Ganaxolone for treating seizures caused by CDKL5 deficiency disorder in people 2 years and over [ID3988]

A Single Technology Appraisal

EAG Review of Company's Response to Technical Engagement Response

---

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

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.

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 TE response is linked to ERG report 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]. Peninsula Technology Assessment Group (PenTAG), 2023.

1. INTRODUCTION

This document provides the External Assessment Group's (EAG's) critique of the company's response to the key issues contained within the EAG's report, within the technical engagement (TE) period as part of the National Institute for Health and Care Excellence (NICE) appraisal of ganaxolone for treating seizures caused by CDKL5 deficiency disorder in people 2 years and over [ID3988].

The company has provided updated clinical effectiveness results from the single-arm, open-label extension (OLE) of its pivotal trial, Marigold (with data to 2-years), and has made several modifications to its economic model. A summary of the company's response is provided in Section 2. Each of the issues outlined in the EAG's report are discussed in further detail in Section 3. The EAG's critique of any additional evidence is provided in Section 4. Finally, the EAG's revised base-case analysis is described in Section 5.

2. Overview of company's technical engagement response

The company presented an updated economic model including a revised base-case analysis. The company's revised base-case analysis is presented in Section 2.2.

In its updated base-case, the company has accepted several errors highlighted by the EAG and has not provided any response or rebuttal on these points (i.e., the company's revised base-case analysis has been integrated within the EAG's edited model, including switches implemented by the EAG). The EAG therefore assumes that any changes implemented by the EAG that were not explicitly discussed in the company's response have been accepted, and so are not discussed further within the EAG's response.

2.1. Additional evidence provided by the company

In summary, the company has included the following within its response in relation to the key issues described in the EAG's report:

- **Key issue 1:** Analysis based on longer-term data from the MARIGOLD open-label extension (OLE) study
- **Key issue 2:** Further justification and evidence in support of the company's chosen model structure
- **Key issue 3:** Clarification on the definition of SF as measured in MARIGOLD, and therefore the meaning of the treatment effect estimates
- **Key issue 4:** Updated approach to modelling the health-related quality of life (HRQoL) impact of reducing SF
- **Key issue 5:** Discussion and analysis concerning duration of treatment effect and wastage
- **Key issue 6:** Further justification for the severity modifier relevant to this appraisal

The company's response also highlighted some additional issues that the EAG considered necessary to provide commentary to assist the committee as part of its decision making:

- The company has introduced a stopping rule for GNX, which assumed that all people who do not exhibit a 30% reduction in SF relative to their baseline SF by 6 months would discontinue treatment (see Section 4.1)

- The company has provided additional evidence regarding the length of inpatient stays (see Section 4.2)

All other changes made by the company to its preferred base-case analysis are relatively minor, and therefore are not discussed further in the EAG's response.

Alongside the company's response, the company also provided an updated cost-effectiveness model. The EAG appreciates the efforts made by the company to maintain the functionality implemented by the EAG to inform its report. However, the EAG was unable to fully revert the company's revised base-case analysis back to the EAG's preferred base-case analysis per its report. This is because the company has implemented several changes that compromise the original functionality of the model (e.g., changing specific input values without implementing a switch).

Nevertheless, the EAG was able to revert to its previous base-case analysis with a few small tweaks to specific input cells/ formulae, based on the following edits:

- Costs parameters, cells K34:K39 (revised costs)
- Clinical parameters, cells Q31, AA95, and AA96 (number of caregivers and utility values)
- Trace for GNX, cells M9 and M10 (rounding error)

2.2. Updated company cost-effectiveness results

The updated company base-case ICER is £[redacted]. However, as noted in the EAG's report, there are a number of important assumptions made by the company to obtain this ICER. Therefore, the EAG highlights the following ICERs associated with different settings and/or assumptions that were previously discussed within the EAG's report (but maintaining all other elements of the company's updated base case):

- Without applying the severity modifier to caregiver utilities, the ICER is [redacted]
- Without applying the stopping rule the ICER is £[redacted]
- With no severity modifier for caregivers or the stopping rule the ICER is £[redacted]

- Using Auvin *et al.* utilities instead of the Lo *et al.* utilities the ICER is £[redacted], and when using the bootstrapped combined average utilities (per company Attachment 3), the ICER is £[redacted].
- With the combined average bootstrapped utilities and no caregiver severity modifier, the ICER is £[redacted]

The EAG's updated base-case ICER is presented in Section 5, alongside a description of the changes made.

3. EAG REVIEW OF KEY ISSUES

In this section, each of the key issues described within the EAG's report are discussed alongside the company's TE response.

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

Summary of the key issue

The EAG considered there to be uncertainty in the clinical effects reported from the Marigold OLE (i.e., all clinical data greater than 17-weeks following treatment) due to a high rate of missing data and a risk of regression to the mean following treatment initiation.

Summary of the company response

In its response to TE, the company provided data for the 28-day change in major motor seizure frequency (MMSF) at 2-years in the ITT population using imputation of missing data. In response to the EAG concerns about a potential regression to the mean effect, the company stated that it did not have access to historical seizure data in participants in Marigold to provide more insight into if / how many participants were experiencing an increase in SF prior to participation in the trial. However, the company provided two additional justifications for the absence of this concern:

- Firstly, the company noted that the baseline period was six weeks, which it argued would mitigate the risk that participants were experiencing a sudden increase in SF.
- Secondly, the company suggested that those participants who switched from placebo to GNX after 17-weeks showed a similar pattern in a reduction in SF, which supported the absence of a regression to the mean effect.

EAG response

The EAG considered that the analysis provided by the company to account for missing data showed that outcomes in the OLE were being affected by attrition bias, and that it was likely that this would affect all OLE outcomes reported in the original CS. The company did not provide updated analyses for other OLE trial outcomes after imputing missing data and therefore the EAG considered that the other OLE outcomes reported in the original CS that did not account for missing data should be considered flawed.

In the original CS, the difference in 28-day MMSF between GNX and placebo was reported to increase over time, from a difference of -27.1% during the 17-week double blind phase to more than a 50% reduction from baseline after 12 months in the OLE. In this updated submission, the company reported the change in seizure frequency after imputing data for missing participants using a last observation carried forward (LOCF) approach (i.e., the last available measurement of SF assessed before the participant discontinued the trial was used at all subsequent timepoints). The results showed that the difference in SF did not increase but was reasonably consistent with the difference recorded at 17-weeks (-29.3%). These data therefore suggested that the median reduction in MMSF shown at 17-weeks could be maintained for up to 2-years. As other anti-seizure medication (ASM) used to treat people with CDKL5 deficiency typically only results in a reduction lasting several months, the stability of the GNX treatment effect could therefore be much improved. However, the EAG cautioned that the LOCF approach may be considered an optimistic approach, for example if any waning of the treatment effect was not evident in participants' last observation or if people were experiencing a benefit of treatment and discontinued for other reasons (e.g. toxicity). In such cases, the treatment effect measured in the last observation was assumed to be maintained throughout the OLE follow-up (i.e. up to 2-years), which may not reflect reality. It was therefore plausible that the MMSF reported using the LOCF approach may be optimistic.

In the original CS, the company reported the number of participants who experienced ≥25% and ≥50% reduction in 28-day MMSF. In the updated submission, the company reported the mean change in MMSF in those who exhibited at least a 30% reduction in MMSF and used this threshold in a new stopping rule for GNX (see Section 4.1). The company did not provide a rationale for the use of this threshold, and the analysis appeared to be post-hoc. As noted in the EAG report, clinical advice to the EAG was that a threshold of 50% was more typically used in epileptic conditions. In the new addendum, the company reported that amongst those participants who experienced ≥30% reduction in MMSF, the median reduction in MMSF was [redacted]% (95% confidence intervals or another measure of variance were not reported). The MMSF in those who did not experience a 30% reduction in MMSF was not reported, and presumably included people with no change, no clinically meaningful change, or an increase in seizure frequency.

With regard to the risk that data were affected by a regression to the mean effect, the EAG did not consider that the company had been able to resolve this issue with the available data. The company argued that a 6-week baseline period could have reduced the risk of a regression to

the mean effect as acute increases in seizure frequency prior to trial entry may have resolved before baseline; however, as noted in the EAG report, this would depend on the typical duration of exacerbations in seizures, and the EAG was unaware of any data to inform this. The EAG understood that the duration of exacerbations may vary greatly across people with CDKL5 deficiency, and so a 6-week period may not be sufficient time for some.

The company further argued that those in the PBO/GNX arm experienced a decrease in MMSF after treatment that was comparable with the GNX/GNX arm, and that this suggested that there was no regression to the mean effect. However, the EAG disagreed and did not consider the single-arm design of the OLE allowed for this be demonstrated. During the DB phase, both arms showed a reduction in MMSF and the difference between arms could be considered to represent the treatment effect of GNX. Without a control arm during the OLE, an unknown proportion of the reduction in MMSF could be caused by factors other than the treatment effect, including a regression to the mean effect. The timing of any regression to the mean effect, including whether this is more likely earlier or later in the OLE follow-up, is related to the typical duration of SF exacerbations, which as noted is currently unknown. The EAG also considered that the calculation of MMSF used in the CS, which converted absolute SF into a median percentage reduction over a 28-day period, made it difficult to interpret any effect of time on SF.

The EAG conclusion on the clinical effectiveness of GNX remained similar to that in the EAG report; i.e. a minority of people with CDKL5 deficiency may experience a meaningful reduction in MMSF following treatment, and new data suggested that this benefit may be sustained for 2 years, which was substantially longer than other ASMs. However, the magnitude of this benefit was somewhat uncertain, given the potential for a natural regression to the mean effect after treatment and the possibility that the missing data analysis in the OLE may be optimistic. The evidence did suggest that the majority of people who receive GNX would not experience a meaningful benefit in seizure frequency. Finally, on the basis of the missing data analysis provided by the company, the EAG considered that other outcomes measured in the OLE that did not account for missing data were flawed, due to the now known attrition bias.

Key Issue 2: The company's model structure

Summary of the key issue

The company's model is a simple Markov state-transition model with two primary health states (alive and dead) which may not capture the full impact of the disease or treatment and may be considered atypical for NICE technology appraisals of genetic epileptic syndromes. The EAG

considered that other model structures could have been considered, but it was unclear to what extent an alternative structure might influence cost-effectiveness results.

Summary of the company response

The company acknowledges that its model is different to models developed for other conditions considered 'similar' to CDD (accepting that CDD has a number of unique features which differentiate it from other conditions, such as TSC, LGS, or DS). However, it explained that the sample size of the Marigold study precluded its ability to reliably construct a model similar to those used for other ('proxy') conditions. Relatedly, the company explained that clinical trials for these other conditions typically recruit larger samples compared with CDD.

In addition, the company explained that specifying a model structure that grouped patients into health states defined by SF (using bounds from other cost-effectiveness analyses in proxy conditions) would also be challenging. This is because the bounds of SF for proxy conditions would not translate well to a CDD population, since a considerable proportion of patients would fall into either the lowest or highest SF categories.

EAG response

The structure of the model remains similar following the company's changes. The previous comments made by the EAG on the company model structure therefore still apply, though the EAG acknowledges the limited data available to inform an alternative structure in the context of this appraisal in CDD.

A major modification to the company's model structure was the introduction of a stopping rule, centred on a response threshold of a 30% reduction in SF. The EAG noted that the definition of response used by the company did not align with that of the MARIGOLD study secondary endpoint (≥50% decrease in SF), or the additional analyses presented in the study CSR (25% and 75% thresholds). Further, the company provided no clinical justification for or clinical testimony in support of a 30% threshold. Therefore, although the EAG supported the use of a stopping rule in principal following clinical advice noted in the EAG report, it had some concerns, which are discussed further in Section 4.1. The EAG was also concerned that an analysis of the HL shift for patients that did not achieve a 30% reduction versus the placebo arm has not been presented. If this showed a non-zero or even worsening shift among those patients, then the model was biased in favour of GNX.

Key Issue 3: Application of seizure frequency

Summary of the key issue

The EAG identified a number of assumptions imposed by the company to reflect SF within its model. These included the decision to capture primary seizures only (i.e., secondary seizures excluded), that the distribution from Marigold reflects UK clinical practice, would not change over time, a treatment effect would apply instantaneously, and that SF distribution was well represented by a lognormal distribution. In addition, the EAG highlighted an apparent error in the application of the treatment effect based on the product rule of logarithms.

Summary of the company response

The company provided analysis of additional data from the OLE of the Marigold study to further support the estimation of SF in the long-term (discussed further in the EAG's response to Key Issue 1). With respect to the instantaneous application of treatment effects, the company agreed with the EAG that the titration and maintenance periods should be modelled separately, but preferred to apply these effects from cycle 2 in the model (i.e., start of the 'maintenance period'), as opposed to from cycle 4 (i.e., approximately Week 17, per the Marigold outcome measure).

Outside of these points, the company provided further information which related mostly to different aspects of the model (e.g., treatment duration) and so these are discussed separately (see Section 4.1).

EAG response

The company did not explicitly confirm in its response if it accepted each of the changes imposed by the EAG within its model linked to SF. However, inferring from the company's revised base-case analysis, the EAG understood that the company accepted its revision of the following settings within the model related to SF:

- Normalised SF distribution densities
- Corrected application of treatment effect (product rule of logarithms)
- Use of EAG's area-under-the-curve function to estimate SF distribution

As noted previously, the company suggested that the full estimated treatment effect should be applied from cycle 2 (week 8) rather than linearly interpolated from baseline to week 16 (to get as close as possible to week 17 per the MARIGOLD evidence). The company explained that

this fit the way that the trial endpoints were calculated more accurately. As stated by the company in its response to key issue 3, the SF quantity at 17 weeks was in fact calculated as *total* seizures over a 17-week period, divided by days ($17 * 7 = 119$ when there is data for each day) and then multiplied by 28. Consequently, the data on which the treatment effect was estimated was in fact the % change in *total* seizures over a 17-week period (with a multiplier of $28/(17 * 7) = 0.235$ applied to it for complete daily data), and *not* the expected change in 28-day SF.

In light of this explanation by the company, the EAG agreed with the company that application of the full treatment effect from cycle 2 was likely to be more appropriate considering that the underlying data was for total seizures over 17 weeks (with a $28/119$ multiplier applied to it) and not per 28-day period.

Key Issue 4: Utility values

Summary of the key issue

Utility values were used to inform estimates of QALYs within the company's model, taken from published vignette studies in proxy conditions. These studies were subject to a number of limitations and were important drivers of the cost-effectiveness results since GNX was modelled to only impact quality and not length of life. Utility values impacted estimates of QALYs for both patients and their caregivers.

Summary of the company response

The company explained that the most suitable source for utility values was the study that best reflected the experience of the CDD population, regardless of how consistent this source was with the other aspects of the company's model (e.g., resource use). Ultimately, the company maintained its preference for the utility values reported by Lo *et al.*, which it considered to be the most suitable source to inform the model. This was based on the following key points:

- The types of seizures experienced by patients with TSC was expected to reflect the experience of a CDD population more closely, versus the participants considered by Auvin *et al.* (people with DS and LGS)
- Estimates of SF for LGS patients were based only on drop seizures in the study by Auvin *et al.*, whereas participants in the Marigold study reported different types of seizures

- The caregiver utilities reported by Auvin *et al.* represented a relatively small number of SF ranges (80 or 110 seizures per month), versus Lo *et al.* which reported four different categories)

In the company's response, it stated: *"*[Company] *have added accuracy to the baseline seizure-free-day (SFD) distribution, to match the Marigold patient level data (previously it was assumed patients are gathered around the mean within the same SFD class)."* (Company's response to Key Issue 2, p.6). While this was not fully explained by the company within its TE response, the EAG understood that the company had undertaken the following analysis:

- In the Auvin *et al.* study, utility values were reported based on SF and the number of SFD within a 30-day period. The SFD categories ranged from 1 (i.e., at least one seizure per day except for 1 day within a 30-day period) to 30 (i.e., no seizures within a given 30-day period)

- Previously, the company assumed all patients with SF between 45 and 130 would have 9 SFD per 30 days. Patients with SF of 20 were assumed to have 12 SFD per 30 days (i.e., 30 – 12 = 18 which is less than 20, compared with 30 – 9 = 21 which is greater than 20), and patients with SF of 0 were assumed to have 30 SFD per 30 days

- The company revised its application by calculating a weighted average of SFD per SF category to re-estimate utility values based on Auvin *et al.* for use within the cost-effectiveness model

- In addition, the company included the option to extract the average number of additional SFD for patients receiving GNX relative to ECM ([redacted]), and estimated a weighted average based on these patients obtaining the highest utility value, whereas all other patients were assigned a value based on SF per the company's original approach

Finally, the company's response also described a further analysis performed to produce alternative utility values: *"In consideration of the EAG´s request we now also present the results from an alternative modelling approach based on microsimulation (bootstrapping) with individual patient data. However, of note even this approach is likely highly conservative for ganaxolone due to the high ceiling and floor effects mentioned above. QALY gains are minimal and in many cases state definitions preclude any QALY gains being demonstrated, even where clinically significant reduction in seizures was experienced. Despite this, the bootstrapping approach largely supports a lower, more narrow range of ICERs."* (Company's response to Key Issue 2,

p.6-7). Some further description of the analysis undertaken was provided in Attachment 3 alongside the company's TE response.

EAG response

The EAG acknowledged the points raised by the company concerning the most suitable source of data to populate the utility values within the model. However, for completeness, the Lo *et al.* study also suffered from a number of limitations, which the EAG explained within its report. Notably, Lo *et al.* did not include an estimate of utility for patients that achieve seizure-free days (SFD) which the EAG considered an important differentiator between the two sources (with SFD noted as an important driver of utility by clinical experts that advised the EAG).

The EAG considered that both options (Lo *et al.*, and Auvin *et al*.) were subject to important limitations, and that neither study exhibited preferred characteristics 'across the board' when considering their applicability to the cost-effectiveness model used in this appraisal. Put another way, each study had its own merits, and both may be suitable to aid decision making. As the EAG stated within its report, utility values from a CDD population specifically would be preferred, as would utility values not based on a vignette study. Nevertheless, the EAG acknowledged the challenges associated with eliciting utility values for people with CDD and their caregivers and considered the ability to explore different options to be helpful for the committee's decision making.

The company's revised approach to considering SFDs represented a re-analysis of the utility values from Auvin *et al.*, which the EAG noted led to broadly similar values – the lower values decreased slightly, whereas the higher values increased slightly. However, the overall impact on the ICER was that the QALY gain was increased by a relatively large amount. This was because the 'poorer' health states (determined by SF) were subjected to a lower utility (favouring GNX, relative to the previous approach), and the 'better' health states (again, determined by SF) were subjected to a higher utility (again, favouring GNX, relative to the previous approach). In other words, making this edit led to a greater QALY gain and therefore a lower ICER, versus the original use of the Auvin *et al.* utilities.

The EAG noted that the company's method assumed that SFD were essentially independent of SF. For example, the same distribution of SFD was used to determine a weighted average utility for the 130 SF category as per the 45 SF category. As per its previous approach, the company accounted for implausible or unlikely combinations (e.g., a patient could not have 20 seizures per month, but also have only 1 SFD per month). If data permitted, the EAG would have

preferred to see a different distribution calculated for each SF category (e.g., one would expect patients in the 130 SF category to be mostly grouped towards the lower end of the SFD distribution, versus the 40 SF category, for which patients may be spread more uniformly across the categories).

With regards to the weighted average approach to account for patients that achieved SFD while receiving GNX, the EAG highlighted that this approach was (to an extent) inconsistent with the company's choice to select a lognormal distribution to model SF, since the support for a lognormal model was $x \in (0, +\infty)$. In other words, the lognormal model cannot estimate a proportion of patients with SF = 0. As such, the company's revised approach represented a somewhat crude adjustment to account for this, as the area-under-the-curve for SF was still estimated to be 100% by virtue of specifying a strictly positive parametric model.

Acknowledging the limited detail provided by the company in its TE response concerning the source data for the distribution of SFD from Marigold (including for GNX responders), and that its revised application was still subject to a number of limitations, the EAG tentatively accepted this alternative approach as a likely more accurate (but still imperfect) estimation of utility values via the Auvin *et al.* study.

In relation to the bootstrapping analysis, while the EAG appreciated the efforts made to provide further analysis of the utility data, it was unable to determine the precise motivation for undertaking the analysis, or how its findings should be interpreted with respect to populating the model. This was because despite the additional work undertaken, the company maintained its preference for the utility values derived from Lo *et al.* Therefore, for this reason, and due to the EAG not having a clear understanding of why the analysis was undertaken, the EAG did not consider these values further.

With the above in mind, the EAG maintained its preference for the study by Auvin *et al.*, but considered that scenarios using either source may be helpful for decision making. The EAG's preference for Auvin *et al.* centred on two main reasons: (i) its arguments set out in its previous report (namely, that this option promotes consistency with the medical resource use estimates and mortality), and (ii) that out of both options this source yields the most conservative estimate of the incremental QALY gain, which given the extent of the structural uncertainty was prudent.

Key Issue 5: Miscellaneous model errors and unsubstantiated assumptions

Summary of the key issue

The EAG identified a number of model errors and assumptions that were unsubstantiated as part of its review, details of which were provided in its report. Where possible, the EAG addressed these by implementing fixes within the company's model, and by eliciting clinical expert opinion to sense check and update key assumptions that it felt were not adequately justified by the company in its submission.

Summary of the company response

The company specifically commented on two components of this key issue: (i) the application and durability of treatment effect, and (ii) the application of wastage for GNX. The first of these points was covered across Key Issues 1 and 3. As such, focus was placed here on the latter point concerning wastage. The company also updated its approach to capturing treatment discontinuation, which the EAG agreed with (i.e., that using exposure time as a basis to calculate discontinuation was appropriate), and so has no further comment.

In short, the company argued that the EAG's clinical expert derived wastage estimate of 10% is inappropriate, since it was not based on empirical evidence (i.e., it is based solely on clinical expert opinion). The company then provided what it considered to be a more realistic estimate of wastage in the region of 0.47% (which, given the lack of a cited source, the EAG understood to be its own opinion). However, because of there being no reports/evidence of spitting/ redosing issues in clinical trials or practice, and because of guidance in the SPC advising against redosing; the company maintained its preference for no wastage within the model.

EAG response

As the company did not provide any further commentary on the other aspects of this key issue, the EAG limits its response to the issue of wastage (with its views on treatment effects covered in its responses to Key Issues 1 and 3). The EAG acknowledged that its base-case assumption was not based on empirical evidence, and that the clinical expert opinion was not based on personal use of GNX (given that it is not currently available in routine NHS practice). However, the EAG also highlighted that the company's opinion was also not based on empirical evidence. Therefore, while the EAG deferred to the clinical opinion it received which suggested that in a 'real-world' setting some wastage was expected, the EAG acknowledged the company's view on this issue.

Ultimately, the EAG considered that this key issue requires further clinical insight to be resolved, given the importance of accurately capturing drug costs on cost-effectiveness results. In the interim, the EAG maintained its preference for 10% wastage, but considered further clinical insight to be critical to determining the most suitable assumption for decision making.

Key Issue 6: Application of a severity modifier

Summary of the key issue

The company applied a severity multiplier to both patient and caregiver 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 be applicable to caregivers. This has implications for the cost-effectiveness results since severity modifiers can substantially impact the magnitude of the overall QALYs gained.

Summary of the company response

The company reaffirmed its position that both patients and caregivers are *"living with the condition"* (terminology used within the NICE methods guide). Further information concerning the burden of disease on both patients are caregivers was provided to clarify the substantial impact CDD has.

EAG response

The EAG highlighted that the decision of how the severity modifier applies within a cost-effectiveness analysis submitted to NICE ultimately sits with NICE and not the EAG. However, it is the responsibility of the EAG to highlight any potential deviations from the NICE methods guide within a company's submission or cost-effectiveness analysis. A webinar hosted by NICE in early 2023 explained that the modifier should only be estimated based on patient shortfall and should only be applied to patients.[1] However, this precise wording is (to the EAG's knowledge) not explicitly stated within a documented report by NICE. Fundamentally, while the EAG accepted evidence presented by the company that carers of people with CDKL5 deficiency experience significant burden, it considered this to be a separate issue to the design and application of QALY modifiers, which are tied to evidence of societal preferences for spending. The EAG also maintained its position that one caregiver should be used to calculate shortfall,

[1] Centre for Health Technology Evaluation Methods Seminar 2023, hosted by NICE. Relevant section 1:06:33 onward. Available at https://www.youtube.com/watch?v=TVz7pT6DM-U

and that shortfall should be calculated separately from the patient's shortfall as a caregiver is a separate entity from a patient. The EAG could not comment further on this key issue since it required input from NICE.

4. EAG CRITIQUE OF ADDITIONAL ISSUES

Based on the company's response, the EAG has highlighted a number of additional aspects of the company's submission and/or model that warrant attention.

4.1. Treatment discontinuation and stopping rule

As part of its response, the company explained that its model had been updated to include a treatment stopping rule which was not included within its original model. In summary, the model now includes a decision point at six months where only patients that 'respond' are permitted to continue treatment, and those that do not respond are assumed to immediately discontinue treatment. The criterion for response is that patients must achieve at least a 30% reduction in SF at the end of the double-blind phase of the Marigold study (versus baseline). Implementation of this stopping rule in the company's model leads to a notable reduction in the incremental costs for the GNX arm, an increase in the QALYs gained, and therefore a reduction in the ICER.

The company does not provide clear justification for, or a detailed description of, the clinical decision-making mechanism surrounding treatment continuation criterion. The EAG notes that any stopping rule introduced for cost-effectiveness purposes means that some patients may have treatment withdrawn even though they are deriving a modest benefit, and therefore such rules must not be considered lightly. The formulation of a rule to take an effective treatment away from a patient, in the EAG's opinion, should always include careful discussion, refinement, validation and agreement with leading clinical experts to determine whether and how it would work in clinical practice.

Despite this, given the palpable uncertainty associated with modelling CDD, and the clear heterogeneity in both baseline SF and treatment effect at the individual level based on data from the Marigold study, a stopping rule introduced to improve cost effectiveness would appear reasonable. Clinical advisers to the EAG suggested a stopping rule may be suitable (but no specifics were discussed concerning a given rule, beyond the fact that 6 months would seem a reasonable time point). However, this is only if this rule is fully supported by clinicians and patient groups. In addition, it is also important to consider how the stopping rule could reasonably be adhered to in NHS practice.

In addition to the lack of evidence for a clinical consensus for the stopping rule, there are several issues with its implementation in the company's updated economic model. Within the

model, patients that do not respond are assumed to have the same SF distribution as the ECM arm. The EAG does not consider this to be appropriate, since some patients on the ECM arm may theoretically achieve a reduction in SF in accordance with the response criterion defined above. Therefore, an investigation of the SF in the ECM arm should be conducted to identify how many (if any) ECM patients achieved the 30% SF reduction, and the expected SF reduction among that group. If it is found that some patients in the ECM arm do in fact respond according to this criterion, then the implications for the structure of the cost-effectiveness model need to be carefully considered and the current model structure is likely biased in favor of GNX.

The EAG also highlights that response at 17 weeks from Marigold is applied at 28 weeks in the company's model. The EAG suspects this is unintentional, as it would seem inappropriate for a clinician to establish non-response according to response at week 17 and then wait another 11 weeks before discontinuing treatment, irrespective of whether or not the patient achieves response by week 28. Equivalently, the EAG does not understand why a patient achieving response at week 17 and then losing it before 28 weeks would, according to the company's rule, continue treatment despite being a non-responder under the company's definition.

Next, the EAG notices that the company have not presented HL shift estimates for those patients in the GNX arm that did not achieve a 30% reduction in SF at week 17 vs baseline. This could potentially bias the model results in favour of GNX because in reality non-responders per the 30% definition could have *increased* their SF over time. The company would need to perform this analysis for the week 17 data and also for the OLE for 17 week 30% non-responders to provide evidence that those non-responders have an approximately 0% shift in SF over time (to match the unsubstantiated assumption currently being imposed in the cost-effectiveness model).

The EAG notices that the inclusion of the stopping rule has led to an increase in the total QALYs gained for GNX (enabled: + QALYs, disabled: + QALYs), while also leading to a substantial reduction in incremental costs (enabled: +£, disabled: +£). This is despite discontinuing approximately 50% of patients at 28 weeks, leading to and drop in HRQoL (to for patients and to for caregivers, which is then multiplied by 1.8 to , a larger drop than for people with CDD). The underlying reason for this is that the company's updated base case associates a % reduction in SF (the reduction being applied to the SF reduction 30% responders at 28 weeks onwards) with a increase in HRQoL for patients (from to) and a

[redacted] increase in HRQoL for caregivers (from [redacted] to [redacted]) which is then multiplied by 1.8 by the company to [redacted] (i.e., a greater HRQoL increase than for people with CDD). As approximately [redacted]% of patients are discontinued at this point, and the magnitude of the utility gain from the higher effect is larger than the utility loss from losing all effect for non-responders, the net effect is that more QALYs are generated on the population level. This is demonstrated in Figure 1, which shows the expected HRQoL of one patient and one carer over time alongside the proportion of survivors remaining on treatment (in the company's revised base-case).

Figure 1: Expected utility and time on treatment over time - company revised base case

[redacted]

The EAG would expect that inclusion of the stopping rule should theoretically lead to a reduction in the overall lifetime QALYs accrued by patients in the GNX arm, since some patients that achieve some benefit but not enough to be considered 'responders' would discontinue treatment and derive no further benefit (in this case, reverting to the ECM level). Ultimately, the EAG expects this finding is at least partly due to the assumption concerning outcomes achieved by non-responders being assumed to be equivalent to the full ECM arm. This is because the effect for 30% responders is [redacted]%, whilst the effect for all patients is 29.31%, implying that the

effect in 30% non-responders (approximately [redacted]% of the population) is very small, if not negative.

Finally, the larger HL shift in 30% responders at 17 weeks is an indicator that the *shape* of the SF distribution could be affected by the stopping rule. This may then affect how appropriate the previous decision by the company to model SF using a lognormal distribution remains. To alleviate this concern, the company would need to investigate and report the shape of the SF distribution among the subgroup of 17 week 30% responders (should that ultimately be determined to be an appropriate threshold for discontinuing treatment by clinical experts).

Within the timeframe available for the EAG to perform its critique of the company's revised model, it was not possible for the EAG to produce an alternative application of the company's stopping rule. However, based on an exploratory analysis, the EAG presents a heat map to investigate this relationship further (Figure 2).

Figure 2: Heat map for impact of stopping rule on incremental costs and QALY gain

In conclusion, when considering the apparent lack of face validity exhibited by the company's stopping rule application, the EAG does not consider this scenario to be suitable to inform

decision making. The EAG highlights that clinical expert opinion and input from NHS England is required to understand the feasibility of implementing the proposed stopping rule. Without the stopping rule, the ICER in the company's updated base case (without any other EAG alterations or corrections and without the split discontinuation rate post-stopping rule) is [redacted]. Therefore, if the stopping rule proposed by the company is not followed in clinical practice, GNX is unlikely to represent a cost-effective use of NHS resources.

4.2. Justification for long-stay hospitalisations

The company provided additional evidence from the international CDD registry, reporting that the median length of stay (LOS) related to CDD hospitalization events was [redacted] days. When examining the distribution of LOS, the company also report that [redacted]% of CDD related hospital stays had LOS ≥ 2 days, which the company stated constitutes a long-stay in the NHS reference costs. Responding to this additional evidence, the company then updated its base case to calculate a weighted average of short stay and long stay hospitalization costs.

The EAG accepts the updated company's updated approach to hospitalization costs based on the new evidence it has provided. The approach to incorporating this new evidence appears reasonable and the decision to calculate a weighted average appears fair. However, there are two issues which the EAG would like to raise before considering this issue wholly resolved.

The first potential issue is whether the international CDD registry data is representative of the way that CDD patients are treated in the UK. For instance, if the UK approach involves facilitating more at-home care than other countries, including the means to prevent longer inpatient stays or discharge patients from hospital more quickly. Consequently, the EAG suggests that the company presents simple subgroup analysis of the ICDD hospitalization data specifically for UK patients and investigates whether this is different from the full international dataset. The full dataset appears large enough to permit subgroup analyses, and this may alleviate the concern that the EAG has that care for CDD may differ between countries.

The second and relatively minor issue is with NHS reference costs source. The NHS reference cost document used by the company does not report the median or mean length of stay associated with the codes applied by the company in the cost-effectiveness model (PRO2A, PRO2B and PRO2C). It is therefore unknown whether the long-stay codes in NHS reference costs are appropriate for long stays with a median duration of [redacted] days. However, as this is

unlikely to have a large impact on the cost-effectiveness results, the EAG accepts the approach the company has taken and the codes used.

In conclusion, the EAG is satisfied with the additional evidence provided by the company and subject to the presentation of UK subgroup analysis results for LOS accepts the revised company approach to incorporating hospital stays into the cost-effectiveness model.

5. EAG'S REVISED BASE-CASE ANALYSIS

Following the EAG's appraisal of the new evidence submitted by the company, and as explained in the appraisal of the key issues in the previous section, the EAG made several changes to the revised company basecase (Table 1).

Consistent with the EAG report, ICERs are presented both with and without the severity modifier for caregivers, pending NICE advice. Under the updated EAG base case, the absolute and proportional QALY shortfalls calculated for caregivers (i.e., separately to patients) were insufficient to warrant any severity modification, so the exclusion of a severity modifier for carers did not affect the (deterministic) results.

The EAG accepted the use of a treatment discontinuation rule in principle, though it was hesitant to accept the implementation within the cost-effectiveness model for the reasons discussed in Section 4.1. Consequently, ICERs are presented both with and without the stopping rule active: this resulted in an EAG base case ICER of £[redacted] with a stopping rule and £[redacted] without it (note: this was after fixing an error in columns BQ and BR of the "Trace Gan" sheet when using Auvin et al utilities, which was increasing utility for patients per the stopping rule even when the stopping rule was switched off). In both cases, it was unlikely that GNX represented a cost-effective treatment option for people with CDD at the relevant willingness to pay threshold.

Table 1: EAG adjustments to revised company base-case

Change made	Justification	ICER
Revised company base-case	N/A	£[redacted]
Change 1: Auvin *et al.* utilities	Insufficient justification to adjust the EAG position, Lo et al does not take seizure-free days into account, consistency of disease with mortality and HCRU model components	£[redacted]
Change 2: 10% wastage	Company provided no evidence to substantiate its claim of no wastage. EAG position remains the same as it is supported by clinical expert opinion.	£[redacted]
Change 3: Without severity modifier for caregivers	No published clear position from NICE on whether disease severity	£[redacted]

Change made	Justification	ICER
	modification applies to caregivers or not, and on what basis it should be calculated	
Change 4: Caregiver severity modifier calculated based on caregiver QALY shortfall not patient QALY shortfall	Caregivers are separate entities from patients, and so have their own QALY shortfall. Therefore, severity modification (as with all cases) should be based on their own HRQoL	£[redacted]
Change 5: Caregiver QALY shortfall based on 1 caregiver not multiplied by 1.8	QALY shortfall is per expected individual affected by the condition, not the total shortfall of a group of people	£[redacted]
Change 6: No stopping rule	Unclear whether the stopping rule was appropriate, timing issues (stopping rule at week 28 but evaluation at week 17), ubiquitous use of stopping rule assumed, lack of analysis on non-responders.	£[redacted]
EAG base case	Combining changes 1, 2, 4 and 5 is the revised EAG base-case with severity modification. As this results in a severity modifier of 1x for caregivers, the ICER is the same as the scenario without severity modification for caregivers. The scenario combining changes 1, 2, 4, 5 and 6 is the EAG base case with no stopping rule applied.	(1+2): £[redacted] (1+2+3): £[redacted] (1+2+4+5): £[redacted] (1+2+4+5+6): £[redacted]

Abbreviations: EAG, external assessment group; HRQoL, health-related quality of life; ICER, incremental cost effectiveness ratio; QALY, quality-adjusted life-year