Belzutifan for treating tumours associated with von Hippel-Lindau disease

Technology appraisal guidance
Published: 16 October 2024

www.nice.org.uk/guidance/ta1011

Your responsibility

The recommendations in this guidance represent the view of NICE, arrived at after careful consideration of the evidence available. When exercising their judgement, health professionals are expected to take this guidance fully into account, alongside the individual needs, preferences and values of their patients. The application of the recommendations in this guidance is at the discretion of health professionals and their individual patients and do not override the responsibility of healthcare professionals to make decisions appropriate to the circumstances of the individual patient, in consultation with the patient and/or their carer or guardian.

All problems (adverse events) related to a medicine or medical device used for treatment or in a procedure should be reported to the Medicines and Healthcare products Regulatory Agency using the Yellow Card Scheme.

Commissioners and/or providers have a responsibility to provide the funding required to enable the guidance to be applied when individual health professionals and their patients wish to use it, in accordance with the NHS Constitution. They should do so in light of their duties to have due regard to the need to eliminate unlawful discrimination, to advance equality of opportunity and to reduce health inequalities.

Commissioners and providers have a responsibility to promote an environmentally sustainable health and care system and should assess and reduce the environmental impact of implementing NICE recommendations wherever possible.

Contents

1 Recommendations

1.1 Belzutifan is recommended with managed access as an option for treating von Hippel-Lindau (VHL) disease in adults:

- who need treatment for VHL-associated renal cell carcinomas, central nervous system hemangioblastomas or pancreatic neuroendocrine tumours, and

- when localised procedures are unsuitable or undesirable.

  It is only recommended if the conditions in the managed access agreement for belzutifan are followed.

1.2 This recommendation is not intended to affect treatment with belzutifan that was started in the NHS before this guidance was published. People having treatment outside this recommendation may continue without change to the funding arrangements in place for them before this guidance was published, until they and their NHS clinician consider it appropriate to stop.

Why the committee made these recommendations

VHL disease is a genetic condition that severely affects the quality of life of people with it, and their families and carers. The condition increases the risk of certain tumours developing, and surgery is the main treatment option. There are no licensed medicines for the underlying causes of VHL.

Clinical-effectiveness evidence from a small study suggests that belzutifan reduces tumour size. It also suggests that it increases the amount of time people have before their condition gets worse, but by how much is uncertain.

There are also uncertainties in the economic model, as well as assumptions that likely favour belzutifan. So, it is not clear what the most likely cost-effectiveness estimates are for belzutifan, and it cannot be recommended for routine use.

Belzutifan has the potential to be cost effective, but more evidence is needed to reduce the uncertainties. More evidence from the trial and NHS practice could help address

these. So, belzutifan is recommended for use with managed access.

2 Information about belzutifan

Marketing authorisation indication

2.1   Belzutifan (Welireg, Merck Sharp & Dohme) is indicated for 'the treatment of adult patients with von Hippel-Lindau (VHL) disease who require therapy for VHL associated renal cell carcinoma (RCC), central nervous system (CNS) hemangioblastomas, or pancreatic neuroendocrine tumours (pNET), and for whom localised procedures are unsuitable or undesirable'.

Dosage in the marketing authorisation

2.2   The dosage schedule is available in the <u>summary of product characteristics for belzutifan</u>.

Price

2.3   The list price for a 90-tablet pack of 40-mg belzutifan is £11,936.70 (excluding VAT; BNF online, accessed November 2023).

2.4   The company has a <u>commercial arrangement</u>. This makes belzutifan available to the NHS with a discount. The size of the discount is commercial in confidence.

3 Committee discussion

The [evaluation committee](#) considered evidence submitted by Merck Sharp & Dohme, a review of this submission by the external assessment group (EAG), and responses from stakeholders. See the [committee papers](#) for full details of the evidence.

Clinical management

The condition

3.1   Von Hippel-Lindau disease (from now, VHL) is caused by a mutation in the VHL gene. This gene is responsible for producing a protein that controls cell growth. A mutation in the gene can cause cells to grow abnormally, leading to cysts or tumours developing in different parts of the body, such as the kidneys, brain and pancreas. This can lead to renal cell carcinoma (RCC), central nervous system haemangioblastomas (CNS Hbs) and pancreatic neuroendocrine tumours (pNETs). The patient experts explained that the experience of living with VHL varies from person to person. Some people with the condition might only develop 1 or a few tumours in their whole life, while others might have multiple tumours in different organs. There are also a wide variety of debilitating symptoms depending on tumour sites. These include constant pain, loss of balance and motor skills, loss of vision, breathlessness, coughing, headaches, confusion, severe nausea and fatigue. Scans before appointments can cause anxiety and people also worry about disability caused by surgery. The clinical experts said that, with more effective treatment, there is potential for people with the condition to live longer and have a better quality of life. Also, the patient experts' statement highlighted that caring for a family with VHL is emotionally challenging and has a psychological effect. The patient experts explained that carers are often the family members of people with VHL. They also live with the constant worry that they carry the gene for VHL and may pass it on to their children and grandchildren. The committee understood that people with VHL often have difficulty doing day-to-day tasks, and fear surgery. Also, the condition can have a negative effect on self-esteem and cause relationship difficulties. Survival with VHL has improved over time. Life expectancy for men with VHL is about 67 years and for women is about 60 years. The committee noted that living with the

condition and caring for people with VHL is physically and emotionally challenging. It concluded that VHL is a highly heterogeneous condition, and has considerable physical and emotional effects from repeated surgeries and anxiety from scans.

Unmet need

3.2   Surgery and other localised procedures are the main treatment options for people with VHL, but sometimes they are not appropriate. The clinical and patient experts explained that there is an unmet need for effective new treatments for people with VHL when surgery is unsuitable. They explained that people often have repeated surgeries throughout their life, and this is primary or most common way to remove VHL tumours. Because of this, people may lose organ function including in their eyes, a part of or a whole kidney, or their pancreas. They may also develop neurological issues such as paralysis after spine or brain surgery or may need lifelong medical intervention such as dialysis if both kidneys are removed from multiple surgeries. The patient experts explained that avoiding multiple surgeries could greatly improve physical and mental wellbeing, and improve quality of life for both people with VHL and their carers. The committee noted that there is an unmet need for treatments that could improve outcomes and quality of life for people with VHL. It was aware that VHL-associated retinal hemangioblastomas, adrenal lesions, endolymphatic sac tumours, epididymal cystadenomas and other pancreatic lesions are also a significant challenge in VHL. But they are not included in the marketing authorisation of belzutifan. The committee concluded that the current unmet need could be partly addressed by belzutifan. This is because it has the potential to preserve or delay the loss of organ function and the associated morbidity in some people with VHL.

Existing treatment

3.3   The patient and clinical experts explained that there are no treatments that address the underlying cause of VHL. It is a heterogeneous condition and clinical management is usually done by a multidisciplinary team. They explained that VHL management starts with diagnosis through molecular genetic testing. This can be done through a known family history of VHL disease or after a new presentation

of a manifestation of the disease. Genetic counselling is a key component of this process. This is followed by regular surveillance by genetics services. Surveillance aims to detect tumours and monitor tumour growth. MRI or ultrasound examinations of the abdomen are done every 12 months for RCC tumours and pNETs. For CNS hemangioblastomas, MRI scans of the head are done every 12 to 36 months. If VHL associated tumours need treatment, interventions are generally based on a specific threshold. The clinical experts explained that RCC tumours are kept under surveillance until they reach a diameter of 3 cm, and pNETs until they have a diameter of 2 cm. At this stage, the risk of metastasis exceeds the benefit of organ preservation and surgery can be recommended. Surgery for CNS Hbs is normally needed when the tumours grow to a size that causes symptoms. The patient experts explained that standard care varies and focuses on preventing tumour growth and metastasis while preserving organ function. The clinical experts explained that surgery is highly effective in most cases for all VHL tumours, with the most benefit for VHL associated RCC. But surgery can result in organ loss after multiple surgeries or morbidity, depending on the primary VHL associated tumour. The clinical experts also explained that:

- VHL tumours can grow to an extent that localised procedures are unsuitable

- accumulated organ loss from repeated surgeries mean that more surgeries are no longer an option

- surgery may also be undesirable because of the risks involved, particularly for CNS tumours.

  When people would generally consider surgery to be unsuitable or undesirable depends on what the primary tumour site is, and the likely outcome or risk from the procedure(see section 3.4). The committee noted that managing VHL is highly complex, heterogeneous and individualised. It noted that the decision to have surgery, or to delay surgery and have active surveillance, is based on several factors. These factors include the type of tumour, its location, the individual's overall health, suitability, and need or desire for surgery. The committee also understood that belzutifan should only be prescribed if the clinician and person at risk agree that the benefits of treatment and the potential delay of surgery outweigh the risks.

Belzutifan marketing authorisation and positioning

3.4   The committee noted that the indication for belzutifan (see section 2.1) involves some subjectivity, especially for 2 criteria:

- when treatment is needed

- when localised procedures are unsuitable or undesirable.

    The committee added that this introduces challenges in clearly defining the population who should be eligible for belzutifan treatment, and at which stage of VHL. At the first meeting, the company clarified the definition of its indication and positioning in the treatment pathway. It explained that:

- when 'treatment is needed' refers to a need for surgery or a related procedure when tumours reach a certain size (for RCC tumours, a more than 3 cm diameter; for pNETs, a more than 2 cm diameter; for CNS tumours, that they cause symptoms)

- 'unsuitable or undesirable' refers to when localised procedures (surgery and radiotherapy procedures) would result in organ loss or severe functional deficits (see section 3.3).

    The clinical experts clarified that people who are fit enough and whose VHL is eligible for surgery would normally have the surgery because it is an effective option. They also explained that belzutifan will provide an option for people when a tumour reaches the treatment threshold, and they are waiting for the surgery. The experts explained that, in clinical practice, surgeries are not done immediately except for rare cases of certain CNS Hbs. This is because people and their tumours are monitored closely over time. The committee noted that there may be an interval between a tumour reaching the treatment threshold and the decision to proceed with surgery. It also noted that belzutifan is positioned for when people have had surgery for tumours that have reached the treatment threshold, and are having active surveillance for high-risk tumours that could lead to organ loss. At the first meeting, the committee noted that the company's marketing authorisation and positioning of belzutifan in the treatment pathway was subject to interpretation. The mismatch between MK-6482-004 and the marketing authorisation meant that MK-6482-004 could not be used to guide

implementation. So, the positioning of belzutifan within its marketing authorisation is likely to evolve with clinical experience.

After consultation, because of the lack of data available, the company did an expert elicitation survey to address the committee's concerns about population misalignment between MK-6482-004 and the marketing authorisation. Its clinical experts in the survey agreed that belzutifan would be used for people with VHL at risk of organ failure or loss, and for whom surgical interventions are no longer suitable. One clinical expert at the second committee meeting said that they consider belzutifan for treating VHL when the risks of further surgery outweigh the benefits. This would particularly be the case for CNS Hb tumours, when surgery or stereotactic radiosurgery is no longer a viable option. This is because of factors such as worsening risk and benefit ratio, and tumours at the craniocervical junction in CNS Hbs. The second expert thought that the trial population in MK-6482-004 with no immediate need for surgery (see section 3.7) was representative of the expected population in NHS clinical practice. The committee thought that the marketing authorisation had narrowed the population to people with the highest unmet need. This population could split into 3 subgroups on the precipice of organ failure:

- RCC: people about to lose their kidneys, which would result in full bilateral nephrectomy, end-stage renal disease and dialysis

- pNETs: people about to lose their pancreas, which would result in full pancreatectomy, brittle or type 3c diabetes and complications

- CNS Hbs: people with tumours when surgery is difficult or when there is a low or no chance of successful treatment without significant neurological complications or death.

The EAG highlighted some limitations of the company's expert elicitation survey. These included how the experts were selected, and the format and methods used for aggregating the response. The committee noted this and the differing opinions between clinical experts on when belzutifan will be used. It also noted considerable uncertainty around the interpretation of the belzutifan eligible population in light of the marketing authorisation. The committee concluded that this uncertainty could be resolved with further

data collection through managed access in the Cancer Drugs Fund to assess how belzutifan is used in clinical practice (see section 3.20).

Relevant population

3.5 The company's trial population was at a different position in the treatment pathway than that in NICE's final scope on belzutifan for treating tumours associated with VHL disease and the marketing authorisation. The marketing authorisation population consists of adults with VHL who need treatment for VHL-associated RCC, CNS Hbs, or pNETs, and for whom localised procedures are unsuitable or undesirable. In contrast, the company's main clinical trial (MK-6482-004) included people who:

- had at least 1 more measurable VHL-associated RCC greater than 3 cm (could have other tumours), and

- did not need imminent surgery and may have had VH-associated tumours in other organs.

  The company compared belzutifan with standard care. This comprised surgery for RCC, CNS Hb and pNET cohorts. The EAG noted that MK-6482-004 excluded people who had an immediate need for surgery for tumour treatment. It explained that this meant there was a misalignment between the marketing authorisation and MK-6482-004. The EAG further explained that there could be important differences between the trial, marketing authorisation and comparator populations. That is, people for whom surgery is deemed suitable (the comparator) are likely to be fitter than the marketing authorisation population. It also explained that people for whom surgery is deemed to be needed may have a greater tumour burden than people recruited to MK-6482-004. The committee noted that MK-6482-004 represented a population with different needs than the population in the marketing authorisation. The committee recognised that designing a trial in the population of interest would have been difficult for ethical reasons. But it thought that this issue severely limited the generalisability and applicability of the clinical-effectiveness evidence. So, at the first meeting, the committee was cautious in interpreting the results from

MK-6482-004.

After consultation, the company provided individual patient data from MK-6482-004, including baseline characteristics, medical history, tumours at baseline and previous treatments. It explained that it examined, in particular, baseline characteristics and prior treatment history to assess to what extent the population of the MK-6482-004 aligned with the marketing authorisation. It explained that 80% of people in the trial had both RCC and CNS tumours. The clinical experts' responses at consultation also suggested that CNS Hb involvement is highly influential. Also, it is present in around 70% to 80% of people with VHL, although few people would need immediate surgery for the CNS Hb involvement. The EAG noted that the populations outlined by both the company and clinical experts appeared identical. But it noted that people were not on the precipice of organ failure. The committee concluded that there were challenges in defining the relevant population. It noted that the uncertainty would persist until clinicians fully understand how belzutifan would be used in clinical practice. But it thought that it was appropriate to consider using belzutifan for people with VHL for whom surgery is unsuitable. The committee reiterated that the relevant population could be identified through managed access in the Cancer Drugs Fund.

Use of belzutifan

3.6 At the second meeting, the committee noted that belzutifan should be used in accordance with its marketing authorisation. This stipulates that belzutifan should be continued until disease progression or unacceptable toxicity. The committee thought that people with VHL may present with multiple primary tumours during their lifetime and may have multiple surgeries. In contrast to other cancers, if a person stops belzutifan because of disease progression in 1 tumour site, this will not mean that later belzutifan would not be effective for a new tumour in the same or different organs. The NHS England Cancer Drugs Fund highlighted that there was potential for problems with implementation if the criteria for initiating treatment on belzutifan were overly restrictive. For example, it could change decisions about surgery if belzutifan could only be tried in each patient once for each tumour. The committee concluded that it was appropriate to consider retreatment for new primary tumours. To avoid additional surgery or other local

procedures, the committee thought that it would be reasonable to continue belzutifan treatment in someone with multiple tumours if some tumours were responding, even if a previously treated tumour had progressed on belzutifan (see section 3.5).

Clinical-effectiveness evidence

The MK-6482-004 trial

3.7   The clinical-effectiveness evidence for belzutifan came from MK-6482-004. This was a multicentre single-arm open-label phase 2 study. It included 61 people with VHL with at least 1 measurable RCC tumour. Fifty of them also had CNS Hbs and 22 also had pNETs. The primary outcome of MK-6482-004 was the objective response rate (complete response or partial response). The secondary outcomes were disease control rate, duration of response, time to response, progression-free survival and time to surgery. At the latest data cut (April 2022), the objective response rate was 63.9% for RCC (95% confidence interval [CI] 50.6% to 75.8%), 44% for CNS Hbs (95% CI 30.0% to 58.7%) and 90.9% for pNETs (95% CI 70.8% to 98.9%). This was assessed using the Response Evaluation Criteria in Solid Tumours 1.1 criteria. Disease control rate (complete response, partial response or stable disease) was 98.4% for RCC (95% CI 91.2% to 100.0%), 90.0% for CNS Hbs (95% CI 78.2% to 96.7%) and 100% for pNETs (95% CI 84.6% to 100.0%). The median time to response for RCC was 11.1 months. Progression-free survival results are considered confidential by the company and cannot be reported here. The committee concluded that belzutifan was likely to be clinically effective in reducing tumour size and so the need for surgery. It noted that there was some uncertainty about how tumour size relates to symptom burden in CNS Hbs. The committee also noted that the time needed for the response was 11.1 months for RCC. It considered this in terms of the positioning of belzutifan and the fact that assessing whether surgery is needed may be challenging to predict in clinical practice. The committee noted there is another ongoing phase 2 study (MK-6482-015) that will include some people with VHL.

Comparator data

3.8   Because MK-6482-004 was a single-arm study, the company used data from a VHL natural history study to inform the comparative effectiveness of the standard-care comparator. This study was a retrospective non-interventional study of existing medical records. It also included supplemental electronic medical record data abstraction and a review of abdominal imaging scans done during routine clinical care in a cohort of people in the US. It included people with at least 1 VHL-associated RCC tumour measured during the study period. They also had to meet other VHL natural history study eligibility criteria that were identified and followed until the end of the assessment window (31 July 2004 to 30 June 2020). The EAG noted that comparative effectiveness results derived from VHL natural study data did not:

- represent the population in belzutifan's marketing authorisation

- collect the appropriate data to address the population of interest.

     The committee noted that the lack of comparator data meant that the company used a matching adjusted indirect comparison (MAIC) method to compare belzutifan with standard care (see section 3.10). The committee noted that the VHL natural history study was well conducted but was US based, so potentially it may not be generalisable if care guidelines are different. The committee highlighted that the VHL natural history study was not aligned with the decision problem. After consultation, the company updated its model using the surgery rates in the standard-care arm from the pretreatment phase of MK-6482-004 instead of the natural history study. The EAG considered using this phase of the trial as a valuable alternative to the natural history study to explore the uncertainty. It noted that this change was only available for RCC, which had a 15% weighting in the model, and that the pretreatment phase was not used for either of the other tumour types. The committee concluded that using the same source of data ensured internal validity. So, it agreed with its use in RCC despite various limitations in the evidence. But, overall, it thought that the comparative effectiveness evidence was particularly weak. It said that it would expect to see additional information and analysis on the natural history of the condition, specific to the target population, when belzutifan exits the Cancer Drugs Fund.

Outcomes

3.9 The company focused on collecting outcomes such as time to response, progression-free survival, time to surgery and overall survival. MK-6482-004 and the VHL natural history study were used to compare the outcomes of treatment with belzutifan with the outcomes for standard care to inform the model. The committee noted that these outcomes were different from those used in standard NICE cancer topic evaluations. The committee noted that time to surgery (a key model transition) represented a highly heterogeneous outcome that depended on several factors, such as:

- size and location of tumours

- symptom development

- extent of previous surgery or potential impact of surgery

- need or desire for surgery

- the overall health of the person with VHL.

    It thought that the outcome of any given surgery would not necessarily directly correlate with any permanent step change in morbidity and mortality expected with VHL, such as loss of organ or neurological function. The committee noted that there was considerable uncertainty in the eligibility criteria for MK-6482-004 and the VHL natural history study compared with the population of interest. It agreed that more information may be needed on outcomes that more closely match loss of organ or neurological function (for example, need for dialysis, paralysis). The committee concluded that it would have preferred to see a model that did not use such a heterogeneous outcome as the main transition. It added that it would like to see a model that aligns more reliably with transitions and health states of greater importance in the natural progression of VHL in individuals when belzutifan exits the Cancer Drugs Fund.

Establishing relative treatment effect

3.10 In its initial submission, the company estimated the relative treatment effects of

belzutifan compared with standard care from an indirect treatment comparison (ITC) using the propensity-score weighting-based MAIC methods. This used individual patient data from the VHL natural history study to match the baseline characteristics of MK-6482-004. After matching, outcomes were compared between treatment with belzutifan and standard care. In the company's cost-effectiveness model, the treatment effects of belzutifan were compared with standard care using data from MK-6482-004 and the reweighted VHL natural history study data respectively. It did this by selecting a subgroup population from the VHL natural history study that matched MK-6482-004's inclusion and exclusion criteria. The committee noted that, for the comparison with standard care, the company did a series of adjustments, specifically:

- Kaplan–Meier curves were fitted to the VHL natural history study (standard care) and the MK-6482-004 trial data.

- The fitted Kaplan–Meier curves from the VHL natural history study data were then adjusted using MAIC to match the population in MK-6482-004 based on variables such as age, gender, previous surgeries and tumour size.

- Time to surgery, second surgery and metastasis in the VHL natural history study were adjusted to reflect a less active surveillance. This was because the company thought that the standard care seen in the VHL natural history study cohort may have been better than that routinely provided in UK clinical practice.

- An additional assumption that 90% of people with RCC or pNETs, and 50% of people with CNS Hbs, have immediate surgery was then applied.

  The committee noted that the relative treatment effect was highly uncertain because of the assumptions needed to convert from MK-6482-004's population to the marketing authorisation population. It thought that the assumption that 90% of people having standard care would proceed to immediate surgery made the MAIC adjustment of the Kaplan–Meier curve relatively unimportant. This was because it only applied the residual 10% of the standard-care population. The committee also thought that this immediate surgery assumption was too simplistic and not evidence based. It noted that the adjustment for the population was only applied to the standard-care arm. The assumption in the belzutifan arm was a return to VHL natural history baseline rates of surgery. This implicitly modelled some return

of organ function after using belzutifan if representing the same population. So, the committee did not think that the treatment arms would be equal if 1 arm had immediate surgery and the other arm modelled a delay to a different rate. The committee understood that this would have substantially biased the comparison in belzutifan's favour.

After consultation, the company updated its base case for the RCC comparison using data from the pretreatment phase of MK-6482-004 to inform the relative effects (see section 3.8). The company also explained that it had removed the immediate surgery assumption from the model and implemented a 4-month delay to surgery in the standard-care arm. The EAG did not think that this addressed the concern of different assumptions applied to the modelled populations. To address this, the EAG provided a scenario analysis removing the assumption of immediate surgery from the model. The committee noted that the issue extended to the fact that it was concerned that the model of people having belzutifan and surgery did not reflect the same population. The committee thought that belzutifan was clinically effective at delaying tumour progression and so the need for surgery (see section 3.7). But it was aware that any attempt to compare the populations was limited by this core issue and lack of data. It was cautious of introducing bias by accepting assumptions when there was a clear lack of evidence. But it was also aware of the challenges of evidence generation for VHL. The committee thought that the company's base case modelled the outcomes for belzutifan from MK-6482-004 without adjustment but that the outcomes for standard care were driven by assumptions. It thought that this was inappropriate and lacked face validity. The committee wanted to choose an analysis that adjusted both populations equally by either:

- adjusting time to surgery in the belzutifan arm to account for the incorrect assumption that people in the belzutifan arm recovered organ function to the rate seen in the VHL natural history study, or

- did not arbitrarily adjust any of the data for the need of immediate surgery (using the rates seen in the VHL natural history study).

The committee requested an analysis with these structural changes to establish the relative treatment effect. It took into account the internal validity of the RCC time to surgery, and the clinical expert and company

comments that the belzutifan population may be representative. It thought that the second option most closely matched expected clinical practice. It also thought that this was more appropriate because it provided a better approximation of relative effect for the MK-648-004 population who were not in need of surgery (see section 3.7). So, it provided a better comparison of the available data rather than a comparison largely driven by assumption. But the committee noted that, if clinical practice favoured later use of belzutifan in the pathway only on the precipice of organ loss, then the first option was plausible. But it concluded that this option had substantially less evidence to populate an economic model with and resulted in an extremely uncertain relative effect.

Economic model

The company's model structure and outputs

3.11 The company used a Markov model structure to estimate the cost effectiveness of belzutifan compared with standard care. The model included 5 health states: before surgery, during surgery, event-free after surgery, metastatic disease (preprogression and postprogression) and death. The model had a lifetime horizon (59 years) and a weekly cycle length. In the company's model, people started at age 41 years in the presurgery health state reflecting the treatment decision point. From there, they could transition to surgery, metastatic disease or death health states. The EAG thought that the company's model structure was only appropriate for the RCC cohort, noting a high level of uncertainty, specifically:

- There was an overlap in the data informing input parameters for the 3 cohorts, which did not seem appropriate given the heterogeneity of the condition.

- The rate of surgeries (moving from presurgery to surgery) was based on surgeries for the primary tumour. But it was not clear from the trial and the VHL natural history study whether the data used to specify these rates related to the treatment of primary tumours.

- Including people for whom surgery was not suitable but who were on the precipice of organ failure in the standard-care arm only was not appropriate.

  At the first committee meeting, the committee thought that the model had substantial structural uncertainties that made it unreliable for decision making. After consultation, the company updated its model using the pretreatment phase of MK-6482-004 (see section 3.10) and combined the cohorts into 1 weighted cohort. But no other significant structural changes to the model were made. The committee noted that many of its concerns and the EAG's concerns with the model structure still applied after consultation. The most important issues included:

- issues arising from the population mismatch between MK-6482-004, VHL natural history study and the target population, and the assumptions used to attempt to account for this mismatch (including starting age, mortality, time horizon, how treatment effect was applied and time on treatment)

- the key model transition or surgery not matching key step-changes in health states (see section 3.9), and lack of information provided on any of those key transitions (that is, time to nephrectomy or pancreatectomy, or neurological complications)

- the modelling and assumptions needed to populate 3 separate cohorts with different primary tumours rather than the totality of the disease and concurrent tumours

- modelling assumptions based on arbitrary assumptions that lacked face validity such as its use of immediate surgery, time on treatment and treatment waning.

  The committee did not think that the company's structure was appropriate for decision making. But, to address the structural uncertainty, it attempted to use a range of structural scenarios to minimise the issues. It considered 2 different structural scenarios to reflect the range of positioning of belzutifan (see section 3.10). One had no adjustment for immediate surgery and the other had adjustment to the belzutifan arm for immediate surgery. It thought that these scenarios attempted to capture the nature of the condition and began to resolve some of the uncertainties inherent in the

model. The committee would have preferred an analysis that used the population without any immediate surgery in either arm. This was because many of the uncertainties were then only linked to the question of generalisability of the treatment to clinical practice. Also, many of the uncertainties were limited because the evidence base was significantly more appropriate for this population and treatment setting. The alternative scenario assuming immediate surgery in both arms, with a delay caused by belzutifan, was less appropriate but plausible if belzutifan was positioned later in clinical practice. The committee thought that the most appropriate way to apply greater flexibility for a higher degree of uncertainty was to consider it in the context of the substantial structural uncertainty associated with the model structure (see section 3.18). It did not think that the structural uncertainty would be resolved with further changes to the model unless more information was known about the population and how belzutifan will be used in clinical practice. So, it thought that the structural scenarios presented were acceptable for decision making on whether belzutifan had plausible potential to be cost effective. But this was only on the condition that there is more consideration of the robustness of the modelling assumptions and how to populate the model on exit from the Cancer Drugs Fund.

Proportion of people in each cohort

3.12   The company provided analysis for 3 separate cohorts with different primary tumours (see section 3.11). It thought that the proportion of people in each of the cohorts would be based on people within its interpretation of the marketing authorisation, so people on the precipice of organ loss or loss of neurological function. The model outputs were weighted dependent on the primary tumour type and based on clinical expert opinion. The updated base case assumed 80% with CNS Hbs, 15% with RCC and 5% with pNETs. The committee thought that this may have overestimated the proportion of people with CNS Hbs if it is used differently in clinical practice than suggested by the company. This is because RCC tumours resulting in reduced or complete organ function may be more prevalent. The committee considered scenarios adjusting the proportion. But it did not have evidence of the distribution in clinical practice, nor the importance of concurrent tumour diagnosis to treatment choice and key outcomes. It thought that this evidence, along with how belzutifan will be used (see section 3.6), could

be identified through managed access in the Cancer Drugs Fund.

Time on treatment

3.13   The company assumed that people would stay on belzutifan until VHL progressed or until they had side effects. It used time-on-treatment data from MK-6482-004 to model time on treatment with belzutifan. The committee noted that almost half of the people who stopped belzutifan in the trial were reported to have done so through choice. A minority who were reported as stopping stated reasons were progression or side effects. The company explored different parametric fits to the patient-level time-on-treatment data (exponential, Gompertz, log-logistic, log-normal, generalised gamma and Weibull) using patient-level data from MK-6482-004. The company preferred using the Gompertz-based model that was based on statistical fit, visual inspection and clinical relevance in its base case. The company also explored the effect of using the second-best (Weibull) model in the scenario analysis. The EAG explained that there was uncertainty in the long-term extrapolations of this data. The committee noted that it was unclear why a high proportion of people stopped treatment for the stated reason of choice. It thought that this would not be generalisable to the population of interest in the company's model. This was because the alternative for people stopping treatment would be surgery resulting in organ loss, as assumed by the company in its model. So, the committee concluded it would have preferred to see the modelling using belzutifan continued until progression or until side effects because this would more closely match the target population.

After consultation, the company provided scenarios using progression-free survival as a proxy for time on treatment. The company explained that it did not think that modelling time on treatment until disease progression was appropriate. This was because, in VHL, progression is non-linear, and the condition is characterised and affected more by surgical outcomes than metastases. It explained that modelling time on treatment until side effects was not feasible because of the timeframe provided. The committee questioned the company about why a high proportion of people stopped treatment with belzutifan in MK-6482-004 by choice. The company explained that 23 (37.7%) people stopped treatment:

- 6 (9.8%) stopped because of disease progression

- 2 (3.3%) had adverse events

- 2 (3.3%) died

- 1 (1.6%) became pregnant

- 11 (18.0%) chose to stop because of a range of non-clinical reasons, including travel and costs related to participation in the trial.

The committee thought that, in clinical practice, people with VHL on the precipice of an organ failure would not stop belzutifan by choice (for non-clinical reasons) if the next option was organ failure with associated profound long-term complications. The committee noted that modelling based on time on treatment or progression-free survival had a minor effect on the results. But these scenarios did not address the nature of the uncertainty related to time on treatment because both attempted to fit the data from MK-6482-004. In the absence of appropriate evidence, the committee accepted the evidence from MK-6482-004, fitting a Weibull extrapolation curve to the data available. But it thought that this was highly uncertain because treatment costs could substantially increase in clinical practice given the significant complexity of the condition. So, it thought that the possibility of retreatment and longer treatment duration should be captured in the analysis. The committee thought that the uncertainty could be resolved through further data collection through managed access in the Cancer Drugs Fund to see treatment duration and prescribing practices in NHS clinical practice.

Treatment waning

3.14   The committed noted that, because of uncertainties in the modelled time on treatment, the company's model incorporated a treatment-effect waning using 'off-treatment' health states. Transition probabilities from the 'off-treatment' health states were assumed to gradually converge to the respective values in the standard-care arm based on data from the VHL natural history study. The company assumed treatment waning occurs gradually over 2.71 years (the

amount of time until the largest RCC tumour reaches the baseline levels of growth) from the end of the maximum follow up (3.84 years). It also assumed that tumour growth would return to baseline level at an average rate of 3.52 mm per year after stopping. The same assumption was used for CNS Hbs and pNETs. The EAG explained that the duration of assumed residual benefit over a period of 2.71 years might be appropriate for RCC but could be different for CNS Hbs and pNETs. The committee noted that this may have overestimated the benefit because treatment does not stop at best response but at progression. The committee noted that treatment waning might be appropriate because of the tumour size reduction seen in MK-6482-004. But it reiterated that there were significant concerns around the generalisability of the data for the population of interest. At the first meeting, the committee highlighted that it would like to have seen extensive sensitivity analyses and testing alternative assumptions on the treatment waning of effect across the tumour types. After consultation, the company explained that the treatment-effect waning period for CNS Hb and pNET cohorts was assumed to be equivalent to the RCC cohort. This was because of the small number stopping treatment in the CNS Hb and pNET subgroups in MK-6482-004 who had an available CNS Hb and pNET measurement near to the time of stopping treatment. The company explained that their clinicians thought that its treatment waning approach was plausible for CNS Hbs and pNETs. It presented further sensitivity analyses around the treatment-effect waning in the CNS Hb and pNET cohorts. The committee agreed that the treatment effect of belzutifan may be substantially different depending on which part of the body part is affected. For instance, the impact of CNS Hbs may not be as related to the size and growth rate as the positioning of the tumour. It thought that it was plausible that people treated with belzutifan would not immediately need surgery on progression because of the reduced tumour size. But it noted that there was very limited evidence for this assumption in the population of interest. The committee concluded that the treatment waning assumptions could not be appropriately explored with the available evidence and was based on somewhat arbitrary assumptions. It thought that there could be additional evidence collected on tumour growth rates and real-world relapse rates on belzutifan to resolve the uncertainty.

Health-related quality of life

3.15　　In MK-6482-004, no health-related quality-of-life data was collected. The committee noted that the company used health-state utility values in the economic model. These were derived from the VHL real-world quality-of-life disease burden study and a trial of pembrolizumab in the adjuvant treatment of RCC (KEYNOTE-564). It thought that the utility values were broadly appropriate but noted that these were not necessarily measured in the population of interest. The company's model also applied disutilities from tumour-associated surgeries and surgical complications (including loss of organ function). The company used a variety of sources to derive disutility values for long-term and short-term complications, including the VHL real-world quality-of-life disease burden study and published literature. The committee noted that the company had used an additive approach to adjust disutility values in the model. It thought that the disutility values adopted by the company for long-term complications of surgery may have lacked face validity. The committee also noted that the multiplicative method is a preferred approach for combining disutilities, in line with NICE's health technology evaluations manual. It thought that the outputs of the effective utility loss for people after surgery were uncertain. The committee was aware that some of the utility decrements the company used to represent the long-term consequences of surgery were large. It noted that people having dialysis had a utility decrement of -0.527. This is much larger than the estimates seen in the literature and the values used in NICE's guidelines on renal replacement therapy and conservative management and chronic kidney disease. The committee understood that the company had derived the disutility values by deducting estimates of quality of life on dialysis from 1. The committee thought that this approach was not appropriate because it effectively assumes that people not on dialysis would be in perfect health. It noted that some of the health-related quality-of-life impact was potentially double counted, for example, separate disutilities for 'stroke' and 'neurological complications' were applied in the model. The committee thought that a more appropriate approach would be to calculate a relative utility impact. This would be done by comparing an absolute estimate of the utility on dialysis with an age- and sex-matched expectation from the general population. The committee highlighted that the total effective utility loss after surgery should have been calculated using a multiplicative method. It added that this should have been validated against literature for similar outcomes (such as end-stage renal disease, post pancreatectomy and neurological damage). It

should also have taken account of any additional consideration for people with VHL.

After consultation, the company updated its model using a multiplicative method. It replaced the disutility value for end-stage renal disease or dialysis based on an absolute estimate of the utility of dialysis with an age- and sex-matched expectation from the general population. The company explained that it could not verify utility or disutility value for chronic kidney disease. The committee questioned the company application of the multiplicative approach. It was aware that the methods used to estimate utility values for complex conditions can produce different utility estimates. The committee reiterated that some of the health-related quality-of-life impact was double counted and potentially implausible. It noted that there were limited attempts to validate against the literature for similar outcomes (such as stroke, end-stage renal disease and postpancreatectomy health states). For example, the committee thought that using a disutility of -0.37 for stroke was too high in the model. It noted that Joundi et al. (2022) reported EQ-5D-3L for people more than 3 months after a stroke at 0.65. This is 20% below that in the general population. So, it preferred to use the lower disutility value of -0.16 for stroke in its preferred assumption. The committee concluded that it would have preferred to see utility values measured directly using evidence similar to the VHL disease burden survey, or values that had been validated against literature values for similar outcomes. But it broadly accepted that the utility values could match the target population. It thought that further analysis of utility values would be needed based on the population having belzutifan in NHS clinical practice, such as re-evaluating the generalisability of the population in the VHL disease burden survey. It noted that this population would be established when belzutifan exits the Cancer Drugs Fund.

Severity

3.16   The committee may apply a greater weight to quality-adjusted life years (QALYs; a severity modifier) if technologies are indicated for conditions with a high degree of severity. So, the committee considered the severity of VHL, that is, the future health lost by people living with the condition and having standard care in the NHS. The company provided absolute and proportional QALY shortfall estimates based on its modelled population for all VHL cohorts, in line with NICE's health

technology evaluations manual. In its analyses, the proportional QALY shortfall was above 0.95 for RCC and CNS Hbs and below 0.95 for pNETs. So, a severity weight of 1.7 was applicable for RCC and CNS Hbs, and of 1.2 for pNETs. The company explained that, in clinical practice or the real world, not all VHL cohorts are distinct from each other. People in MK-6482-004 had more than 1 tumour manifestation (see section 3.5), and pNETs are associated with high mortality and morbidity because of pancreatic surgery. So, the company applied a QALY weighting of 1.7 to all 3 VHL cohorts. Based on the QALYs generated from the company's model, the company and EAG agreed that, for RCC and CNS Hbs, the QALY weighting 1.7 was applicable. But the EAG thought that it was inappropriate to apply the QALY weighting of 1.7 for pNETs. The committee acknowledged the substantial impact of VHL. It also noted that both the company's and the EAG's analyses were subject to a high degree of uncertainty because of the underlying assumptions adopted for modelling standard care. The committee was unable to apply an appropriate severity weight based on the calculations presented by the company at the first meeting because of uncertainty in its underlying assumptions.

At the second meeting, the company proposed that a severity weighting of 1.7 was more appropriate despite its model suggesting a severity weighting of 1.2 should be used. This was because of the rarity of VHL and poor outcomes faced by people with the condition (see section 3.2). It thought that its model was unable to fully capture QALY benefits. For example, people with CNS Hbs also had a full nephrectomy. Both the company and the EAG agreed that, based on the absolute and proportional QALY shortfall calculations, the appropriate severity weight was 1.2. The committee explained that it could consider the substantial effect of VHL on people, and their families and carers. It also explained it could take into account in its decision making other uncaptured benefits, but it could not use this to justify a quantitative severity weighting of 1.7. The committee reiterated that both the company's and the EAG's analyses were subject to a high degree of uncertainty because of the underlying assumptions adopted for modelling standard care. The committee considered the substantial uncertainty with the shortfall calculations. But, based on the totality of the evidence, it thought that the severity weight of 1.2 applied to the QALYs was appropriate. The committed considered this in its decision making.

Cost-effectiveness estimates

The committee's preferred assumptions

3.17   The committee did not think that the company base case had face validity (see sections 3.10 to 3.16). So, it considered the deterministic incremental cost-effectiveness ratio (ICER) for belzutifan compared with standard care using structural scenario analysis. Because of confidential commercial arrangements for belzutifan, the exact cost-effectiveness results cannot be reported here. The committee was aware that the EAG was unable to define its base case because of uncertainties in the evidence and assumptions made in the model. The committee noted considerable structural uncertainty, but within that analysis, its preferred cost-effectiveness assumptions informing the estimate included:

- using the surgery rates in the standard-care arm from the pretreatment phase of the MK-6482-004 for the RCC cohort

- using the surgery rates in the standard-care arm from the VHL natural history study estimations for the CNS Hb and pNET cohorts

- removing the assumption of immediate surgery and associated organ loss in the standard-care arm

- a disutility value of -0.16 for stroke in the CNS Hb cohort

- using a multiplicative approach for utilities

- a 1.2 QALY modifier for severity.

    The committee thought that this ICER represented a reasonable scenario for establishing the cost effectiveness of the probable population with no adjustment to the standard-care arm for immediate surgery. It also noted that it was substantially greater than £30,000 per QALY gained. The alternative scenario with adjustment to belzutifan for immediate surgery was also plausible if belzutifan was only used at the precipice of organ failure or loss of neurological function. This resulted in an ICER of between £20,000 to £30,000 per QALY gained, but the committee thought that this was less likely than its preferred scenario. Given the substantial structural uncertainty, the committee concluded that it could not recommend belzutifan for routine use

in the NHS.

Other factors

3.18   Because of the rarity of VHL, the committee recognised difficulties in the ability to collect or generate clinical evidence on belzutifan's comparative effectiveness. It also noted that the uncertainty about the natural history of VHL in the marketing authorisation population contributed to significant uncertainty in the decision making. This was a direct result of the inability to generate the appropriate clinical evidence. It also noted that belzutifan is likely to be an innovative treatment for VHL because of its new biological mechanism in a complex and heterogeneous treatment pathway. This also affected the ability to collect clinical evidence. The committee also noted that there may be other factors that had not been included in the analyses. These included the potential of belzutifan to reduce fear and anxiety because of frequent scans, and affect tumours that were not included in the marketing authorisation such as retinal hemangioblastomas. The committee thought that, because of these factors, it would be able to apply greater flexibility in accepting a higher degree of uncertainty, as described in section 6.2.34 of NICE's health technology evaluations manual. The committee chose to exercise that flexibility by considering how:

- the structural uncertainties inherent in the modelling related to the inability to generate the evidence (see section 3.11 and section 3.17 and section 3.20)

- the evidence could be collected with managed access in the Cancer Drugs Fund.

Acceptable ICER

3.19   NICE's health technology evaluations manual notes that, above a most plausible ICER of £20,000 per QALY gained, judgements about the acceptability of a technology as an effective use of NHS resources will take into account the degree of certainty around the ICER. The committee will be more cautious about recommending a technology if it is less certain about the ICERs presented. The

committee thought that, because of the structural uncertainty, there was a large degree of uncertainty in the cost-effectiveness estimates. But, despite the substantial structural uncertainty, the committee thought that there may also be uncaptured benefits (see section 3.17). It also agreed that many of the uncertainties were a result of difficulties in evidence generation. For this reason, the committee agreed that an acceptable ICER would be around £30,000 per QALY gained. This is at the upper end of the range normally considered a cost-effective use of NHS resources (£20,000 to £30,000 per QALY gained).

Managed access

3.20   Having concluded that belzutifan could not be recommended for routine use in the NHS, the committee then considered whether it could be recommended with managed access in the Cancer Drugs Fund. It discussed the arrangements for the Cancer Drugs Fund agreed by NICE and NHS England in 2016, noting NICE's Cancer Drugs Fund technology appraisal process and methods guide (addendum). The committee was aware that the company had expressed an interest in being considered for funding through the Cancer Drugs Fund in its submission. This was because of acknowledged uncertainties and a lack of data directly relevant to the decision problem. It also noted that managed access in the Cancer Drugs Fund may provide the opportunity to collect additional data to address some uncertainties about belzutifan's efficacy in clinical practice. The committee clarified that the Cancer Drugs Fund does not normally accept models needing refinement. It thought that the model structure could be reconceptualised after a period of proposed data collection for key model concepts and structure. This was because of the nature of the uncertainties mostly relating to identification of the population, which could likely only be resolved through use in clinical practice. It noted that it had not seen evidence from MK-6482-015, which is still ongoing and may provide additional evidence, alongside further data cuts from MK-6482-004. The committee was aware that a period in the Cancer Drugs Fund may not fully resolve uncertainties such as relative efficacy, the natural history of the standard-care arm (see section 3.10), utility values (see section 3.15) and the model structure. But it thought that there would still be benefits in:

- characterising people with VHL who would have belzutifan in clinical practice

(eligible population in line with marketing authorisation and percentage of tumour in each type) from SACT data

- belzutifan's longer-term efficacy from MK-6482-004

- treatment waning from MK-6482-004 and SACT data

- time on treatment from SACT.

The key issue of identifying the precipice population may also assist with partially resolving some of the key uncertainties and assumptions that would not be directly informed by this data collection. The committee then considered whether any of the cost-effectiveness estimates had plausible potential to be cost effective. It thought that scenarios that modelled the population on the precipice of organ failure or neurological issues (see section 3.10) could plausibly be correct if clinical practice prefers using belzutifan, as in the company's suggested positioning. This was not the committee's preferred analysis. But the committee thought that it was within the range of plausible assumptions that did result in ICERs within the range normally considered a cost-effective use of NHS resources, despite uncertainty in the validity of model inputs and assumptions. The committee concluded that belzutifan met the criteria to be considered for managed access in the Cancer Drugs Fund as an option for treating VHL in adults who need treatment for RCC, CNS hemangioblastomas or pNETs, when localised procedures are unsuitable or undesirable.

Equalities

3.21   The committee noted that, because VHL is a genetic condition, some families are disproportionately affected. The condition can affect people when they are very young. It also noted that people with language, learning or cultural barriers, or disabled people may be at a disadvantage with accessing treatment. The committee did not think that these would be equality issues because, if belzutifan were to be recommended, the recommendation would not restrict access for some people over others. No other equality issues were identified.

Conclusion

Recommendation

3.22   The committee recalled the uncertainties in the evidence for this technology (see
section 3.16) and the other factors involved in its decision making (see
section 3.18). Taking these into account, the ICERs based on assumptions were
higher than what NICE normally considers a cost-effective use of NHS resources.
So, it concluded that belzutifan could not be recommended for routine use. But
the committee acknowledged that, despite significant uncertainty, belzutifan is
very likely to offer improved clinical outcomes for some people in clinical practice.
The committee thought that it was appropriate to apply greater flexibility in
accepting a higher degree of uncertainty because of the nature of the challenges
in evidence generation for this condition. The committee thought that belzutifan
does have plausible potential to be cost effective in some situations, and that
some of the uncertainties may be resolved with further data collection (see
section 3.20). So, belzutifan is recommended for use with managed access in the
Cancer Drugs Fund for treating VHL in adults who need treatment for RCC,
CNS Hbs or pNETs, when localised procedures are unsuitable or undesirable.

4 Implementation

4.1   When NICE recommends a treatment as an option for use with managed access, NHS England will make it available according to the conditions in the managed access agreement. This means that, if a patient has tumours associated with von Hippel-Lindau disease and the doctor responsible for their care thinks that belzutifan is the right treatment, it should be available for use, in line with NICE's recommendations and the criteria in the managed access agreement.

4.2   Chapter 2 of Appraisal and funding of cancer drugs from July 2016 (including the new Cancer Drugs Fund) – A new deal for patients, taxpayers and industry states that for those drugs with a draft recommendation for use in the Cancer Drugs Fund, interim funding will be available (from the overall Cancer Drugs Fund budget) from the point of marketing authorisation, or from release of positive draft guidance, whichever is later. Drugs that are recommended for use in the Cancer Drugs Fund will be funded in line with the terms of their managed access agreement, after the period of interim funding. The NHS England Cancer Drugs Fund list provides up-to-date information on all cancer treatments recommended by NICE since 2016. This includes whether they have received a marketing authorisation and been launched in the UK.

4.3   The Welsh ministers have issued directions to the NHS in Wales on implementing NICE technology appraisal guidance when the drug or treatment, or other technology, is approved for use with managed access. When a NICE technology appraisal guidance recommends the use of a drug or treatment, or other technology, for use with managed access, the NHS in Wales must usually provide funding and resources for it within 2 months of the first publication of the final draft guidance or agreement of a managed access agreement by the NHS in Wales, whichever is the later.

5 Evaluation committee members and NICE project team

Evaluation committee members

The 4 technology appraisal committees are standing advisory committees of NICE. This topic was considered by committee B.

Committee members are asked to declare any interests in the technology being evaluated. If it is considered there is a conflict of interest, the member is excluded from participating further in that evaluation.

The minutes of each evaluation committee meeting, which include the names of the members who attended and their declarations of interests, are posted on the NICE website.

Chair

Charles Crawley
Chair, technology appraisal committee B

NICE project team

Each evaluation is assigned to a team consisting of 1 or more health technology analysts (who act as technical leads for the evaluation), a technical adviser and a project manager.

Harsimran Sarpal
Technical lead

Adam Brooke
Technical adviser

Vonda Murray
Project manager

Richard Diaz
Associate director

ISBN: 978-1-4731-6518-2