
in collaboration with:

Erasmus School of Health Policy & Management

Maastricht University

# Belzutifan for treating tumours associated with Von Hippel-Lindau disease [ID3932]

## EAG response to TE submission including cost effectiveness sections

**Produced by** Kleijnen Systematic Reviews (KSR) Ltd. in collaboration with Erasmus University Rotterdam (EUR) and Maastricht University

**Authors**
Susan O'Meara, Reviews Manager, KSR Ltd, United Kingdom (UK)
Venetia Qendri, Health Economist, Erasmus School of Health Policy & Management (ESHPM), EUR, the Netherlands (NL)
Eline Krijkamp, Health Economist, ESHPM, EUR, NL
Jiongyu Chen, Heath Economist/Systematic Reviewer, KSR Ltd, UK
Xiaoyu Tian, Health Economist/Systematic Reviewer, KSR Ltd, UK
Mubarak Patel, Systematic Reviewer, KSR Ltd, UK
Rachel Croft, Information Specialist, KSR Ltd, UK
Lisa Stirk, Senior Information Specialist, KSR Ltd, UK
Nigel Armstrong, Health Economics Manager, KSR Ltd, UK
Isaac Corro Ramos, Health Economics Researcher, Institute for Medical Technology Assessment (iMTA), NL
Robert Wolff, Managing Director, KSR Ltd, UK

**Correspondence to** Susan O'Meara, Kleijnen Systematic Reviews Ltd
Unit 6, Escrick Business Park
Riccall Road, Escrick
York, YO19 6FD
United Kingdom

**Date completed** 25/10/2023

**Source of funding**: This report was commissioned by the National Institute for Health Research (NIHR) Evidence Synthesis Programme as project number STA 13/56/85.

**Declared competing interests of the authors** None.

Commercial in confidence (CiC) data are highlighted in blue throughout the report.
Academic in confidence (AiC) data are highlighted in yellow throughout the report.
Confidential comparator prices are highlighted in green throughout the report.
Any de-personalised data are highlighted in pink throughout the report.

# Table of Contents

**EAG's comments on company's Stakeholder Engagement Response From**

***Key Issue 1, 2 and 6: Implication of differences between intervention and comparator populations given interpretation of the MA that standard of care for most patients is immediate surgery, Misalignment between the decision problem and MK-6482-004 study populations; and between the latter and the UK target population, and There is mismatch between the population in the economic analyses and the population included in the sources of evidence used to inform such analyses***

The EAG note that the company recognise the difference between the MA population, which they maintain should be the same as the decision problem population, and the population of the MK-6482-004 study. They also, in the history of obtaining MA, show that a population description that would have been consistent with MK-6482-004 is "not requiring immediate surgery". As stated in the EAG report, this shows the large discrepancy, as opposed to "slightly different": Not requiring immediate surgery implies a much less severe stage of disease than surgery being "unsuitable or undesirable". As also stated in the EAG report, this also highlights the discrepancy between the populations implied by immediate surgery only occurring with standard care i.e., more like the DP or the MA, as opposed to no immediate surgery with belzutifan i.e., more like the MK-6482-004 study.

The company go on to argue that no surgery is required if belzutifan is given because belzutifan is "an effective therapy" and that the "onset of efficacy is rapid". This might make sense if no substantial harm might befall patients who have to wait to see if belzutifan is effective. However, as the EAG report stated, the company stated that surgery was "the only treatment option available to keep patients alive…". As stated in the EAG report, it cannot be true that such immediate surgery is required without belzutifan, but not with it. Indeed, since the population in the MK-6482-004 study is those not requiring immediate surgery, then active surveillance as comparator seems to be, in contrast to the company's assertion, entirely plausible. This would imply a change to the DP, which might not be consistent with the MA: if this is the case then, as stated in the EAG report, perhaps there needs to be some differentiation between patients who need lifesaving surgery and those who need it for symptom relief or progression prevention.

The company also refer to an expert elicitation exercise, a summary of which is presented in Appendix 2 and in response to which the EAG have already presented an addendum. Although the experts' responses have not been provided, this summary reveals nothing that would change the EAG critique. Therefore, key issues 1, 2 and 6 remain relevant.

Finally, the company discuss potential belzutifan benefits that are not captured in the current economic model. While these could be relevant, the lack of evidence remains a main concern. In this respect, the EAG agrees with the company that additional data collection might resolve or reduce some of the uncertainties associated to the remaining key issues.

***Key Issue 3: Potential risk of study selection bias resulting in possible omission of relevant comparator studies***

It is unclear to the EAG how "…a specific treatment that alone would not be representative of overall standard of care in UK clinical practice…" would not be informative, particularly given the lack of evidence for the clinical effectiveness for what the company are regarding as

standard of care i.e., immediate ablative procedures. Indeed, the only clinical effectiveness evidence that was submitted was the VHL Natural History study, where no immediate surgery or its sequalae were observed. This therefore remains a key issue.

***Key Issue 4: Lack of clear link between clinical and cost-effectiveness sections in relation to sources of data for comparator***

The company continue to misunderstand the purpose of the clinical effectiveness evidence and any ITC within that i.e., its importance regardless of its use in the economic analysis. Also, although the additional clarification regarding the various data sources is helpful, it is still unclear why the rates of events from the pre-surgery state could not be estimated from the Optum study. The company states: "…the outcomes calculated from this study (i.e., rates of events after 1st surgery) are not the same as the outcomes collected from the MK-6482-004 study (i.e. rates of events from pre-surgery)…", but it does not provide any further explanation. Therefore, this remains a key issue.

***Key Issue 5: Limitations in the indirect treatment comparison hinder the assessment of the effectiveness of belzutifan compared to standard of care***

The lack of pooling of all IPD remains a problem, particularly in the context of a method of adjusting for confounding that has been identified by the NICE DSU in TSD 18 as very likely to be unsatisfactory.

The company continue to not provide any objective evidence as to the most important variables to adjust for in the MAIC. They also make the spurious argument that there is unlikely to be confounding because the treatment effect estimated is too large: a very large treatment effect might suggest that no treatment effect is unlikely, but can say nothing about the size or direction of any confounding. Indeed, it is concerning that this very large treatment effect is estimated based on natural history data that does not include the sequalae of immediate surgery, which the company assume is standard of care, and which is at least party intended to reduce the risk of disease progression.

Limitation in the ITC therefore remain a key issue.

***Key Issue 7: The comparator data might not be representative for the UK***

Additional data collection from UK clinical practice might reduce the uncertainties associated to the comparator data, but, until then, this remains a key issue.

***Key Issue 8: Data to inform effectiveness in the belzutifan arm (MK-6482-004 trial) are either immature or unavailable***

The EAG would like to refer to the corresponding FAC response: Key Issue 8 refers to immaturity in the sense that there is uncertainty in the long-term extrapolations of treatment effectiveness in general. Again, this key issue might be resolved with additional data collection from UK clinical practice.

In addition, we would like to emphasize that statements such as "the alternatives [parametric distributions] explored by the EAG produce results that show belzutifan to remain consistently

cost-effective with the provided PAS" made by the company should be considered with extreme caution. Based on the data shown in Section 5.3.2.2 of the EAG report for example, the EAG would conclude that none of the parametric distributions should be considered reliable enough to support such statements.

***Key Issue 9: There is uncertainty in the derivation of the transition probabilities in the standard of care arm***

The EAG would like to thank the company for the additional clarification and scenarios. As above, there remains the issue that the data were not collected on the decision problem population.

***Key Issue 10: There is uncertainty in the implementation of time on treatment and treatment effect waning***

The EAG would like to thank the company for the additional clarification and refer to the corresponding FAC response. This remains a key issue.

***Key Issue 11: There is uncertainty in the derivation and implementation of health-related quality of life in the model***

The EAG would like to thank the company for the additional clarification and refer to the corresponding FAC response. The mismatch between the decision problem population and evidence used to inform health-related quality of life in the model remains a key issue.

Regarding the immediate health-related quality of life benefit for the belzutifan arm, the EAG would also like to refer to the corresponding FAC response: The EAG considers that this issue might have been resolved by including time to treatment response in the model and by linking the objective response level to time to response to calculate utility values in the pre-surgery, surgery, and event-free after surgery states. The EAG explored a scenario in which a fixed cut-off at the median time to treatment response was included to the QALY calculation (please refer to Scenario analyses set 6 in Section 6.1.2.6). The company also explored the impact of using fixed proportions at each response level in response to clarification question B21c.4 However, these scenarios were only exploratory, and their results should be interpreted with caution. The latter should be emphasised since the company seem to imply that implementing the median time-to-response into the QALY calculation would resolve this issue, whereas the EAG considers that this is not the case. In fact, we believe that in the current version of the model which includes the median time-to-response into the QALY calculation, the immediate effect of belzutifan is still present in the economic model. This can be seen for example by running the model for a time horizon of 0.9 years (approximately 11 months). This would be before the median time to response observed in the RCC cohort (11.11 months). If we remove immediate surgery from the SoC arm, it would be expected that before that time, no differences would be predicted between belzutifan and SoC. However, the model predicts [redacted] incremental QALYs in the RCC cohort in favour of belzutifan when the model is run under these settings.

***Key Issue 12: Cost-effectiveness analyses should be based on subgroup-specific parameters.***

The EAG would like to highlight that our point of view regarding key issue 12, as explained to the company in our responses to the FAC comments and during the TE call, has not changed. For convenience, these are summarized below.

**Subgroup specific parameters**

The EAG understands how the "subgroups" were defined. It is unclear however that the company is stating that the "marketing authorisation population should be assessed as a whole" when the whole submission is based on three different subgroups. Given that the evidence presented suggests that the clinical effectiveness or the disease severity may be different per subgroup, the EAG considers it to be more appropriate to use subgroup-specific parameters in this submission.

**Severity modifier**

We would like to express, again, that our method for calculating the QALY severity weighting is not methodologically flawed. However, if the company think otherwise, we would like to invite the company to formally challenge the methods by Versteegh et al. 2019 published in *PharmacoEconomics*.[1] We acknowledge that peer reviewed publications are not exempt from being flawed, but we also consider that in case it is, it should be formally proven.

We would also like to stress that we consider unacceptable the argument of using a tool based in the Netherlands, as opposed to an UK-based tool, to demerit our approach. As acknowledged by our UK colleagues in their recent publication, they also made available "*an R-Shiny online tool (https://shiny.york.ac.uk/shortfall)* ***inspired*** *by the iDBC platform of Versteegh et al. (https://imta.shinyapps.io/iDBC/)*".[2] Both tools therefore should be considered as appropriate to estimate QALY weighting. One of the main reasons why the QALY Shortfall Calculator tool has also been previously used in other EAG assessments conducted by KSR was that the iDBC tool was being updated to include more recent life tables and discounting, the latter to conform with the latest NICE methods.

In response to FAC comments we already mentioned that one of the differences with respect to the iDBC tool, but not the only one, is the use of Hernández Alava instead of Heijink for the UK value set. While the EAG agrees with the company that using Hernández Alava would be in line with the NICE reference-case, we would like to invite the company to conduct the severity analyses with both tools and check whether results are substantially different. As an example, we compared the company's PSA results obtained with the iDBC and the QALY Shortfall Calculator tools. Note that the only difference would be in the number of QALYs without the disease. These would be 18.02 with the iDBC tool and 18.15 with the QALY Shortfall Calculator tool using Hernandez-Alava (as reported in the company's model). The results for the RCC cohort are as follows:

- RCC cohort with iDBC tool: 42.7% for weight 1.2 and 57.3% for weight 1.7. Weighted ICER £[redacted].

- RCC cohort with QALY Shortfall Calculator tool: 42.2% for weight 1.2 and 57.8% for weight 1.7. Weighted ICER £[redacted].

Therefore, as it can be seen, the impact of using Hernández Alava instead of Heijink for the UK value set is minimal.

The EAG considers that the discussion on the severity of the condition should not be focused on whether one specific tool or another should have been used. We consider this irrelevant. The main EAG issue with the CS relates to the fact that the same severity weights were used for all three subgroups. The EAG believes this is incorrect.

In addition, the EAG considers that, if it is accepted that the estimated QALYs under standard of care are uncertain, then the estimated proportional and absolute shortfall should be considered uncertain as well. Regardless of which tool is used, the EAG considers that a fairer assessment of proportional and absolute shortfall is to account for this uncertainty, as there may be submissions that happen to have a deterministic QALY loss resulting in a QALY multiplier group that is not fitting to the entire sample. While it is understandable that this might cause some resistance with submissions where the deterministic QALY happens to correspond with the upper QALY multiplier, there will equally be cases conceivable where the opposite occurs. Both can be dealt with equally by accounting for the uncertainty in the QALY loss predicted by the model.

Furthermore, the fact that the updated NICE methods manual does not mention a particular methodology, in this case the severity-adjusted probability of being cost effective, should not prevent the EAG from using it. We would also like to clarify that, in the example provided by the company, interpreting the severity-adjusted probability of being cost effective as effectively equating to a QALY weight of 1.48 is indeed incorrect, since as the company correctly indicate, there are only three possible weights. The severity-adjusted probability of being cost effective should be interpreted in relation to the cost-effectiveness thresholds.

**EAG's comments on company's updated cost-effectiveness analyses included in the Stakeholder Engagement Response From**

***New model version***

A new version of the model was submitted after the Technical Engagement meeting. The company indicated that the following changes were made to the model:

- Specifications sheet:
  - Row 96: option to delay utility benefit from response achievement until median TTR – this was also incorporated in the updated company base-case.
  - Row 121: the PAS submitted to PASLU, described below, was added.
- DSA & PSA results sheets:
  - Results have been rerun with the updated base case assumptions & included discount. Both also included the weighted GB MA cohort results.
- Utility sheet:
  - Rows 15-18: the option to delay utility benefit from response achievement until median TTR (as per specifications sheet). The median TTR reported from MK-6482-004 for each cohort is used.

Based on these changes the ICERs were different, as shown below.

***PAS price***

The company submitted a patient access scheme (PAS) in the form of [redacted]. The cost effectiveness results presented by the company after the Technical Engagement meeting are based on the new version of the model, with the changes mentioned above, including thus the PAS for belzutifan.

The EAG would like to emphasise that the inclusion of a PAS price for belzutifan would not resolve the uncertainties associated to the key issues since these mostly relate to the lack of appropriate clinical data.

***Updated results***

Given the time constraints associated with this project, the EAG could not reproduce all tables with the updated cost effectiveness results in this document. Comments regarding the updated results are provided below:

- In general, statements regarding the cost effectiveness of belzutifan should be considered with extreme caution given the remaining uncertainties highlighted in the key issues above.
- We still disagree with the company in the way severity weighting was implemented: we consider that applying a weight of 1.7 for all cohorts is incorrect, regardless the

approach to estimating the proportional and absolute QALY shortfall (deterministic or probabilistic). For the pNET cohort a (deterministic) weight of 1.2 should be applied.

- With the most recent version of the model that the EAG has (NICE ID3932 STA Submission CEA v5.0 (CIC).xlsm) it was possible to replicate the ICERs presented by the company in response to Technical Engagement comments. Base-case ICERs (including belzutifan PAS and QALY weight) in the model: £[redacted], £[redacted], £[redacted] for the RCC, CNS Hb and pNET cohorts respectively.
- The updated base-case ICER for the pNET cohort, including the new PAS for belzutifan, and a deterministic QALY weight of 1.2 is £[redacted].
- The updated base-case ICERs for the RCC, CNS Hb and pNET cohorts, including the new PAS for belzutifan and the EAG probabilistic approach to the proportional and absolute QALY shortfall are £[redacted], £[redacted] and £[redacted], respectively.
- The impact of using the EAG's approach to severity weighting on the ICERs presented by the company in the scenario analyses should be similar to the one observed in the base-case analysis.

**EAG's preliminary comments on Appendix 2 of the company's Stakeholder Engagement Response From**

***General comment***

The EAG acknowledges that the company has presented the "Results from clinical expert elicitation & discussion". However, it appears that the experts' responses to the expert elicitation exercise have not been provided alongside this. This means that the company has just presented a summary of the discussion and the EAG cannot check if this summary indeed corresponds to actual statements made by the experts.

***Key validation point 1***

The experts' answers seem to confirm that the population in the MK trial is not the same as the label population. However, there is no actual mention of the MK trial – this confirms that surgical interventions would be harmful for those who would receive belzutifan. What it doesn't do is state that such surgical interventions should only be given in a world without belzutifan. In fact, it is confirmed under key validation point 2 that SoC without belzutifan would be surgery and that there would be no wait. This undermines the company's assertion that patients can wait until the outcome of belzutifan has been determined.

***Key validation point 2***

The company appears to have misunderstood the earlier EAG critique. The EAG did not suggest delaying immediate surgery in SoC, the issue being that belzutifan patients are also in need of immediate surgery, but they do not get it, since they wait until they respond to belzutifan. The question was more what happens to them until they respond to treatment, (assuming that they do respond). This undermines the company's assertion that one can wait until the outcome of belzutifan has been determined before undergoing surgery.

***Key validation point 3***

Some parts of this point are not clear. The experts indicated that patients would be classified as having stable disease but if patients do not respond, presumably it cannot be assumed that the tumour has stopped growing. The EAG agrees with the company when they said that non-response is accounted for in the transition probability from pre-surgery to surgery, however this is based on the MK trial, where we have a different population. It appears that there may have been a misunderstanding during the expert elicitation process as it makes no sense that patients can wait to see if belzutifan works if without it they need immediate harmful surgery and it makes even less sense that if they don't respond to belzutifan they somehow have been transformed into not needing surgery at all.

***Key validation point 4***

The EAG does not have anything new to add here but wishes to reiterate their previous point that more severe patients (as in the label population) should also have more surgeries also in the belzutifan arm. This continues to highlight the misapprehension that 'belzutifan eligible' patients are not the same as those who get SoC without belzutifan – we need to know what SoC is for the belzutifan eligible patients and if those patients wouldn't get immediate harmful surgery with belzutifan then they wouldn't with SoC.

***Key validation point 5***

The EAG does not have anything new to add here however, would point out again that the QoL study was conducted in a different population.

***Key validation point 6***

The EAG is not clear why the experts chose to refrain from making explicit comments relating to treatment of the metastatic population. Why is the same logic not applicable to the label population? It has only been studied in RCC patients following the MK trial, but conclusions have been generalised to other types of patients. This continues to highlight that belzutifan cannot be considered as a substitute for immediate surgery – it is to delay progression. The only way to reconcile no surgery with belzutifan is to assume that the belzutifan eligible population, which should be that of the DP, is not so severe that immediate surgery is required.

***Final point***

The EAG does not consider that the submitted material has any implications for changing assumptions in the CEA model.

**References**

[1] Versteegh MM, Ramos IC, Buyukkaramikli NC, Ansaripour A, Reckers-Droog VT, Brouwer WBF. Severity-adjusted probability of being cost effective. Pharmacoeconomics 2019; 37(9):1155-1163

[2] McNamara S, Schneider PP, Love-Koh J, Doran T, Gutacker N. Quality-adjusted life expectancy norms for the English population. Value Health 2023; 26(2):163-169