm during the double-blind phase.

HRQoL and functioning

There was no statistically significant difference in quality of life between the two arms of Marigold at the end of the double-blind phase. While four subscales (positive emotions, social interaction, leisure and outdoors, and independence) showed a numerical benefit for GNX, these differences were not statistically significant, and variance was high. Furthermore, there were no clear benefits of GNX over placebo for parenting stress; anxiety, depression and mood; attention; or children's sleep habits. Quality of life and functional outcomes were not reported separately for responders to treatment, and therefore the EAG considered is plausible that some benefits may be shown for those participants who experience a reduction in seizures with treatment. However, the EAG considered that the potential for treating seizures to produce meaningful change in quality of life and function in the context of such a severe disease to be unproven. Clinical experts to the EAG disagreed about whether reducing seizures early in life would lead to later benefits for functioning, with both acknowledging that such an effect was not yet supported by evidence.

Subgroup analyses

In the CSR appendices, seizure outcomes were reported separately for a subgroup of participants based in the UK, Australia, France, Israel and Italy (n=35). For this group, data showed that trial arms differed in baseline major motor seizure frequency, with a higher rate of seizures in the GNX arm (median [IQR]: [redacted] vs. [redacted]). While there was a greater overall reduction in major motor seizure frequency in the GNX arm, this was not statistically significant and a similar number of people in each arm showed a response (GNX [redacted]% and PBO [redacted]%) and were considered by caregivers to have improved (GNX [redacted]% and PBO [redacted]%).

In the trial CSR appendices, the company reported data separately for different types of seizures, including tonic, tonic-clonic, myoclonic, drop, absence, and motor seizures without altered awareness. The EAG noted that GNX was more likely to show an effect for seizures with a major motor feature. As discussed in Section 3.2.2.5, the EAG considered it likely that this may be due to difficulties in detecting an effect in seizures without major motor symptoms. However, the EAG also noted that it was plausible that GNX may have a differential effect across different types of seizures.

Safety

The company provided data for AEs reported in Marigold and its OLE in the CS (Document B, section B.2.10): Table 23 [Marigold] and Table 26 [Marigold OLE]. AE event data for the Phase IIa trial was reported in the trial CSR[19] provided by the company.

During the Marigold trial, rates of overall treatment-emergent adverse events (TEAEs) were comparable between arms, but there was [redacted] rate of treatment-related adverse events in the GNX arm (70.0%) compared to placebo (43.1%). Inspection of the drug-related AEs reported in the trial CSR appeared mild in nature, with no clear pattern of effect aside from an increased risk of somnolence in those receiving GNX.

The vast majority of participants in both arms of Marigold experienced at least one TEAE, though in general these were mild or moderate in nature. There was a trend for those in the GNX arm to experience more moderate than mild TEAEs, and the reverse in the placebo arm. Severe TEAEs were experienced by 2.0% (n=1) and 5.9% (n=3) of participants in the GNX and placebo arms, respectively. Comparison of specific AE types showed that somnolence and pyrexia were more common in the GNX arm than in the placebo arm. All other event rates were low in incidence and a clear pattern was not discernible. There was no clear difference in TEAEs that would be expected to lead to significant healthcare resource use, such as hospitalisation. There was also no clear evidence that GNX was more likely to cause TEAEs leading to permanent or temporary discontinuation, or to a dose reduction.

Rates of TEAEs reported by those who switched to GNX in the Marigold OLE were comparable to those reported for the GNX arm of the double-blind phase. In the CS, the company claimed that a lower rate of TEAEs between the GNX/GNX arm compared to the PBO/GNX arm (reported to the February 2021 cut-off) was suggestive that adverse events occurred early in the treatment and/or reduce over time. However, the EAG did not think there was sufficient evidence to support this claim, considering that there was only a small change in the number of participants receiving GNX in both trial phases who experienced TEAEs (86% in the double-blind phase and 76.7% in the OLE). The company also did not report TEAE data at later timepoints of the OLE, which may have demonstrated whether such a reduction in AEs occurred over time. Moreover, there was a higher rate of discontinuation in the OLE compared to the double-blind phase, meaning that rates of AE may appear artificially low in comparison.

There was one TEAE resulting in death in the OLE trial in the GNX/GNX arm. Though the company reported that the event was unlikely due to the study drug, the EAG highlighted that there was no explanation behind mortality cause in the CS, and there is no detail into how the company determined if the mortality was treatment-related.

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

No indirect comparison was undertaken.

3.4. Additional work on clinical effectiveness undertaken by the EAG

All additional work has been reported throughout.

3.5. Conclusions of the clinical effectiveness section

- GNX was more likely than ECM to reduce seizure frequency for a minority of people with CDD as assessed during the 17-week DB phase of Marigold. The EAG was uncertain about absolute reductions in seizure frequency due to the risk of a regression to the mean effect during the trial (Key issue 1), and evidence was strongest for the impact of treatment on major motor seizures compared to other seizure types.
- All outcomes assessed in the OLE were at risk of this due to the lack of a comparator and the discontinuation of participants due to the treatment outcome (and the absence of a missing data analysis). Overall, the EAG therefore considered that the long-term data showed a promising prolongation of treatment effect for some participants, which may exceed the typical length of time that ASMs show effect for people with CDD. However, the magnitude of the effect and the number of people who may benefit were both considered uncertain, due to limitations in the OLE data.
- Caregivers reported that GNX may have a beneficial effect for seizure duration and/or severity (reported as a combined outcome), however there was no effect of GNX for HRQoL, functioning, or caregiver wellbeing as compared to ECM.
- The EAG considered that it was unclear reductions in seizure frequency shown in the trials would be meaningful to people with CDD and their caregiver and, if so, what impacts these would likely have. All participants in the trials continued to experience regular seizures, and the EAG therefore considered the potential benefit of GNX to be a reduction in the

frequency of these for some people. Clinical experts advised that there was no high-quality evidence to suggest that reducing seizures would have long-term benefits for functioning and wellbeing. One expert considered this to be unproven yet plausible, while another considered that the severe nature of the condition and its impacts on brain development may mean that reducing seizures may have little overall impact. The EAG considered that reducing seizure frequency may have benefits for carer burden, though these may be difficult to measure against the broader carer burden for the condition.

- Overall, the evidence suggested that GNX was a relatively safe treatment option for treating seizures in people with CDD and may therefore be considered as an option alongside existing ASMs and therapies. However, the EAG noted that many people may still not experience a response to treatment, and in the absence of evidence for population effect modifiers, treatment would likely follow a ‘trial and error’ method.

4. COST-EFFECTIVENESS

During the appraisal, the company submitted three versions of their economic model to evaluate the cost-effectiveness analysis: one in the original submission (Model 1; 27/10/2022) and two subsequent versions at clarification (Model 2; 30/11/2022) and following clarification (Model 3; 22/12/2022). Each model version included a distinct company base case. Table 11 provides a top-line summary of the changes to the company's base case over the three versions.

Table 11: Company revisions to their cost-effectiveness model

Model identifier used in subsequent sections	Key differences to previous	Company base-case ICER
Model 1 (27/10/2022)	NA	£20,860
Model 2 (30/11/2022)	• 50% higher mortality for ECM patients for entire time horizon (EAG not notified) • Maintenance efficacy of 29.31% applied (EAG not notified) • Correction of 0.02 patient disutility error (B16) • Correction of applying annual mortality every 28 days (B17) • Increase time horizon to 100 years (B17) • Correction of extrapolating general population mortality based on only males in a predominantly female population (B18) • Correction of incorrect dosing for GNX (B23) • Inclusion of rescue medication costs per arm (B25) • Correction of AE costs for entire follow up being applied every cycle (B26) • Other minor changes	£19,419
Model 3 (22/12/2022)	• Reversion of 50% mortality increase for ECM patients • Reversion from maintenance period HL shift of 29.31% to full Marigold HL shift of 27.08%	£22,200

Abbreviations: AE, adverse event; EAG, External Assessment Group; ECM, established clinical management; GNX, ganaxolone; ICER, incremental cost-effectiveness ratio; OLE, open-label extension.

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

The company conducted a single review to identify all relevant evidence for this submission, including evidence for the clinical effectiveness, cost-effectiveness, health-related quality of life (HRQoL), and cost and resource use. A summary of the EAG's critique of the company's approach to identifying these types of evidence is provided in Table 12.

Ultimately, the company did not identify any cost-effectiveness evaluations of therapies for people with CDD; though it is unclear from the CS whether any relevant studies were identified for other genetic epilepsy populations captured within the inclusion criteria of the searches. Similarly, the company's SLR did not yield any HRQoL or cost and resource use studies in a CDD population specifically, though studies from other populations were considered to serve as a proxy for CDD (for the purpose of informing the company's cost-effectiveness model).

Table 12. Summary of EAG's critique of the methods implemented by the company to identify cost-effectiveness evidence

Systematic review step	Section of CS in which methods are reported	EAG assessment of robustness of methods
Searches	Appendix D, Section D.1.1	The company searched a combination of bibliographic databases, conference websites, clinical trials registries, websites of relevant organisations, google scholar, and reference lists of relevant studies. The strategy used appeared appropriate, although at clarification (question A2) the EAG questioned the search terms used for alternative patient populations. In response, the company re-ran the search using alternative terms to confirm that no studies had been missed. At clarification (question A1), the EAG also requested further details about the strategy used for supplementary searches. In response, the company submitted the terms used to search one such resource, which were appropriate and provided reassurance that other sources were appropriately searched.
Inclusion criteria	Appendix D, Section D.1	Inclusion criteria were not formally defined with respect to cost-effectiveness evidence. However, criteria appeared broad (including non-CDD populations and a range of burden-of-illness studies) and therefore were likely to have captured available evidence if it existed.
Screening	Appendix D, Section D.1.3.1	Dual screening was used at all levels of evidence, with involvement of a third reviewer as needed.
Data extraction	Appendix D, Section D.1.3.1	A DET was discussed but not explicitly presented. A single reviewer extracted data with quality assurance by a second reviewer.

Systematic review step	Section of CS in which methods are reported	EAG assessment of robustness of methods
QA of included studies	Appendix G, Section G.1	There was no apparent QA of cost effectiveness studies in other populations, though their inclusion in the review was unclear. No QA was conducted for HRQoL or cost and resource studies.

Abbreviations: CDD, CDKL5 deficiency disorder; CS, Company Submission; DET, data extraction template; EAG, External Assessment Group; HRQoL, health-related quality of life; QA, quality assessment.

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

4.2.1. NICE reference case checklist

Table 13: NICE reference case checklist

Attribute	Reference case	EAG comment on company's submission
Perspective on outcomes	All direct health effects, whether for patients or, when relevant, caregivers	× Perspective of model captured health effects on both patients and caregivers, but was not exhaustive and was subject to a number of limitations
Perspective on costs	NHS and PSS	✓ All costs included related to patients – no costs included for caregivers
Type of economic evaluation	Cost–utility analysis with fully incremental analysis	✓ Single comparison (GNX + ECM versus ECM alone) presented
Time horizon	Long enough to reflect all important differences in costs or outcomes between the technologies being compared	✓ Lifetime horizon of up to a maximum of 100 years, set to 75 years in original base-case analysis and updated to 100 years following clarification
Synthesis of evidence on health effects	Based on systematic review	✓ Relevant studies identified from systematic review (with scope extended to include proxy conditions given anticipated low number of hits in a CDD-specific population)
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.	✓ Health effects expressed as QALYs, though EQ-5D not used for estimation of all included utility values
Source of data for measurement of health-related quality of life	Reported directly by patients and/or caregivers	× Seizure-related utility based on a vignette study
Source of preference data for valuation of changes in	Representative sample of the UK population	× Vignette study by Lo *et al.*, (2022) used general population valuation, though as 200 participants

Attribute	Reference case	EAG comment on company's submission
health-related quality of life		were included the representativeness of this sample is unclear
Equity considerations	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	**?** Severity weighting of 1.7 applied to QALYs gained by both patients and caregivers
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	✓ Majority of costs sourced from standard NHS and PSS reference material. Some costs were assumed, but these only influenced incremental results when a survival benefit was modelled
Discounting	The same annual rate for both costs and health effects (currently 3.5%)	✓ Costs and QALYs discounted at 3.5% per annum

Key: CDD, Cyclin-depended Kinase-like 5 [CDKL5] Deficiency Disorder; ECM, established clinical management; EQ-5D, EuroQol 5 dimension; GNX, ganaxolone; HRQoL: health-related quality of life; NHS, National Health Service; PSS, Personal Social Services; QALY: quality-adjusted life year; TA: technology appraisal

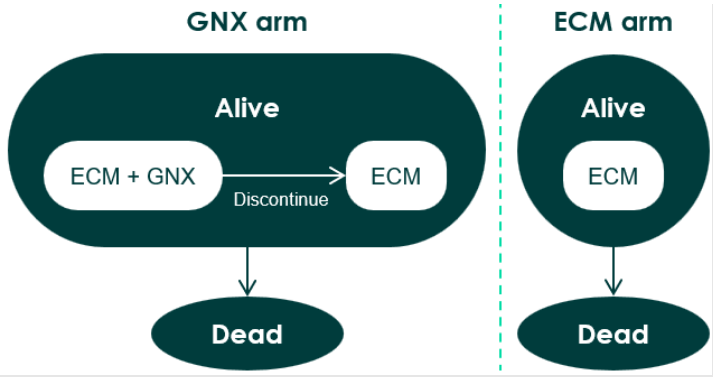
4.2.2. Model structure

The company presented a simple Markov state-transition model with two primary health states (alive and dead). In each 28-day (28d) cycle, patients transitioned from the alive state to the dead state in accordance with an overall survival extrapolation. This extrapolation was derived using a standardised mortality ratio (SMR) applied to general population survival, which asserted several assumptions that are discussed in Section 4.2.6. Patients in the GNX arm could also be on or off GNX treatment, with patients that were alive and on GNX discontinuing GNX at a rate of [redacted]% per 28d estimated based on data from the Marigold study (see Section 4.2.6). This effectively added a health state for the GNX arm for patients that were receiving GNX treatment, and so the EAG provided a revised model schematic to illustrate this (see Figure 1).

Patients treated with GNX were assumed to instantaneously receive the full treatment effect calculated using data from baseline and week 17 in the double-blind phase of the Marigold trial. Mechanically, the distribution of seizure frequency (SF) amongst a cohort treated with GNX was assumed to immediately change from that of the ECM population at baseline in the Marigold trial to a "shifted" distribution using the GNX treatment effect estimate using a Hodges-Lehmann estimation of location shift (hereafter referred to as HL for brevity). [20]

Upon discontinuation of GNX, the treatment effect was assumed to be immediately lost and the distribution of SF for discontinued patients became that of the ECM arm. No change of treatment effect over time was modelled, meaning that the company's model assumed that the 27.08% reduction in SF associated with GNX remains irrespective of the amount of time a patient has been treated with GNX.

Figure 1: Company model structure



Key: ECM, established clinical management; GNX, ganaxolone.
Note: Discontinuation of GNX at a rate of [redacted]% per cycle.

The model structure illustrated in Figure 1 may be considered atypical for NICE technology appraisals of genetic epileptic syndromes like DS or LGS. Previous NICE TAs have included Markov models with discrete SF-based health states (e.g., TA614 in DS) and patient-level microsimulations (e.g., TA808, also in DS). The EAG expected that these structures could potentially have been more appropriate in the case of this decision problem due to a number of potential advantages in this disease area. For instance, the SF-state based Markov model approach was non-parametric and considered changes in the distributional shape of SF in the population over time in both treatment arms. It would have also been possible to calculate the transition probabilities between health states to align with the health state definitions implied by the two available utility studies (see Section 4.2.7 for further details on utility values). The microsimulation approach would have allowed for nuances like seizure-free days (SFD) or repeated GNX treatment periods (see Section 4.2.6.2) to be modelled alongside SF, possibly taking correlation structures and non-linear associations into account. It may have also been possible in a patient-level simulation to simulate the process of response assessment and discontinuation of treatment. Within the timeframe of the appraisal, the EAG was unable to fully investigate whether there would be barriers to using these methods for this appraisal, though it

did not consider that the company had provided sufficient justification for using its chosen structure.

Overall, the EAG believed that a Markov approach in this context was reasonable in principle. However, while the EAG accepted that CDD was a rare condition and data was scarce, it considered that the company's implementation of the approach was heavily simplified. Limitations of this include that the analysis may have failed to consider the full effect of the treatment on this population and may therefore offer either an optimistic or a conservative estimate of the treatment effect and modelled outcomes.

4.2.3. Population

The prevalent population of people with CDD varies in age, symptom burden, and both the frequency and severity of seizures. In practice, people are likely to be treated with several different combinations of concomitant treatment to control seizures and other symptoms of the disease. Seizures in patients with CDD are recognised to be difficult to treat, requiring constant care associated with serious disease burden on both patients and those that provide care to them.[21] As CDD is extremely rare, there are challenges in generating high quality data showing the natural history of the disease under current standards of care. Moreover, as CDD was established as a disease in its own right relatively recently, there is an inevitable absence of long-term data.

The EAG had concerns regarding the population considered in the cost-effectiveness analysis in terms of its representativeness of the expected population who would be treated with GNX in UK clinical practice for two main reasons: (i) the age at which treatment would be initiated in practice (described below), and (ii) how SF was captured by the model (including the baseline distribution and how this may change over time; described in Section 4.2.6.1).

In the company's model, GNX was assumed to be initiated at an average age of [redacted] old. This differed from the expected marketing authorisation of GNX, which was for patients aged two years and over. In the Marigold trial, a small minority of participants only were aged under three years (range 2-19 years, median 6, mean 7.26, IQR 3-10 years).

Pending the marketing authorisation, the EAG considered it likely that, upon introduction into UK clinical practice, GNX would likely be initiated in people younger than seven years of age due to increasing awareness of CDD and facilities for diagnosis. However, the EAG was unclear whether GNX would likely be introduced before or after people with CDD had been prescribed

other ASMs. Uncertainty surrounding the likely starting age of people who receive GNX was significant to the appraisal as dosing of GNX was weight-based and would therefore be affected by age. As no dose-response relationship was assumed in the model (i.e., the treatment effect remained the same regardless of weight/age/dose), a reduction in starting age reduced the ICER for GNX as patients achieved the same SF % reduction for a smaller amount of GNX and therefore cost.

To explore the uncertainty of this, the EAG conducted a scenario analysis setting the baseline age of the modelled population to match that of the Marigold study. This is discussed in Sections 4.2.8.1 and 6.2.

4.2.4. Interventions and comparators

4.2.4.1. Intervention

The intervention modelled by the company was GNX and ECM. GNX was administered via an oral delivery syringe. EAG discussions on dosing and implications for the cost-effectiveness analysis are provided in Section 4.2.8.1. GNX is intended to be used adjunctive to ECM, meaning that unless contra-indicated, alternative treatments for seizures used in people with CDD may continue following GNX initiation.

4.2.4.2. Comparator

The comparator to GNX was ECM without GNX. ECM included a wide variety of different treatment approaches to manage seizures, including ASMs and non-pharmacological therapies. Please see Section 3.2.2.4 for more detail concerning the specification of ECM in the Marigold trial.

The estimation of a treatment effect for ECM was modelled differently across the three versions of the model submitted by the company. In the final submitted model, the company assumed the SF distribution of ECM was time invariant and therefore ASMs on average maintain SF indefinitely (regardless of the likelihood that ineffective treatments will be withdrawn, and new treatments initiated). This is discussed further in Section 4.2.6.1.

For the purposes of cost-effectiveness modelling, only those elements likely to differ between treatment arms (i.e., GNX+ECM and ECM) necessitated inclusion in the model. Some elements of ECM may have theoretically been relevant to the decision problem through an efficacy modifying effect, or potentially through GNX reducing the need for some existing ASMs. For

example, if GNX were to reduce the need for people to receive multiple ASMs to treat seizures, this may have resulted in benefits through reduced negative effects of polypharmacy. The EAG noticed that cannabidiol use in the Marigold study was restricted, and clinical advice to the EAG was that use may be higher in practice. The EAG was unclear whether cannabidiol would be expected to interact with GNX or alter the ECM treatment effect, and this issue was therefore not explored further.

4.2.5. Perspective, time horizon and discounting

4.2.5.1. Time horizon and discounting

The company included a time horizon of 100 years in the company's final model, arguing that this was a lifetime horizon and included the period in which any feasible clinical benefit and cost associated with introducing GNX to the CDD treatment pathway in the UK would be relevant. In principle, the EAG agreed with the company, however, the company did not present any scenarios using alternative time horizons.

The company applied discounting per the NICE reference case, at a rate of 3.5% per annum. The EAG agreed that this was appropriate. However, like the base-case choice of time horizon, the company did not present any scenarios based on discount rates applied.

Due to the above, the EAG introduced several scenario analyses, based on time horizon and discount rates applied within the cost-effectiveness model (see Section 6.2), to further explore the sensitivity of cost-effectiveness results.

4.2.5.2. Perspective on outcomes

The perspective taken throughout the submission was that of patients and caregivers, and outcomes were presented in the form of QALYs. However, the company's model did not fully capture all relevant outcomes which may be affected by the introduction of GNX. This is discussed in Section 4.2.6.1.

4.2.5.3. Perspective on costs

The perspective of the company's cost-effectiveness analysis was NHS and Personal Social Services (PSS). The company sourced the cost inputs for the cost-effectiveness model from a combination of the National Schedule of NHS Costs 2020-2021, and Unit Costs of Health and Social Care 2021 from the Personal Social Services Research Unit (PSSRU). The company also cited UK sources for resource use where available.

4.2.6. Treatment effectiveness and extrapolation

In the company's model, GNX was modelled to impact the estimation of QALYs through (i) SF, (ii) treatment duration, and (iii) mortality. These aspects of the company's model are described in the sub-sections that follow.

4.2.6.1. Seizure frequency (SF)

The company modelled and extrapolated count data on SF per 28-days by applying the estimated treatment effect from Marigold (see Table 31 of the CS) directly to the parameters of a parametric (lognormal) fit to baseline SF pooled across Marigold treatment arms.

The treatment effect of GNX was applied as a percentage reduction directly to the [redacted]. As discussed later in this sub-section, this was a mathematical error due to it violating the [redacted] – i.e., that:

[redacted]

Also see Section 6.1.1 where the EAG corrected this application. However, irrespective of implementation errors, the treatment effect of GNX was to move the SF distribution to the left (i.e., to reduce population average SF) for those remaining on GNX.

The company assumed that patients transitioned from one distribution (the pooled baseline SF from Marigold) to the other (the same distribution with the treatment effect applied) instantaneously upon initiation of GNX, that the treatment effect did not change over time, and that the treatment effect was lost immediately upon discontinuation of GNX. In the cost-effectiveness model, this translated to a simple modelling framework which essentially provided a weighted average SF distribution for patients in the ECM and GNX arms, depending on the GNX treatment effect and the proportion of patients that remained alive and on treatment. Mean SF was not explicitly calculated in the model but was reflected in the proportion of patients that fell into the health-state utility values (HSUVs), which were linked to SF (see Section 4.2.6.3 for further discussion related to utilities).

The lognormal fit to the SF data was not extrapolated or investigated for changes in distributional shape at different time points using the Marigold data, and no alternative candidate distributions were included in the company's model. The EAG asked the company about alternative distributions at clarification stage (see question B9), and the company explained that [redacted]. The EAG did not see this as a

justification for exclusion but agreed that the lognormal distribution provided the best fit from the included distributions.

The company's overall approach to capturing SF incorporated a large number of assumptions which had a considerable impact on modelled patient outcomes and therefore cost-effectiveness results. Many of the assumptions were implemented in the absence of evidence for this condition. An overview of the company's model assumptions for modelling SF and the EAG view on these is provided in Table 14. The importance of each assumption was determined by the potential impact on the cost-effectiveness of GNX. Where necessary, more thorough discussion on each issue is provided in the sub-sections that follow.

Table 14: Summary of key assumptions made by the company on CDD seizure frequency

Assumption	Company evidence and/or justification	EAG position, comments, and importance
Treatment effect maintained provided patients still on treatment	None provided	**High importance:** Disagree based on clinician input • Clinical experts suggested that they would perform an assessment at 6 months from initiation and determine response/discontinuation at this point • In clinical practice, patients are likely to discontinue if they have not or are no longer responding to treatment. This would then mean that amongst those that stay on treatment, the proportion that are responders would increase over time (see e.g., Specchio 2020). A regression to the mean effect may also be expected in those people who initiate treatment following a surge in SF (Key Issue 1). Pending clarity on the way in which treatment with GNX would be initiated and discontinued, the EAG was unable to incorporate a scenario to test the effect of alterations on the cost-effectiveness of GNX. However, the EAG was confident that the cost-effectiveness of GNX would be considerably improved by the implementation of clinically based treatment discontinuation (rather than just based on adverse events).
Secondary and tertiary seizures omitted from model	Secondary/ tertiary seizures not primary endpoint of Marigold, less common, difficult to measure and less impactful	**High importance:** Disagree. True ICER may be between scenarios with "primary seizures only" and "all seizures" • Effect of GNX could potentially differ by seizure type, though this is difficult to establish from limited data and challenges with measurement • All seizures would have been more in keeping with the scope of the appraisal • Company's estimate of HL shift for "all seizures" scenario was likely to be conservative
Baseline SF distribution in Marigold representative of UK clinical practice	None provided	**High importance:** Inconclusive. Current data were extremely scarce. However, there was a published survey which could have provided an alternative scenario • Clinical experts explained that ASM trial inclusion criteria (including Marigold) restrict baseline populations to high SF, which for some participants may be *"at the crest of a wave"* of seizures • Marketing authorisation for GNX was pending and there was uncertainty surrounding the way GNX would be used in clinical practice and if this would be comparable to Marigold (i.e., minimum threshold SF and previous failed ASMs)
Distribution of SF will not change over time	Some limited evidence provided in clarification response	**High importance:** Inconclusive. There was a lack of evidence to show long-term trends in SF, but longer-term comparative follow-up data could have influenced the ICER substantially • The company provided some evidence at clarification (question A9) that supported stable SF over time, but this had limitations (only information on the 17-week double-blind period was supplied and it was understandably difficult to illustrate the data without some clutter in the graphs). The accumulation of events appeared linear but the response was not considered by the EAG to be definitive • The EAG also identified a published survey in people with CDD showing that SF may change over time. [22]

Assumption	Company evidence and/or justification	EAG position, comments, and importance
		• Model results were considered likely to be sensitive to the shape of SF distribution at baseline, so the representativeness of the Marigold SF distribution remained an area of uncertainty
Appropriate to apply HL shift directly to the distributional parameters	None provided	**High importance:** Agree, following the EAG correction (see Section 6.1.1) • The original application was incorrect, which the EAG investigated further through a simulation exercise (presented in Section 6.1.1) • The EAG method generated reductions in mean, median and standard deviation close to 27.08%, whilst the company's method led to approximately [redacted]% reductions. HL shift estimates should generate corresponding changes in mean, median, standard deviation (i.e. 27.08%)
No change in seizure type or severity following introduction of GNX	None provided	**Medium importance:** Unclear. Evidence to the contrary was provided in the Marigold CSR, [14] but there was an unclear impact on cost-effectiveness results • Evidence in the Marigold CSR suggested there could potentially be variation in effect across different types of seizures, suggesting seizure type distribution changed for GNX patients • The EAG considered it plausible that seizure types had distinct utility and resource use impacts, and that GNX patients will then have had different utility and resource use implications per seizure versus ECM
Instantaneous treatment effect	None provided	**Medium importance:** Disagree. Contrary evidence was provided in the Marigold CSR[14] and the CS • The Marigold CSR reported a smaller HL shift estimate for GNX vs PBO during the titration weeks 0-4 vs. weeks 0-17 and 4-17 (-18.70%, -27.08% and -29.31%) • The EAG linearly interpolated the treatment effect between weeks 0, 4, and 16 in their base case (due to impossibility of 17 weeks within model structure, see following sections), which increased the ICER. Scenarios are presented without interpolation for comparison.
SF distribution best modelled with a lognormal distribution	Statistical fit of the distributions included in comparison to Marigold data at baseline (pooled across arms)	**Low importance:** Agree with choice of distribution, but some limitations. • The EAG expanded testing to include count-data distributions (Poisson, binomial, and negative binomial). Lognormal remained statistically best fitting • Lognormal distribution [redacted] • Alternative yet plausible distributions important to consider where possible
SF immediately reverts to baseline distribution after discontinuation of GNX	No evidence provided, but justified as being conservative	**Low importance:** Disagree due to down-titration of GNX per SmPC • SmPC for GNX stated that patients were to be down titrated upon discontinuation, as sudden discontinuation could cause an increase in the frequency of seizures • Clinical experts consulted by the EAG suggested that the down-titration phase of many ASMs would be long, ranging from 2 weeks to several months, depending on context • Discontinued patients would mostly consist of non-responders causing attrition effects, leading to GNX SF reduction moving upwards over time

Assumption	Company evidence and/or justification	EAG position, comments, and importance
		• Cost and efficacy implications of the down-titration period were unclear as there was no evidence, so the EAG were not able to include a scenario to test cost-effectiveness impacts

Key: ASM, anti-seizure medication; CSR, clinical study report; EAG, External Assessment Group; ECM, established clinical management; GNX, ganaxolone; HL, Hodges–Lehmann; ICER, incremental cost-effectiveness ratio; SF, seizure frequency; SmPC, summary of product characteristics.

Types of seizures to include in the cost-effectiveness model

The company elected to include “primary seizures” only within its model (see Section 3.2.2.5 and the Marigold CSR[14] for definitions) using the following justification:

1. Primary (major-motor) seizures were the primary endpoint of the Marigold trial
2. Secondary and tertiary seizures are less frequent and can be difficult to measure
3. Primary seizures tend to be the most impactful on patients and caregivers

The EAG discusses each of these points below.

The restriction of the model to primary seizures only was inconsistent with the scope for this appraisal (Section 2.4), regardless that it was the primary endpoint in the clinical trial evidence. As other seizure types were evaluated in the available clinical trial evidence, the EAG considered that these may have been considered within the company’s model.

The EAG noted that the number of observed secondary and tertiary seizures were lower than the number of primary seizures (see Marigold CSR Section 11.1.1.3.4), and agreed that these may be more challenging to measure (see Section 3.2.2 and Document B Section B.3.3.1). Data points for secondary and tertiary seizures provided in the trial CSR appendices had a high level of variance, though suggested the possibility of numerical differences in treatment effect between primary and secondary seizures type. While the EAG considered the data to be uncertain, it considered that the company had not been able to demonstrate that the treatment effect for GNX would be consistent across seizure types. This was, in the EAG’s opinion, a source of uncertainty surrounding the treatment effect of GNX which warranted consideration.

The total number of seizures in the analysis was based on a larger sample when including more types of seizures, so the company's argument of smaller N for secondary and tertiary seizures held only when analysing secondary and tertiary seizures separately from primary seizures. To clarify, the EAG did not advocate isolating secondary and tertiary seizures but considered that these could be combined within an “all seizures” analysis. However, as the estimated treatment effect of GNX may differ by seizure type, the primary SF distribution and the GNX impact on primary SF was unlikely to be a good proxy for secondary and tertiary SF distributions and the respective effects of GNX. This was complicated further when considering that the proportion of seizures by type may have been impacted following the introduction of GNX, which was not captured by the company’s model.

It was the EAG's view that all direct health effects associated with the introduction of GNX for people with CDD in the NHS were directly relevant to the decision problem at hand, per the final scope issued by NICE. The clinical experts consulted by the EAG both indicated that any seizure (irrespective of type) over five minutes in duration was an emergency, requiring both rescue medication and hospitalisation, and having considerable impact on patients and their caregivers. As people with CDD experience a broad range of seizure types, and the trial evidence for GNX suggested that the treatment effect may vary across types, it did not follow that secondary and tertiary seizures could be considered irrelevant to the NICE decision problem, even if they were expected to be less common and less impactful.

In the EAG's opinion the exclusion of secondary and tertiary seizures from the cost-effectiveness model introduced decision uncertainty and potential bias. Ideally secondary and tertiary seizures should have been incorporated into the model to take into consideration the potentially differential treatment effect. Yet, due to issues with the data on SF for secondary and tertiary seizures, there was inherent uncertainty linked with introducing these additional seizure types within the model.

The EAG expected the most accurate ICER to lie between the two approaches (i.e., only primary seizures vs. all seizures). However, the EAG expected the ICER was likely to be closer to the "primary seizures" scenario ICER, due to the lower incidence of secondary and tertiary seizures, so the EAG continued to use primary seizures within its base case. However, in the EAG's opinion the all-seizures scenario analyses presented contributed considerably to the overall uncertainty surrounding GNX cost-effectiveness.

Baseline SF distribution

Overall, the EAG considered that many of the eligible criteria for the Marigold trial may align with the target population in clinical practice (see Section 3.2.2.2). However, the EAG were uncertain whether the frequency of seizures experienced by trial participants was representative of the whole CDD population. For instance, a published survey of (non-UK) caregivers for people with CDD by Leonard *et al.* reported distributions of SF at two time points ("baseline" and "follow up") [22] which differed from the Marigold sample, as shown in Table 15.

Table 15: CDD seizure frequency from Leonard *et al.*, (2022)

Seizure frequency	Baseline (n, %)	Follow-up (n, %)
None	12 (8.4)	17 (11.9)

Seizure frequency	Baseline (n, %)	Follow-up (n, %)
≤4 per month	14 (9.8)	15 (10.5)
1-6 per week	28 (19.6)	21 (14.7)
1-4 per day	59 (41.3)	38 (26.6)
≥5 per day	30 (21.0)	52 (36.4)

Notably, these data suggested that a considerable proportion of people with CDD may experience periods without seizures. The EAG assumed that some of these patients would not be eligible for GNX and were therefore not relevant to the appraisal. However, it also considered it plausible that some people with CDD may experience periods of time without seizures. A different shaped SF distribution may therefore be applicable to people with CDD in real-world practice versus the Marigold trial.

The company elected to represent the SF distribution at baseline in the Marigold trial via a lognormal distribution. Goodness of fit tests were performed on each candidate distribution explored by the company. The results of this process were reported in Table 30 of the CS, which included Akaike and Bayesian Information Criteria (AIC and BIC, respectively) and "GOF test p-value". These "GOF tests" were different for each distribution, which was not explained in the CS. At clarification stage, the company explained: [redacted] (company's response to question B7). Unfortunately, the EAG was unable to fully interpret the rightmost column in CS Table 30. However, the AIC and BIC values for the lognormal distribution were the smallest by a considerable margin, perhaps indicating superior distributional fit.

Overall, the EAG acknowledged that data on the distribution of SF in CDD populations was limited. However, it was the EAG's opinion that how well the SF distribution in the cost-effectiveness model characterised the SF of patients likely to receive GNX in clinical practice was critical to accurately capturing cost-effectiveness. In addition, while the lognormal distribution appeared to provide a reasonable fit to the Marigold data, this did not necessarily mean that this distribution provided a good fit to the real-world SF distribution.

Application of the treatment effect

The company presented an analysis of the change in SF over time to provide evidence for the efficacy of GNX in CDD. However, the distributional shape of SF was positively skewed, with

more patients having fewer seizures per four-week period (though some patients were shown to have hundreds of seizures). Consequently, the mean became less useful for characterising impacts on patients. Furthermore, SF was considered likely to non-linearly impact patient HRQoL. For example, the impact of one additional seizure for patients experiencing an average of 0 seizures per month was likely to be greater than the impact for patients already experiencing a large number of seizures per month (e.g., 100 increasing to 101 seizures per month). Consequently, simple characterisation of efficacy using the effect of GNX on mean SF was likely to provide biased cost-effectiveness analysis results.

The Marigold statistical analysis investigated changes in SF using individual patient data, including the arithmetic and proportional (percentage) change in SF at baseline and week 17. The arithmetic and percentage changes in SF were calculated as below for individual i:

$$\delta SF_i = SF_{w17,i} - SF_{bl,i}$$

$$\delta SF\%_i = \frac{SF_{w17,i} - SF_{bl,i}}{SF_{bl,i}} \times 100$$

Where SF is seizure frequency, δ is change, $w17$ is week 17, and bl is baseline.

The mean of δSF_i was then the mean of the individual changes in SF in the baseline cohort. In other words, this was one way of characterising the average change in SF or treatment effect. The same would have been true of the median, which may also be more appropriate in non-normally distributed contexts.

The company did not report the distribution of δSF_i or $\delta SF\%_i$, and therefore no evidence that these were non-normal in shape was provided. As δSF_i and $\delta SF\%_i$ were based on differences in SF over time rather than a cross-sectional or time-average estimate of SF itself, it does not follow that δSF_i and $\delta SF\%_i$ must have the same distributional characteristics as SF_i. Consequently, it was not possible for the EAG to examine whether the HL estimate of shift was the most appropriate means to incorporate the effect of GNX on individual patient changes in SF over time into a cost-effectiveness model. It may have been the case that mean difference or some simple mixed-effects regression analysis of SF, δSF_i or $\delta SF\%_i$ was a more appropriate approach for extrapolating SF and the efficacy of GNX long-term in a cost-effectiveness modelling setting.

Generally, the application of HL shift estimates to distributions of SF to estimate the distributional shape of SF for treated patients could have been a reasonable approach to capturing the value of a treatment that reduces seizures. However, applying it in this way assumed that the treatment effect did not include changing the *shape* of the distribution and therefore the effect was not itself in some way dependent on the number of seizures a patient was having at baseline (for instance, exponentially more beneficial for patients with higher initial SF). This application of the treatment effect assumed that all patients were affected in the same way, with the same percentage reduction of their SF.

The company assumed that the full treatment effect at 17 weeks in the Marigold trial applied to patients immediately from week 0. This implied that the first titration dose received by patients was just as effective as the full dose, and that patients immediately experienced the full reduction in SF. Both were optimistic assumptions that likely biased the cost-effectiveness analysis in favour of GNX. Clinical advice suggested that the effect would take time to manifest.

The CS detailed the treatment effect identified at the end of the titration period (4 weeks from baseline). This reported an HL shift estimate that was considerably smaller than the treatment effect at 17 weeks (-18.70% and -27.08%). This evidence directly contradicted the assumption that the treatment effect was instantaneously at its week-17 level. Therefore, the EAG amended this in its base-case cost-effectiveness analysis (see Section 6.3). The EAG interpolated the distributional parameters for modelled primary SF each 4 weeks linearly, using an initial value of 0, a 4-week value matching that of the Marigold trial, and a 16-week value equal to the 17-week value of Marigold. The treatment effect values were half-cycle corrected for fairness. This was slightly optimistic as it assumed the treatment effect reached maturity at week 16 rather than 17. However, due to the confines of the company's model structure, the EAG considered this sufficient to account for the evidence that SF gradually fell in a cohort treated with GNX.

Treatment waning and prolonged efficacy

The company assumed that the treatment effect never waned, meaning a patient was assumed to derive the same benefit from GNX for as long as they continued to receive treatment. Upon questioning about the clinical plausibility of this assumption, the clinical experts consulted by the EAG could not provide a definitive opinion due to a lack of long-term follow up data. Other ASMs typically only provide short-term benefits for SF, and so it is plausible that GNX may also offer only temporary relief. The EAG's opinion was therefore that this may have been an optimistic assumption which could have overestimated the long-term effectiveness of GNX (if,

for example, the effect of GNX reduces over time). However, a clinical expert suggested to the EAG that GNX would be withdrawn in practice from people who had not responded or had lost a response. If this was true, then SF only for those remaining GNX would improve when 'removing' non-responders from the sample until a point of stability (until any treatment effect waning occurred, at which point the on-treatment SF distribution would worsen). This effect is present in the poster by Specchio *et al.* reporting interim results from the Marigold OLE, and showing diminishing N but apparent continued improvement in SF amongst GNX/GNX patients.[23]

While the assumption of permanent treatment efficacy was potentially optimistic, the effect of attrition on the SF of those remaining on treatment was not accounted for within the model. On balance, the EAG expected that the combination of no treatment effect waning but no attrition-driven continued improvement in SF in those continuing to receive GNX to be a preferred approach that was a conservative-yet-uncertain assumption.

Upon discontinuation, the company assumed that the treatment effect was immediately lost. On the surface this seemed a conservative assumption. However, the SmPC for GNX revealed that patients must be down-titrated upon discontinuation from GNX to avoid the risk of an increase in SF. The EAG therefore assumed that in practice patients would continue to receive diminishing doses of GNX beyond discontinuation. This was not represented in the model, but as the EAG did not have any data on which to base a tapering off of the treatment effect and cost upon treatment discontinuation, it was not able to develop a scenario for this. The EAG expected that accounting for this would increase the ICER for GNX (though it was uncertain due to the uncertainty around down-titration duration and lingering treatment effect).

Overall, the application of the treatment effect in the company's base-case model via a HL shift estimate applied to a fitted distribution was heavily simplified, and this led to a mix of optimism and conservatism, the net effect of which was unclear. Where possible the EAG made adjustments and introduced scenarios to correct what it considered to be errors (see Section 6.1) and tested the sensitivity of the model result (see Section 6.2).

4.2.6.2. Duration of treatment

The company used data from the Marigold trial to estimate a discontinuation rate for GNX (see Document B Section B.3.3.1.3). This used what the EAG believed to be the number of discontinuations between the baseline and the end of the OLE (n=[redacted]), the number of patients

that continued to the end of the OLE (n=[redacted]), and the duration from baseline to the end of the OLE (2 years, or 104 weeks in the cost-effectiveness model). The company incorporated the following calculation to obtain the discontinuation rate used in the cost-effectiveness model:

[redacted]

For completeness, the model assumed that once patients discontinue treatment with GNX they will never reinitiate treatment with GNX (i.e., discontinuation was assumed to be permanent).

The EAG considered two important aspects of the company's approach to be important assumptions that required further consideration. These were: (i) that patients were assumed to discontinue GNX at a constant rate over time, and (ii) that patients could receive GNX only once over their lifetime. For brevity, the EAG's agreement or disagreement along with explanation for each of these two important assumptions are summarised in Table 16. Where necessary, more thorough discussion on each issue is provided in the sub-sections that follow.

Table 16: Summary of key assumptions about duration of ganaxolone treatment made by the company

Assumption	Company evidence and/or justification	EAG position, comments, and importance
Patients discontinue GNX at a fixed rate	Based on analysis of discontinuation for any reason in the Marigold study	**High importance:** Disagree based on clinical advice received • Estimation of the discontinuation rate was flawed as it was not based on exposure time • One clinical expert consulted by the EAG suggested that patients would be assessed at 6-months and those that have not experienced sufficient clinical benefit from the treatment would be discontinued from GNX • The EAG agreed that the model should reflect clinical advice and should incorporate clinical assessments for response if this was expected in practice
Patients can only receive GNX once in their lifetime	None provided	**Medium importance:** Disagree based on clinical advice received • One clinical expert consulted by the EAG suggested that patients could be initiated and discontinued from GNX multiple times in their lifetime, as a response to their SF increasing • The EAG considered that accounting for multiple uses of GNX could have influenced the cost-effectiveness of GNX in either direction. However, no evidence was available on repeated provision of GNX to CDD patients, so the EAG could not comment further on the likely impact.

Key: CDD, CDKL5 deficiency disorder; EAG, External Assessment Group; GNX, ganaxolone; SF, seizure frequency.

Constant discontinuation rate

The formula used by the company to estimate the discontinuation rate assumed that [redacted] patients were at risk of discontinuation throughout the DB and OLE phases of Marigold, and that discontinuations happened at a continuous rate (i.e., an exponential model would best characterise time on treatment). This calculation was incorrect, leading to a substantial overestimation of the discontinuation rate for GNX, which then led to an underestimate of GNX treatment cost, biasing the ICER downwards.

In the absence of a robust analysis of time on treatment, the EAG considered this an area of uncertainty which had a considerable impact on the ICER. To capture the GNX discontinuation rate more accurately using the data available to the EAG (i.e., summary-level data from the Marigold trial), the EAG estimated the total person time at risk of discontinuation in the Marigold DB phase (i.e., GNX dosed days) in units of 28 days to match the company's model, using Table 12 in the Marigold CSR.[14] The resulting rate (converted to a probability) was [redacted] per 28 days. This is detailed in Section 6.2.1. The EAG used this discontinuation rate in its base case.

The company model applied discontinuation randomly within the sample, rather than this being based on response to GNX, which the EAG considered implausible. A clinical expert informed the EAG that an assessment at 6 months from baseline would be conducted, at which time a patient would be considered for continuation or discontinuation of treatment. There was no consensus on the threshold of SF reduction which should be used to make this decision, and at clarification the company confirmed that they have not defined specific discontinuation criteria for GNX, and so the EAG was unable to consider this further. However, the EAG considered it plausible that an informal stopping rule would be adopted in practice, which would be associated with considerable improvement to the cost-effectiveness of GNX.

One course of GNX treatment possible during patient lifetime

A clinical expert informed the EAG that people would likely not receive GNX permanently but would stop and re-initiate treatment over their lifetimes. The plausibility of this would be evident with experience of its use in clinical practice, however benefits of this approach would include reduced polypharmacy, which is a major concern for people with CDD. If this occurred, eventually there would be a stable proportion of people being retreated with GNX. On the individual level for cost-effectiveness modelling, this was more complicated and required data

that was not available (i.e., time to event data on time to retreatment with GNX given number of previous rounds of GNX treatment). As these data did not exist, the EAG was unable to factor this into the cost-effectiveness modelling. This therefore remained an area of uncertainty.

4.2.6.3. Mortality

The company modelled the mortality of people with CDD based on a standardised mortality ratio (SMR), which was based on values provided by Chin *et al*'s study on LGS mortality and HCRU.[24] The SMR calculated by the company was 8.33, based on the ratio 5/0.6 deaths per 1,000 person-years. The value of 5 was taken as the midpoint from the statement: *"Results from the present study, using ONS linked data, demonstrate patients with LGS have a crude mortality rate of 4–6 per 1000 person-years"*. This rate was applied to the company's extrapolation of general population overall survival. The result was reported in CS Figure 17.

The key mortality assumptions made by the company, along with a summary of the EAG critique is summarised in Table 17 with further details provided in the subsections below.

Table 17: Summary of key assumptions about mortality made by the company

Assumption	Company evidence and/or justification	EAG position, comments, and importance
No explicit relationship between SF and mortality	None provided	**Medium importance:** Inconclusive. There were several publications linking SF to ORs for SUDEP, but not in a CDD population • SUDEP was a known issue in epileptic conditions and increased SF was highlighted as a risk factor, implying a relationship between SF and mortality • There was some evidence suggesting generalised seizures and ASM use were both associated with increased SUDEP risk in epilepsy[25] • Incorporating this into the model would decrease the ICER for GNX
Unclear derivation of mortality estimate	Limited explanation of source material provided	**Low importance:** EAG agreed with the overall approach taken to base mortality on proxy diseases given lack of data for CDD, but the derivation of the SMR is unclear • Mortality had a small impact on results if no difference is assumed between arms • However, mortality ultimately drove how long carer benefits were modelled, so it was necessary to ensure the approach taken was plausible

Key: ASM, anti-seizure medication; CDD, CDKL5 deficiency disorder; EAG, External Assessment Group; GNX, ganaxolone; ICER, incremental cost-effectiveness ratio; OR, odds ratio; SF, seizure frequency; SMR, standardised mortality ratio; SUDEP, sudden unexpected death in epilepsy.

No explicit link between SF and mortality

The company did not link mortality to SF, a factor which may have led to a slightly conservative analysis. There was a known relationship between seizures and sudden unexpected death in epilepsy (SUDEP), with some papers publishing odds ratios by level of SF. For this reason, it may be the case that reducing SF in a population could be associated with a marginal survival benefit, though the magnitude of this is likely to be small considering the -27.08% HL shift between arms in Marigold.

Determination of mortality in CDD relative to general population

As there is a lack of long-term data for survival amongst people with CDD, mortality estimates used in the company model are highly uncertain. Clinical advisors to the EAG disagreed widely on estimations of survival, and whether estimates for LGS (which is a common diagnosis in people with CDD) may offer a reasonable proxy estimate. Chin *et al.* reported crude mortality rates of 6.12 and 4.17 for confirmed and probable LGS per 1000 person-years, respectively. [24] Therefore, in the absence of data in the target population, the EAG suggested that the average crude mortality rate of the confirmed and probable LGS values reported by Chin *et al* should be used. This was a small change, and the effect on the cost-effectiveness model results was negligible in the company's original base-case analysis

The EAG was unable to reconcile the 0.6 per 1000 person-year value with the citation provided in the company submission, or the corresponding citation from Chin *et al.* The document cited was Death registrations, Populations and Age Standardised Rates, England 1981 to 2018. [26] The reported statistics are per 100,000 population, not per 1000 person-years, and it was unclear to the EAG how these rates could be used to calculate rate of death per person-year without further (and therefore uncited) information.

4.2.7. Health-related quality of life

Within the company's base-case model, utility values for patients were estimated based on a published study by Lo *et al.*, (2022) which allowed for differences in utility to be estimated according to the frequency of seizures within a given model cycle (the company explained that this approach was validated by a Key Opinion Leader (KOL), with reference to a review of types of seizures, their impact, and general comparability to CDD). [27] These utility values were then adjusted according to general population norms to take into consideration the impact of aging. A sensitivity analysis was presented using data from a study by Auvin *et al.*, (2021). [28]

For caregivers, utility values were also taken from Lo *et al.*, (2022), and were based on the estimated frequency of seizures, with each patient having an assumed number of caregivers for whom utility was impacted while patients were still alive. [27] However, no age adjustment was applied for caregivers. In the sub-sections that follow, a critique of these approaches is provided.

4.2.7.1. Patient utility

Lo et al. (2022) was a vignette study that aimed to produce utility values for people with tuberous sclerosis complex (TSC), and their caregivers. From this study, the company extracted values shown in Table 18.

Table 18: Utility values taken from Lo *et al.*, (2022) – patient utility

Label in Lo *et al.*	Value: mean (SE)	Description in Lo *et al.*	Application in company's model
P1	0.7250 (0.0250)	0 GSD^{-1}; 0 FSD^{-1}	[redacted]
P2*	0.5040 (0.0370)	0 GSD^{-1}; 1-2 FSD^{-1}	[redacted]
P3*	0.2820 (0.0530)	0 GSD^{-1}; 3-4 FSD^{-1}	[redacted]
P4*	0.0740 (0.0550)	0 GSD^{-1}; 5-14 FSD^{-1}	[redacted]
P5	0.1830 (0.0570)	1 GSD^{-1}; 0 FSD^{-1}	[redacted]
P6	0.0890 (0.0540)	2 GSD^{-1}; 0 FSD^{-1}	[redacted]
P7	-0.1130 (0.0590)	3-4 GSD^{-1}; 0 FSD^{-1}	[redacted]
P8*	-0.2340 (0.0560)	3-4 GSD^{-1}; 5-14 FSD^{-1}	[redacted]

Key: FSC^{-1}, focal seizures [with impaired awareness] per 28-day model cycle; FSD^{-1}, focal seizures [with impaired awareness] per day; GSC^{-1}, generalised seizures per 28-day model cycle; GSD^{-1}, generalised seizures per day; SE, standard error.

Note(s): *These greyed-out values are not applied in the company's base-case analysis, since focal seizures were not modelled.

Source(s): Values taken from Lo *et al.*, (2022). [27] Company model description based on submitted model file.

An alternative utility source was provided by the company and used in a scenario analysis. This was a different vignette study of people (and caregivers) with LGS or DS by Auvin *et al.*, (2021). The company did not state whether KOL validation was performed on this study (as well as the study by Lo *et al.*) to assess its suitability for use in this modelling context, nor did it explicitly state why Lo *et al.* was chosen in favour of Auvin *et al.* to inform its base-case analysis. The Auvin *et al.* study reported utility values based on the number of seizures within a month versus the number of SFD. The utility values from this study that were applied in scenario analysis ranged from 0.83 (0 seizures per month) to 0.36 (130 seizures per month).

There were some important limitations with the company's approach using Lo *et al.*. First, from Table 18 it could be seen that patients experiencing 0-27 seizures per 28 days were assumed to have the utility of the health state in Lo *et al.* defined as having 0 seizures per day. This assumption was incompatible with the fact that 0% of patients on both treatment arms were modelled to have 0 seizures per 28-day model cycle as a direct consequence of using a lognormal distribution to describe SF (see Section 4.2.6 for further details concerning estimation of SF). No information was available from Lo *et al.* concerning the utility of patients experiencing between 1 and 27 seizures per 28 days (i.e., greater than 0 per day, but less than 1 per day on average). Therefore, the EAG considered that the application of the values from Lo *et al.* may lack accuracy in describing the impact of SF on patient utility, with the expectation that in general, patient utility may be overestimated (e.g., a patient with 0.9 seizures per day is modelled to have the utility of 0 seizures per day, rather than 1 per day or a value between these estimates).

The second limitation was that there was a misalignment of the descriptions of seizures by type used in the study by Lo *et al.* and the company's model. The company's model took data from Lo *et al.* regarding generalised seizures (in its base-case analysis) and assumed these could be used as a proxy to describe the impact on utility for patients that experience primary ("major motor") seizures, excluding the impact of any focal seizures. It was unclear to the EAG whether this meant the impact of seizures on utility is under- or over-estimated by the company's model, considering that Lo *et al.* demonstrated that the addition of focal seizures had a marked impact on utility (i.e., an increase in focal seizures on top of generalised seizures led to a further decline in utility). Furthermore, it was unclear precisely how much overlap there was (in terms of utility impact) for patients that experienced generalised versus primary ("major motor") seizures.

In addition to these limitations, there was additional uncertainty with using data generated from a vignette study in a different patient population to inform utility values within the company's model. The EAG considered the two populations from the vignette studies (TSC [Lo *et al.*] and LGS or DS [Auvin *et al.*]) to be potential proxies for CDD. However, the EAG undertook further exploratory analysis of the utility values used in the company's model to investigate how influential alternative values were on model results (see Section 6.2).

There was a small difference in the percentage of SFD between the PBO and GNX arms in Marigold, both at baseline and at the end of follow up. This difference may have resulted in differences in patient HRQoL in states of the world with and without GNX included in the

treatment pathway for CDD. Clinical experts supported the importance of SFD to patient and caregiver health-related quality of life, further illustrating the merit of Auvin *et al.* to inform utility values within the model.

The utility values estimated by Auvin *et al.* were generally higher than those in Lo *et al.,* but both studies showed that increased SF was associated with considerable disease burden. In Auvin *et al.,* the range in health states was 0.21-0.83 (1 SFD and 130 seizures per month vs seizure-free). However, caregiver utilities were provided in the supplementary materials to the article which were lower than those applied to people with LGS or DS, which the EAG considered to lack face validity (see Section 4.2.7.2).

Overall, the EAG considered Auvin *et al.* to be a more appropriate study to inform utility values and applied the reported outcomes within its preferred base-case analysis for the following reasons:

- Auvin *et al.* utilities are for the same disease as Chin *et al.* for HCRU and mortality[24,28]
- Auvin *et al.* utilities cover more granular health states for SF
- Auvin *et al.* utilities take into account the proportion of SFD patients have

4.2.7.2. Caregiver utility

As per patient utility in the company's base-case analysis, the utility impact for caregivers was based on data from the study by Lo *et al.* (2022).[27] A summary of the corresponding utility values from this study is provided in Table 19.

Table 19: Utility values taken from Lo *et al.*, (2022) – caregiver utility

Label in Lo *et al.*	Value: mean (SE)	Description in Lo *et al.*	Application in company's model
P1	0.9050 (0.0080)	0 GSD^{-1}; 0 FSD^{-1}	[redacted]
P2*	0.7910 (0.0170)	0 GSD^{-1}; 1-2 FSD^{-1}	[redacted]
P3*	0.6380 (0.0370)	0 GSD^{-1}; 3-4 FSD^{-1}	[redacted]
P4*	0.4310 (0.0490)	0 GSD^{-1}; 5-14 FSD^{-1}	[redacted]
P5	0.5460 (0.0390)	1 GSD^{-1}; 0 FSD^{-1}	[redacted]
P6	0.4760 (0.0450)	2 GSD^{-1}; 0 FSD^{-1}	[redacted]
P7	0.3190 (0.0480)	3-4 GSD^{-1}; 0 FSD^{-1}	[redacted]
P8*	0.2210 (0.0530)	3-4 GSD^{-1}; 5-14 FSD^{-1}	[redacted]

Key: FSC^{-1}, focal seizures [with impaired awareness] per 28-day model cycle; FSD^{-1}, focal seizures [with impaired awareness] per day; GSC^{-1}, generalised seizures per 28-day model cycle; GSD^{-1}, generalised seizures per day; SE, standard error.

Note(s): *These greyed-out values are not applied in the company's base-case analysis, since focal seizures were not modelled.

Source(s): Values taken from Lo *et al.*, (2022). [27] Company model description based on submitted model file.

Within the company's model, caregivers were modelled to be separate entities to patients. This assumption entailed caregiver utility falling out of the scope of the NHS and PSS perspective upon the death of the patient being cared for. In a model, this is mechanically identical (though philosophically different) to assuming that the caregiver dies along with their patient. An immediate consequence of how caregiver utility was modelled was that estimates of survival were important drivers of caregiver QALYs, since this drove how long a difference in utility was modelled between the treatment arms (unless there was no difference in modelled survival between arms, per models 1 and 3). This result was exaggerated when considering that patients could have multiple caregivers (in this case, 1.8 caregivers until the age of 18 years, and then 1 after this point in time). At clarification stage, the company explained that values of 1.8 and 1 were chosen "due to the contribution of parental care during childhood and reflecting the average number of parents would be less than 2; after which the average reduces at age 18 due to patients reaching adulthood." (Company's response to clarification question B13). While based on assumption, the EAG considered it plausible that people with CDD under the age of 18 may have multiple caregivers versus adult patients.

In past NICE appraisals of ASMs for Dravet Syndrome (such as TA808 and TA614), the committee's preference was to take a "decrements only" approach to incorporating carer utility within a cost-effectiveness model. For example, the final guidance from TA808 contains a section titled: *"Incorporating carers' quality of life in the model is appropriate but this should be done by applying a carer disutility"* (TA808 technology appraisal guidance, p.23). [29] More specifically, the committee commented that incorporating carer utility – whereby caregivers were modelled to die at the same time as the patient – was unusual and would result in biased results. In the context of the current appraisal, the EAG concurred in principle with the preference of the committee for TA808 and agreed that carer utilities should only be considered in terms of disease burden additive to that of the patient being cared for. However, there are other limitations with using a "decrements only" approach, especially when the disease burden is extremely high as in CDD, due to the perverse incentives the approach may provide. In this appraisal, the EAG did not consider the use of a "decrements only" approach to caregiver

utilities to avoid negative utilities and difficulty with interpretation of results, and thus retained the utility approach used by the company. However, the EAG acknowledged the limitations of this approach, and that this was inconsistent with previous NICE committee preferences.

While not highlighted within the CS or implemented in the cost-effectiveness model, the study by Auvin *et al.*, (2021) also provided utility values for caregivers. These were reported in the supplementary appendix to the main text of this study. Values included:

- 0.38 (130 seizures and 3 SFD in an average month)
- 0.52 (80 seizures and 15 SFD in an average month)
- 0.78 (0 seizures, and 30 SFD in an average month)

The EAG was unclear on why these values were not incorporated into the scenario analysis which makes use of the Auvin *et al.* utility values for patients, and instead the company used values from Lo *et al* to inform the Auvin scenario. Furthermore, the EAG noticed that the company's implementation of Auvin *et al.* calculated the utilities of the states relative to the best state (seizure-free), rather than applying the utility values reported as they are reported. As no justification for or mention of this was provided by the company in its submission, the EAG implemented Auvin *et al.* as (absolute) utility values for both caregivers and patients in its base-case, per NICE methods guidance.

4.2.7.3. General population adjustment

The company applied the study by Ara & Brazier 2010, [30] which was used in previous NICE technology appraisals. However, this publication did not include the variance-covariance matrix required to apply a multivariate normal distribution to simulate the correlation structure between the parameters. Consequently, varying the parameters of the equation published in the article led to an unknown bias in the probabilistic results.

In 2022, the NICE decision support unit (DSU) published updated general population norms, which then updated the preferred source for NICE. [31] This source also provided a variance-covariance matrix which allowed the utilities to be varied according to their correlation structure. However, within the timeframe of the appraisal, the EAG chose not to apply this in the model as it was anticipated that it would have a small impact on model results.

4.2.8. Resources and costs

The company's model included costs that could broadly be considered to fall into one of three categories: (i) drug acquisition and treatment administration, (ii) health-state and resource use, and (iii) resolution and management of adverse reactions. These are discussed in the sub-sections that follow.

4.2.8.1. Drug acquisition and administration

GNX was administered orally three times daily with food, based on the following weight-based dosages:

- For patients weighing 28 kg or less:
  - Maximum dose 63 mg/kg per day (see CS Section B.3.5.4.1)
  - Average dose of [redacted] mg/kg per day (see CS Table 37)
- For patients weighing more than 28 kg:
  - Maximum dose 1,800 mg per day (see CS Section B.3.5.4.1)
  - Average dose of [redacted] mg/kg per day (see CS Table 37)

Initiation of GNX was based on a titration schedule for the first three weeks of treatment, again based on patient body weight (schedule taken from the FDA prescribing information)[32]:

- For patients weighing 28 kg or less:
  - Days 1 to 7: 18 mg/kg per day
  - Days 8 to 14: 33 mg/kg per day
  - Days 15 to 21: 48 mg/kg per day
- For patients weighing more than 28 kg:
  - Days 1 to 7: 450 mg per day
  - Days 8 to 14: 900 mg per day
  - Days 15 to 21: 1,350 per day

In addition to GNX, the company's model included two other types of drug acquisition costs: ECM and rescue medication. ECM was costed at £15 per day, irrespective of SF or treatment assignment. Rescue medication costs were omitted from the company's Model 1, as a cost of £0 per day was attributed to rescue medication. Both costs were applied as a daily cost, and the CS stated that in the Marigold study, *"… patients could receive a broad range of medications and other treatments concomitantly; received by both patients on ECM alone and ECM + GNX"*

and that *"no difference between arms"* was observed (CS Section B.3.5.4.1, Table 37). At Model 2 onwards, the company incorporated these based on a previous NICE TA (see Table 11). However, the implementation was incorrect (See Section 6.1.7).

The company model did not include any administration costs for GNX or other treatments given to patients as part of ECM across both arms.

In the company's model, the titration schedule for GNX was not explicitly modelled (nor was it defined within the CS). While the EAG would have preferred for the titration schedule to be explicitly modelled for accuracy of costings, it did not consider the omission of this likely to have a large impact on the overall acquisition costs of GNX.

Patients were assumed to enter the company's model aged [redacted], with a mean body weight of [redacted] kg (based on data collected in the Marigold trial). Over the course of the model's time horizon, the average weight of the cohort increased as patients aged, and the required dose was adjusted accordingly.

The company's model assumed no wastage in the acquisition cost of GNX. It was the EAG's understanding that GNX would be available in 110 mL (50 mg/mL) bottles, containing a total dose of 5,500 mg. Taking the maximum daily dose of 1,800 mg as an example, this meant that one bottle would provide at least three days treatment with some remaining (3 x 1,800 mg = 5,400 mg, with 100 mg remaining). The EAG considered it plausible that some wastage would occur both while administering a dose to a patient and in the changeover between bottles. Clinical advice suggested that around 10% of each bottle may be wasted in real-world practice.

The EAG considered the inclusion of ECM at a simple cost of £15 per day to be reasonable though arguably unnecessary given that no difference to ECM was expected across arms, and that this cost therefore had no impact on incremental costs in the model unless there is a difference in overall survival between arms.

Several errors with the company's implementation of rescue medication costs were identified by the EAG. These are addressed in Section 6.1.7. As noted above, at Model 2 onwards (see Table 11), the company incorporated rescue medication, based on NICE TA614 (mislabelled as ID1211 in the "CostParams" sheet). The EAG presumed that the values entered into the cost-effectiveness model were based on the values presented in Tables 29 and 30 of the CS in the TA614 committee papers. [33] Yet the implementation in the cost-effectiveness model does not match the align with the values those tables.

The company's model did not include any administration costs for GNX. Given that GNX was anticipated to be administered three times daily in the community setting, and that there was no difference in modelled ECM costs, the EAG considered the omission of an administration cost to be appropriate.

The impact of addressing the discordance in the anticipated dosing regimen for GNX versus the application in the company's model is explored further in Section 6 of the EAG's report, and a revised application was incorporated within the EAG's preferred base-case analysis (see Section 6.3). In addition, an alternative approach to account for potential wastage was also considered in Section 6 of the EAG's report.

4.2.8.2. Health-state and resource use

In the company's base-case analysis, health-state and resource use costs were included on the basis of a study by Chin *et al.*, (2021).[24] This was a retrospective linkage cohort study using data from the UK Clinical Practice Research Datalink (CPRD) GOLD database of patients with LGS. This study reported estimated frequencies of specific resource use items per patient year, stratified by whether patients were aged <12 or ≥12 years of age. In its model, the company applied the estimated frequencies for patients aged <12 years of age with 'confirmed' LGS (excluding those with 'probable' LGS). No explanation was provided in the CS concerning the restriction to patients aged <12 years of age or those with confirmed LGS.

The following costs were captured within the model: General practitioner (GP) consultation, GP home visit, GP phone call, nurse consultation, nurse home visit, nurse phone call, hospital outpatient visits, hospital inpatient admissions (all cause), hospital inpatient admissions (epilepsy related), and accident and emergency (A&E) visits. However, the latter two of these (i.e., epilepsy related inpatient admission and A&E visits) were assumed to differ between treatment arms – all other items were assumed to have the same frequency for both treatment arms for the full model time horizon.

The company acknowledged that use of inputs from Chin *et al.*, (2021) represented a non-CDD population, and as such inputs from other proxy conditions could have also reasonably been included. Therefore, the company provided an alternative option using inputs from a study by Lagae *et al.*, (2019).[34] This study comprised a survey of mostly European patients with DS and their caregivers, with total costs reported in USD, but results were presented on a subgroup analysis of UK patients only. While not explicitly stated in the CS, costs appeared to have been

converted from 2016 USD into 2021 GBP using a ratio of approximately 0.811. The specific categories included are presented in Table 40 of the CS (Section B.3.5.5.2, p.110).

The CS stated that in using this alternative approach, it was assumed that *"… there was no difference in the healthcare resource use between GNX as adjunctive therapy and ECM alone arms"* (CS, Section B.3.5.5.2, p.110). However, this was incorrect as similar to the base-case approach, GNX was assumed to lead to a 27.08% reduction in emergency visits and ambulance calls.

Overall, this alternative approach led to a smaller difference in the per-cycle resource use costs across both arms, as illustrated in Table 20.

Table 20: Comparison of resource use costs per 28 days

Arm	Chin *et al.* (2021) cost per 28 days	Lagae *et al.*, (2019) cost per 28 days
ECM alone	[redacted]	[redacted]
GNX + ECM	[redacted]	[redacted]
Difference	[redacted]	[redacted]

Abbreviations: ECM, established clinical management; GNX, ganaxolone.

Owing to a lack of data to the contrary, the EAG tentatively accepted the company's base-case approach to use the study by Chin *et al.* to inform resource use estimates that did not vary by treatment arm, with the understanding that there may have been differences in real-world practice (possibly in favour of GNX, if resource use was related to SF). Instead, the EAG focused its critique on the two items that were assumed to differ between treatment arms and therefore impact the model results.

In Chin *et al.*, (2021), 1.50 admissions associated with epilepsy were estimated per patient year (<12 years of age with confirmed LGS), whereas for GNX patients a 27.08% reduction in hospital admissions was assumed (using the point estimate of reduction in SF, discussed further in Section 4.2.6.1 of the EAG's report), resulting in 1.09 admissions per patient year. The same approach to capture the difference between arms was used to adjust the number of A&E admissions in the company's model: 0.85 for ECM, reducing to 0.62 for GNX (i.e., a reduction of 27.08%). The CS stated that this assumption *"was validated by the clinical KOL consulted"* (Section B.3.5.5.1).

The main assumption inherent in this approach was that a reduction in SF was perfectly positively correlated with the number of epilepsy-related hospital admissions and A&E visits. No empirical evidence was presented in support of this assumption, though the EAG acknowledged that limited data were expected to be available within the context of a CDD population. While this assumption was potentially plausible, the EAG highlighted that not all seizures would result in hospitalisation or an A&E visit. As such, a better proxy for the difference in resource use could potentially have been based on only including severe seizures (or at least specific types of seizures known to be linked with hospitalisation). The potential impact of this on the model results remained unclear and could have plausibly led to a lesser or greater reduction in resource use costs associated with GNX.

The unit cost used for an epilepsy-related hospital admission was £6,545.75, based on NHS reference costs 2020/21. A weighted average by the recorded number of Finished Consultant Episodes (FCEs) for the codes PRO2A, PRO2B, and PRO2C (paediatric epilepsy syndrome), as a non-elective long-stay admission. The EAG noted that the assumption of a long-stay admission was somewhat at odds with the Chin *et al.*, (2021) study which reported an average length of stay for an epilepsy-related hospital admission of 2.48 days (<12 years of age with confirmed LGS). However, the CS cited a study by Mangatt *et al.*, (2016) to support the assumption of a long-stay admission in a CDD population, citing an average length of stay of 27.4 days. [35] The exact quote from Mangatt *et al.*, (2016) was: *"For the children of the 69/91 families with seizure-related admissions who provided sufficient detail on these, the mean number of days in hospital due to seizure-related events was 27.4 (median 19 days, range 1 day to 4.9 months)"*. [35] The EAG highlighted that it was unclear from this whether the value of 27.4 days referred to an average length of stay per admission, or an overall average length of stay in hospital over an extended period of time potentially covering multiple admissions. As such, the EAG explored an alternative analysis wherein a non-elective short-stay admission was used in place of a long-stay cost (reducing the cost from £6,545.75 to £1,036.71). This cost was used in the EAG's base-case analysis.

4.2.8.3. Resolution and management of adverse events

The company's model also included costs associated with the resolution/ management of AEs. The approach used to capture these costs was relatively simple. The proportion of patients across both arms in the Marigold trial that experienced any AE requiring or prolonging

hospitalisation ([redacted] out of 101, [redacted]%) were assigned the cost of an inpatient stay (£1,182) at each model cycle for the full model time horizon.

The EAG highlighted two potential issues with the approach taken to capture costs associated with AEs. First, the EAG considered it inappropriate to apply this proportion at each model cycle across the full model time horizon (i.e., that it was unlikely that [redacted]% of patients would require an inpatient stay every 28 days). In Section 6 of the EAG's report, a revised approach was proposed to incorporate this adjusting for the duration of the Marigold trial (see Section 6.3). However, this change had no impact on the model results since no difference in the occurrence of AEs by treatment arm was modelled.

Second, the risk of AEs was assumed to be symmetrical in the GNX and ECM arms, which the EAG considered to lack face validity. While GNX was not associated with a major increase in AEs within the Marigold trial, some differences in AEs were noted by the EAG (see Section 3.2.3.1). Due to uncertainty about the generalisability of rates of AEs in the trial (caused by the small sample size and low event rate of AEs), and the expectation that AE costs would have little impact on model results, the EAG elected not to change this assumption in the model, despite its limitations.

5. COST-EFFECTIVENESS RESULTS

5.1. Company's cost-effectiveness results

5.1.1. Base case results

The results reported by the company are shown in Table 21 (based on Model 3, see Table 11). The deterministic and probabilistic results were associated with ICERs of £22,200 and £23,139 per QALY gained, respectively. However, the EAG identified errors in the company base-case analysis, and the corrected company base case results are presented in Section 6.1. Of note, the EAG highlighted that a severity modifier of 1.7 was applied both to patient and caregiver incremental QALYs. The severity modifier is discussed further in Section 7.

Table 21: Company base case results (model 3)

	Discounted costs	Discounted QALYs*	Incremental discounted costs	Incremental discounted QALYs*	Cost per QALY gained
Company deterministic base case					
ECM	[redacted]	[redacted]	-	-	-
GNX + ECM	[redacted]	[redacted]	[redacted]	[redacted]	£22,200
Company probabilistic base case					
ECM	[redacted]	[redacted]	-	-	-
GNX + ECM	[redacted]	[redacted]	[redacted]	[redacted]	£23,139

Abbreviations: ECM, established clinical management; GNX, ganaxolone; QALYs, quality adjusted life years;.

Note: *QALYs presented are adjusted to account for a severity modifier of 1.7, which is applied to the QALYs gained by both patients and their caregivers. Numerical results differ to those contained within the original company submission owing to edits made post-submission (see Table 11).

5.2. Company's sensitivity analyses

5.2.1. One-way sensitivity analysis

To explore the impact of changing key model parameters on the ICER, the company undertook a deterministic one-way sensitivity analysis (OWSA). The results of this analysis are provided in Figure 2 in the form of a tornado plot. The main parameters shown to influence the ICER were related to the dosing of GNX, utility values (including the number of caregivers), medical resource use, and the average age of patients entering the model.

Figure 2: Company's one-way sensitivity analysis tornado plot (model 3)

Abbreviations: FS, focal seizures; GS, generalized seizures; ICER, incremental cost-effectiveness ratio.

The EAG noted that all parameters were varied based on taking values equivalent to plus or minus 20% of the base-case value, regardless of any available information concerning parameter uncertainty (e.g., standard error [SE]) or skew within the distribution for each parameter. This also meant that the uncertainty expressed within the OWSA was misaligned with the uncertainty feeding into the probabilistic sensitivity analysis (PSA).

To illustrate this issue with an example, the primary measure of treatment effect took a base-case value of 27.08% and was varied at bounds of 21.66% and 32.50% in the OWSA. However, this parameter was sampled according to a Beta distribution within the PSA, using a SE of approximately 0.0969. If 2.5th and 97.5th percentiles were drawn from a Beta distribution using this information, the equivalent bounds would be 10.55% and 47.92%. This would be more closely aligned with the original 95% CI reported in the CS of 9.95% to 47.92% (CS, Section B.3.3.1.2, Table 31). Taking these values of the HL location shift, the ICER range was estimated to be [redacted] (lower bound) and [redacted] (upper bound). Therefore, it was the EAG's view that the OWSA did not adequately reflect the 'true' parameter uncertainty inherent within the company's model, and did not provide a reliable basis on which to determine which parameters appear to have the greatest influence on results, or the magnitude of impact on results.

5.2.2. Probabilistic sensitivity analysis

In addition to the OWSA, the company also undertook a probabilistic sensitivity analysis (PSA) to further explore parameter uncertainty. To do this, 1,000 iterations of the model results were produced informed by sampled parameters. The results of this analysis are provided in [redacted]

in the form of a scatterplot, demonstrating the incremental costs and QALYs for the comparison of GNX + ECM versus ECM alone. Per the company's base-case analysis, the probability that GNX + ECM was cost-effective at willingness-to-pay thresholds of £20,000 and £30,000 per QALY gained (taking into consideration severity weighting*) is [redacted] and [redacted], respectively.

[redacted]

Abbreviations: ECM, established clinical management; GNX, ganaxolone; QALY, quality-adjusted life year.

Note: Scatterplot re-formatted for ease of interpretation by presenting incremental scatterplot, changing colours, adding reference lines for willingness-to-pay thresholds, average results, and adjusting dimensions of plot for clarity of presentation within this report. Numerical results are unchanged from company's model re-submitted in response to clarification questions. Numerical results differ to those contained within the original company submission owing to edits made post-submission.

The EAG noted with respect to [redacted] that there was a large spread in the results outputted by the PSA, with incremental QALYs ranging from [redacted] to [redacted], and incremental costs ranging from [redacted] to [redacted] (the deterministic and mean probabilistic results were also similar).

In its submission, the company speculated that the average probabilistic ICER was slightly higher than the deterministic ICER due to *"a 'floor effect' introduced by attempts to*

* Severity weighting applied per company's base-case approach.

conservatively model the left-skewed [SF] *data from the Marigold study"* (CS, Section B.3.10.1.3, p.120). However, the EAG noted that the average ICER presented in the CS was taken as the average across each iteration, rather than basing this on the ratio of the average incremental costs and QALYs in each instance. When re-calculating the mean probabilistic ICER (and taking into consideration the edits made by the company following clarification questions), the results were broadly aligned (see Table 21). The EAG therefore did not consider the company's comment regarding a 'floor effect' to be of material impact to decision making. Overall, the EAG did not identify any major concerns with the PSA undertaken by the company. However, owing to the number of assumptions made to inform the model, the EAG noted that the results of the PSA may underestimate the true uncertainty associated with the model results.

5.2.3. Scenario analyses

In its original submission (i.e., using model 1), the company presented several deterministic scenario analyses to further test model settings and assumptions.

Scenario analyses were not updated following submission of model 2 or 3. For completeness, the EAG attempted to re-produce all the scenarios using model 3, and available results are shown in Table 22. The EAG was unable to re-produce the results of scenarios B and C provided in the CS as changing the related settings in the model did not replicate the results presented by the company, and so Table 22 includes the ICERs the EAG calculated when changing the relevant model settings.

Overall, the EAG highlighted that the range of scenarios presented by the company was limited in number. Other scenarios of potential interest included exploration of alternative discontinuation rates, inclusion/exclusion of rescue medication costs, and alternative assumptions related to mortality. Where feasible within the timeframe available to the EAG to conduct its review, further analyses were undertaken and are reported in Section 6.2 of this report.

Table 22: Summary of company scenario analyses

Scenario*	CS ICER	EAG comment	EAG calculated ICER post-CQs†
Base case	[redacted]	-	[redacted]

Scenario*	CS ICER	EAG comment	EAG calculated ICER post-CQs†
A	[redacted]	Switch included within the company's model, which functions as intended.	[redacted]
B	[redacted]	No switches provided but could re-produce manually and in automated scenario analysis incorporated by the EAG. Results do not match CS as base-case has changed at Models 2 and 3 (Table 11), and the company have not provided any results errata or addenda.	[redacted]
C	[redacted]		[redacted]
D	[redacted]		[redacted]
E	[redacted]		[redacted]
F	[redacted]		[redacted]

Abbreviations: CQs, clarification questions; CS, company submission; EAG, External Assessment Group; ICER, incremental cost-effectiveness ratio.

Note: *For scenario labels, please refer to CS Table 50; †ICERs presented here are aligned with the company's base-case results provided at clarification stage. **Note that this matches the ICER of Model 3 sent to the EAG, see Table 11.

5.3. Model validation and face validity check

The company did not present any information concerning model validation. While the company stated that no published economic evaluations of treatments for CDD were identified in the SLR (CS, Section B.3.13.1, p.123), the EAG did not consider this to be sufficient justification in accordance with NICE methods. [36] The CS stated that steps were taken to test the proposed data and conditions used as proxies via validation with a clinical KOL, but no further details were provided. As such, the EAG was unable to critique the company's approach taken to model validation and assessing the face validity of results.

6. EXTERNAL ASSESSMENT GROUP'S ADDITIONAL ANALYSES

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

This section is organised as follows: Section 6.1 details the impact of errors identified in the EAG's validation of the company model. Section 6.2 details a series of scenario analyses exploring the robustness of the cost-effectiveness results to specific assumptions and additional uncertainties identified by the EAG. These analyses were conducted within the company corrected base-case analysis.

In Section 6.3, the EAG base-case is presented based on a combination of the exploratory analyses presented in Section 6.2, and taking into consideration the corrections made in Section 6.1.

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

The EAG identified errors in the original model submitted by the company (model 1). In its response to clarification (questions B16, B17, B23, B26, C1, C2), the company resolved several of these, which were then considered to be resolved and not discussed further in this section. However, a number of errors were remaining in model 3. These are summarised in Table 23, and where necessary, more detail is then provided in the sections that follow.

Table 23: Errors found in Company's cost-effectiveness model

Error found and section (if necessary)	Importance and explanation	EAG solution
Error 1: The application of the HL shift estimate to the distribution of SF was mathematically incorrect ([redacted]), considerably overestimating the treatment effect. Section 6.1.1	**High**. The ICER was sensitive to SF and the error results in a modelled ~67.5% reduction in mean, median and SD SF, not 27.08% or 29.31% as per HL estimates	The EAG followed [redacted] correct the error. This resulted in reductions in mean, median and standard deviation in line with the HL estimates reported (via simulation).
Error 2: Caregivers were simulated to be ageless and their utilities were not age-adjusted	**High**. Caregiver utility was implemented as constant, leading to an overestimate of caregiver QALYs	The EAG used ONS data[37,38] on the distribution of age at parenthood and the baseline age of CDD patients in Marigold to estimate the age of caregivers at baseline, then used this to age-adjust their utility values using Ara & Brazier 2010

Error found and section (if necessary)	Importance and explanation	EAG solution
Error 3: The company omitted the caregiver utility values reported in the supplementary materials of Auvin *et al.* 2021 from the model. Instead, the company linked to the caregiver utilities from Lo *et al.* 2021. Section 4.2.7.2	**High** when Auvin et al. was used as utility source (no effect on company base-case, but considerable effect on EAG base-case)	The EAG incorporated the omitted caregiver utilities from Auvin *et al.*
Error 4: The implementation of rescue medication was incorrect, overestimating the cost and therefore the cost reduction of GNX. Section 6.1.7	**High** cost of rescue medication was considerably overestimated, which disproportionately benefitted GNX	The EAG corrected the error by calculating the proportion of patients in each state at each timepoint and calculating a weighted average cost using the correctly inflated rescue medication cost value
Error 5: The parameters were not varied correctly in the one-way sensitivity analysis. Section 5.2	**High**. The company's tornado plot was misleading and does not appropriately reflect the true sensitivity of the ICER to changes in model parameters	This issue increased uncertainty in model results
Error 6: Absolute utility values estimated via Ara & Brazier 2010 were applied to patient utilities as multipliers.	**Moderate**. This should be relative to general population age- and sex- adjusted utility at baseline	The EAG calculated age-adjusted utilities relative to their value at baseline
Error 7: The maximum SF per 28 days included in the model was 400, meaning the total density in each distribution was not the same and the area under the curve did not approach 1. Section 6.1.4	**Moderate**. The 400 limit underestimated QALY benefit of GNX	The EAG expanded SF to 1000/28d to ensure that >99% of the density was included for both treatment arms and the distributions could be more consistently compared
Error 8: The values from Lo et al were not applied correctly due to a small overlap in days (28 days). Section 6.1.2	**Low**. ICER effect was small	The EAG fixed this error and included this in its automated scenario analysis
Error 9: The SMR based on Chin *et al.* was incorrectly based on rounded values. Section 6.1.5	**Low**. ICER effect was small	The values 6.12 and 4.17 from Chin *et al.,* are used by the EAG instead of 6 and 4.
Error 10: The scenario analyses presented in the CS could not be replicated accurately by the EAG due to model version changes and lack of automation. Section 5.2	**Unclear**.	The EAG have built automated scenario analysis into the cost-effectiveness model to ensure consistency. The EAG uses their own scenario results for inference

Error found and section (if necessary)	Importance and explanation	EAG solution
Error 11: The probabilistic ICER was calculated incorrectly as the mean of the probabilistic ICERs, rather than the mean of incremental costs divided by the mean of incremental QALYs	**Low**. Affects uncertainty estimations	The EAG included a single cell in the PSAcalcs sheet which calculated the correct probabilistic ICER

Abbreviations: EAG, external assessment group; CDD, CDKL5 deficiency disorder; DS, Dravet's syndrome; HL, Hodges–Lehmann; ICER, incremental cost-effectiveness ratio; LGS, Lennox-Gastaut Syndrome; ONS, Office for National Statistics; QALY, quality-adjusted life year; SF, seizure frequency; SMR, standardised mortality ratio.

Note: *For scenario labels, please refer to CS Table 50; †ICERs presented here are aligned with the company's base-case results provided at clarification stage.

6.1.1. The treatment effect was applied incorrectly

The company argued for a lognormal distribution to characterise the distribution of SF in CDD patients. The EAG expanded testing of potential distributions, and agreed with the company that lognormal was likely the most appropriate distribution (see Section 4.2.6.1).

The EAG investigated the application of treatment effect in the model. A full discussion is provided in Appendix A. To summarise, the company implementation was incorrect and resulted in a large overestimate of the impact of the estimated HL shift associated with GNX treatment on a lognormal distribution. This was investigated further by the EAG via a simulation study, which showed that the company's implementation resulted in an effect of around 67%, whilst the EAG-corrected implementation results in an effect of approximately the HL shift observed in the Marigold study.

6.1.2. Lo et al. implementation error

In Cells Q87:R88 in "ClinicalParams", the days included in the two rows both include 28. This was a simple implementation error, which the EAG corrected. The EAG implemented the switch "EAG_corr_loTopRow" so that the company can easily toggle the fix.

6.1.3. Age adjustment for caregivers

The company did not implement age adjustment for caregivers, assuming them to be ageless which was incorrect. The EAG considered this to be an implementation error biasing the ICER in favour of GNX due to the overestimate of incremental caregiver QALYs that resulted from not adjusting for the age caregivers over time.

To correct this, an estimated age of caregiver at baseline was calculated by the EAG using data from ONS[37,38] and the Marigold study. ONS data on the frequency of maternity by age was used to calculate a weighted average age at birth of child (calculations were provided in the "Settings" sheet of the EAG's modified company cost-effectiveness model). This resulted in a value of 30.41 years. The model then simply added age at baseline from the cost-effectiveness model (assumed [redacted] in the company's base-case) to this to provide expected age of parent at the time of GNX initiation. This age value was then applied to the equation in Ara & Brazier to produce a utility for age and sex matched general population utility corresponding to caregiver characteristics at baseline.

The EAG implemented age adjustment for caregivers to align with the age adjustment applied for patients. See Section 6.1.6 for the discussion on implementation.

6.1.4. SF distribution is truncated at 400 seizures / 28-days

In the 'seizure model' sheet within the cost-effectiveness model, the company presented a table providing the density associated with each SF value from 0 to 400. The lognormal distribution does not have an upper bound and therefore it was impossible to have an integral of 1 without an upper bound of infinity. The usual course of action would be to select an upper bound which covered at least 99% of the distribution in both arms to reduce bias and ensure a reasonably accurate estimate of the mean value. However, the company did not do this, which resulted in the total density of the ECM arm summing to 96.06% whilst the total density of the GNX arm summed to 99.83%. This introduced bias into the cost-effectiveness model. For instance, the mean estimated in the ECM arm in the company base-case was [redacted], which was considerably less than the [redacted] estimated when setting the upper bound to 5000, or the true mean of [redacted]. Consequently, expected SF in the ECM arm was underestimated by a larger percentage than in the GNX arm. This error biased the model *against* GNX through underestimating incremental QALYs.

To correct this error, the EAG made the following changes:

1) The EAG increased the upper limit of SF to 1000 to include more of the density
2) The EAG designed a VBA function to approximately calculate area under the curve between two integer bounds (default 0 and 1000), allowing either expected value or proportion to be produced (provided in the cost-effectiveness model)

3) The EAG incorporated this function into the Excel model, allowing both interpolation of the treatment effect over time and efficiently increasing a larger upper bound to SF

6.1.5. Chin *et al.* LGS mortality rate incorrectly calculated

As discussed in Section 4.2.6.3, the company used values from Chin *et al.* to provide an estimated SMR to apply to general population mortality. This was based on the values 4 and 6, which were rounded. Elsewhere in the article, Chin *et al.* reported slightly different values (6.12 and 4.17). Therefore, the EAG took the average of those (5.145), instead of 5 per the company's original base-case.

6.1.6. Age adjustment for patients

The EAG implemented a revised approach to age-adjusted utility values for patients, via a multiplicative approach rather than an additive approach. Two absolute utility values cannot be multiplied as the resulting value has no basis, whereas an absolute utility value and a relative utility can be multiplied, with the result having a basis in the absolute value. The company's original implementation of age adjustment was to calculate what they refer to as a "base utility" value for the age- and sex-matched general population. This base value was an absolute utility value of the age- and sex-matched general population. This value was then multiplied by the health state utility value for CDD given SF per either Lo *et al.* or Auvin *et al.* This was incorrect as the result of multiplying two absolute utility values had no meaning.

To amend this, the EAG calculated general population utility relatively to its value at baseline, and then applied this relative utility to the absolute utility of CDD given SF. This method was used by default for age adjustment of caregiver utility (See Section 6.1.3).

6.1.7. Correction to the implementation of rescue medication costs

As discussed in Section 4.2.8.1, at Model 2 (see Table 11), the company incorporated rescue medication costs into the cost-effectiveness model. These were based on NICE TA614. However, the EAG identified several errors with this implementation and have corrected them.

The company multiplied the proportion of patients with SF 0-28 per 28-days by £204 and those 28+ per 28 days by £408 using the proportions fitting into the Lo et al health states for reasons the EAG did not understand. This was incorrect for several reasons:

- The underlying cost value had not been inflated from 2018 values to the most recent available using the PSSRU[39]

- The calculation should be based on the states in Table 29 of the TA614 committee papers (i.e., $SF \leq 8$, $8 < SF \leq 25$ and $SF > 25$ with SF monthly). However, only under- and over-28 were used by the company, and the wrong values were used for this (e.g., 24 uses per year * £34 per use = £816 per year for those between 8 and 25 seizures per month, but only £204 and £408 used by company without any explanation in the report or response to clarification)

- Months were not translated into 28-day cycles, so the time unit was mismatched between the source material and cost-effectiveness model

- The rates reported in Table 30 of the committee papers were uses of rescue medication per year (the table in the TA614 committee papers is titled “annual rates”). Yet, the company used costs calculated based on these annual rates at every 28 days (See “Trace Gan” and “Trace SoC” sheets column AP). This led to an estimated lifetime rescue medication cost of around £266,000 for ECM patients, which translated to 7824 uses of rescue medication for the average patient lifetime, or 137.52 times per year of life (i.e., 10.6 times per 28-day cycle, around 2-3 times per week).

The EAG incorporated a correction (controlled in the cost-effectiveness model using the toggle “EAG_corr_rescueMed”), which used the VB function described in Section 6.1.4 to estimate the proportion of patients in each of the states corresponding to TA614 at each time point in the model (See Sheet “EAG_util_and_RM”). These proportions were used to calculate a weighted average cost of rescue medication for patients on ECM+GNX and ECM (over time when the treatment effect is interpolated). For instance, the estimated rescue medication cost per cycle for ECM patients was £112.20, which corresponded to a per cycle use of rescue medication of 3.14/28d. This was on the high end for the TA614 health states, as the worst state is 25+ seizures per month which corresponded to 75.05% of patients at baseline in Marigold (for the lognormal fit).

6.1.8. EAG-corrected company base-case analysis

To summarise, the corrections made to the company’s cost-effectiveness model were:

- Correction 1 (Error 1): The mathematically incorrect treatment effect application

- Correction 2 (Error 8): Minor error in implementation of Lo *et al.* utilities
- Correction 3 (Error 2): Age adjustment applied to caregivers
- Correction 4 (Error 9): Correction of the SMR calculated using Chin *et al.*
- Correction 5 (Error 7): Function for area under lognormal and increase upper bound to 1000
- Correction 6 (Error 6): Correction of age adjustment for patients
- Correction 7 (Error 4): Errors in the implementation of rescue medication costs
- Other corrections (Errors 3, 5): Inclusion of Auvin *et al.* caregiver utilities, use of absolute utilities reported from Auvin *et al.*

Note, errors 10 & 11 related to issues with the sensitivity analyses and did not affect the company deterministic base-case.

Table 24 reports the individual and cumulative impacts of these corrections on the estimated ICER. Notably, the correction with the largest impact was the application of the HL shift estimate to the lognormal distributional fit to the Marigold baseline SF data discussed in Section 6.1.1. This considerably increased the ICER because the company's incorrect implementation substantially over-applied the treatment effect, leading to the GNX cohort experiencing approximately a 67% reduction in SF rather than the company's intended base-case reduction of 27.08%.

Aside from the correction to the application of the treatment effect, the other corrections were less impactful. Corrections 5 and 6 reduced the ICER, corrections 2, 3, and 7 increased the ICER, and correction 4 had a negligible effect due to the lack of any mortality benefit for GNX in the company's base-case. The net impact of the other changes to the model was to reduce the ICER slightly from that with only correction 1. This was because the ICER-reducing impact of corrections 5 and 6 were larger than the combined increasing effects of corrections 2, 3, 4, and 7. Note that correction 7 was made following Model 2, which reintroduced rescue medication following clarification questions.

Table 24: Individual and cumulative impact of corrections made to errors in the Company's model

Preferred assumption	Section in EAG report	Incremental costs	Incremental QALYs	ICER £/QALY	+/- company base case
Company base-case	5	[redacted]	[redacted]	[redacted]	
Impact of individual EAG corrections					
Correction 1: Incorrectly implemented treatment effect	6.1.1	[redacted]	[redacted]	[redacted]	[redacted]
Correction 2: Implementation of Lo *et al.* utilities	6.1.2	[redacted]	[redacted]	[redacted]	[redacted]
Correction 3: Age adjustment for caregivers	6.1.3	[redacted]	[redacted]	[redacted]	[redacted]
Correction 4: SMR based on wrong values from Chin et al	6.1.5	[redacted]	[redacted]	[redacted]	[redacted]
Correction 5: Using EAG AUC function and increasing SF upper limit to 1000	6.1.4	[redacted]	[redacted]	[redacted]	[redacted]
Correction 6: Age adjust patients	6.1.6	[redacted]	[redacted]	[redacted]	[redacted]
Correction 7: Rescue medication	6.1.7	[redacted]	[redacted]	[redacted]	[redacted]
Cumulative impact of EAG corrections					
Correction 1+2		[redacted]	[redacted]	[redacted]	[redacted]
Correction 1+2+3		[redacted]	[redacted]	[redacted]	[redacted]
Correction 1+2+3+4		[redacted]	[redacted]	[redacted]	[redacted]
Correction 1+2+3+4+5		[redacted]	[redacted]	[redacted]	[redacted]
Correction 1+2+3+4+5+6		[redacted]	[redacted]	[redacted]	[redacted]
Correction 1+2+3+4+5+6+7		[redacted]	[redacted]	[redacted]	[redacted]

Note: In the corrections, the severity modifier used was calculated to be 1.7x for caregivers and patients throughout. This was primarily due to the use of the Lo *et al.* SF-based utility values. See Section 7 for discussion.

Table 25 provides the EAG's corrected version of the company's base-case analysis results. Two base-cases are presented to show the results both with and without the severity modifier for caregiver utilities.

Table 25: EAG-corrected company base case results

	Discounted costs	Discounted QALYs	Incremental discounted costs	Incremental discounted QALYs	Cost per QALY gained
EAG corrected company deterministic base case (**With** severity modifier* for caregivers)					

	Discounted costs	Discounted QALYs	Incremental discounted costs	Incremental discounted QALYs	Cost per QALY gained
ECM	[redacted]	[redacted]	-	-	-
GNX+ECM	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
EAG corrected company deterministic base case (**Without** severity modifier* for caregivers)					
ECM	[redacted]	[redacted]	-	-	-
GNX+ECM	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
EAG corrected company probabilistic base case (**With** severity modifier* for caregivers)					
ECM	[redacted]	[redacted]	-	-	-
GNX+ECM	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
EAG corrected company probabilistic base case (**Without** severity modifier* for caregivers)					
ECM	[redacted]	[redacted]	-	-	-
GNX+ECM	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Abbreviations: QALYs, quality adjusted life years

Note: In the EAG-corrected company base-case, the severity modifier used was calculated as 1.7x. See Section 7.

Figure 3 provides an updated cost-effectiveness plane incorporating the severity modifier for patients and caregivers, showing that only a small minority of probabilistic iterations were cost effective at a willingness to pay threshold of £30,000 ([redacted]%). When not applying the severity modifier to caregivers this probability fell to [redacted]% (Figure 4).

Figure 3: Cost-effectiveness plane for corrected company base-case with 1.7x severity modifier applied to incremental caregiver QALYs

Note: The severity modifier used was calculated to be 1.7x. See Section 7 for discussion.

Figure 4: Cost-effectiveness plane for corrected company base-case with 1x severity modifier applied to incremental caregiver QALYs

Note: See Section 7 for discussion around severity modifiers.

6.2. Exploratory and sensitivity analyses undertaken by the EAG

The scenario analyses presented in this section focus on the following issues and uncertainties:

- Changes made and included within the EAG's preferred base-case:
  - Discontinuation rates
  - Use of Marigold OLE efficacy estimate (for GNX/GNX cohort)
  - Removal of mortality benefit (not required upon receipt of Model 3, see Table 11)
  - Utility sources, and use of absolute utility values rather than relative to best state
  - Dynamics of the treatment effect
  - Inclusion of wastage
  - Adjustment of hospitalisation cost
  - The applicability of disease severity modification to caregivers
- 'Standard' scenarios requested in NICE methods guidance but not presented by the company:
  - Discounting scenarios
  - Time horizon scenarios
- EAG exploratory/robustness scenarios:
  - Seizure types to consider
  - Analysis time points to consider
  - Patient age at baseline
  - Number of caregivers
  - Caregivers for adult patients

6.2.1. Discontinuation rates

The company applied a value of [redacted]% per 28-days, which was calculated using discontinuations divided by patients at the end of follow up. In the Marigold trial, there were [redacted] GNX discontinuations (CSR section 10.1), and baseline GNX population at risk was [redacted] patients. The median exposure time was [redacted] treated days (CSR Table 12). An estimated total would be [redacted] days at risk of GNX discontinuation. Converting to 28-day cycles gave [redacted] 28-day cycles at risk of discontinuation from GNX. [redacted] produces a *rate*, r, of [redacted] GNX discontinuations per 28-day cycle at risk of discontinuation (i.e., on GNX treatment). Using $P = 1 - e^{-rt}$ to assume a continuous exposure (i.e., exponential), the resulting 28-day cycle probability of discontinuation from GNX was [redacted]%. This was applied in the EAG's preferred base-case analysis.

6.2.2. Efficacy data used

At Model 2 (see Table 11), the company changed its base-case to use the Marigold maintenance period HL shift estimate to power the model (29.31%). At Model 3, this was reverted to the estimate for the whole DB period (27.08%) without any explanation from the company.

In principle, the EAG agreed with Model 3 – that in the case that the treatment effect was applied from baseline, the efficacy for the DB period should be used. However, owing to the EAG's stance on the dynamics of the treatment effect (see Section 6.2.5) the EAG considered it more appropriate to apply the maintenance period reduction in SF from Marigold when interpolating the treatment effect over time. This then considered the difference between the treatment effect before and after titration of GNX. Consequently, in its base-case the EAG preferred to use the HL shift estimate from the maintenance period (29.31%).

6.2.3. Mortality assumptions

In Model 2, the company introduced the assumption that ECM patients were exposed to 50% more mortality than GNX patients, irrespective of whether they were on or off GNX at the time. The company labelled this as 'hypothetical' in the model file, though it featured within its revised base-case analysis. In Model 3 (see Table 11), this assumption was revoked. The EAG agreed with the removal of this assumption from the base case.

6.2.4. Utility assumptions

6.2.4.1. Auvin *et al.* 2021 utilities

In its model, the company included two different utility sources – both of which were vignette studies of potential proxy diseases to CDD in terms of disease burden. The EAG preferred Auvin *et al.* over Lo *et al.* for the following reasons (discussed in more detail in Sections 4.2.7.1 and 4.2.7.2):

- The intersection between LGS and CDD
- Granularity of SF health states
- Factoring in of SFD
- Consistency with basis of other modelling areas (i.e., LGS as a proxy disease for CDD)
  - Mortality data on LGS patients reported by Chin *et al.*
  - HCRU data on LGS patients reported by Chin *et al.*

The company implementation of Auvin *et al.* omitted the caregiver utilities that were reported in the supplementary materials (Appendix A; see Sections 4.2.7.1 and 4.2.7.2). Furthermore, supplementary data file 1 contained a full report of the caregiver vignette study, with a more detailed breakdown of the mapping exercises. The EAG considered this to be an error by omission, as no justification was provided for linking to the caregiver utilities in Lo *et al.* instead of simply using the caregiver utilities reported in Auvin *et al.* As the company did not use Auvin *et al.* in its base case, this did not affect the company corrected base-case ICER but did influence the results when using Auvin *et al.* as a utility source.

The company implemented the utilities from Auvin *et al.* as relative utilities (relative to utility in the seizure-free health state). This set patient utility in the seizure-free health state to 1, which the EAG saw as unrealistic considering patients would still be affected by the broader impacts of CDD. Therefore, the EAG preferred to apply the utilities from Auvin *et al.* as absolute values.

The EAG made the following adjustments to the cost-effectiveness model for its base case:

1) Use of Auvin *et al.* over Lo *et al.*, as discussed in Sections 4.2.7.1 and 4.2.7.2
2) Inclusion of caregiver utilities reported in Auvin *et al.*, as discussed in Section 4.2.7.2
3) Calculation of Auvin *et al.* utilities as absolute rather than relative values

6.2.4.2. Application of disease severity modifiers to caregivers

The relevant wording in the NICE methods guide on QALY shortfall and severity modifiers (methods guide section 6.2.12) is as follows:

> *"The committee will consider the severity of the condition, defined as the future health lost by people living with the condition with standard care in the NHS (including use of other available treatments, diagnostics, or best supportive care). The extent of unmet health need is reflected within the severity definition"*

The EAG considered this to apply to those who have the condition, and not those taking care of those with the condition. However, as the term "living with the condition" was used, the EAG considered that this could feasibly be interpreted to be ambiguous towards patients and their caregivers (who the EAG see as 'living with those that are living with the condition'). The EAG contacted NICE for clarification on whether the application of a severity modifier for caregivers would be considered to be consistent with the guidance, and at the time of submission of this report, the issue was under discussion within the NICE team.

To allow for a pending decision on the use of a severity modifier for carers, the EAG presents two separate base-case ICERs – one with caregiver severity modification and one without. This, and the applicability of the modifier are discussed in more depth in Section 7.

6.2.5. Treatment efficacy interpolation

As discussed throughout Section 4.2.6, the company presented three different HL shift estimates for primary seizures at three different time points. These were:

- -18.70% at 4 weeks from baseline in Marigold (titration period)
- -27.08% at 17 weeks from baseline in Marigold (DB period)
- -29.31% weeks 4-17 in Marigold (maintenance period)

The EAG considered this evidence that the treatment effect of GNX was not instantaneously the -29.31% effect estimated by the company for the maintenance period within Marigold or the -27.08% effect for the double-blind period. This suggested that it was potentially optimistic to assume the full effect immediately from baseline. To resolve this, the EAG linearly interpolated the treatment effect of GNX (see Section 6.2.5). This then ensured that the cost-effectiveness

model followed the clinical evidence on SF distributional change in a GNX treated cohort (as would be expected of a Markov model).

On the other hand, in other decision modelling settings, such as oncology, a treatment effect (e.g., a time-invariant hazard ratio) may be calculated using the full follow-up data and then applied from baseline for those on treatment. The company approach of applying an instantaneous treatment effect followed this convention as all patients instantaneously have their SF reduced by 27.08% conditional on GNX treatment (analogously to hazard being reduced according to a hazard ratio whilst on a treatment). Overall, the EAG preferred to reflect the dynamics of the treatment effect at different times reported according to the clinical data reported in Marigold and the OLE. Yet, to reflect the convention of instantaneous rather than cumulating treatment effect the EAG also presents scenarios reflecting the EAG's base case *without* interpolation of the treatment effect (Table 27).

The function was used to estimate the proportion of patients in the Lo *et al.* and Auvin *et al.* health states over time up until [redacted] weeks (the extent of the follow up in the OLE).

6.2.6. Drug wastage

Two clinicians consulted by the EAG indicated that it may be likely that some GNX product would be wasted in practice. One expert suggested that drug wastage of approximately 10% may be expected and would seem a reasonable estimate to inform the model. This was incorporated into the model as a simple 10% increase to the cost of GNX per cycle. While the value of 10% was palpably uncertain, the EAG highlighted that the assumption of zero wastage was misaligned with the clinical advice received by the EAG, and so this estimate was preferred over the company's base-case analysis which included no wastage.

6.2.7. Resource use costs

As discussed in Section 4.2.8.2, the company implemented a long-stay cost when the data on LGS patients from Chin *et al.* suggested that hospital stays tended to be short. [24] Therefore, the EAG preferred to use a short-stay cost in its base-case.

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

The EAG made the changes described in the sub-sections above. Each change was made individually and was combined within the EAG's preferred base-case analysis (see Section 6.3).

The results of the EAG's exploratory analyses are provided in Table 26. All scenarios presented in the table are based on the EAG corrected company base-case. The individual changes are ordered descending in terms of impact on the corrected company base-case ICER.

Table 26: EAG's exploratory analyses

Scenario / change to cost-effectiveness model	Section in EAG report	Incremental costs	Incremental QALYs	ICER £/QALY	+/- corrected company base case
Impact of individual changes					
EAG corrected company base-case*	6.1	[redacted]	[redacted]	[redacted]	[redacted]
Use of Auvin *et al.* (with absolute values and caregiver utilities)*, ***	4.2.7 6.2.4.1	[redacted]	[redacted]	[redacted]	[redacted]
Age 7.26 years at baseline (Marigold age)	4.2.3	[redacted]	[redacted]	[redacted]	[redacted]
1 caregiver	4.2.7.2	[redacted]	[redacted]	[redacted]	[redacted]
No severity modifier for caregivers**	6.2.4.2	[redacted]	[redacted]	[redacted]	[redacted]
Interpolation of the treatment effect*	4.2.6.1 6.2.5	[redacted]	[redacted]	[redacted]	[redacted]
Hospitalisation short stay based on Chin *et al.**	4.2.8.2 6.2.7	[redacted]	[redacted]	[redacted]	[redacted]
Including 10% wastage*	4.2.8.1 6.2.6	[redacted]	[redacted]	[redacted]	[redacted]
Discontinuation rate based on exposure time*	4.2.6.2 6.2.1	[redacted]	[redacted]	[redacted]	[redacted]
No caregivers 18+	4.2.7.2	[redacted]	[redacted]	[redacted]	[redacted]
Use of the maintenance HL*	6.2.2	[redacted]	[redacted]	[redacted]	[redacted]
"Standard" scenarios					
No discounting	Standard	[redacted]	[redacted]	[redacted]	[redacted]
No discounting - costs	Standard	[redacted]	[redacted]	[redacted]	[redacted]
No discounting - QALYs	Standard	[redacted]	[redacted]	[redacted]	[redacted]
TH 10 years	Standard	[redacted]	[redacted]	[redacted]	[redacted]
TH 20 years	Standard	[redacted]	[redacted]	[redacted]	[redacted]
TH 50 years	Standard	[redacted]	[redacted]	[redacted]	[redacted]
Selected combined scenarios					

Scenario / change to cost-effectiveness model	Section in EAG report	Incremental costs	Incremental QALYs	ICER £/QALY	+/- corrected company base case
Marigold age, primary seizures, caregiver severity 1x	Exploratory	[redacted]	[redacted]	[redacted]	[redacted]
Marigold age, primary seizures, caregiver severity 1x, all seizures	Exploratory	[redacted]	[redacted]	[redacted]	[redacted]

Abbreviations: EAG, External Assessment Group; ICER, incremental cost-effectiveness ratio; QALY, quality adjusted life year; TH, time horizon

Notes: *Used in the EAG base-cases, ** Included/excluded in the EAG base-cases, ***1.2x severity modifier calculated using the ECM arm patient flow sheet in the company cost-effectiveness model

The most impactful individual changes were those affecting the cost of GNX (e.g., baseline age) and those affecting patient utility (e.g., use of Auvin *et al.*). Other notably impactful scenarios include interpolation of the treatment effect, which then interacted with those scenarios affecting patient HRQoL given SF.

Overall, none of these scenarios suggested that GNX had an ICER at or below £30,000/QALY gained, even when accounting for disease severity and applying a severity modifier to caregivers.

6.3. EAG's preferred assumptions

The EAG preferred base case ICERs were £868,980 without the severity modifier for caregivers and £783,900 with a (1.2x) modifier for caregivers. Table 27 shows the individual and cumulative impact of the changes selected by the EAG.

In preparation of the final preferred EAG base case, the EAG opted not to include scenarios shown in the top section of Table 26 that were considered conservative. The EAG therefore consider the final reported ICERs to be a balanced estimate of the cost effectiveness of introducing GNX into clinical practice. The results of relevant scenarios are presented in Table 28 for completeness.

Table 27: EAG's preferred model assumptions

Preferred assumption	Section in EAG report	Cumulative ICER £/QALY
Company base-case	5	[redacted]
EAG corrected company base-case	6.1	[redacted]

Preferred assumption	Section in EAG report	Cumulative ICER £/QALY
EAG 1: Discontinuation rate based on exposure time	4.2.6.2 6.2.1	[redacted]
EAG 2: Use of the Marigold maintenance HL	6.2.2	[redacted]
EAG 3: Use of Auvin et al (with absolute values and caregiver utilities) (Note: affects severity modifier)*	4.2.7 6.2.4.1	[redacted]
EAG 4: Interpolation of the treatment effect	4.2.6.1 6.2.5	[redacted]
EAG 5: Including 10% wastage	4.2.8.1 6.2.6	[redacted]
EAG 6: Hospitalisation short stay cost	4.2.8.2 6.2.7	[redacted]
EAG 7: Severity modifier applied to patients only	6.2.4.2	[redacted]
EAG 1 + 2	4.2.6.2 6.2.1 6.2.2	[redacted]
EAG 1 + 2 + 3*	4.2.7 6.2.4.1	[redacted]
EAG 1 + 2 + 3 + 4*	4.2.6.1 6.2.5	[redacted]
EAG 1 + 2 + 3 + 4 + 5*	4.2.8.1 6.2.6	[redacted]
EAG 1 + 2 + 3 + 5 + 6*	6.2.5	[redacted]
EAG 1 + 2 + 3 + 6*	6.2.6	[redacted]
EAG base-case (EAG 1 + 2 + 3 + 4 + 5 + 6)*: applying caregiver severity modifier	6.3	[redacted]
EAG base-case (EAG 1 + 2 + 3 + 4 + 5 + 6 + 7)*: Not applying caregiver severity modifier	6.3	£868,980

Abbreviations: EAG, External Assessment Group; ICER, incremental cost-effectiveness ratio; QALY, quality adjusted life year

Note: *1.2x severity modifier calculated using the ECM arm patient flow sheet in the company cost-effectiveness model, and applied to incremental QALYs between GNX and ECM arms.

Table 28: Additional exploratory scenarios not included in the EAG base-case (based on the EAG's base-case)

Scenario / change to cost-effectiveness model	Section in EAG report	Incremental costs	Incremental QALYs	ICER £/QALY	+/- corrected company base case
Marigold age, primary seizures, applying caregiver severity modifier, maintenance efficacy	Exploratory	[redacted]	[redacted]	[redacted]	[redacted]
Marigold age, primary seizures, not applying caregiver severity modifier, maintenance efficacy	Exploratory	[redacted]	[redacted]	[redacted]	[redacted]
Marigold age, all seizures, applying caregiver severity	Exploratory	[redacted]	[redacted]	[redacted]	[redacted]

Scenario / change to cost-effectiveness model	Section in EAG report	Incremental costs	Incremental QALYs	ICER £/QALY	+/- corrected company base case
modifier, maintenance efficacy					
Marigold age, all seizures, not applying caregiver severity modifier, maintenance efficacy	Exploratory	[redacted]	[redacted]	[redacted]	[redacted]

Abbreviations: EAG, External Assessment Group; ICER, incremental cost-effectiveness ratio; QALY, quality adjusted life year

Note: *1.2x severity modifier calculated using the ECM arm patient flow sheet in the company cost-effectiveness model, and applied to incremental QALYs between GNX and ECM arms.

6.4. Conclusions of the cost-effectiveness section

- The company's model adopted a simple structure, revolving around health states of 'alive' and 'dead'. The EAG considered that this structure constituted an over-simplification of a complex disease, and in turn meant that interpretation of the cost-effectiveness results for GNX based on the model were subject to substantial uncertainty. In addition, a number of miscellaneous model errors and unsubstantiated assumptions were identified as part of the EAG's review, further adding to the uncertainty associated with the results generated from the model.

- While the Marigold trial suggested that GNX may reduce SF for some people with CDD compared to ECM, the long-term treatment effect of GNX was highly uncertain and this also impacted the results of the cost-effectiveness model. The application of the reduction in SF within the company's model was flawed from multiple perspectives, and so the EAG addressed this as far as was possible with the available data within its preferred analysis.

- Capturing the association between SF and HRQoL was challenging, especially considering that no utility values could be generated from the Marigold trial. In lieu of this, the company sought data from vignette studies, each of which were associated with notable uncertainty. The choice of study to populate the model had a large impact on cost-effectiveness results, impacting both patients and their caregivers.

- There were a number of outstanding issues associated with the cost-effectiveness modelling that the EAG was unable to address within the scope of its appraisal. These

included the potential for re-treatment with GNX over a lifetime horizon, the possibility of a different model structure better reflecting the impact of GNX on patient outcomes, and a lack of data available for a CDD-specific population to populate a number of model parameters (i.e., mortality, resource use, and quality of life).

- Overall, after correcting for errors in the modelling, the ICER for GNX appeared to be in excess of the range of £20,000 and £30,000 per QALY gained. This finding was based on what the EAG considered to be a highly optimistic corrected company base-case. When making what the EAG considered to be reasonable changes to the company's base-case analysis, the ICER increased substantially beyond the NICE willingness to pay threshold.

7. SEVERITY MODIFIER

The company applied a disease severity modifier to both CDD patients *and* their caregivers of 1.7x the incremental QALYs between the GNX and ECM arms. The company did not include any scenarios exploring different modifiers, or the applicability of those modifiers to caregivers and/or patients.

Using the mortality data provided by Chin *et al.* on patients with LGS, [24] and the utility estimates from Lo *et al.*, [27] the company calculated the expected lifetime discounted QALYs for a patient with CDD treated with ECM from aged [redacted] to be [redacted]. This compared to an age- and sex-matched general population discounted QALY estimate of [redacted] QALYs. As the absolute discounted QALY shortfall was more than 18, the corresponding severity modifier was 1.7x. In the EAG base-case, the expected lifetime discounted absolute QALYs for ECM patients was [redacted], leading to an absolute shortfall of 15.51 discounted QALYs, hence a severity weighting of 1.2x.

As discussed in Section 6.2.4.2, the EAG considered the guidance for the applicability of severity modification to caregivers as ambiguous. Further, the EAG considered that if the severity modifier were to be applicable to caregivers, then the determination of the severity modifier applied should be based on their distinct shortfall. That is, the amount of QALYs caregivers would be expected to accrue during their time (relevant to the NICE decision problem) compared to the equivalent period if they were not caring for a person living with the condition. Within the context of this decision problem, this would be when imposing the overall survival of CDD patients (estimated to be [redacted] years in the cost-effectiveness model) to general population HRQoL and comparing this to the equivalent for those caregivers in the ECM arm. From this, absolute and proportionate shortfalls could be calculated.

When this exercise was conducted, the absolute and relative QALY shortfalls based on discounted QALYs in the EAG corrected company base case were [redacted] QALYs and [redacted]% respectively ([redacted] QALYs and [redacted]% respectively in the EAG's base case). These were insufficient to meet either the 1.2x or 1.7x severity modification thresholds. Therefore, if caregivers were considered for disease severity modifiers based on their shortfall (i.e., treated as separate entities), the severity modification would not apply to them in this case as they would not meet the criteria. However, as it remained unclear whether the severity modification

based on patients should be used for caregiver utilities, the EAG presented results both including and excluding the modifier.

The QALY gain in the company's original (uncorrected) deterministic analysis reduced from [redacted] to [redacted] if the severity modifier was applied only to patients (i.e., removed for caregivers), which reduced further to [redacted] if the severity modifier was removed altogether (i.e., removed for both patients and caregivers). The corresponding (deterministic) ICERs for these scenarios were £[redacted] (severity modifier for patients and caregivers), £[redacted] (severity modifier for patients only), and £[redacted] (no severity modifier) in the original un-corrected company base-case. In the EAG corrected company base case the difference grew larger, and then larger again in the EAG base-case. The other scenarios are presented throughout Sections 5 and 6 and inclusion/exclusion of the severity modifier to caregivers had a similar effect of substantially affecting the ICER for GNX+ECM versus ECM.

References

1. Jakimiec M, Paprocka J, Śmigiel R. CDKL5 Deficiency Disorder—A Complex Epileptic Encephalopathy. Brain Sciences. 2020;10(2):107.

2. Kothur k, Holman K, Farnsworth E, et al. Diagnostic yield of targeted massively parallel sequencing in children with epileptic encephalopathy. Seizure. 2018;59:132-40.

3. Liang J-S, Huang H, Wang J-S, Lu JF. Phenotypic manifestations between male and female children with CDKL5 mutations. Brain and Development. 2019;41(9):783-9.

4. Kalscheuer VM, Tao J, Donnelly A, et al. Disruption of the Serine/Threonine Kinase 9 Gene Causes Severe X-Linked Infantile Spasms and Mental Retardation. American Journal of Human Genetics. 2003;72(6):1401-11.

5. Lyst MJ, Bird A. Rett syndrome: a complex disorder with simple roots. Nature Reviews Genetics. 2015;16:261-75.

6. Fehr S, Wong K, Chin R, et al. Seizure variables and their relationship to genotype and functional abilities in the CDKL5 disorder. Neurology. 2016;87(21).

7. Fehr S, Downs J, Ho G. Functional abilities in children and adults with the CDKL5 disorder. American Journal of Medical Genetics. 2016;170(11):2860-9.

8. Leonard H, Downs J, Benke TA, et al. CDKL5 deficiency disorder: clinical features, diagnosis, and management. The Lancet Neurology. 2022;21(6):563-76.

9. Mori Y, Downs J, Wong K, Heyworth J, Leonard H. Comparing Parental Well-Being and Its Determinants Across Three Different Genetic Disorders Causing Intellectual Disability. Journal of Autism and Developmental Disorders. 2018;48(5):1651-65.

10. Porter RJ, Meldrum BS. Antiseizure Drugs. In: Katzung BG, Masters SB, Trevor AJ, editors. Basic & Clinical Pharmacology (12th Edition). USA: McGraw-Hill; 2012.

11. Olson HE, Daniels CI, Haviland I, et al. Current neurologic treatment and emerging therapies in CDKL5 deficiency disorder. Journal of Neurodevelopmental Disorders. 2021;13(40).

12. Dale T, Downs J, Wong K, Leonard H. The perceived effects of cannabis products in the management of seizures in CDKL5 Deficiency Disorder. Epilepsy & Behavior. 2021;122(108152).

13. Lim Z, Wong K, Downs J, et al. Vagus nerve stimulation for the treatment of refractory epilepsy in the CDKL5 Deficiency Disorder. Epilepsy Research. 2018;146:36-40.

14. Marinus Pharmaceuticals. Clinical Study Report for 1042-CDD-3001. A double-blind, randomized, placebo-controlled trial of adjunctive Ganaxolone treatment in children and young adults with cyclin-dependent kinase-like 5 (CDKL5) deficiency disorder (CDD) followed by long-

term open-label treatment. Document date: Version 1.0: 12 Jan 2021. Version 2.0: 28 May 2021. 2021.

15. Pestana-Knight EM, Amin S, Bahi-Buisson N, et al. Safety and efficacy of ganaxolone in patients with CDKL5 deficiency disorder: results from the double-blind phase of a randomised, placebo-controlled, phase 3 trial. The Lancet Neurology. 2022;21(5):417-27.

16. Leonard H, Junaid M, Wong K, Demarest s, Downs J. Exploring quality of life in individuals with a severe developmental and epileptic encephalopathy, CDKL5 Deficiency Disorder. Epilepsy Research. 2021;169(106521).

17. Lamb YN. Ganaxolone: First Approval. Drugs. 2022;82:933-40.

18. Telethon Kids Institute. QI-Disability. Available at: https://www.telethonkids.org.au/our-research/brain-and-behaviour/disability/child-disability/qi-disability/ (last accessed 13 January 2023) 2023.

19. Marinus Pharmaceuticals. Clinical Study Report for 1042-0900. A multicenter, 26 week open-label proof-of-concept trial of Ganaxolone in children with PCDH19 female pediatric epilepsy and other rare genetic epilepsies followed by 52 week open-label treatment. 2019.

20. Hodges Jr JL, Lehmann EL. Estimates of Location Based on Rank Tests. The Annals of Mathematical Statistics. 1963;34(2):598-611.

21. Grabenstatter H, Leonard H, Kiriakopoulos E. CDKL5 Deficiency Disorder. Published 24 June 2022. Available at: https://www.epilepsy.com/causes/genetic/cdkl5-disorder#:~:text=Seizures%20in%20CDKL5%20deficiency%20disorder,received%20FDA%20approval%20for%20CDD (last accessed: 5 Jan 2023) 2022.

22. Leonard H, Junaid M, Wong K, et al. Influences on the trajectory and subsequent outcomes in CDKL5 deficiency disorder. Epilepsia. 2022;63(2):352-63.

23. Specchio N, Amin S, Aimetti A, Hulihan J. Extended Duration Safety and Efficacy of Ganaxolone for the Treatment of CDKL5 Deficiency Disorder: Preliminary Open-Label Extension Analysis (Marigold Study). Poster. American Epilepsy Socity, Virtual; 4-8 December.2020.

24. Chin RFM, Pickrell WO, Guelfucci F, Martin M, Holland R. Prevalence, healthcare resource utilization and mortality of Lennox-Gastaut syndrome: retrospective linkage cohort study. Seizure. 2021;91:159-66.

25. Hesdorffer DC, Tomson T, Benn E, et al. Do antiepileptic drugs or generalized tonic–clonic seizure frequency increase SUDEP risk? A combined analysis. Epilepsia. 2011;53(2):249-52.

26. Office for National Statistics. Death registrations, Populations and Age Standardised Rates, England 1981 to 2018. Release date: 23 January 2020. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/adho

cs/11168deathregistrationspopulationsandagestandardisedratesengland1981to2018 (last accessed 5 Jan 2023). . 2020.

27. Lo SH, Marshall J, Skrobanski H, Lloyd A. Patient and Caregiver Health State Utilities in Tuberous Sclerosis Complex. PharmacoEconomics Open. 2022;6(1):105-21.

28. Auvin S, Damera V, Martin M, et al. The impact of seizure frequency on quality of life in patients with Lennox-Gastaut syndrome or Dravet syndrome. Epilepsy & Behavior. 2021;123: 108239.

29. National Institute for Health and Care Excellence (NICE). TA808: Fenfluramine for treating seizures associated with Dravet syndrome. Published 8 July 2022. Available at: https://www.nice.org.uk/guidance/ta808 (last accessed 5 Jan 2023). 2022.

30. Ara R, Brazier JE. Populating an Economic Model with Health State UtilityValues: Moving toward Better Practice. Value in Health. 2010;13(5):509-18.

31. Hernández Alava M, Pudney S, Wailoo A. Estimating EQ-5D by Age and Sex for the UK. Published on 12 January 2022. Available at: https://www.sheffield.ac.uk/nice-dsu/methods-development/estimating-eq-5d (last accessed 5 Jan 2023). NICE Decision Support Unit. University of Sheffield; 2022.

32. U.S. Food and Drug Administration (FDA). FDA prescribing information for ZTALMY (ganaxolone) . Issued 3/2022. Reference ID: 4955025. Available at: https://www.accessdata.fda.gov/drugsatfda_docs/label/2022/215904s000lbl.pdf (last accessed 6 Jan 2023).

33. National Institute for Health and Care Excellence (NICE). TA614: Cannabidiol with clobazam for treating seizures associated with Dravet syndrome. Committee papers. Published 11 November 2019. Available at: https://www.nice.org.uk/guidance/ta614/documents/committee-papers (last accessed 6 Jan 2023). 2019.

34. Lagae L, Irwin J, Gibson E, Battersby A. Caregiver impact and health service use in high and low severity Dravet syndrome: A multinational cohort study. Seizure. 2019;65:72-9.

35. Mangatt M, Wong K, Anderson B, et al. Prevalence and onset of comorbidities in the CDKL5 disorder differ from Rett syndrome. Orphanet Journal of Rare Diseases. 2016;11:39.

36. National Institute for Health and Care Excellence (NICE). NICE health technology evaluations: the manual. Published 31 January 2022. Available at: https://www.nice.org.uk/process/pmg36/chapter/introduction-to-health-technology-evaluation (last accessed 5 Jan 2023). 2022.

37. Office for National Statistics. Birth characteristics in England and Wales: 2017. Release date: 10 January 2019. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/bul

letins/birthcharacteristicsinenglandandwales/2017#average-ages-of-mothers-and-fathers-of-all-babies-have-continued-to-rise (last accessed 6 Jan 2023). 2019.

38. Office for National Statistics. Dataset Births by parents' characteristics. Release date: 13 January 2022. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/datasets/birthsbyparentscharacteristics (last accessed 6 Jan 2023). 2022.

39. Jones K, Burns A. Unit Costs of Health and Social Care 2021, Personal Social Services Research Unit, University of Kent, Canterbury. Available at: https://www.pssru.ac.uk/project-pages/unit-costs/unit-costs-of-health-and-social-care-2021/ (last accessed 6 Jan 2023). 2021.

40. R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.r-project.org/ (last accessed: 6 Jan 2023).

Appendix A: Detailed summary of HL shift implementation error

In Microsoft Excel, the function [redacted] was used by the company to simulate the distribution of SF for ECM patients (by pooling across arms at baseline, see Section 4.2.6.1). The [redacted] function in excel takes arguments for the desired value, [redacted]. The company entered these parameters into the function in a table ranging from 0 to 400 seizures (see Section 6.1.4 for the EAG's amendments to this).

To then simulate the distribution of SF in the GNX-treated cohort, the company [redacted]. However, applying the HL shift as a multiplier directly to [redacted] was not the same as applying it [redacted]. The HL estimate was based on [redacted], and so did not apply in this manner. The % HL estimate can, however, be applied correctly to lognormal distributional parameters using the [redacted]:

[redacted]

The HL shift estimate represents a 'shift' or compression/expansion of the SF distribution in the horizontal direction [redacted] and should therefore be associated with that same change in mean, median and standard deviation. That is, the mean, median and standard deviation of a lognormal distribution should all be reduced by approximately 27.08% using the marigold 17-week HL, or 29.31% using the Marigold maintenance period HL. In simple terms and functional form, the GNX distribution should simply be based on SF values with the % reduction applied:

$$f(SF_{ECM}) = f(SF)$$
$$f(SF_{GNX}) = f\big(SF * (1 - HL)\big)$$

So, for a lognormal distribution, it follows that:

[redacted]

Simulations were performed in the statistical software *R*, [40] using one million iterations of a lognormal distribution with the parameters provided by the company. An HL value of 27.08% (per the company's original base case analysis) was used and compared the SF distribution:

- For the ECM arm
- With the company's application of the 27.08% HL
- With the EAG corrected application of the 27.08% HL

The distributional characteristics of the simulation results were then compared to the Marigold baseline data and the results of the HL shift estimate. Note that no upper limit was placed on SF in these draws from the distribution. To align values with those in the company submitted cost-effectiveness model, the resulting draws could be filtered down to only those of 400 or under and the process repeated (to truncate the distribution as it has been truncated by the company, see Section 6.1.4). A simulation exercise conducted by the EAG demonstrated that the company's implementation led to an unambiguous overestimated treatment effect.

Box 1: Simulation exercise proving applicability of product rule to lognormal distribution

```
# simulate the distribution of SF per ECM with 10^6 iterations [redacted]
its         <- 1E6
ecm_meanlog <- [redacted]
ecm_sdlog   <- [redacted]
hl          <- [redacted]
set.seed(987321)

# ecm distribution and characteristics:
ecm_sim    <- [redacted]
ecm_mean   <- mean(ecm_sim  ) # [redacted]
ecm_sd     <- sd(ecm_sim  )   # [redacted]
ecm_median <- median(ecm_sim) # [redacted]

# apply treatment effect per company:
gnx_sim_company    <- [redacted]
gnx_company_mean   <- mean(gnx_sim_company  ) # [redacted]
gnx_company_sd     <- sd(gnx_sim_company  )   # [redacted]
gnx_company_median <- median(gnx_sim_company) # [redacted]

# apply treatment effect per EAG (i.e. product rule):
gnx_sim_eag    <- [redacted]
gnx_eag_mean   <- mean(gnx_sim_eag  ) # [redacted]
gnx_eag_sd     <- sd(gnx_sim_eag  )   # [redacted]
gnx_eag_median <- median(gnx_sim_eag) # [redacted]

# Calculate percentage changes to demonstrate alignment with HL shift estimate.
# Simple function to report tidy % change results to desired decimal places:
f_pr_chng <- function(new, orig, dp=2) {
  change <- new - orig
  return(paste0(round((change / orig) * 100,dp),"%"))
}

# company implementation of treatment effect. Highly optimistic:
f_pr_chng(gnx_company_mean  , ecm_mean)   # [redacted]% change in mean SF
f_pr_chng(gnx_company_sd    , ecm_sd)     # [redacted]% change in s.d. SF
f_pr_chng(gnx_company_median, ecm_median) # [redacted]% change in median SF

# EAG corrected implementation of treatment effect. Slightly optimistic:
f_pr_chng(gnx_eag_mean  , ecm_mean)   # [redacted]% change in mean SF
f_pr_chng(gnx_eag_sd    , ecm_sd)     # [redacted]% change in s.d. SF
f_pr_chng(gnx_eag_median, ecm_median) # [redacted]% change in median SF

# The EAG corrected method is therefore within 1% of HL estimate on all measures,
# whilst the company implementation more than doubles the treatment effect.
```