Article:
Drummond, Mike orcid.org/0000-0002-6126-0944, Nicod, Elena, Meregaglia, M. et al. (3 more authors) (2022) Consideration of quality of life in the health technology assessments of rare disease treatments. European Journal of Health Economics. pp. 645-669. ISSN 1618-7601

https://doi.org/10.1007/s10198-021-01387-w

The European Journal of Health Economics

Consideration of quality of life in the health technology assessments of rare disease treatments

--Manuscript Draft--

Manuscript Number:	EJHE-D-21-00236R1
Full Title:	Consideration of quality of life in the health technology assessments of rare disease treatments
Article Type:	Original Research
Keywords:	patient-reported outcome; rare disease; orphan medicinal products; health-state utility value; Health technology assessment; reimbursement
Corresponding Author:	Elena Nicod, PhD Universita Bocconi Milano, ITALY
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Universita Bocconi
Corresponding Author's Secondary Institution:	
First Author:	Elena Nicod, PhD
First Author Secondary Information:	
Order of Authors:	Elena Nicod, PhD
	Michela Meregaglia
	Amanda Whittal
	Sheela Upadhyaya
	Karen Facey
	Michael Drummond
Order of Authors Secondary Information:	
Funding Information:	Horizon 2020 Framework Programme (779312) — Dr. Elena Nicod Dr Michela Meregaglia Dr Amanda Whittal Mrs Sheela Upadhyaya Dr Karen Facey Prof Michael Drummond
Abstract:	Objectives: Challenges with patient-reported outcome (PRO) evidence and health state utility values (HSUVs) in rare diseases exist due to small, heterogeneous populations, lack of disease knowledge and early onset. To better incorporate quality of life (QoL) into Health Technology Assessment, a clearer understanding of these challenges is needed. Methods: NICE appraisals of non-oncology treatments with an EMA orphan designation (n=24), and corresponding appraisals in the Netherlands, France, and Germany were included. Document analysis of appraisal reports investigated how PROs/HSUVs influenced decision-making and was representative of QoL impact of condition and treatment. Results: PRO evidence was not included in 6/24 NICE appraisals. When included, it either failed to demonstrate change, capture domains important for patients, or was uncertain. In the other countries, little information was reported and evidence largely did not demonstrate change. In NICE appraisals, HSUVs were derived through the collection of EQ-5D data (7/24 cases), mapping (6/24), vignettes (5/24), and published literature or other techniques (6/24). The majority did not use data collected alongside clinical trials. Few measures demonstrated significant change due to lack of sensitivity

	or face validity, short-term data, or implausible health states. In 8/24 NICE appraisals, patient surveys or input during appraisal committee meetings supported the interpretation of uncertainty or provided evidence about QoL. Conclusions: This study sheds light on the nature of PRO evidence in rare diseases and associated challenges. Results emphasise the need for improved development and use of PRO/HSUVs. Other forms of evidence and expert input are crucial to support better appraisal of uncertain or missing evidence.
Response to Reviewers:	Dear Editor and Reviewers, Thank you for providing your suggestions and opportunity to improve the quality of this manuscript. We have revised it in accordance with the suggestions as detailed in the point-by-point response included in the submission. We've also included two versions of the manuscript: one with all changes tracked for your reference, and another clean version. We are happy to make any further amendments should they be required. Many thanks again. Best wishes, Elena Nicod

Point-by-point response

Submission ID: EJHE-D-21-00236

Consideration of quality of life in the health technology assessments of rare disease treatments

Dear Editor and Reviewers,

Thank you for providing your suggestions and opportunity to improve the quality of this manuscript. We have revised it in accordance with the suggestions as detailed in the point-by-point response below. We are happy to make any further amendments should they be required.

Reviewer #1

Very well written and important paper that I highly recommend getting published. Minor comments to address.

Reviewer's comment	**Author's response**
1) where "lack of knowledge" is mentioned in the abstract and text, change to "lack of knowledge of natural history" for clarity	Thank you for spotting this. We have changed this to "lack of disease knowledge", which would be broader than natural history (e.g. patient population, treatment pathways, etc). We hope this is ok
2) pg 2 line 8- sentence starting with Quality of Life...is not true across the board. Rephrase to say Quality of Life of patients living with a rare disease is often poor due to the disease impacting multiple aspects of functioning".	Changed as suggested, thank you
3) Next sentence- issues with diagnosis are mainly diagnosis delays (add delays);	Changed as suggested
4) og 2, line 49- drop "more nuanced" that isn't needed	Changed as suggested
5) pg 5 line 11- the symptoms, not "they symptoms" (typo)	Thank you for spotting

6) Figure 2 is really hard to understand what it's trying to convey and another type of figure (or table) should be considered	We've changed the figures to tables to make them more easier to understand
7) The table 2 is great but very lengthy for a manuscript and the organization wasn't clear/ could be improved to show some type of pattern	We feel that Table 2 is important to include as it gives the "data" that is the basis for our analysis, so it is more for reference than detailed review. The text and new preceding summary table draws out the key elements.
8) in the discussion more solutions could be mentioned including early scientific advice with the EU agencies, and working with disease registries early on to get natural history data	Thank you for this suggestion. We have included in the discussion, under how QoL information should be better used for HTA, the following paragraph: "This could be achieved through greater involvement in early multi-stakeholder dialogues and early scientific advice to better align across HTA bodies and agree on what QoL evidence would be accepted, and a greater acceptance of registry data to leverage early on data on natural history on the disease."

Reviewer #2

The concept of the study is interesting, and the authors have clearly undertaken a huge amount of work in analysing so many appraisals. However, with so much information presented (in the results and discussion particularly), it is difficult for me to understand exactly what the authors did, what they found and what the implications are. I believe that part of my difficulty stems from the methods used – I am not overly familiar with thematic analyses and suggest that more detail in this section would be beneficial. I would also suggest that the authors reconsider the presentation of the results as neither the text, tables, not figures are particularly intuitive.

I have provided comments on specific sections below, but overwhelmingly feel that the paper needs substantial revisions and restructuring to make what should be a useful and thought-provoking analysis much more accessible to readers.

We have provided responses, point by point, below.

Reviewer's comment	**Author's response**
I believe that part of my difficulty stems from the methods used – I am not overly familiar with thematic analyses and suggest that	We have provided more clarifications on what was done, how and have referenced a key resource on thematic analysis + a source on the methodological framework

more detail in this section would be beneficial.	previously developed and used as a basis to structure the data collection and analysis – see methods section
I would also suggest that the authors reconsider the presentation of the results as neither the text, tables, not figures are particularly intuitive.	The figures have been redrawn (Figure 1) and new tables developed to ease interpretation (Tables 1, 3 and 5)
Introduction line 5 – should be referenced to EMA or similar, not reference 1 (is ref 1 the 2020 publication in cost effectiveness and resource allocation)?	Thank you for spotting this – the reference was not correct. We now referenced this to the webpages from the EC on orphan medicinal product and EURORDIS on what is a rare disease
Intro line 43 – HSUVs can be lower than 0	We've added the worse than dead state as negative HSUV values
Intro line 24 – please change to active voice rather than passive (throughout)	The choice of active or passive voice is a stylistic matter. If the journal has a strong preference for the active voice, we can make the change. Meanwhile, we have proof read the paper but have not explicitly changed the voice
The objective is not entirely clear…to understand challenges, how QoL was appraised, and whether aligned with expected disease impact? What does this actually mean?	We have now made the objectives more specific: "To better incorporate QoL evidence into HTA decision-making, a clearer understanding of the challenges encountered when using PRO evidence and HSUVs in rare diseases is needed. Hence this research explored how QoL evidence has been used in appraisal of non-oncology rare disease treatments in a selection of countries using different HTA approaches."
Methods – why not other countries (e.g. Australia, Canada, Sweden)?	We selected two countries per HTA approach (e.g. use of cost-effectiveness estimates, or use of added clinical benefit). EU countries were preferred over non-EU as this work was conducted as part of an EU project. We have added the following clarification in the "study sample" section in the methods: "Considering the depth of the analysis conducted, the inclusion of four countries was considered sufficient to understand the nuances between one HTA approach and another, and the differences within each approach."

Methods – data collection and analysis. How were these criteria chosen? Are they established elsewhere? They could appear fairly arbitrary otherwise. Why would being heterogenous indicate a higher burden of disease? Why would the administration mode affect QoL?	We have provided references and clarifications in the methods section. This work leveraged an existing methodological framework developed, that allows for a structured collection and analysis of decision-making processes. Further explanation about thematic analysis and how this was done have also been added (and referenced). We hope this is sufficient. We have swapped the order of what has been collected – and linked the data collected to how it was analysed. Hopefully it is more clear now.
Methods – what does thematic analysis entail? What published techniques were used?	This has been better explained in the methods section and references have been provided
Methods – data collection and analysis line 23. Latter means last of two, replace with last.	Changed as suggested
Really this is a study of NICE appraisals with a couple of paragraphs about other HTA bodies. Is it valuable to include these other HTA bodies – what do they add with the level of information available? Is there a useful comparison to NICE here?	The intention was to compare all countries equally but little information was found from other countries. Hence it is this that hampered the review. This is now explained more clearly in the text. Some elements of interest can be drawn from the other countries, particularly France and Germany that consider added benefit and we would not wish to miss the opportunity to share those findings.
In NICE appraisals, how are you differentiating between company and ERG work and the committee's preferred assumptions?	The focus has been on the appraisal process, meaning the interpretation of the evidence by the Committee. The ERG's model/assumptions/results on HSUVs are included when discussed by the Committee and accounted for in the decision – this is described in Table 4 under the HSUV technique column, and last column on the decision and reasons for decision does specify when ERG's assumptions have been preferred. This is explained in the methods

What is meant by "patient-based evidence" and how is this different from PROs/HSUVs?	We now refer to all other evidence relating to QoL as patient evidence – this is now defined in the methods.
Is patient-based evidence really a separate category that is comparable to PROs and HSUVs? Surely preferences on administration mode are very different to patient information being used to derive HSUVs?	We have clarified in the text that we mean evidence related to QoL that does not arise from PROs or HSUVs. This could include feedback from patient focus groups, or patients' contributions to the discussions during the Committee discussions.
I'm confused about the numbers for patient submissions – there are 5 appraisals cited where patient surveys were used, but only four sources (two company submissions and two patient submissions). What was the source of the patient and clinician input for the other 4 appraisals?	Thank you for spotting that, we have double checked all the data and updated the results accordingly – see results section The sources are the appraisal reports, where the interpretation of this evidence was commented on. In the methods section, we reference the appraisal reports and mention that we refer to these appraisal reports when discussing the NICE appraisals of the individual drugs in the text
Would it be useful to distinguish between PROs in addition to HSUVs? Otherwise it doesn't make sense to say that PRO wasn't reported, or PROMs and results weren't discussed.	In the PRO analysis, we distinguished when the PRO data were used to derive HSUVs used to inform the decision, versus when the PRO data were discussed separately or in addition to the HSUVs and had an influence on the decision. This distinction highlights cases when, in cost-effectiveness approaches, the interpretation of the HSUV data was complemented by the PRO data. Hopefully, this is now clear in the paper
Conclusion – is patient-based evidence preferable to PROMs/HSUV data though? Were they preferred by committees to the HSUV/PROM data from studies? I can see that they might be used to supplement this, but I disagree with the "most importantly" conclusion	We now clarify that patient evidence is additional information, which can be discussed in the absence of, or additional to, HSUV or PROM data. We have changed 'Most importantly' to 'Additionally'
The paper needs a comparison to non-rare diseases. I'm not expecting the authors to do the same analysis for all NICE appraisals	The objective of our study was to explore the issues with interpreting and using quality of life data in rare diseases. It was

that aren't in rare diseases, but is there something published that they could refer to? Otherwise how will readers know that these results and conclusions are specific to rare diseases?	our prior view that this could be particularly problematic, for the reasons we give. Although it was not in our remit to undertake a comparison, it is possible that some of the same problems could exist in the HTA of treatments for common diseases, even given the increased likelihood of having validated PROMs and HSUVs in the literature. We are not aware of an equivalent study in common diseases that we could refer to. Therefore, we now mention this in the limitations, and highlight that a comparison with non-rare diseases could be an area for further research. "Finally, this study highlights some of the nuances in considering QoL evidence in rare diseases. It is possible that some of the same issues could arise in the HTA of more common diseases, Further research would be needed, comparing the results from this analysis with those from a similar analysis of HTAs for common non-cancer treatments."
Figure 1 – please replace with a grouped bar chart. Centring at zero (as per a tornado diagram) seems strange.	We have changed the chart as suggested – hopefully more clear now.
Figure 2 – there are circles for PRO evidence reported and PRO evidence not reported. Some appraisals don't have either circle in different areas – how are we to interpret these? It seems like they neither had PRO evidence reported nor didn't have PRO evidence reported. This figure suggest that some of the appraisals not have any QoL evidence at all (eg Holoclar, strimvelis), but in the text it says other QoL were considered for 2 studies – what about the other 4 (obethocholic acid, strimvelus, burosumab, pirfenidone)?	We have now changed Figures 2-3 to Tables 1 and 3, respectively. In Table 1, the legend explains the "not reported" cases, which relate to PRO results not being reported in the appraisal report, but mentioned as having been collected. We assumed these had no influence on the decision. Indeed there are a number of cases that did not consider any PRO data – in the table, there is no sign corresponding to any of the PROMs. These included: strimvelis, burosumab, obeticholic acid, holoclar,

	pirfenidone, elosulfase alfa. In the second paragraph of the results section, under “use and influence of PRO evidence in NICE appraisals”, it states that 6/24 didn’t consider any PRO data and explains why it was the case for the 6 drugs. For the remaining 4, no PROM evidence was considered and HSUV were derived based on other data / using other techniques, e.g. strimvelis - published literature, burosumab/vignettes, obeticholic acid/published literature. No information on the quality of life data accounted for pirfenidone was a re-assesssment and no new QoL was provided (this is specified in the paragraph). We have added the following clarification in the paragraph: “QoL evidence for the remaining drugs without PRO data was based on HSUVs derived from vignettes or published evidence (Table 3, discussed in the next section on HSUVs).” In the table we added the following clarifications: *No PRO data was provided, but patient evidence or PRO data from the literature was used instead (Elosulfase alfa, Holoclar) ** No PRO data was provided, QoL evidence was derived from HSUVs from vignettes or published literature (Strimvelis, Burosumab, Obeticholic acid, Holoclar) *** No PRO data was provided as this is a re-assessment and no new QoL evidence was presented (Pirfenidone) We hope that it is now more clear
Similar, in figure 3 – did some appraisals not use HSUVs?	Figure 3 is now Table 3 and hopefully more clear. All drugs apart from Pirfenidone considered HSUVs. We have added that clarification in the footnotes: “No HSUV results were considered as this is a re-assessment and no new QoL evidence was provided“

The circle size suggests that the influence was the same between HSUVs and other evidence – this is clearly not true when HUSVs are used to generate ICERs.	Figures changed to Tables, signs used and explained in legend
Conclusion – is patient-based evidence preferable to PROMs/HSUV data though? Were they preferred by committees to the HSUV/PROM data from studies? I can see that they might be used to supplement this, but I disagree with the "most importantly" conclusion	Repeat of question above

Reviewer #3

Results are worth being published as they raise an important and often neglected aspect of HTA.

Reviewer's comment	**Author's response**
The paper can benefit from a review of the text to make it less wordy and bring up the key results.	Text has been cut, particularly in the intro and methods and new tables developed to succinctly present the results
The text also needs to explain some key distinctions (PROs vs HSUVs) and why HTA such as NICE focus on the latter.	This has now been clarified in the text
Finally, there is a need to point out that the issues identified might be relevant also in more common diseases, particularly in NICE appraisals where the focus is on cost per QALY.	We have not added this as we only have data for rare diseases. In the "Limitations' section we mention that a similar study should be undertaken of HTAs for common diseases.
Line 8-9 page 2: the article quoted refers to more prevalent chronic conditions not only more prevalent. Please check the reference and adjust the wording accordingly.	We have changed this based on reviewer #1's comment to read "QoL is poor in rare disease patients due to multiple aspects affecting functioning" – and took away the comparison with common conditions
Line 10-11 page 2: Is the HSUV that is not sufficiently sensitive? Is it not an issue with the descriptive measure, the PRO, rather than the utilities?	Thank you for spotting this. In the referenced paper by Pearson et al, they refer to QALY's not being sensitive to disease severity of the population. We've also added some references.

Overall, the introduction on page 1 and 2 can be shortened, particularly around the details of the HTA approach. If it is assumed that the audience knows about them, the text of the section can be reduced. There are some word repetitions that can be edited to improve readability. What it would be useful to know is the difference between PROs and HSUV, which currently is not explained very clearly (patient completing the questionnaire vs eliciting public preferences on the health states). Finally is HSUV used at all in added benefits countries?	The intro has been shortened and we've clarified the difference between PRO and HSUVs In added benefit countries, HSUVs are not being used – e.g. in the IQWIG methods guidelines, there is a provision for something akin to disease-specific QALYs, but they have never been used (to our knowledge). We have not added anything in the manuscript, to keep the two approaches distinct and avoid confusion
Line 57 page 3: the reasons for excluding cancer treatments are not explained "above" as stated in the text.	We've added the explanation: "…because added benefit often relies on survival gains, and many rare cancers are subsets of more common cancers for which a validated PRO often exists"
Line 44: it is administration of EQ5D (in the trial?) AND use of the UK value set.	We have specified how the EQ5D data to derive HSUVs was collected: "within a trial (4/7) or from a registry or cohort study (3/7)"
How HSUV is derived from EQ5D is not clear from the paper. Most readers should know but given that it is the focus of the paper it should be explained	A brief description of how HSUVs are derived from EQ-5D has been added in the introduction.
Line 50 page 7: unclear what HSUVs for adverse events or administration mode is?	We've clarified this in the text: "… to determine the HSUV estimated to measure the impact of adverse events on QoL included in the submission" We've also clarified that these were derived and considered alongside the HSUVs
In figure 3, what is patient-based evidence? How can you use patient evidence to obtain utilities (usually based on societal values)?	Thanks for pointing this out, we've changed it to patient evidence and clarified in the text that we mean evidence related to QoL that does not arise from PROs or HSUVs. This could include feedback from patient focus

	groups, or patients' contributions to the discussions during the Committee discussions.
Table 2 page 8: the information in table 2 is far too detailed for a paper. In addition, no reference to the decision outcome and the cost per QALY is made in the text and it is not stated how this information was extracted. I suggest excluding the table or presenting a much reduced version of it.	We feel that Table 2 is important to include as it gives the "data" that is the basis for our analysis, so it is more for reference than detailed review. The text and new preceding summary table draws out the key elements. One option could be to make the table supplementary material, but we would prefer not to as we would lose this level of detail which is important. We would also prefer to include the cost/QALY information, even if it is not explicitly discussed, the table does highlight when the PRO evidence made a difference in accepting (or not) a cost/QALY on the higher end. However, we would accept the Editor-in-Chief's guidance on whether Table 2 should be moved to the supplementary material We've included the appraisal report references for each of the drugs
Line 29 page 9: I can only see table 1 and 2. Table 3 is missing from the text.	Table 4 should have been Table 3 – apologies, now changed in the manuscript. Table 5 also now included (it hadn't been uploaded in the initial submission – apologies!)
Line 34 page 10. Is the sentence misleading? It sounds like in all cases improvement in QoL was the main objective of the treatment and so presumably included as primary endpoints. Is this the case? Line 16 to 23 on page 5 can help in explaining this.	This sentence has been re-worded as follows, and is hopefully more clear: "For all of the treatments investigated, their added benefit was also considered to improve QoL."
Line 47 to 55 page 10: I tend to agree with the missed opportunity to consider rich PRO data in HTA. However, is this only an issue in rare diseases? I believe this happens for common disease too because the remit of HTA bodies, in particular NICE. The key decision criterion is cost effectiveness and hence committee will primarily consider HSUVs via the incremental cost effectiveness ratio. The paragraph at the	We agree it is worth emphasising the distinction depending on the approach used. We have added the following sentences: "Results illustrate that different QoL evidence would be considered depending on the HTA approach. For cost-effectiveness oriented approaches, HSUVs are considered within the incremental cost-effectiveness ratio and are derived from

end of page 10 partially explained this. In countries focusing on added benefits, HSUVs will not be considered (and this should be stated explicitly in the paper). Another point that should be raised is around the reasons for not considering PROs in added benefits countries: are they all related to the quality of the data provided (as explained on page 9) rather than the intention of not considering the data?	PRO data using indirect techniques (e.g. generic preference-based instruments, mapping), or measured directly from patient responses using direct preference-based techniques (e.g. time-trade-off) [18]. In countries with an added benefit assessment approach, the PRO data would be considered and interpreted as is without being derived into a numerical HSUV. To help with the comparability and interpretation of the PRO data, generic PROMs are often preferred." We also specified that "It was not clear from the appraisal reports why this evidence was not reported or accounted for."
Line 37 to 49 page 13: the sentence about the role of patients and clinical input might not need to be reported in the conclusion. The sentence starting "patients should be better informed" should be deleted as it is unclear how the authors can draw these conclusions and it seems to be in contrast with the previous one. Further research are not very clear and informative so they either need to be expanded in the discussion section or be deleted.	We have clarified how patient evidence and patient / clinical input can help in the process, and deleted the point about patients needing to be better informed about what type of evidence and input can be meaningful. We've also deleted the last sentence on further research, which is more focused on methodological aspects of HSUV techniques
References do not include the NICE appraisals consulted.	References now included

Consideration of quality of life in the health technology assessments of rare disease treatments

Elena Nicod[1], PhD, Michela Meregaglia[1], PhD, Amanda Whittal[1], PhD, Sheela Upadhyaya[2], Karen Facey[3], PhD, Michael Drummond[4], PhD

[1]SDA Bocconi School of Management, Research Centre on Health and Social Care Management (CERGAS), Milan, Italy

[2]National Institute for Health and Care Excellence, London, UK

[3]University of Edinburgh, Usher Institute for Population Health Sciences & Informatics, Edinburgh, UK

[4]University of York, Centre for Health Economics, York, UK

Corresponding author:

Elena Nicod
Centre for Research on Health and Social Care Management (CERGAS)
SDA Bocconi School of Management
Via Sarfatti 10
20136 Milan, Italy
Email: elena.nicod@unibocconi.it
Tel: +39 373 769 9945

Declarations:

Funding: This research was funded by the European Commission's Horizon 2020 research and innovation programme and was undertaken under the auspices of IMPACT-HTA (grant # 779312). The results presented here reflect the authors' views and not the views of the European Commission. The European Commission is not liable for any use of the information communicated. No funds, grants or other support was received.

Conflicts of interest: Michael Drummond and Karen Facey have received funding and consultancy fees from manufacturers of treatments for rare diseases outside of this work. Elena Nicod and Amanda Whittal are part-time employed by Dolon Ltd and have no conflicts with this work.

Availability of data and material (data transparency): Data was extracted from publicly available reports and literature

Authors' contributions: All authors made substantial contributions to the design of the work; EN and AW collected the data, EN conducted the data analysis and drafted the work; all authors revised it

critically at several occasions for important intellectual content; all authors approved the version being submitted; all authors agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Key words: patient-reported outcome, rare disease, orphan medicinal products, health-state utility value, health technology assessment, reimbursement

Acknowledgments: We would like to thank Dr. Andrew Lloyd for his time and feedback on the results.

ABSTRACT

Objectives: Challenges with patient-reported outcome (PRO) evidence and health state utility values (HSUVs) in rare diseases exist due to small, heterogeneous populations, lack of disease knowledge and early onset. To better incorporate quality of life (QoL) into Health Technology Assessment, a clearer understanding of these challenges is needed.

Methods: NICE appraisals of non-oncology treatments with an EMA orphan designation (n=24), and corresponding appraisals in the Netherlands, France, and Germany were included. Document analysis of appraisal reports investigated how PROs/HSUVs influenced decision-making and was representative of QoL impact of condition and treatment.

Results: PRO evidence was not included in 6/24 NICE appraisals. When included, it either failed to demonstrate change, capture domains important for patients, or was uncertain. In the other countries, little information was reported and evidence largely did not demonstrate change. In NICE appraisals, HSUVs were derived through the collection of EQ-5D data (7/24 cases), mapping (6/24), vignettes (5/24), and published literature or other techniques (6/24). The majority did not use data collected alongside clinical trials. Few measures demonstrated significant change due to lack of sensitivity or face validity, short-term data, or implausible health states. In 8/24 NICE appraisals, patient surveys or input during appraisal committee meetings supported the interpretation of uncertainty or provided evidence about QoL.

Conclusions: This study sheds light on the nature of PRO evidence in rare diseases and associated challenges. Results emphasise the need for improved development and use of PRO/HSUVs. Other forms of evidence and expert input are crucial to support better appraisal of uncertain or missing evidence.

Consideration of quality of life in the health technology assessments of rare disease treatments

ABSTRACT

Objectives: Challenges with patient-reported outcome (PRO) evidence and health state utility values (HSUVs) in rare diseases exist due to small, heterogeneous populations, lack of disease knowledge and early onset. To better incorporate quality of life (QoL) into Health Technology Assessment, a clearer understanding of these challenges is needed.

Methods: NICE appraisals of non-oncology treatments with an EMA orphan designation (n=24), and corresponding appraisals in the Netherlands, France, and Germany were included. Document analysis of appraisal reports investigated how PROs/HSUVs influenced decision-making and was representative of QoL impact of condition and treatment.

Results: PRO evidence was not included in 6/24 NICE appraisals. When included, it either failed to demonstrate change, capture domains important for patients, or was uncertain. In the other countries, little information was reported and evidence largely did not demonstrate change. In NICE appraisals, HSUVs were derived through the collection of EQ-5D data (7/24 cases), mapping (6/24), vignettes (5/24), and published literature or other techniques (6/24). The majority did not use data collected alongside clinical trials. Few measures demonstrated significant change due to lack of sensitivity or face validity, short-term data, or implausible health states. In 8/24 NICE appraisals, patient surveys or input during appraisal committee meetings supported the interpretation of uncertainty or provided evidence about QoL.

Conclusions: This study sheds light on the nature of PRO evidence in rare diseases and associated challenges. Results emphasise the need for improved development and use of PRO/HSUVs. Other forms of evidence and expert input are crucial to support better appraisal of uncertain or missing evidence.

Introduction

Rare diseases are conditions affecting a small number of patients (e.g. less than 1/2,000 people in Europe), which are life-threatening and/or chronically debilitating, frequently genetic and with an early onset [1, 2]. Quality of life (QoL) of patients living with a rare disease is often poor due to multiple aspects affecting functioning [3]. This is partly explained by issues around diagnostic delays, and/or a lack of knowledge about the disease, its treatment pathways or treatment options [3]. Given the severity of these conditions and paucity of curative treatments, understanding their impact on QoL is crucial, particularly when assessing the benefit of a new treatment.

Health Technology Assessment (HTA) aims to assess the value of a treatment to inform decisions on whether it should be provided routinely to the relevant patient population. The assessment generally relies on clinical and patient-reported outcome (PRO) endpoints, which provide evidence about health outcomes and impact on patients' wellbeing [4]. In the latter case, PRO evidence is collected directly from patients or proxies using patient-reported outcome measures (PROMs) [5]. PROMs are intended to capture aspects that matter most to patients about the impact of disease and treatment on symptoms, QoL or health status [6].

HTA relies on the critical assessment of added benefit or cost-effectiveness of a treatment. This is then appraised by a Committee, taking account of other relevant factors, who decide on reimbursement (and pricing in some cases). Added benefit is assessed by considering the magnitude and certainty of treatment benefit over existing therapies based on the clinical and PRO evidence presented. The level of benefit is then generally ranked into categories as, for example, in France, where the added benefit (ASMR) is ranked between I and V. In cost-effectiveness assessments, an economic evaluation is conducted that models the progression through health states along the care pathway, with and without the new treatment under review. In order to assist cost-effectiveness assessments, techniques have been developed to translate PRO evidence into numerical values called health state utility values (HSUVs). HSUVs represent individual preferences for given health states measured on a scale between 0, representing dead and 1, full health (with negative values implying states considered to be worse than dead). These are then merged with survival data (e.g., length of life) into a composite measure called quality-adjusted life-year (QALY). HSUVs represent the utility value associated with the different models' health states, for both treatment and comparator arms [7]. The most common way of deriving HSUVs is currently using indirect techniques, e.g. preference-based instruments such as EQ-5D that are accompanied by an algorithm (or a set of tariffs) providing HSUVs. Tariffs are pre-determined at individual country level by a sample of the general population that uses direct techniques (e.g. time trade-off) to express preferences for a subset of health states derived from the combination of instrument's dimensions and levels.

Challenges exist when developing and using PRO evidence and HSUVs for HTAs of rare diseases treatments due to the small and heterogeneous nature of the patient populations, and frequent lack of knowledge about the disease [8, 9]. Additionally, patients are often children or infants who cannot

self-report and who may be cognitively impaired or unable to communicate. There may also be distinct challenges around capturing meaningful outcomes, including: difficulty achieving concept validity through concept saturation, use of methods that may not capture aspects important for patients, or selecting the appropriate PROM when natural history is poorly understood. There are also few validated disease-specific PROMs for rare diseases, probably due to the amount of time and resources needed to develop these instruments, which are further complicated by the nature of these diseases [10].

Additional challenges frequently encountered when deriving and using HSUVs for rare diseases include the need for a large number of respondents to minimise random measurement errors (e.g. person-trade-off, development of mapping algorithms), identification of appropriate values corresponding to the model's health states from the existing literature, and QALYs being insufficiently sensitive to disease severity or changes that are important for patients [11–14].

Although these challenges in measuring QoL are common to all rare disease treatments, they are most important in treatments of non-oncological diseases, since in cancer the main value of treatment is often increased survival, and many rare disease treatments in cancer are for sub-populations of a more common cancer for which validated QoL measures may be available.

To better incorporate QoL evidence into HTA decision-making, a clearer understanding of the challenges encountered when using PRO evidence and HSUVs in rare diseases is needed. Hence this research explored how QoL evidence has been used in appraisal of non-oncology rare disease treatments in a selection of countries using different HTA approaches.

Methods

Study sample

For this EU Horizon2020 project, European countries were selected to represent those that make decisions based on added clinical benefit and those that focus on cost-effectiveness, and who have publicly available reports. Those selected were England (National Institute for Health and Care Excellence, NICE) and the Netherlands (National Health Care Institute, ZIN) as users of the cost-effectiveness approach, and France (Haute Autorité de Santé, HAS) and Germany (Federal Joint Committee, G-BA) as users of the added benefit approach. Considering the depth of the analysis conducted, the inclusion of four countries was considered sufficient to understand the nuances between one HTA approach and another, and the types of contrasts within one approach. As only the reports from NICE presented detailed information about the committee deliberations of the QoL evidence, the analysis focused on the NICE appraisals, and other countries' appraisals were used as a contrast to highlight any different approaches.

All treatments with a European Medicines Agency (EMA) orphan medicinal product designation and appraised by NICE within its Technology Appraisal (TA) or Highly Specialised Technologies (HST) programmes before 1 June 2020 were selected (n=50). Cancer treatments (26) were excluded

because added benefit often relies on survival gains, and many rare cancers are subsets of more common cancers for which a validated PRO often exists [4, 15]. This left 24 rare disease treatments (12 TA and 12 HST) for analysis [16, 17, 26–35, 18, 36–39, 19–25] In the results section, the information reported for the individual NICE sample was extracted from these reports.

Data collection and analysis

Information about QoL, PROs, HSUVs and other evidence from patients about their QoL (such as patient group submissions and patient expert input) was extracted from NICE's appraisal reports. These were sufficiently detailed to enable documentation of the source of the evidence, results, issues highlighted by the appraisal committee and the influence on the decision. If needed, supporting documentation such as manufacturer submissions and Evidence Review Group (ERG) reports were reviewed. A published framework was used to extract these key aspects of the appraisal in a structured way [40]. PRO evidence was categorised on the basis of the type of PROM used (generic, disease-group (developed for a range of conditions), disease-specific or symptom-specific). HSUVs were categorised on the basis of the technique used to derive them (e.g. collection using an instrument such as EQ-5D, or mapping from other PROMs).

Thematic analysis was undertaken to identify issues arising in appraisals and their influence on decision-making based on the researchers' interpretation of the discussion reported in the published documents. The identification of themes was done iteratively and was continuously refined while the researchers familiarised themselves with the data and grouped the data in a logical way to allow for a better understanding of the decision process [41]. Once the themes were identified and categorised, the researchers assessed the level of influence of PRO evidence or patient evidence on decisions. This was categorised as "influence" when the Committee explicitly recognised and accounted for a change in QoL in their decisions, "possible influence" when PRO evidence or patient evidence was explicitly reported but considered limited by the Committee and it was therefore unclear whether it influenced the decision, and "no influence" when PRO evidence or patient evidence was reported, but inconclusive or failed to demonstrate change and did not influence the Committee's decision. HSUVs were considered in all cases to have influence on the decision, since the Committee always took note of the incremental cost per QALY ratio. The interpretation of the HSUV evidence was distinguished as "accepted" when the Committee recognised the evidence presented was acceptable, "not commented" when no issues were raised with the HSUVs presented and therefore there was a high likelihood that it was accepted, and "uncertain" when a number of issues around the HSUV were highlighted, which rendered their interpretation challenging.

The second set of extracted information related to the burden of disease and treatment impact: infant/childhood onset; progressive; heterogeneous; multi-systemic; debilitating; life-threatening; supportive care; and regarding the impact of treatment on QoL: length of life improved, QoL improved from reduced symptoms, daily living, families/carers, compared to current treatment, administration mode. The burden of disease and intended impact of the treatments reported in the HST and TA programmes, respectively, were extracted and compared.

The analysis aimed to understand how QoL was appraised, and the extent to which PRO evidence and/or HSUVs were considered appropriate. The possible influence of the nature of the rare diseases on PRO evidence and HSUV estimates was also explored to generate a better understanding of what was feasible in the different contexts. In the cross-country analysis, the information reported by the other countries was scarce. The focus was therefore on the PRO evidence and HSUVs considered and their influence on the decision.

Results

Impact of disease and treatment on quality of life

Most of the diseases undergoing the TA and HST processes were life-threatening and/or debilitating (Figure 1). The burden of disease, however, was greater in the diseases undergoing the HST programme compared with the TA in that these diseases affect children, have a heterogeneous and progressive nature, or affect multiple organs. With the exception of the prophylaxis treatment letermovir, the symptoms of all of the diseases analysed affect patients' daily living and QoL. No previous treatments were available for 58% (7/12) and 17% (2/12) of those undergoing the HST and TA processes, respectively.

In terms of the intended effects of treatment, 67% (8/12) of HST and 83% (10/12) of TA treatments aim to improve length of life, while all improve patients' daily living and QoL by reducing symptoms (with the exception of letermovir). Six of these aim solely to improve QoL. Furthermore, all of the HST and 83% (10/12) of the TA treatments aim to improve patients' daily living and QoL over standard of care. In 50% of all cases (12/24), QoL improvement is linked to a different administration mode.

All conditions appraised by HST and half of those by TA were considered to affect carer QoL. In all cases, with the exception of letermovir, the treatment intends to improve their QoL.

The estimated yearly number of patients to be treated in England ranged between 1-50 for 10 of the 12 HST treatments [1-7 patients for 3 treatments, 20-35 for 2 treatments, 50-100 for 3 treatments, and 140-150 for 2 treatments]. No details about patient numbers were provided in all other cases.

<Figure 1. Proportion of appraisals for which various items of burden of disease and treatment impact are relevant in NICE Highly Specialised Technology and Technology Appraisal programmes (n=24)>

Use and influence of PRO evidence in NICE appraisals

In NICE appraisals, PROMs can have an influence either through being considered directly and/or through their use in generating HSUVs. Across the 24 treatments appraised by NICE, 28 different PROMs were reported. This included 10 generic PROMs considered across 14 treatments, seven disease-group PROMs across seven treatments, three disease-specific PROMs across five treatments, and eight symptom-specific PROMs across seven treatments (Table 1). Several PROMs

could considered for the same treatments. Examples of disease-group PROMs include the Paediatric Outcomes Data Collection Instrument (PDOCI) measuring functional outcomes in paediatric orthopedics [42], and the St George's Respiratory Questionnaire (SGRQ) measuring overall health, daily life, and perceived well-being in patients with obstructive airways disease [43]. The three disease-specific PROMs considered were for cystic fibrosis (Cystic Fibrosis Questionnaire Revised questionnaire, CFQ-R), recurrent angioedema (Angioedema Quality of Life questionnaire, AE-QoL), and neuronal ceroid lipofuscinosis type 2 (Neuronal Ceroid Lipofuscinosis Type 2 Quality of Life Instrument, CLN2-QoL). The seven symptom-specific PROMs related to pain, gastro-intestinal symptoms, diabetic neuropathy, fatigue, asthma, and anxiety and depression (Table 2).

Forty-two percent (10/24) of submissions did not include any generic PRO, and 25% (6/24) no PRO evidence at all. Reasons for the latter included PRO evidence not collected in trials (elosulfase alfa, obetocholic acid, holoclar), collected but limited (strimvelis), being collected and not reported (burosumab), or not presented given it was a re-assessment based on new clinical evidence (pirfenidone). In two of the cases without PRO evidence, other QoL evidence was considered, such as observational studies and cross-sectional surveys involving patients and families (elosulfase alfa), and visual acuity data from the literature used to derive HSUVs (holoclar) (Table 1). QoL evidence for the remaining drugs without PRO data was based on HSUVs derived from vignettes or published evidence (Table 3, discussed in the next section on HSUVs).

<Table 1. Types and influence of PRO evidence considered in NICE TA and HST appraisals of non-oncology rare disease treatments (n=24)>

<Table 2. Use and influence of PRO evidence in NICE TA and HST appraisals of non-oncology rare disease treatments (n=24)>

Further exploration of the influence of PRO evidence on NICE decisions suggested that beyond those used to derive HSUVs, few of them had any influence on the decisions (Tables 1 and 2).

Of the 14 appraisals considering generic PRO evidence, eight were used to derive HSUVs and the remaining six had unclear or no influence on the decisions. For asfotase alpha, the EQ-5D data collected in a patient survey may have been considered by clinicians when developing the vignette's health states, but it is not discussed in the report. For cerliponase, it was inconclusive due to the lack of correspondence between EQ-5D and the model's health states, and short trial duration for the Pediatric Quality of Life Inventory (PedsQL). For ataluren, no significant improvements in the PedsQL were shown, despite the positive trend in the functioning subscale. For the remaining treatments (migalastat, letermovir, lanadelumab), the SF36 and EQ-5D collected did not show any significant improvements and were not considered.

With the exception of one disease-group PROM used to derive the economic model's HSUVs, their inclusion had limited influence. This was the case for mepolizumab, where SGRQ data, suggesting improved QoL due to fewer exacerbations and improved symptom control and lung function, was

mapped to EQ-5D to obtain HSUVs. In the other cases, the PODCI data collected for ataluren showed improvements on two dimensions, but was considered uncertain due to the short trial duration. In all other cases (letermovir, asfotase alfa, voretigene, darvadstrocel and nintedanib), the disease-group PROMs, Functional Assessment of Cancer Therapy - Bone Marrow Transplantation (FACT-BMT), PODCI, Visual Function Questionnaire (VFQ), perianal disease activity index (PDAI), SGRQ or Shortness of Breath Questionnaire (SOBQ) either did not show a significant improvement or were not reported.

A similar situation was seen for the disease-specific PROMs. For only one case, colistimethate sodium and tobramycin DPI, the CFQ-R was mapped to HSUVs and used for the decision. However, it did not show any improvement in QoL relating to administration mode (dry powders for inhalation versus nebuliser) given a non-inferiority trial design was adopted. For three treatments, the PRO evidence was uncertain and thus the influence on the decision was unclear. The data collection period of CLN2-QoL for cerliponase was considered too short, and the CFQ-R data collected for mannitol dry and lumacaftor-ivacaftor did not show a statistically significant improvement. Results from the AE-QoL data collected for lanadelumab were not commented on in the appraisal report.

Of the six treatments that considered symptom-specific PROMs, one of them influenced and another possibly influenced the decision. For patisiran, the Neuropathy Impairment Score (NIS) and Norfolk Quality of Life Questionnaire - Diabetic Neuropathy (Norfolk QoL-DN) data collected was statistically improved and contributed to recognising treatment effectiveness. For eliglustat, no significant improvements were demonstrated for the Fatigue Severity Scale (FSS) and Brief Pain Inventory (BPI), and it was unclear whether they were used to determine the HSUV estimated to measure the impact of adverse events on QoL included in the submission. In the remaining cases, there was either no demonstration of change with BPI and Gastrointestinal Symptoms Rating Scale (GSRS) for migalastat and with Norfolk QoL-DN for inotersen, or results were not reported (Cough and Sputum Assessment Questionnaire, CASA-Q for nintedanib, Asthma Control Questionnaire, ASQ for mepolizumab and Hospital Anxiety Depression Scale, HADS for lanadelumab).

For eight of these drugs, determination of QoL impact was influenced by patient evidence (Table 2). First, patient surveys provided information about impact of QoL on patients and carers (eculizumab, ataluren), preferences for administration mode (eliglustat, colistimethate sodium and tobramycin DPI), or whether it was used to derive HSUVs (elosulfase alfa). Respondents were patients and in one case also family members, three formed part of the company submissions and the other two, patient submissions. Second, patients and clinicians provided input about the dimensions not captured in the model (patisiran), about impact on QoL (letermovir), effect on tolerability (nintedanib), and administration mode (migalastat, eliglustat, colistimethate sodium and tobramycin DPI).

Use and influence of HSUV estimates in NICE appraisals

The most frequently used technique to derive HSUVs in NICE appraisals was through EQ-5D data (7/23) collected within a trial (4/7) or from a registry or cohort study (3/7), followed by mapping (6/23), vignettes (5/23), published literature (3/23), Health Utility Index Mark 2 (HUI2) (1/23) and other (1/23)

(Table 3). No HSUVs were reported for one treatment (pirfenidone) given it was a re-assessment; therefore, it was excluded from this analysis, which focused on the 23 remaining treatments. The mapping technique was more frequently used in the TA, and vignettes in the HST process. Additional HSUVs were derived to measure the impact on QoL of adverse events (9/23), of the administration mode (4/23), of carer burden (7/23) or other (7/23) and considered alongside the HSUV derived.

< Table 3. Techniques used to derive HSUVs in NICE TA and HST appraisals of non-oncology rare disease treatments (n=23) >

The detail and summary of the individual appraisals are summarised in Table 4. Seven treatments used EQ-5D, two of which collected EQ-5D 3L in trials and the remaining collected EQ-5D 5L (mapped to 3L) or foreign EQ-5D datasets converted using the UK tariff. In one case, the HSUVs included in the model were considered acceptable by the TA Committee (lanadelumab). For mannitol, the generic Health Utility Index Mark 2 (HUI2) was used to derive HSUV estimates. Even if EQ-5D would have been preferred, the HUI2 was accepted by the relevant committee. For all remaining cases, a number of issues were raised by the relevant committees, which included benefits (eculizumab, migalastat) or long-term effects not captured (letermovir), measure insensitive to change (nintedanib), uncertain duration (patisiran), or possible implausible health states (inotersen).

<Table 4. Use and influence of HSUVs in NICE TA and HST appraisals of non-oncology rare disease treatments (n=23) >

Mapping was used in six cases, in one of which (lumacaftor-ivacaftor) the applicant developed a new algorithm, while in the others published functions were used. Source measures included lung function and pulmonary exacerbation (lumacaftor–ivacaftor), SF36 (eliglustat), PedsQL (nursinersen), CFQ-R (colistimethate sodium and tobramycin DPI), SGRQ (mepolizumab) and visual acuity (holoclar); all were converted to EQ-5D-3L. The results were considered acceptable in only one case (mepolizumab), or not commented on (likely acceptable) in two cases (eliglustat, holoclar). The issues raised regarding the remaining cases included: ceiling effects and little change captured even though it was collected in the largest existing cystic fibrosis trial (lumacaftor-ivacaftor), limited face validity resulting in expert elicitation being used to estimate the HSUVs (nusinersen), or limited methodological approach (colistimethate sodium and tobramycin DPI).

Vignettes were used in five cases. Reasons for their use over more conventional approaches included a lack of correspondence between QoL data collected in the clinical trial and model health states (cerliponase), lack of negative values when deriving the PedsQL being considered unrealistic considering the condition's severity (cerliponase), or QoL data not collected in trial (darvadstrocel, burosumab, voretigene). The health states were developed by patient and clinical experts (voretigene), or only clinicians (cerliponase, asfotase alpha, burosumab). Respondents included clinicians (voretigene, cerliponase, asfotase alpha, burosumab), or patients and public

(darvadstrocel). The QoL measure included was EQ-5D-5L (cerliponase, asfotase alpha, burosumab), and HUI2 and EQ-5D (voretigene).

A number of issues were raised about the vignettes. For voretigene, poor convergent validity between EQ-5D and HUI2 and preference for EQ-5D (considered to better capture overall QoL over HUI2) were highlighted. For alfotase alfa, trial data would have been preferred over vignettes by the appraisal committee; however, QoL results from the vignette were compared to results from a patient survey and considered aligned. Additionally, given the health states were based on the surrogate outcome "six-minute walking test" (6MWT), all of the relevant symptoms that would produce lower HSUVs in the more severe states may not have been captured (likely underestimate). The HST Committee was also concerned with clinicians responding to the vignettes instead of patients (burosumab). Furthermore, there was concern about the uncertain robustness of the vignettes given an unclear association of other elements (e.g., pain) to health states (cerliponase).

Published literature was used in three cases. This was because QoL was not measured in the trials (strimvelis, obeticholic acid) or the available mapping algorithm was conducted on a healthy population and thus unsuitable (ataluren). No detail on the published literature was provided for strimvelis and ataluren, whereas for obeticholic acid, values from an analogue disease (Hep C) were used.

In one case (elosulfase alfa), HSUVs were derived by converting improvement in 6MWT and forced vital capacity (FVC) collected in natural history studies and combining these with the correlation observed between 6MWT, FVC and QoL from the patient and families survey. For each additional benefit reported by patients not captured in 6MWT or FVC, an HSUV increment was derived from the literature. The HST Committee highlighted that the data were not collected within a trial but recognised the challenges in collecting QoL data from children alongside the lack of validated PROMs.

Use and influence of PRO evidence and HSUV estimates in HAS, G-BA and ZIN appraisals

Comparing the appraisal of PRO evidence by NICE with those by ZIN, G-BA and HAS, a number of observations arose (Table 5). First, a proportion of the appraisal reports did not include any detail about QoL evidence (38% for ZIN, 61% for HAS, and 16% for G-BA). Second, a vast majority of those that did report QoL data were deemed inconclusive. The main reasons were the lack of statistical significance (ZIN, HAS, G-BA), the exploratory nature of the evidence, e.g. secondary endpoint (HAS), the non-inclusion of a hierarchical test (HAS), the lack of validated or non-clinically relevant endpoint (G-BA). Third, in the few cases when QoL was considered to be improved by treatment in one country, a different outcome was determined in the other countries. Only one treatment appraised by HAS (inotersen) was considered to provide a moderate improvement in QoL, as it was one of the trial's co-primary endpoints; whereas no meaningful clinically relevant change was recognised by NICE and G-BA. Two treatments appraised by G-BA, patisiran and lanadelumab, were considered to provide some benefit as they were both validated and clinically relevant endpoints. For ZIN, it was

unclear whether the PRO evidence had any influence on the decisions and the HSUVs appraised for three treatments were considered very uncertain.

<Table 5. Use and influence of PRO evidence in HAS, G-BA and ZIN appraisals of non-oncology rare disease treatments>

Impact on carers

Eighteen of the treatments were considered to have an impact on carers (Table 1), whereas evidence on carer impact was considered for only nine of these by NICE (8 HST and 1 TA). Impact of disease and treatment on carers was considered either qualitatively or quantitatively through HSUVs. In the former case, the relevant committees discussed the burden on carers during the deliberative process (mepolizumab, strimvelis, asfotase alfa), and in other cases, considered evidence from patient/carer surveys (eculizumab, elosulfase alfa). In the latter cases, HSUVs were derived from various sources (e.g. published literature, number of carers affected, report on challenges from living and caring for a sick child, or cross-sectional surveys). Some of the HSUVs submitted were changed so as to better align with previous appraisals (patisiran), to only include HSUVs for children (voretigene), to reflect a shorter timeframe (cerliponase), or to reflect a different number of carers (ataluren). In four of these cases, carer disutility was uncertain (also in the decision). Carer QoL was not reported in the other countries.

Discussion

This study explored the appraisal of QoL in all the non-oncology rare disease treatments considered by NICE. It is the first study of this type, which furthers our understanding of the nature of QoL evidence and the nuances of its use in HTA of rare disease treatments.

Our results primarily enable a better understanding of whether the QoL evidence was actually considered. The vast majority of conditions investigated, particularly in NICE's HST programme, are life-threatening and/or debilitating. For all of the treatments investigated, their added benefit was also considered to improve QoL. Measuring their impact on QoL is therefore critical in determining their added benefit, particularly for those treatments aiming solely to improve QoL. This, however, is not reflected in our results. PRO evidence was not reported for a large number of treatments across all of the study countries, and when reported, most of the PROMs and results were not discussed (and therefore we assume not accounted for). In the other study countries, no PRO evidence was reported in 16%, 38% and 61% respectively in Germany, the Netherlands and France. It was not clear from the appraisal reports why this evidence was not reported nor accounted for. When PRO evidence was reported, it was limited to one or two PROMs (versus more in the NICE reports).

Overall, a large amount of QoL data was collected, but these data were barely reported or referred to in the appraisal reports across the different study countries. However, the 28 different PROMs identified and collected in the trials are most likely covering concepts important for patients [5]. Their lack of use points either to a loss of valuable information on the patient perspective, issues in capturing meaningful change in rare diseases, or issues in accounting for all of these PROMs within the HTA approach adopted. Results illustrate that different QoL evidence would be considered depending on the HTA approach. For cost-effectiveness oriented approaches, HSUVs are considered within the incremental cost-effectiveness ratio and are derived from PRO data using indirect techniques (e.g. generic preference-based instruments, mapping), or measured directly from patient responses using direct techniques (e.g. time-trade-off) [44]. In countries with an added clinical benefit assessment approach, the PRO data would be considered and interpreted as is without being derived into a numerical HSUV. To help with the comparability and interpretation of the PRO data, generic PROMs are often preferred. Consideration should therefore be given to how this information could be better used in HTA. This could be achieved through greater involvement in early multi-stakeholder dialogues and early scientific advice to better align across HTA bodies and agree on what QoL evidence would be accepted, and a greater acceptance of registry data to leverage early on data on natural history on the disease.

Second, our results enable a better understanding of whether the QoL evidence actually considered was impactful. Results point to a limited influence of PRO evidence in general. In the NICE appraisals, this was because QoL is mainly measured by HSUVs used in the economic models. PRO evidence was considered to support the interpretation of HSUVs included in the model in one case, and potentially in a few other cases; but overall, its influence was fairly limited. Just over 1/3 of the HSUVs were accepted, even if, in some cases, they were recognised as not ideal. In the remaining cases, the HSUVs were highly uncertain and in most cases the relevant committee recognised that all benefits were not captured. In these cases, interpretation was informed by information from patient and clinicians in four cases, and a patient survey in one case.

Only three disease-specific PROMs were reported, but their consideration had a limited influence on the decision and in only one case, it was mapped to derive HSUVs. This confirms the issue of a lack of validated disease-specific preference-based PROMs and their conversion into HSUVs through mapping [8, 45, 46]. Disease-group PROMs were more frequently used and may constitute a suitable alternative for rare diseases; however, their influence was also limited. A similar situation was seen around the use of symptom-specific PROMs. By contrast, there were a number of cases where the relevant committees recognised that the QoL evidence did not capture the full range of dimensions important to patients. These related to improvements in QoL, such as the ability to return to work, to perform daily activities, to have a social life, to maintain independence and dignity, improving in walking, better tolerability profile, reduced dosing frequency, or improved patient choice, as well as decrements in QoL, such as the impact from relying on wheelchairs, or adverse events not captured. However, considering that many of these domains are typically covered in PROMs, the issue may be more around the lack of sensitivity of these measures rather than domains not being captured.

In Germany and France, most of the PRO evidence was considered inconclusive due to the frequent lack of a statistically significant improvement and/or they did not meet the country-specific evidentiary requirements. In Germany, PROMs need to be validated and PRO evidence clinically relevant (based on a minimally important difference (MID)). However, a treatment failing to meet the MID criterion does not imply lack of improvement across all patients, where there may be some patients improving above the MID and others under [47]. This may be more frequent in heterogeneous and small patient populations [48]. In France, the PRO endpoint should be a significant one (e.g. primary endpoint). In only one case in France and two cases in Germany was QoL considered improved, and this concerned different treatments. Similarly in the Netherlands, the PRO evidence was generally inconclusive and the HSUVs reported in three cases were considered very uncertain.

Overall, findings suggest that a big proportion of the PRO evidence and HSUVs appraised are either not considered or provide inconclusive uncertain outcomes. The main contrast between NICE and the other countries is their willingness to account for other forms of evidence, such as patient surveys or expert input to provide additional and complementary information on QoL impact. They also appear to be more flexible when interpreting QoL evidence, e.g. in recognising that all benefits are not captured by the measures used, and account for that when making their decisions.

We then tried to understand whether the issues highlighted by the relevant committees related to nature of rare disease treatments. One main distinction seen in NICE's HST Programme is a greater likelihood of treatments targeting children/infants or treating heterogeneous and/or multi-systemic conditions. In the 15 NICE appraisals affecting children, only three considered children-specific PROMs (PedsQL) and none considered any proxy-reported PRO evidence. This confirms the frequent lack of validated measures in children [10], but does not reflect the common reliance on proxy-reported data [9]. In only one case was the PedsQL mapped to EQ-5D, but results were limited and the challenges in collecting data from children recognised (together with another case).

The extent to which it may be more difficult to capture meaningful and generalisable outcomes in heterogenous populations and conditions affecting multiple organs [9, 10, 49] was not entirely clear from the results. There were, however, cases where evidence on QoL was lacking to estimate the HSUVs required by the model (ataluren), to capture all relevant symptoms (asfotase alfa), or to deal with multiple co-morbidities (mannitol).

Patient numbers for three HST treatments were small (an incidence of 1-7 patients/year in England), possibly resulting in uncertain aggregated results [9]. In one case (cerliponase), the HST Committee recognised an initial improvement in QoL based on PRO evidence. However, vignettes were used to derive HSUVs due to the lack of correspondence of PRO evidence with health states. The other two cases either did not report (asfotase alfa), nor collect (strimvelis) any PRO evidence, and published literature was used to derive HSUVs.

No existing treatments were available for almost 60% of the 12 HST and 17% of the 12 TA treatments (in total, 9 of 24 treatments). Current standards of care for these diseases require multi-disciplinary specialised services and are considered burdensome for patients and their carers. They generally entail monitoring of disease, management of symptoms, complications or disability, and/or supportive

care (e.g. counselling, occupational therapy, physiotherapy, social care, palliative care, etc.). This may create additional challenges in identifying the relevant domains of QoL to measure in the comparative arm [50].

Three quarters of the conditions appraised affect QoL of families and carers, and the treatments were considered to improve their QoL. None of the PRO evidence collected and reported related to carer burden. However, the NICE Committees did account for the impact on carers either qualitatively or in cases where impact on carer's QoL was collected within a patient and carer survey (eculizumab). On the other hand, carer HSUVs were estimated in only eight cases for which more than half the data were uncertain or inconclusive. This further emphasises the tendency for inconsistent inclusion of carer HSUVs and the variety of approaches used for their measurement [51]. There is a need for methodological guidance on when and how to include carer HSUVs in QALY and non-QALY approaches to HTA [52]. Considering that 80% of rare diseases affect children, and are often severe and disabling, including carer QoL is crucial in determining the added benefit of a new treatment.

Limitations

This study is not without limitations. First, it relies on information from a small number of appraisals, which is unavoidable given the small number of RDTs (excluding oncology treatments) considered each year. Secondly, it relies on official reports, which may not comprehensively depict the full appraisal process. This was more pronounced for some study countries that do not provide detail of their appraisal of the evidence. Based on expectations around transparency, we considered that the items documented in the HTA reports included the most important determinants of decisions. Further, there may have been some limitations relating to language barriers given the use of google translator for some of the countries. However, no inconsistencies across countries were identified that could indicate missing or misinterpreted information. Additionally, our document analysis was qualitative and as a result, we may have missed or misinterpreted some aspects leading to the decision. Given the complexity of some of these appraisals, it was challenging to identify explanations for some of the limitations highlighted, and how they related to the nature of rare diseases. However, we attempted to identify some possible explanations and examples on some of the implications. Finally, this study highlights some of the nuances in considering QoL evidence in rare diseases. It is possible that some of the same issues could arise in the HTA of more common diseases. Further research would be needed to compare the results from this analysis with those from a similar analysis of HTAs for common non-cancer treatments.

Conclusions

This study highlights some of the limitations and challenges in appraising PRO evidence and HSUVs to understand the impacts of a rare condition and treatments on QoL, and the influence of these aspects on determination of value. In many cases, PRO evidence did not have a major influence in

HTA decisions, as it often did not demonstrate meaningful change or was inconclusive. The HSUVs were often very uncertain due to numerous reasons, such as being insensitive to change, ceiling effects, limited face validity, not capturing all domains important to patients, lack of long-term data or methodological issues. This emphasises the need for improved development, testing, use and reporting of PRO evidence, and use of HSUVs that are better adapted to rare disease specificities, such as small sample sizes. HTA bodies would also benefit from greater flexibility in accepting less conventional techniques to derive HSUVs, for example, using vignettes, but there is a need to develop methodologies that support their robust development and application. Additionally, patient evidence, including patient surveys, focus groups, interviews, and expert testimony, have shown to be crucial for providing information about the burden of illness, treatment benefits including outcomes that matter most, and in supporting the interpretation of uncertain aspects of the QoL evidence considered important for the decision.

REFERENCES

1. European Commission: Orphan medicinal products, https://ec.europa.eu/health/human-use/orphan-medicines_en
2. Eurordis: About rare diseases, https://www.eurordis.org/about-rare-diseases
3. Bogart, K.R., Irvin, V.L.: Health-related quality of life among adults with diverse rare disorders. Orphanet J. Rare Dis. (2017). https://doi.org/10.1186/s13023-017-0730-1
4. Drummond, M., Sculpher, M., Claxton, K., Stoddart, G., Torrance, G.: Measuring and valuing health effects. In: Methods for economic evaluation in health care. pp. 123–180. Oxford University Press, Oxford (2015)
5. FDA: Guidance for industry - Patient-reported outcome measures: use in medical product development to support labeling claims. (2009)
6. Kingsley, C., Patel, S.: Patient-reported outcome measures and patient-reported experience measures. BJA Educ. 17, 137–144 (2017). https://doi.org/10.1093/bjaed/mkw060
7. Brazier, J., Ara, R., Azzabi, I., Busschbach, J., Chevrou-Séverac, H., Crawford, B., Cruz, L., Karnon, J., Lloyd, A., Paisley, S., Pickard, A.S.: Identification, Review, and Use of Health State Utilities in Cost-Effectiveness Models: An ISPOR Good Practices for Outcomes Research Task Force Report. Value Heal. (2019). https://doi.org/10.1016/j.jval.2019.01.004
8. Whittal, A., Meragaglia, M., Nicod, E.: The use of patient-reported outcome measures (PROMs) in rare diseases and implications for HTA. Patient. under revi, (2020)
9. Benjamin, K., Vernon, M.K., Patrick, D.L., Perfetto, E., Nestler-Parr, S., Burke, L.: Patient-Reported Outcome and Observer-Reported Outcome Assessment in Rare Disease Clinical Trials: An ISPOR COA Emerging Good Practices Task Force Report. Value Heal. (2017). https://doi.org/10.1016/j.jval.2017.05.015
10. Slade, A., Isa, F., Kyte, D., Pankhurst, T., Kerecuk, L., Ferguson, J., Lipkin, G., Calvert, M.: Patient reported outcome measures in rare diseases: A narrative review, (2018)
11. Pearson, I., Rothwell, B., Olaye, A., Knight, C.: Economic Modeling Considerations for Rare Diseases. Value Heal. 21, 515–524 (2018). https://doi.org/10.1016/j.jval.2018.02.008
12. Towse, A., Garau, M.: Appraising ultra-orphan drugs: is cost-per-QALY appropriate? A review of the evidence. (2018)
13. Annemans, L., Aymé, S., Le Cam, Y., Facey, K., Gunther, P., Nicod, E., Reni, M., Roux, J.-L., Schlander, M., Taylor, D., Tomino, C., Torrent-Farnell, J., Upadhyaya, S., Hutchings, A., Le Dez, L.: Recommendations from the European Working Group for Value Assessment and Funding Processes in Rare Diseases (ORPH-VAL). Orphanet J. Rare Dis. 12, 50 (2017). https://doi.org/10.1186/s13023-017-0601-9
14. Gutierrez, L., Patris, J., Hutchings, A., Cowell, W.: Principles for consistent value assessment

and sustainable funding of orphan drugs in Europe. Orphanet J. Rare Dis. 10, 53 (2015). https://doi.org/10.1186/s13023-015-0269-y

15. Kleijnen, S., Leonardo Alves, T., Meijboom, K., Lipska, I., De Boer, A., Leufkens, H.G., Goettsch, W.G.: The impact of quality-of-life data in relative effectiveness assessments of new anti-cancer drugs in European countries. Qual. Life Res. (2017). https://doi.org/10.1007/s11136-017-1574-9

16. National Institute for Health and Care Excellence.: Asfotase alfa for treating paediatric-onset hypophosphatasia. HST6. (2017)

17. National Institute for Health and Care Excellence.: Eculizumab for treating atypical haemolytic uraemic syndrome. HST1. 2015.

18. National Institute for Health and Care Excellence.: Colistimethate sodium and tobramycin dry powders for inhalation for treating pseudomonas lung infection in cystic fibrosis. TA276. (2013)

19. National Institute for Health and Care Excellence.: Migalastat for treating Fabry disease. HST4. (2017)

20. National Institute for Health and Care Excellence.: Inotersen for treating hereditary transthyretin amyloidosis. HST9. (2019)

21. National Institute for Health and Care Excellence.: Voretigene neparvovec for treating inherited retinal dystrophies caused by RPE65 gene mutations. HST11. (2019)

22. National Institute for Health and Care Excellence.: Mannitol dry powder for inhalation for treating cystic fibrosis. TA266. (2012)

23. National Institute for Health and Care Excellence.: Lumacaftor–ivacaftor for treating cystic fibrosis homozygous for the F508del mutation. TA398. (2016)

24. National Institute for Health and Care Excellence.: Elosulfase alfa for treating mucopolysaccharidosis type IVa. HST2. (2015)

25. National Institute for Health and Care Excellence.: Burosumab for treating X-linked hypophosphataemia in children and young people. HST8. (2018)

26. National Institute for Health and Care Excellence.: Strimvelis for treating adenosine deaminase deficiency–severe combined immunodeficiency. HST7. (2018)

27. National Institute for Health and Care Excellence.: Nusinersen for treating spinal muscular atrophy. TA588. (2019)

28. National Institute for Health and Care Excellence.: Letermovir for preventing cytomegalovirus disease after a stem cell transplant. TA591. (2019)

29. National Institute for Health and Care Excellence.: Mepolizumab for treating severe refractory eosinophilic asthma. TA431. (2017)

30. National Institute for Health and Care Excellence.: Patisiran for treating hereditary transthyretin

amyloidosis. HST10. (2019)

31. National Institute for Health and Care Excellence.: Darvadstrocel for treating complex perianal fistulas in Crohn's disease. TA556. (2019)
32. National Institute for Health and Care Excellence.: Eliglustat for treating type 1 Gaucher disease. HST5. (2017)
33. National Institute for Health and Care Excellence.: Holoclar for treating limbal stem cell deficiency after eye burns. TA467. (2017)
34. National Institute for Health and Care Excellence.: Lanadelumab for preventing recurrent attacks of hereditary angioedema. TA606. (2019)
35. National Institute for Health and Care Excellence.: Obeticholic acid for treating primary biliary cholangitis. TA443. (2017)
36. National Institute for Health and Care Excellence.: Cerliponase alfa for treating neuronal ceroid lipofuscinosis type 2. HST12. (2019)
37. National Institute for Health and Care Excellence.: Pirfenidone for treating idiopathic pulmonary fibrosis. TA504. (2018)
38. National Institute for Health and Care Excellence.: Nintedanib for treating idiopathic pulmonary fibrosis. TA379. (2016)
39. National Institute for Health and Care Excellence.: Ataluren for treating Duchenne muscular dystrophy with a nonsense mutation in the dystrophin gene. HST3. (2016)
40. Nicod, E., Kanavos, P.: Developing an evidence-based methodological framework to systematically compare HTA coverage decisions: A mixed methods study, (2016)
41. Bryman, A.: Social Research Methods. , Oxford (2004)
42. Lerman, J.A., Sullivan, E., Barnes, D.A., Haynes, R.J.: The Pediatric Outcomes Data Collection Instrument (PODCI) and functional assessment of patients with unilateral upper extremity deficiencies. J. Pediatr. Orthop. (2005). https://doi.org/10.1097/01.bpo.0000149866.80894.70
43. Jones, P.W., Quirk, F.H., Baveystock, C.M.: The St George's Respiratory Questionnaire. Respir. Med. (1991). https://doi.org/10.1016/S0954-6111(06)80166-6
44. Meregaglia, M., Nicod, E., Drummond, M.: The estimation of health state utility values in rare diseases: overview of existing techniques. Int. J. Technol. Assess. Health Care. (2020). https://doi.org/10.1017/S0266462320000665
45. Guidance to submitting companies for completion of new product assessment form (NPAF). Supplement for medicines for extremely rare conditions (ultra-orphan medicines). , Glasgow (2019)
46. Beusterien, K., Leigh, N., Jackson, C., Miller, R., Mayo, K., Revicki, D.: Integrating preferences

into health status assessment for amyotrophic lateral sclerosis: The ALS Utility Index. Amyotroph. Lateral Scler. 6, 169–176 (2005). https://doi.org/10.1080/14660820410021339

47. Rüther, A., Elstein, D., Wong-Rieger, D., Guyatt, G.: Aspects of patient reported outcomes in rare diseases: A discussion paper. Int. J. Technol. Assess. Health Care. 32, 126–130 (2016). https://doi.org/10.1017/S0266462316000271

48. Schulz, S., Passon, A., Kulig, M., Perleth, M., Matthias, K.: Orphan drug benefits asssessments at the Federal Joint Committee in Germany. In: HTAi conference (2019)

49. Bell, J.A., Galaznik, A., Pompilus, F., Strzok, S., Bejar, R., Scipione, F., Fram, R.J., Faller, D. V., Cano, S., Marquis, P.: A pragmatic patient-reported outcome strategy for rare disease clinical trials: application of the EORTC item library to myelodysplastic syndromes, chronic myelomonocytic leukemia, and acute myeloid leukemia. J. Patient-Reported Outcomes. (2019). https://doi.org/10.1186/s41687-019-0123-4

50. Morel, T., Cano, S.J.: Measuring what matters to rare disease patients - Reflections on the work by the IRDiRC taskforce on patient-centered outcome measures. Orphanet J. Rare Dis. (2017). https://doi.org/10.1186/s13023-017-0718-x

51. Goodrich, K., Kaambwa, B., Al-Janabi, H.: The inclusion of informal care in applied economic evaluation: A review. Value Heal. (2012). https://doi.org/10.1016/j.jval.2012.05.009

52. Pennington, B., Wong, R.: Modelling carer health-related quality of life in NICE Technology Appraisals and Highly Specialised Technologies. (2019)

Consideration of quality of life in the health technology assessments of rare disease treatments

ABSTRACT

Objectives: Challenges with patient--reported outcome (PRO) evidence and health state utility values (HSUVs) in rare diseases exist due to small, heterogeneous populations, lack of disease knowledge ~~of disease~~ and early onset. To better incorporate quality of life (QoL) into Health Technology Assessment ~~(HTA)~~, a clearer understanding of these challenges is needed.

Methods: NICE appraisals of non-oncology treatments with an EMA orphan designation (n=24), and corresponding appraisals in the Netherlands, France, and Germany were included. Document analysis of appraisal reports investigated how PROs/ ~~and~~ HSUVs influenced decision-making and was representative of QoL impact of condition and treatment.

Results: PRO evidence was not included in 6/24 NICE appraisals. When included, it either failed to demonstrate change, capture domains important for patients, or was uncertain. In the other countries, little information was reported~~,~~ and evidence largely did not demonstrate change. In NICE appraisals, HSUVs were derived through the collection of~~using~~ EQ-5D data (~~in~~ 7/24 cases), mapping ~~in~~ (6/24), vignettes ~~in~~ (5/24), and published literature or other techniques (6/24)~~in the remainder~~. The majority did not use data collected alongside clinical trials. Few measures demonstrated significant change due to lack of sensitivity or face validity, short-term data, or implausible health states. In 8/24 NICE appraisals, patient surveys or input during appraisal committee meetings supported the interpretation of uncertainty or provided evidence about QoL.

Conclusions: This study sheds light on the nature of PRO evidence in rare diseases and associated challenges. Results emphasise the need for improved development and use of PRO/HSUVs. Other forms of evidence and expert input are crucial to support better appraisal of uncertain or missing evidence.

Introduction

Rare diseases are conditions affecting a small number of patients (e.g. less than 1/2,000 people in Europe), which are ~~most often~~ life-threatening and/or ~~severely~~ chronically debilitating, frequently genetic and with an early onset [1, 2]. Quality of life (QoL) of patients living with a rare disease is often poor~~er than those living with more prevalent conditions~~ due to multiple aspects affecting functioning [3]. This is partly explained by issues around diagnostic delays~~is~~, and/or a lack of knowledge about the disease, its treatment pathways or treatment options [3]. Given the severity of these conditions and paucity of curative treatments, understanding their impact on QoL is crucial, particularly when assessing the benefit of a new treatment.

Health Technology Assessment (HTA) aims to assess the value of a treatment to inform decisions on whether it should be provided routinely to the relevant patient population. The assessment generally relies on clinical and patient-reported outcome (PRO) endpoints, which provide evidence about health outcomes and impact on patients' wellbeing ~~g~~[4]. In the latter case, PRO evidence is collected directly from patients or proxies using patient-reported outcome measures (PROMs) [5]. PROMs are intended to capture aspects that matter most to patients about the impact of disease and treatment on symptoms, QoL or health status [6].

HTA relies on the critical assessment of added benefit or cost-effectiveness of a treatment. This is then appraised by a Committee, taking account of other relevant factors, who decide on reimbursement (and pricing in some cases). Added benefit is assessed ~~through~~by considering the magnitude and certainty of treatment benefit over existing therapies based on the clinical and PRO evidence presented. The level of benefit is then generally ranked into categories as, for example, in France, where the added benefit (ASMR) is ranked between I and V. In cost-effectiveness assessments, an economic evaluation is conducted that models the progression through ~~possible~~ health states along the care pathway, with and without the new treatment under review~~, e.g. decision tree~~. ~~For each health state, treatment effect and probability of that health state occurring are estimated.~~ In order to assist cost-effectiveness assessments, techniques have been developed to translate PRO evidence into numerical values called health state utility values (HSUVs). HSUVs represent individual preferences for given health states ~~are values~~ measured on a scale between 0, representing dead ~~(or negative values when ranked as worse than dead),~~ and 1, full health (with negative values implying states considered to be worse than dead). These are then merged with survival data (e.g. length of life) into a composite measure called quality-adjusted life-year (QALY). HSUVs represent the utility value associated with the different models' health states, for both treatment and comparator arms [7]. The most common way of deriving HSUVs is currently using indirect techniques, e.g. preference-based instruments such as EQ-5D that are accompanied by an algorithm (or a set of tariffs) providing HSUVs. Tariffs are pre-determined at individual country level by a sample of the general population that uses direct techniques (e.g. time trade-off) to express

preferences for a subset of health states derived from the combination of instrument's dimensions and levels.

~~More nuanced c~~Challenges exist when developing and using PRO evidence and HSUVs for HTAs of rare diseases treatments due to the small and heterogeneous nature of the patient populations, and frequent lack of knowledge about ~~natural history~~ the disease [8, 9]. ~~This leads to challenges around data collection, often requiring multi-country trials, which raise additional challenges to achieve psychometric validation.~~ Additionally, patients are often children or infants who cannot self-~~-~~report and~~,~~ who may ~~also~~ be cognitively impaired or unable to communicate. There may also be distinct challenges around capturing meaningful outcomes, including: difficulty achieving concept validity through concept saturation, use of methods that may not capture aspects important for patients, or selecting the appropriate PROM when natural history is poorly understood. There are also few validated disease-specific PROMs for rare diseases, probably due to the amount of time and resources needed to develop these instruments, which are further complicated by the nature of these diseases [10].

Additional challenges frequently encountered when deriving and using HSUVs for rare diseases include the need for a large number of respondents to minimise random measurement errors (e.g. person-trade-off, development of mapping algorithms), identification of appropriate values corresponding to the model's health states from the existing literature, and QALY~~HSUV~~s being insufficiently sensitive to disease severity or changes that are important for patients [11–14].

Although these challenges in measuring QoL are common to all rare disease treatments, they are most important in treatments of non-oncological diseases, since in cancer the main value of treatment is often increased survival, and many rare disease treatments in cancer are for sub-populations of a more common cancer for which validated QoL measures may be available.

To better incorporate QoL evidence into HTA decision-making, a clearer understanding of the challenges encountered when using PRO evidence and HSUVs in rare diseases is needed. Hence ~~The objective of~~ this research explored ~~is to better understand~~ how QoL evidence has been used in appraisal of non-oncology rare disease treatments in ~~is appraised across~~ a selection of countries using different HTA approaches.

Methods

Study sample

For this EU Horizon2020 project, European c~~C~~ountries were selected to represent those that make decisions based on added clinical benefit and those that focus on cost-effectiveness, and who ~~approaches in HTA with~~have publicly available reports. Those selected were~~:~~ England (~~NICE -~~ National Institute for Health and Care Excellence, NICE) and the Netherlands (~~ZIN –~~ National Health Care Institute, ZIN) as users of~~for~~ the cost-effectiveness approach, and France (~~HAS –~~ Haute Autorité de Santé, HAS) and Germany (~~G-BA –~~ Federal Joint Committee, G-BA) as users of ~~for~~ the added

benefit approach. Considering the depth of the analysis conducted, the inclusion of four countries was considered sufficient to understand the nuances between one HTA approach and another, and the types of contrasts within one approach. ~~In Germany, the Institute for Quality and Efficiency in Health Care (IQWiG) conducts the assessment and G-BA the appraisal[15]. At HAS in France, the Transparency Committee conducts assessments of added benefit. In England, rare disease treatments can be considered either in NICE's general Technology Appraisal (TA) programme or its Highly Specialised Treatments (HST) programme. Due to the greater availability of detailed~~As only the reports from NICE presented detailed information about the committee deliberations of the QoL evidence, the analysis ~~mainly~~ focused on the NICE appraisals, and ~~reflections were made on how these contrast with the~~ other countries' appraisals were used as a contrast to highlight any different approaches.

All treatments with a European Medicines Agency (EMA) orphan medicinal product designation and appraised by NICE within its Technology Appraisal (TA) or Highly Specialised Technologies (HST) ~~either~~ programmes before 1 June 2020 were selected (n=50). Cancer treatments (26) were excluded ~~for the reasons outlined above~~because added benefit often relies on survival gains, and many rare cancers are subsets of more common cancers for which a validated PRO often exists [4, 15]. This left ~~Twenty-four~~24 rare disease treatments ~~were selected~~ (12 TA and 12 HST) for analysis [16, 17, 26–35, 18, 36–39, 19–25]~~.~~ In the results section, the information reported for the individual NICE sample was extracted from these reports.

Data collection and analysis

~~Data collection consisted of extracting all relevant i~~Information about QoL, PROs, HSUVs and other evidence from patients about their QoL (~~patient evidence,~~ such as~~e —~~ patient group submissions and ~~,~~ patient expert input~~t, etc)~~ was extracted from NICE's appraisal reports. These ~~are~~were sufficiently detailed to enable documentation of the~~ing~~ ~~the~~ source of the evidence, results, issues highlighted by the appraisal committee and the influence on the decision. ~~and, i~~If needed, ~~from the~~ supporting documentation ~~available on their website (e.g.,~~such as ~~committee papers,~~ manufacturer submissions and Evidence Review Group (ERG) ~~assessment group~~ reports were reviewed. ~~)~~A published ~~. Two sets of information about (1) the appraisal of QoL were extracted, and (2) the burden of disease and impact of treatment on QoL, and (2) the appraisal of QoL were extracted. The following information was collected regarding the burden of disease: infant/childhood onset; progressive; heterogeneous; multi-systemic; debilitating; life-threatening; supportive care; and regarding the impact of treatment on QoL: length of life improved; QoL improved from reduced symptoms, daily living, families/carers, compared to current treatment, administration mode. The information extracted about the appraisal of QoL was based on an existing methodological~~ framework was used to extract ~~in a structured way~~ these key aspects of the appraisal~~-~~ in a structured way ~~-~~[40]~~.~~ ~~included: (1) PRO evidence, HSUVs or other patient-based evidence considered, (2) their source (e.g. trial, patient/clinical input), (32) their outcome (e.g. meaningful change), (43) the issues highlighted by the Committee, and (45) their influence on the decision~~ PRO evidence was categorised on the basis of the type of PROM used

(generic, disease-group (developed for a range of conditions), disease-specific or symptom-specific). ~~c), and~~ HSUVs were categorised on the basis of the technique used to derive them (e.g.~~,~~ collection using a~~n~~t instrument such as EQ-5D, or mapping from other PROMs). ~~Patient-based evidence was defined as any other evidence on patient QoL that is not PRO or HSUV evidence. This includes evidence from surveys (e.g. on patient and families QoL impact, or administration modes preferences), and patient or clinical input supporting the interpretation of the QoL evidence during the appraisal process.~~

Thematic analysis was undertaken to identify ~~the lastlatter two items (~~issues arising in appraisals and their influence on decision-making~~)~~ based on the researchers' interpretation of the discussion reported in the published documents. ~~This consisted in identifying the themes arising from the information extracted for each bucket outlined in the previous paragraph, and categorising these.~~ The identification of themes was done iteratively and was continuously refined while the researchers familiarised themselves with the data and grouped the data in a logical way to allow for a better understanding of the decision process [41]. Once the themes were identified and categorised, the researchers assessed the ~~The~~ level of influence of PRO evidence or patient evidence on decisions. This ~~was~~ was categorised as "influence"~~,~~ when the Committee explicitly recognised and accounted for a change in QoL in their decisions, "possible influence"~~,~~ when PRO evidence or patient~~.~~ evidence was explicitly reported but considered limited by the Committee and it was therefore unclear whether it influenced the decision, and~~;~~ "no influence"~~,~~ when PRO evidence or patient evidence was reported, but inconclusive or ~~failing~~ failed to demonstrate change and did not influence the Committee's decision. HSUVs were considered in all cases to have influence on the decision, since the Committee always took note of the incremental cost per QALY ratio. The interpretation of the HSUV evidence was distinguished as "accepted" when the Committee recognised the evidence presented was acceptable, "not commented" when no issues were raised with the HSUVs presented and therefore there ~~is~~ was a high likelihood that it was accepted, and "uncertain" when a number of issues around the HSUV were highlighted, which rendered their interpretation challenging.

The second set of ~~information~~ extracted information related to ~~The following information was collected regarding~~ the burden of disease and treatment impact: infant/childhood onset; progressive; heterogeneous; multi-systemic; debilitating; life-threatening; supportive care; and regarding the impact of treatment on QoL: length of life improved~~;~~ QoL improved from reduced symptoms, daily living, families/carers, compared to current treatment, administration mode.

The burden of disease and intended impact of the treatments reported in the HST and TA programmes, respectively, were extracted and compared.

The analysis aimed to understand how QoL was appraised, and the extent to which PRO evidence and/or HSUVs were considered appropriate. The possible influence of the nature of the rare diseases on PRO evidence and HSUV estimates was also explored to generate a better understanding of what ~~is~~ was feasible in the different contexts. In the cross-country analysis, the information reported by the

other countries was scarce. The focus was therefore on the PRO evidence and HSUVs considered and their influence on the decision.

Results

Impact of disease and treatment on quality of life

Most of the diseases undergoing the TA and HST processes ~~are~~ were life-threatening and/or debilitating (Figure 1). The burden of disease, however, was greater in the diseases undergoing the HST programme compared with~~to~~ the TA ~~in terms~~ in that ~~of~~ these diseases ~~affecting~~ affect children, ~~their~~ have a ~~heterogeneity~~ heterogeneous and progressive nature, or ~~fact that they~~ affect multiple organs. With the exception of the prophylaxis treatment letermovir, the~~y~~ symptoms of all of the diseases analysed affect patients' daily living and QoL. No previous treatments were available for 58% (7/12) and 17% (2/12) of those undergoing the HST and TA processes, respectively.

In terms of the intended effects of treatment, 67% (8/12) of HST and 83% (10/12) of TA treatments aim to improve length of life, while all improve patients' daily living and QoL by reducing symptoms (with the exception of letermovir). Six of these aim solely to improve QoL. Further more, all of the HST and 83% (10/12) of the TA treatments aim to improve patients' daily living and QoL over standard of care. In 50% of all cases (12/24), QoL improvement is linked to a different administration mode.

All conditions appraised by HST and half of those by TA were considered to affect carer QoL. In all cases, with the exception of letermovir, the treatment intends to improve their QoL.

The estimated yearly number of patients to be treated in England ranged between 1-50 for 10 of the 12 HST treatments [1-7 patients for 3 treatments, 20-35 for 2 treatments, 50-100 for 3 treatments, and 140-150 for 2 treatments]. No details about patient numbers were provided in all other cases.

<Figure 1. Proportion of appraisals for which various items of burden of disease and treatment impact are relevant in NICE Highly Specialised Technology and Technology Appraisal programmes (n=24)>

Use and i~~I~~nfluence of PRO evidence in NICE appraisals

In NICE appraisals, PROMs can have an~~d~~ influence either through being considered directly~~,~~ and/or through ~~its~~their use in generating HSUVs. Across the 24 treatments appraised by NICE, 28 different PROMs were reported. This included 10 generic PROMs considered across 14 treatments, seven disease-group PROMs across seven treatments, three disease-specific PROMs across five treatments, and eight symptom-specific PROMs across ~~six~~ seven treatments (Table ~~Figure 2~~1). Several PROMs ~~may have been~~ could considered for the same treatments. Examples of disease-group PROMs include the Paediatric Outcomes Data Collection Instrument (PDOCI) measuring functional outcomes in paediatric orthopedics [42], and the St George's Respiratory Questionnaire (SGRQ) measuring overall health, daily life, and perceived well-being in patients with obstructive

airways disease [43]. The three disease-specific PROMs considered were for cystic fibrosis (Cystic Fibrosis Questionnaire Revised questionnaire, CFQ-R), recurrent angioedema (Angioedema Quality of Life questionnaire, AE-QoL), and neuronal ceroid lipofuscinosis type 2 (Neuronal Ceroid Lipofuscinosis Type 2 Quality of Life Instrument, CLN2-QoL). The seven symptom-specific PROMs related to pain, gastro-intestinal symptoms, diabetic neuropathy, fatigue, asthma, and anxiety and depression (Table 2~~1~~).

Forty-two percent (10/24) of submissions did not include any generic PRO, and 25% (6/24) no PRO evidence at all. Reasons for the latter included PRO evidence not collected in trials (elosulfase alfa, obetocholic acid, holoclar), collected but limited (strimvelis), being collected and not reported (burosumab), or not presented given it was a re-assessment based on new clinical evidence (pirfenidone). In two of the cases without PRO evidence, other QoL evidence ~~were~~ was considered, such as observational studies and cross-sectional surveys involving patients and families (elosulfase alfa), and visual acuity data from the literature used to derive HSUVs (holoclar) (~~Figure 2~~Table 1). QoL evidence for the remaining drugs without PRO data was based on HSUVs derived from vignettes or published evidence (Table 3, discussed in the next section on HSUVs).

<~~Figure~~ Table 1~~2~~. Types and influence of PRO evidence considered in NICE TA and HST appraisals of non-oncology rare disease treatments (n=24)>

<Table ~~1~~2. Use and influence of PRO evidence in NICE TA and HST appraisals of non-oncology rare disease treatments (n=24)>

Further exploration of the influence of ~~the~~ PRO evidence on NICE decisions suggested~~s~~ that beyond those used to derive HSUVs, few of them had any influence on the decisions (Tables 1 and 2~~Figure 2, Table 1~~).

Of the 14 appraisals considering generic PRO evidence, eight were used to derive ~~the economic models'~~ HSUVs and the remaining six had unclear or no influence on the decisions. For asfotase alpha, the EQ-5D data collected in a patient survey may have been considered by clinicians when developing the vignette's health states, but it is not discussed in the report. For cerliponase, it was inconclusive due to the lack of correspondence between EQ-5D and the model's health states, and short trial duration for the Pediatric Quality of Life Inventory (PedsQL). For ataluren, no significant improvements in the PedsQL were shown, despite the positive trend in the functioning subscale. For the remaining treatments (migalastat, letermovir, lanadelumab), the SF36 and EQ-5D collected did not show any significant improvements and were not considered.

With the exception of one disease-group PROM used to derive the economic model's HSUVs, their inclusion had limited influence. This was the case for mepolizumab, where SGRQ data, suggesting improved QoL due to fewer exacerbations and improved symptom control and lung function, was mapped to EQ-5D to obtain HSUVs. In the other cases, the PODCI data collected for ataluren showed improvements ~~in~~ on two dimensions, but ~~were~~ was considered uncertain due to the short trial

duration. In all other cases (letermovir, asfotase alfa, voretigene, darvadstrocel and nintedanib), the disease-group PROMs, Functional Assessment of Cancer Therapy - Bone Marrow Transplantation (FACT-BMT), PODCI, Visual Function Questionnaire (VFQ), perianal disease activity index (PDAI), SGRQ or Shortness of Breath Questionnaire (SOBQ) either did not show a significant improvement or were not reported.

A similar situation was seen for the disease-specific PROMs. For only one case, colistimethate sodium and tobramycin DPI, the CFQ-R was mapped to HSUVs and used for the decision. However, it did not show any improvement in QoL relating to administration mode (dry powders for inhalation versus nebuliser) given a non-inferiority trial design was adopted. For three treatments, the PRO evidence ~~were~~ was uncertain and thus ~~their~~ the influence on the decision was unclear. The data collection period of CLN2-QoL for cerliponase was considered too short, and the CFQ-R data collected for mannitol dry and lumacaftor-ivacaftor ~~was not statistically improved~~did not show a statistically significant improvement. Results from the AE-QoL data collected for lanadelumab were not commented on in the appraisal report.

Of the six treatments that considered symptom-specific PROMs, one of them influenced and another possibly influenced the decision. For patisiran, the Neuropathy Impairment Score (NIS) and Norfolk Quality of Life Questionnaire - Diabetic Neuropathy (Norfolk QoL-DN) data collected was statistically improved and contributed to recognis~~z~~ing treatment effectiveness. For eliglustat, no significant improvements were demonstrated for the Fatigue Severity Scale (FSS) and Brief Pain Inventory (BPI), and it was unclear whether they were used to determine the HSUV estimated to measure the impact of ~~on QoL of treatment s for~~ adverse events, on QoL included in the submission. In the remaining cases, there was either no demonstration of change with BPI and Gastrointestinal Symptoms Rating Scale (GSRS) for migalastat and with Norfolk QoL-DN for inotersen, or results were not reported (Cough and Sputum Assessment Questionnaire, CASA-Q for nintedanib, Asthma Control Questionnaire, ASQ for mepolizumab and Hospital Anxiety Depression Scale, HADS for lanadelumab).

For eight of these drugs, determination of QoL impact was influenced by patient ~~based~~ evidence (Table 2). First, patient surveys provided information about impact of QoL on patients and carers (eculizumab, ataluren), preferences for administration mode (eliglustat, colistimethate sodium and tobramycin DPI), or ~~were~~ whether it was used to derive HSUVs (elosulfase alfa). Respondents were patients and in one case also family members, three~~wo~~ formed part of the company submissions and the other two, patient submissions. Second, patients and clinicians provided input about the dimensions not captured in the model (patisiran), about impact on QoL (letermovir), effect on tolerability (nintedanib), and administration mode (migalastat, eliglustat, colistimethate sodium and tobramycin DPI).

Use and i~~I~~nfluence of HSUV estimates in NICE appraisals

The most frequently used technique to derive HSUVs in NICE appraisals was through ~~the~~ ~~the administration of~~ EQ-5D data (7/23) collected within a trial (4/7) or from a registry or cohort study

(3/7), followed by mapping (6/23), vignettes (5/23), published literature (3/23), Health Utility Index Mark 2 (HUI2) (1/23) and other (1/23) (~~Figure~~ Table 3). No HSUVs were reported for one treatment (pirfenidone) given it was a re-assessment; therefore, it was excluded from this analysis, which focuse~~d~~s on the 23 remaining treatments. The mapping technique was more frequently used in the TA, and vignettes in the HST process. Additional HSUVs were derived to measure the impact on QoL of ~~relating to~~ adverse events (9/23), of the administration mode (4/23), of carer burden~~s~~ (7/23) or other (7/23) and ~~were~~ considered alongside the HSUV derived.

< ~~Figure~~ Table 3. Techniques used to derive HSUVs in NICE TA and HST appraisals of non-oncology rare disease treatments (n=23) >

The detail and summary of the individual appraisals are summarised in Table 4~~2~~. Seven treatments used EQ-5D, two of which collected EQ-5D 3L in trials and the remaining collected EQ-5D 5L (mapped to 3L) or foreign EQ-5D datasets converted using the UK tariff. In ~~only~~ one case, the HSUVs included in the model were considered acceptable by the TA Committee (lanadelumab). ~~In only one other case,~~ For mannitol, the generic Health Utility Index Mark 2 (HUI2) was used to derive HSUV estimates. Even if EQ-5D would have been preferred, the HUI2 was accepted by the relevant committee. For all remaining cases, a number of issues were raised by the relevant committees, which included benefits (eculizumab, migalastat) or long-term effects not captured (letermovir), measure insensitive to change (nintedanib), uncertain duration (patisiran), or possible implausible health states (inotersen). ~~In only one other case, mannitol, the generic Health Utility Index Mark 2 (HUI2) was used to derive HSUV estimates. Even if EQ-5D would have been preferred, the HUI2 was accepted by the relevant committee.~~

<Table 4~~2~~. Use and influence of HSUVs in NICE TA and HST appraisals of non-oncology rare disease treatments (n=23) >

Mapping was used in six cases, in one of which (lumacaftor-ivacaftor) the applicant developed a new algorithm, while in the others published functions were used. Source measures included lung function and pulmonary exacerbation (lumacaftor–ivacaftor), SF36 (eliglustat), PedsQL (nursinersen), CFQ-R (colistimethate sodium and tobramycin DPI), SGRQ (mepolizumab) and visual acuity (holoclar); all were converted to EQ-5D-3L. The results were considered acceptable in only one case (mepolizumab), or not commented on (likely acceptable) in two cases (eliglustat, holoclar). The issues raised regarding the remaining cases included: ceiling effects and little change captured even though it was collected in the largest existing cystic fibrosis trial (lumacaftor-ivacaftor), limited face validity resulting in expert elicitation being used to estimate the HSUVs (nu~~r~~sinersen), or limited methodological approach (colistimethate sodium and tobramycin DPI).

Vignettes were used in five cases. Reasons for their use over more conventional approaches included ~~the~~ a lack of correspondence between QoL data collected in the clinical trial and model health states

(cerliponase), lack of negative values when deriving the PedsQL being considered unrealistic considering the condition's severity (cerliponase), or QoL data not collected in trial (darvadstrocel, burosumab, voretigene). The health states were developed by patient and clinical experts (voretigene), or only clinicians (cerliponase, asfotase alpha, burosumab). Respondents included clinicians (voretigene, cerliponase, asfotase alpha, burosumab), or patients and public (darvadstrocel). The QoL measure included was EQ-5D-5L (cerliponase, asfotase alpha, burosumab), and HUI2 and EQ-5D (voretigene).

A number of issues were raised about the vignettes. For voretigene, poor convergent validity between EQ-5D and HUI2 and preference for EQ-5D (considered to better capture overall QoL over HUI2) were highlighted. For alfotase alfa, trial data would have been preferred over vignettes by the appraisal committee; however, QoL results from the vignette were compared to results from a patients survey and considered aligned. Additionally, given the health states were based on the surrogate outcome "six--minute walking test" (6MWT), all of the relevant symptoms that would produce lower HSUVs in the more severe states may not have been captured (likely underestimate). The HST Committee was also concerned with clinicians responding to the vignettes instead of patients (burosumab). Furthermore, Tthere was also concern about the uncertain robustness of the vignettes given an unclear association of other elements (e.g., pain) to health states (cerliponase).

Published literature was used in three cases. This was because QoL was not measured in the trials (strimvelis, obeticholic acid) or the available mapping algorithm was conducted on a healthy population and thus unsuitable (ataluren). No detail on the published literature was provided for strimvelis and ataluren, whereas for obeticholic acid, values from an analogue disease (Hep C) were used.

In one case (elosulfase alfa), HSUVs were derived by converting improvement in 6MWT and forced vital capacity (FVC) collected in natural history studies and combining these with the correlation observed between 6MWT, FVC and QoL from the patient and families survey. For each additional benefit reported by patients not captured in 6MWT or FVC, an HSUV increment was derived from the literature. The HST Committee highlighted that the data were not collected within a trial, but recognised the challenges in collecting QoL data from children alongside the lack of validated PROMs.

Use and influence of PRO evidence and HSUV estimates in HAS, G-BA and ZIN appraisals

Comparing the appraisal of PRO evidence by NICE with those by ZIN, G-BA and HAS, a number of observations arose (Table 45). First, a proportion of the appraisal reports do did not include any detail about QoL evidence (38% for ZIN, 61% for HAS, and 16% for G-BA). Second, a vast majority of those that did report QoL data were deemed inconclusive. The main reasons were the lack of statistical significance (ZIN, HAS, G-BA), the exploratory nature of the evidence, e.g. secondary endpoint (HAS), the non-inclusion of a hierarchical test (HAS), the lack of validated or non-clinically relevant endpoint (G-BA). Third, in the few cases when QoL was considered to be improved by treatment in one country, a different outcome was determined in the other countries. Only one treatment,

~~inotersen,~~ appraised by HAS (inotersen) was considered to provide a moderate improvement in QoL, as it was one of the trial's co-primary endpoints; whereas no meaningful clinically relevant change was recognised by NICE and G-BA. Two treatments~~, patisiran and lanadelumab,~~ appraised by G-BA, patisiran and lanadelumab, were considered to provide some benefit as they were both validated and clinically relevant endpoints. For ZIN, it was unclear whether the PRO evidence had any influence on the decisions and the HSUVs appraised for three treatments were considered very uncertain.

<Table 5~~4~~. Use and influence of PRO evidence in HAS, G-BA and ZIN appraisals of non-oncology rare disease treatments>

Impact on carers

Eighteen of the treatments were considered to have an impact on carers (Table 1), whereas evidence on carer impact was considered for only nine of these by NICE (8 HST and 1 TA). Impact of disease and treatment on carers was considered either qualitatively or quantitatively through HSUVs. In the former case, the relevant committees discussed ~~during the deliberative process~~ the burden on carers during the deliberative process (mepolizumab, strimvelis, asfotase alfa), and in other cases, considered evidence from patient/carer surveys (eculizumab, elosulfase alfa). In the latter cases, HSUVs were derived from various sources (e.g. published literature, number of carers affected, report on challenges from living and caring for a sick child, or cross-sectional surveys). Some of the HSUVs submitted were changed~~s~~ so as to better align with previous appraisals (patisiran), to ~~include~~ only include ~~children~~ HSUVs for children (voretigene), to reflect a shorter timeframe (cerliponase), or to reflect a different number of carers (ataluren). In four of these cases, carer disutility was uncertain (also in the decision). Carer QoL was not reported in the other countries.

Discussion

This study explored the appraisal of QoL ~~of~~in ~~a sample of~~all the non-oncology rare disease treatments considered by NICE. It is ~~a~~the first study of this type, which furthers our understanding of the nature of QoL evidence and ~~its~~the nuances of its use in ~~for~~ HTA ~~in~~of rare disease treatments.

Our results primarily enable a better understanding of whether the QoL evidence was actually considered. The vast majority of conditions investigated, particularly in NICE's HST programme, are life-threatening and/or debilitating. For all of the treatments investigated, their added benefit was also considered to improve QoL. ~~All of the treatments investigated aimed to improve QoL and, for 75% of these, also length of life.~~ Measuring their impact on QoL is therefore critical in determining their added benefit, particularly for those treatments aiming solely to improve QoL. This, however, is not reflected in our results. PRO evidence was not reported for a large number of treatments across all of the study countries, and when reported, most of the PROMs and results were not discussed (and therefore we

assume not accounted for). In the other study countries, no PRO evidence was reported in 16%, 38% and 61% respectively in Germany, the Netherlands and France. It was not clear from the appraisal reports why this evidence was not reported nor accounted for. When PRO evidence was reported, it was limited to one or two PROMs (versus more in the NICE reports).

Overall, a large amount of QoL data was collected, but these data were barely reported or referred to in the appraisal reports across the different study countries. However, the 28 different PROMs identified and being collected in the trials are most likely covering concepts important for patients [5]. Their lack of use points either to a loss of valuable information on the patient perspective, issues in capturing meaningful change in rare diseases, or issues in accounting for all of these PROMs within the HTA approach adopted. Results illustrate that different QoL evidence would be considered depending on the HTA approach. For cost-effectiveness oriented approaches, HSUVs are considered within the incremental cost-effectiveness ratio and are derived from PRO data using indirect preference-based techniques (e.g., generic preference-basedmulti-attribute utility instruments, mapping), or measured directly from patient responses using direct preference-based techniques (e.g., time-trade-off), vignettes) [44]. In countries with an added clinical benefit assessment approach, the PRO data would be considered and interpreted as is without being derived into a numerical HSUV. To help with the comparability and interpretation of the PRO data, generic PROMs are often preferred. Consideration should therefore be given to how this information could be better used in HTA. This could be achieved through greater involvement in early multi-stakeholder early dialogues and early scientific advice to better align across HTA bodies and agree on what QoL evidence would be accepted, and a greater acceptance of registry data to leverage early on data on natural history on the disease.

Second, our results enable a better understanding of whether the QoL evidence actually considered was impactful. Results point to a limited influence of PRO evidence in general. In the NICE appraisals, this was because QoL is mainly measured by HSUVs used in the economic models. PRO evidence was considered to support the interpretation of HSUVs included in the model in one case, and potentially in a few other cases; but overall, its influence was fairly limited. Just over 1/3 of the HSUVs were accepted, even if, in some cases, they were recognised as not ideal. In the remaining cases, the HSUVs were highly uncertain and in most cases the relevant committee recognizsed that all benefits were not captured. In these cases, interpretation was informed by information from patient and clinicians in four cases, and a patient survey in one case.

Only three disease-specific PROMs were reported, but their consideration had a limited influence on the decision and in only one case, it was mapped to were included in only one case to derive HSUVs. This confirms the issue of a lack of validated disease-specific (preference-based) PROMs and their conversion into HSUVs through mapping [8, 45, 46][45][8][46]. Disease-group PROMs were more frequently used and may constitute a suitable alternative for rare diseases; however, their influence was also limited. A similar situation was seen around the use of symptom-specific PROMs. By contrast, there were a number of cases where the relevant committees recognizsed that the QoL

evidence did not capture the full range of dimensions important to patients. These related to improvements in QoL, such as the ability to return to work, to perform daily activities, to have a social life, to maintain independence and dignity, improving in walking, better tolerability profile, reduced dosing frequency, or improved patient choice, as well as decrements in QoL, such as the impact from relying on wheelchairs, or adverse events not captured. However, considering that many of these domains are typically covered in PROMs, the issue may be more around the lack of sensitivity of these measures rather than domains not being captured.

In Germany and France, most of the PRO evidence was considered inconclusive due to the frequent lack of a statistically significant improvement and/or ~~as~~ they did not meet the country-specific evidentiary requirements. In Germany, ~~the~~ PROMs need to be validated and ~~the~~ PRO evidence clinically relevant (based on a minimally important difference (MID)). However, a treatment failing to meet the MID criterion does not imply lack of improvement across all patients, where there may be some patients improving above the MID and others under [47]. This may be more frequent in heterogeneous and small patient populations [48]. In France, the PRO endpoint should be a significant one (e.g.~~,~~ primary endpoint). In only one case in France and two cases in Germany was QoL considered improved, and this concerned different treatments. Similarly in the Netherlands, the PRO evidence was generally inconclusive and the HSUVs reported in three cases were considered very uncertain.

Overall, findings suggest that a big proportion of the PRO evidence and HSUVs appraised are either not considered or provide inconclusive uncertain outcomes. The main contrast between NICE and the other countries is their willingness to account for other forms of evidence, such as patient surveys or expert input to provide additional and complementary information on QoL impact. They also appear to be more flexible when interpreting ~~the~~ QoL evidence, e.g.~~,~~ in recognis~~z~~ing that all benefits are not captured by the measures used, and account for that when making their decisions.

We then tried to understand whether the issues highlighted by the relevant committees related to nature of rare disease treatments. One main distinction seen in NICE's HST Programme is a greater likelihood of treatments targeting children/infants~~,~~ or treating heterogeneous and/or multi-systemic conditions. In the 15 NICE appraisals affecting children, only three considered children-specific PROMs (PedsQL) and none considered any proxy-reported PRO evidence. This confirms the frequent lack of validated measures in children [10], but does not reflect the common reliance on proxy-reported data [9]. In only one case was the PedsQL mapped to EQ-5D, but results were limited and the challenges in collecting data from children recognis~~z~~ed (together with another case).

The extent to which it may be more difficult to capture meaningful and generalisable outcomes in heterogenous populations and conditions affecting multiple organs [9, 10, 49] was not entirely clear from the results. There were, however, cases where evidence on QoL was lacking to estimate the HSUVs required by the model (ataluren), to capture all relevant symptoms (asfotase alfa), or to deal with multiple co-morbidities (mannitol).

Patient numbers for three HST treatments were small (an incidence of 1-7 patients/year in England), possibly resulting in uncertain aggregated results [9]. In one case (cerliponase), the HST Committee

recognised an initial improvement in QoL based on PRO evidence. However, vignettes were used to derive HSUVs due to the lack of correspondence of PRO evidence with health states. The other two cases either did not report (asfotase alfa), nor collect (strimvelis) any PRO evidence, and published literature was used to derive HSUVs.

No existing treatments were available for almost 60% of the 12 HST and 17% of the 12 TA treatments (in total, 9 of 24 treatments). Current standards of care for these diseases require multi-disciplinary specialised services and are considered burdensome for patients and their carers. They generally entail monitoring of disease, management of symptoms, complications or disability, and/or supportive care (e.g. counselling, occupational therapy, physiotherapy, social care, palliative care, etc.). This may create additional challenges in identifying the relevant domains of QoL to measure in the comparative arm [50].

Three quarters of the conditions appraised affect QoL of families and carers, and the treatments were considered to improve their QoL. None of the PRO evidence collected and reported related to carer burden. However, the NICE Committees did account for the impact on carers either qualitatively or in cases where impact on carer's QoL was collected within a patient and carer survey (eculizumab). On the other hand, carer HSUVs were estimated in only eight cases for which more than half the data were uncertain or inconclusive. This further emphasises the tendency for inconsistent inclusion of carer HSUVs and the variety of approaches used for their measurement [51]. There is a need for methodological guidance on when and how to include carer HSUVs in QALY and non-QALY approaches to HTA [52]. Considering that 80% of rare diseases affect children, and are often severe and disabling, including carer QoL is crucial in determining the added benefit of a new treatment.

Limitations

This study is not without limitations. First, it relies on information from a small number of appraisals, which is unavoidable given the small number of RDTs (excluding oncology treatments) considered each year. Secondly, it relies on official reports, which may not comprehensively depict the full appraisal process. This was more pronounced for some study countries that do not provide detail of their appraisal of the evidence. Based on expectations around transparency, we considered that the items documented in the HTA reports included the most important determinants of decisions. Further, there may have been some limitations relating to language barriers given the use of google translator for some of the countries. However, no inconsistencies across countries were identified that could indicate missing or misinterpreted information. Additionally, our document analysis was qualitative and as a result, we may have missed or misinterpreted some aspects leading to the decision. Given the complexity of some of these appraisals, it was challenging to identify explanations for some of the limitations highlighted, and how they related to the nature of rare diseases. However, we attempted to identify some possible explanations and examples on some of the implications. Finally, this study highlights some of the nuances in considering QoL evidence in rare diseases. It is possible that some of the same issues could arise in the HTA of more common diseases. Further research would be

needed~~, comparing~~ to compare the results from this analysis with those from a similar analysis of HTAs for common non-cancer treatments.

Conclusions

This study highlights some of the limitations and challenges in ~~using~~appraising PRO evidence and HSUVs to understand the impacts of a rare condition and treatments on QoL, and the influence of these aspects on determination of value ~~in appraising rare disease treatments~~. In many cases, PRO evidence did not ~~failed to~~ have a major influence in HTA decisions, as it~~they~~ often ~~failed to~~ did not demonstrate meaningful change or ~~evidence~~ was inconclusive. The HSUVs were often very uncertain due to ~~various~~ numerous reasons, such as being insensitive to change, ceiling effects, limited face validity, ~~failed to capture~~ not capturing all domains important to patients, lack of long-term data or methodological issues. This emphasises the need for improved development, testing, use and reporting of PRO evidence, and use of HSUVs that are better adapted to rare disease specificities, such as~~e.g.~~ small sample sizes. HTA bodies would also benefit from greater flexibility in accepting ~~use of~~ less conventional techniques to derive ~~QoL evidence, such as~~ HSUVs, for example, using vignettes, but there is a need to develop methodologies that support their robust development and ~~use~~application. ~~vignettes or patient surveys, which are currently often not accepted. However, there is a need to better understand how to develop methodologically sound vignettes and other less conventional evidence.~~ Additionally, ~~patient~~ patient ~~based~~ evidence, including patient surveys, focus groups, interviews, and ~~and~~ expert testimony, ~~etc~~ ~~and patient and clinical~~ ~~and clinical input~~ have shown to be crucial ~~in~~ for providing information about the burden of illness, treatment benefits including outcomes ~~those~~ that matter most, and in ~~helping in~~ supporting the interpretation of uncertain aspects of the QoL evidence considered important for the decision. ~~Patients should be better informed about the different types of evidence and input that could be useful for decision making and be involved throughout the process. Future research could compare the techniques used to derive HSUVs in the HTA appraisals with those available in the published research to identify any additional learnings from the application of these techniques in specific disease areas.~~

REFERENCES

1. European Commission: Orphan medicinal products, https://ec.europa.eu/health/human-use/orphan-medicines_en
2. Eurordis: About rare diseases, https://www.eurordis.org/about-rare-diseases
3. Bogart, K.R., Irvin, V.L.: Health-related quality of life among adults with diverse rare disorders. Orphanet J. Rare Dis. (2017). https://doi.org/10.1186/s13023-017-0730-1
4. Drummond, M., Sculpher, M., Claxton, K., Stoddart, G., Torrance, G.: Measuring and valuing health effects. In: Methods for economic evaluation in health care. pp. 123–180. Oxford University Press, Oxford (2015)
5. FDA: Guidance for industry - Patient-reported outcome measures: use in medical product development to support labeling claims. (2009)
6. Kingsley, C., Patel, S.: Patient-reported outcome measures and patient-reported experience measures. BJA Educ. 17, 137–144 (2017). https://doi.org/10.1093/bjaed/mkw060
7. Brazier, J., Ara, R., Azzabi, I., Busschbach, J., Chevrou-Séverac, H., Crawford, B., Cruz, L., Karnon, J., Lloyd, A., Paisley, S., Pickard, A.S.: Identification, Review, and Use of Health State Utilities in Cost-Effectiveness Models: An ISPOR Good Practices for Outcomes Research Task Force Report. Value Heal. (2019). https://doi.org/10.1016/j.jval.2019.01.004
8. Whittal, A., Meragaglia, M., Nicod, E.: The use of patient-reported outcome measures (PROMs) in rare diseases and implications for HTA. Patient. under revi, (2020)
9. Benjamin, K., Vernon, M.K., Patrick, D.L., Perfetto, E., Nestler-Parr, S., Burke, L.: Patient-Reported Outcome and Observer-Reported Outcome Assessment in Rare Disease Clinical Trials: An ISPOR COA Emerging Good Practices Task Force Report. Value Heal. (2017). https://doi.org/10.1016/j.jval.2017.05.015
10. Slade, A., Isa, F., Kyte, D., Pankhurst, T., Kerecuk, L., Ferguson, J., Lipkin, G., Calvert, M.: Patient reported outcome measures in rare diseases: A narrative review, (2018)
11. Pearson, I., Rothwell, B., Olaye, A., Knight, C.: Economic Modeling Considerations for Rare Diseases. Value Heal. 21, 515–524 (2018). https://doi.org/10.1016/j.jval.2018.02.008
12. Towse, A., Garau, M.: Appraising ultra-orphan drugs: is cost-per-QALY appropriate? A review of the evidence. (2018)
13. Annemans, L., Aymé, S., Le Cam, Y., Facey, K., Gunther, P., Nicod, E., Reni, M., Roux, J.-L., Schlander, M., Taylor, D., Tomino, C., Torrent-Farnell, J., Upadhyaya, S., Hutchings, A., Le Dez, L.: Recommendations from the European Working Group for Value Assessment and Funding Processes in Rare Diseases (ORPH-VAL). Orphanet J. Rare Dis. 12, 50 (2017). https://doi.org/10.1186/s13023-017-0601-9
14. Gutierrez, L., Patris, J., Hutchings, A., Cowell, W.: Principles for consistent value assessment

and sustainable funding of orphan drugs in Europe. Orphanet J. Rare Dis. 10, 53 (2015). https://doi.org/10.1186/s13023-015-0269-y

15. Kleijnen, S., Leonardo Alves, T., Meijboom, K., Lipska, I., De Boer, A., Leufkens, H.G., Goettsch, W.G.: The impact of quality-of-life data in relative effectiveness assessments of new anti-cancer drugs in European countries. Qual. Life Res. (2017). https://doi.org/10.1007/s11136-017-1574-9

16. National Institute for Health and Care Excellence.: Asfotase alfa for treating paediatric-onset hypophosphatasia. HST6. (2017)

17. National Institute for Health and Care Excellence.: Eculizumab for treating atypical haemolytic uraemic syndrome. HST1. 2015.

18. National Institute for Health and Care Excellence.: Colistimethate sodium and tobramycin dry powders for inhalation for treating pseudomonas lung infection in cystic fibrosis. TA276. (2013)

19. National Institute for Health and Care Excellence.: Migalastat for treating Fabry disease. HST4. (2017)

20. National Institute for Health and Care Excellence.: Inotersen for treating hereditary transthyretin amyloidosis. HST9. (2019)

21. National Institute for Health and Care Excellence.: Voretigene neparvovec for treating inherited retinal dystrophies caused by RPE65 gene mutations. HST11. (2019)

22. National Institute for Health and Care Excellence.: Mannitol dry powder for inhalation for treating cystic fibrosis. TA266. (2012)

23. National Institute for Health and Care Excellence.: Lumacaftor–ivacaftor for treating cystic fibrosis homozygous for the F508del mutation. TA398. (2016)

24. National Institute for Health and Care Excellence.: Elosulfase alfa for treating mucopolysaccharidosis type IVa. HST2. (2015)

25. National Institute for Health and Care Excellence.: Burosumab for treating X-linked hypophosphataemia in children and young people. HST8. (2018)

26. National Institute for Health and Care Excellence.: Strimvelis for treating adenosine deaminase deficiency–severe combined immunodeficiency. HST7. (2018)

27. National Institute for Health and Care Excellence.: Nusinersen for treating spinal muscular atrophy. TA588. (2019)

28. National Institute for Health and Care Excellence.: Letermovir for preventing cytomegalovirus disease after a stem cell transplant. TA591. (2019)

29. National Institute for Health and Care Excellence.: Mepolizumab for treating severe refractory eosinophilic asthma. TA431. (2017)

30. National Institute for Health and Care Excellence.: Patisiran for treating hereditary transthyretin

amyloidosis. HST10. (2019)

31. National Institute for Health and Care Excellence.: Darvadstrocel for treating complex perianal fistulas in Crohn's disease. TA556. (2019)

32. National Institute for Health and Care Excellence.: Eliglustat for treating type 1 Gaucher disease. HST5. (2017)

33. National Institute for Health and Care Excellence.: Holoclar for treating limbal stem cell deficiency after eye burns. TA467. (2017)

34. National Institute for Health and Care Excellence.: Lanadelumab for preventing recurrent attacks of hereditary angioedema. TA606. (2019)

35. National Institute for Health and Care Excellence.: Obeticholic acid for treating primary biliary cholangitis. TA443. (2017)

36. National Institute for Health and Care Excellence.: Cerliponase alfa for treating neuronal ceroid lipofuscinosis type 2. HST12. (2019)

37. National Institute for Health and Care Excellence.: Pirfenidone for treating idiopathic pulmonary fibrosis. TA504. (2018)

38. National Institute for Health and Care Excellence.: Nintedanib for treating idiopathic pulmonary fibrosis. TA379. (2016)

39. National Institute for Health and Care Excellence.: Ataluren for treating Duchenne muscular dystrophy with a nonsense mutation in the dystrophin gene. HST3. (2016)

40. Nicod, E., Kanavos, P.: Developing an evidence-based methodological framework to systematically compare HTA coverage decisions: A mixed methods study, (2016)

41. Bryman, A.: Social Research Methods. , Oxford (2004)

42. Lerman, J.A., Sullivan, E., Barnes, D.A., Haynes, R.J.: The Pediatric Outcomes Data Collection Instrument (PODCI) and functional assessment of patients with unilateral upper extremity deficiencies. J. Pediatr. Orthop. (2005). https://doi.org/10.1097/01.bpo.0000149866.80894.70

43. Jones, P.W., Quirk, F.H., Baveystock, C.M.: The St George's Respiratory Questionnaire. Respir. Med. (1991). https://doi.org/10.1016/S0954-6111(06)80166-6

44. Meregaglia, M., Nicod, E., Drummond, M.: The estimation of health state utility values in rare diseases: overview of existing techniques. Int. J. Technol. Assess. Health Care. (2020). https://doi.org/10.1017/S0266462320000665

45. Guidance to submitting companies for completion of new product assessment form (NPAF). Supplement for medicines for extremely rare conditions (ultra-orphan medicines). , Glasgow (2019)

46. Beusterien, K., Leigh, N., Jackson, C., Miller, R., Mayo, K., Revicki, D.: Integrating preferences

into health status assessment for amyotrophic lateral sclerosis: The ALS Utility Index. Amyotroph. Lateral Scler. 6, 169–176 (2005). https://doi.org/10.1080/14660820410021339

47. Rüther, A., Elstein, D., Wong-Rieger, D., Guyatt, G.: Aspects of patient reported outcomes in rare diseases: A discussion paper. Int. J. Technol. Assess. Health Care. 32, 126–130 (2016). https://doi.org/10.1017/S0266462316000271

48. Schulz, S., Passon, A., Kulig, M., Perleth, M., Matthias, K.: Orphan drug benefits asssessments at the Federal Joint Committee in Germany. In: HTAi conference (2019)

49. Bell, J.A., Galaznik, A., Pompilus, F., Strzok, S., Bejar, R., Scipione, F., Fram, R.J., Faller, D. V., Cano, S., Marquis, P.: A pragmatic patient-reported outcome strategy for rare disease clinical trials: application of the EORTC item library to myelodysplastic syndromes, chronic myelomonocytic leukemia, and acute myeloid leukemia. J. Patient-Reported Outcomes. (2019). https://doi.org/10.1186/s41687-019-0123-4

50. Morel, T., Cano, S.J.: Measuring what matters to rare disease patients - Reflections on the work by the IRDiRC taskforce on patient-centered outcome measures. Orphanet J. Rare Dis. (2017). https://doi.org/10.1186/s13023-017-0718-x

51. Goodrich, K., Kaambwa, B., Al-Janabi, H.: The inclusion of informal care in applied economic evaluation: A review. Value Heal. (2012). https://doi.org/10.1016/j.jval.2012.05.009

52. Pennington, B., Wong, R.: Modelling carer health-related quality of life in NICE Technology Appraisals and Highly Specialised Technologies. (2019)

Figure 1

Click here to access/download;Figure;Figure 1.png

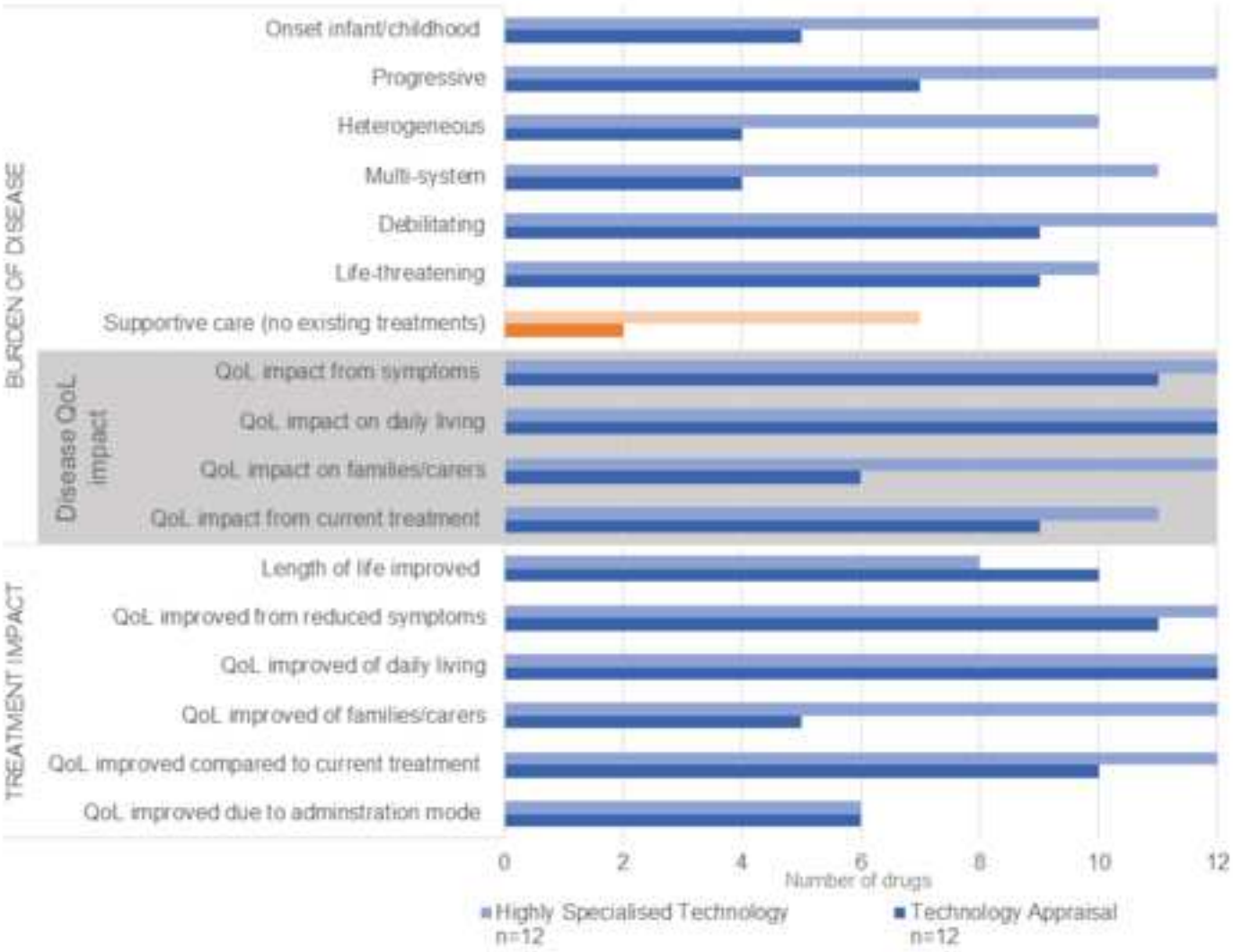

Table 1. Types and influence of PRO data and patient evidence considered in NICE TA and HST appraisals of non-oncology rare disease treatments (n=24)

		PRO data considered in NICE appraisals by PROM type and level of influence				**Patient evidence**
		Generic	**Disease-group**	**Disease-specific**	**Symptom-specific**	
NICE Highly Specialised Technology	Asfotase alfa	++	NR			
	Eculizumab	HSUV				+++
	Patisiran	HSUV			+++	+++
	Voretigene		NR			
	Cerliponase	++		++		
	Elosulfase alfa*					+++
	Ataluren	++	++			
	Migalastat	≈			≈	+++
	Eliglustat	HSUV			++	+++
	Strimvelis**					
	Burosumab**					
NICE Technology Appraisal	Inotersen				≈	
	Mannitol	HSUV		++		
	Colistimethate sodium and tobramycin dry powders for inhalation (DPI) [=antibiotics]			HSUV		+++
	Nintedanib	HSUV	NR		NR	++
	Lumacaftor–ivacaftor	HSUV		++		
	Mepolizumab	HSUV	HSUV		NR	
	Obeticholic acid**					
	Holoclar *					
	Pirfenidone***					
	Darvadstrocel		NR			
	Nusinersen	HSUV				
	Letermovir	≈	≈			+++
	Lanadelumab	≈		≈	NR	

Legend: HSUV: Health State Utility Value; NICE: National Institute for Health and Care Excellent; PRO: Patient Reported Outcome; PROM: Patient Reported Outcome Measure

*No PRO data was provided, but patient based evidence or PRO data from the literature was used instead

** No PRO data was provided, QoL evidence was derived from HSUVs from vignettes or published literature

*** No PRO data was included in the appraisal report as this is a re-assessment and no new QoL evidence was presented

+++	**Influence** – PRO results / patient evidence were influential in the final decision
++	**Possible influence** - PRO results / patient evidence were reported and suggested some type of benefit, but no discussion about these results was reported

≈	**No influence** - PRO results were reported, but did not show any benefit and no discussion about these results was reported
NR	**Not reported** / **no influence** - PRO results not reported, but the report listed the PROM as having been collected
HSUV	**Used to derive HSUVs** – PRO results was used to derive HSUVs (further discussed in Table X)

Table 2

Table 1. Use and influence of PRO evidence in NICE TA and HST appraisals of non-oncology rare disease treatments (n=24)

MEDICINE Generic name - indication	PROM (by type) Instrument (source)	PRO EVIDENCE Description of results	APPRAISAL Influence of PRO evidence on decision
Asfotase alfa Paediatric-onset hypophosphatasia	**G** - EQ-5D (EU patient survey) **G** - CHAQ, LEFS (small trials)	EQ-5D - children 0.76 treatment arm~~x~~ vs 0.43 no ~~tx~~treatment - adults 0.39 no ~~tx~~treatment - scores varied depending on walking ability [-0.24 to 0.73 in children, -0.01 to 0.51 in adults] Other trial PRO data in academic confidence	EQ-5D results not used to derive HSUVs but may have been considered by clinicians when developing the vignette's health states based on 6MWT severity levels. Specifically, mental health and pain domains
Eculizumab Atypical haemolytic uraemic syndrome (aHUS)	**G** - EQ-5D (2 phase II prospective, open-label, non-randomised, single arm trials, n=37) **P** - Survey (patient submission, n=37)	EQ-5D: mean improvement = 0.208 Survey: burden of disease and of current treatment on patients, carers/families	EQ-5D used to derive HSUVs. Survey shows greatly impaired QoL of patients and carers from living with aHUS
Patisiran Hereditary transthyretin amyloidosis	**G** - EQ-5D-5L (RCT, n=255) **D** - NIS, Norfolk-DN (RCT, n=255) **P** - Patient and clinical ~~l~~input	All PRO evidence significantly improved. NIS was trial's primary endpoint Patient input: factors not captured in model important for patients, e.g. ability to walk	EQ-5D-5L used to derive HSUVs. Effective based on significantly improved outcomes. PRO evidence captures most relevant treatment impacts, except for ability to return to work, daily activities, social life, impact on carers and families. Higher ICER accepted given effect size and aspects not captured.
Voretigene Inherited retinal dystrophies	**D** - VFQ (patient survey)	Results not reported as confidential	The committee highlighted preference for QoL~~l~~ collected from trials

MEDICINE Generic name - indication	PROM (by type) Instrument (source)	PRO EVIDENCE Description of results	APPRAISAL Influence of PRO evidence on decision
Cerliponase Nuronal ceroid lipofuscinosis type 2	**G -** EQ-5D 5L, PedsQL, PedsQL-FM (pivotal trial, single-arm, open-label, n=23, children 3-16 years) **D -** CLN2-QoL (pivotal trial)	QoL evidence: improvement in initial treatment phase (only short term data)	Recognition of limited QoL evidence due to short term data. Unclear if PRO data influenced committee discussions
Elosulfase alfa Mucopolysaccharidosis type IVa	**P** - ~~c~~Cross-sectional survey from patient and family, company submission, 63 patients + 56 families **O -** Observational study on natural history (n = 325 people, up to 10 years)	No PROMs collected in trials Survey~~s~~: QoL impact related to reliance on wheelchair, endurance, pulmonary function and height. Impact on carers up to 15 hours/day Observational study: decline in endurance, restricted growth, limitations in daily living	Survey used to derive HSUVs Observational study: supported interpretation of impact on QoL and HSUVs, including aspects not captured in HSUV
Ataluren Duchenne muscular dystrophy	**G -** PedsQL (phase IIb) **G -** PODCI, ADLQ (RCT, confirmatory trial) **P** - Survey of carers (company submission)	PRO results not reported. PODCI, ADLQ confidential Survey: impact on multiple aspects of life, e.g. emotional wellbeing, mental health, personal care, ability to maintain relationships. Caregivers felt tired, depressed, anxious. In many cases, at least another family member in addition to both parents were involved in giving care (for example, siblings and grandparents)	QoL~~l~~ data (all): underestimate due to short trial duration (48 weeks too short to capture impact on ability to walk) PedsQL: results not aligned with patient statements on meaningful stabilisation or improvement in walking, or ability to conduct daily activities Survey: unclear influence , possibly considered in estimating extent of impact on caregivers, but not reported

MEDICINE Generic name - indication	PROM (by type) Instrument (source)	PRO EVIDENCE Description of results	APPRAISAL Influence of PRO evidence on decision
Migalastat Fabry disease	**G** - SF-36 physical and mental health components (open-label, non-inferiority RCT (ATTRACT) and RCT (FACETS)) **D** - BPI, GSRS (ATTRACT, FACETS) **P -** P~~p~~atient and clinical input (oral administration)	QoL data: inconclusive (no change), except for change in GSRS Input: benefits of oral administration	Patient input confirmed benefit of oral administration over infusion PRO data not discussed in report, nor used to derive HSUVs
Eliglustat Gaucher disease	**G -** SF-36 general health, physical and mental components (open-label trial (ENCORE) and RCT (ENGAGE)) **D -** FSS, BPI (ENCORE and ENGAGE) **P -** ~~p~~Patient survey and patient submission (oral administration)	QoL: maintained with treatment FSS: fatigue > placebo (not statistically significant) SF-36, BPI: no change Patient input: preference for oral administration	SF-36 used to derive HSUV Unclear influence of PRO evidence. Adverse event HSUVs included, not clear if influenced by FSS or BPI Advantage of oral administration as key driver for decision (patient survey)
Strimvelis Adenosine deaminase deficiency–severe combined immunodeficiency	None	No QoL evidence presented. Data being collected within trial (not reported)	Not reported
Burosumab for X-linked hypophosphataemia	None	No QoL evidence presented. Data being collected within trial (not reported)	Not reported

MEDICINE Generic name - indication	PROM (by type) Instrument (source)	PRO EVIDENCE Description of results	APPRAISAL Influence of PRO evidence on decision
Inotersen Hereditary transthyretin amyloidosis	**D** – Norfolk QoL-DN (RCT)	Norfolk QoL-DN: no change in treatment arm, decrease in placebo arm	Not reported
Mannitol for cystic fibrosis	**G** - HUI2 (RCT, trial 301) **D -** CFQ-R (RCT, trial 302)	HUI2: no significant change CFQ-R: no significant change, improvement in respiratory, physical and vitality domains, but not significant	HUI2 used to derive HSUVs. No ideal measures to capture the QoL impact, including adverse events from current treatments, e.g. unpleasant taste or sensations, as reported by patients
Colistimethate sodium and tobramycin dry powders for inhalation (DPI) [=antibiotics] Pseudomas lung infection in cystic fibrosis	**D** - CFQ-R (open-label RCT) **P -** Treatment satisfaction questionnaire: administration mode, manufacturer submission **P -** Patient input	**Colistimethate sodium DPI** CFQ-R from non-inferiority trial **Tobramycin DPI** N~~n~~o QoL data collected in trial, relied on treatment satisfaction questionnaire and patient input	**Colistimethate sodium DPI** CFQ-R: no improvement since non-inferiority trial **Tobramycin DPI** Questionnaire: higher values for DPI over nebuliser Limited influence of QoL data on decision and interpretation of economic model. Recognition of improved speed and adherence with DPI based on patient input and questionnaire
Nintedanib Idiopathic pulmonary fibrosis	**G -** EQ-5D, PGI-C (RCT) **D -** SQRQ, SOBQ, CASA-Q **P -** ~~p~~Patient input (tolerability)	PRO data: not reported Patient input: better tolerability profile impacting QoL~~l~~, ability to go outdoors due to less photo sensibility	EQ-5D used to derive HSUVs

MEDICINE Generic name - indication	PROM (by type) Instrument (source)	PRO EVIDENCE Description of results	APPRAISAL Influence of PRO evidence on decision
Lumacaftor–ivacaftor Cystic fibrosis	**G** - EQ-5D (RCTs TRAFFIC and TRANSPORT) **D -** CFQ-R (TRAFFIC and TRANSPORT)	EQ-5D: high baseline values due to patients perception of life as "normal", difficult to capture improved QoL (ceiling effect, common in cystic fibrosis). No significant difference [mean difference 0.0095 (TRAFFIC) and -0.0009 (TRANSPORT)] CFQ-R: mean difference of 2.2 < 4 MID	CFQ-R: other studies with similar severity levels showed greater changes compared to trial results EQ-5D: no evidence on reasons for being inappropriate. EQ-5D usually captures most important aspects in cystic fibrosis based on expert input
Mepolizumab Severe refractory eosinophilic asthma	**G -** EQ-5D (RCT DREAM) **D** - SGRQ, ASQ (RCTs MENA and SIRIUS)	SGRQ: QoL increase due to fewer exacerbations AND improved symptom control and lung function	EQ-5D used to derive HSUVs SGRQ: possible confounding (exacerbation reduction ~ fewer symptoms). Improved symptoms recognised (beyond those from fewer exacerbations). SGRQ also mapped to derive HSUVs
Obeticholic acid for primary biliary cholangitis	None	No PRO data collected in trial	
Holoclar for limbal stem cell deficiency after eye burns	None	No PRO data collected in trial	HSUVs derived from impact on visual acuity
Pirfenidone Idiopathic pulmonary fibrosis	None	Re-submission to extend indication to patients >80% FVC. Quality of life data not discussed (as did not change from initial submission, for which the report was no longer available)	No PROMs reported, no impact on decision (apart from QoL data captured in model).
Darvadstrocel Crohn's disease	**D -** PDAI (RCT ADMIRE)	PDAI results not reported	PDAI does not capture QoL impact (only symptoms) => preference for EQ-5D trial data

MEDICINE Generic name - indication	PROM (by type) Instrument (source)	PRO EVIDENCE Description of results	APPRAISAL Influence of PRO evidence on decision
Nusinersen Spinal muscular atrophy	**G -** PedsQL (RCT CHERISH)	PedsQL results not reported in appraisal report, only in committee papers. Data kept confidential, likely due to the challenges to collect data from babies and children for SMA	PedsQL mapped to EQ-5D
Letermovir Cytomegalovirus	**G** - EQ-5D (RCT PN001) **D** - FACT-BMT (PN001) **P** – Patient and clinical input (on QoL from preventing CMV)	PN001 trial: not powered to show changes QoL, no improvements Results confounded by mix of patients who have had CMV reactivation and started pre-emptive therapy and those who have not	Trial limitations and challenges to capture change recognized Patient and clinical experts input on QoL impact from preventing CMV accounted for in decision (ICER likely to be lower due to this, which lead to a positive decision)
Lanadelumab for hereditary angioedema	**G -** EQ-5D-5L, SF12, WPAI:GH (RCT HELP-03, open-extension HELP-04) **D -** AE-QoL~~l~~, HADS (HELP-03 + 04)	EQ-5D-5L: no change due to lack of sensitivity in condition (timing of response - only two responses during attacks captured) AE-QoL~~l~~: statistically improved Other PROMs not reported in appraisal report or committee papers	EQ-5D-5L data used to derive HSUVs Other results not commented

Legend: G: generic patient reported outcome measure; D: disease, disease-group or symptom-specific patient reported outcome measure; P: patient ~~based~~ evidence; NA: no report available; MID: minimal important difference; CHAQ: childhood health assessment questionnaire; LEFS: lower extremity functional scale; NIS: neuropathy impairment score; Norfolk-DN: Norfolk quality of life-diabetic neuropathy; VFQ: visual function questionnaire; PedsQL: Paediatric Quality of Life Inventory - Parent Report for Toddlers; PedsQL-FM: PedsQL family impact module; CLN2-Qol: CLN2 quality of life instrument; PODCI: paediatric outcomes data collection instrument; ADLQ: activities of daily living questionnaire; FSS: fatigue severity scale; BPI: brief pain inventory; CFQ-R: cystic fibrosis questionnaire revised; HUI2: Health Utility Index Mark 2; SGRQ: St George Respiratory Questionnaire; SOBQ: University of California San Diego shortness of breath questionnaire; CASA-Q: cough and sputum assessment questionnaire; PGI-C: patient global impression of change; ASQ: asthma control questionnaire; PDAI: perianal disease activity index; FACT-BMT: functional assessment of cancer therapy; AE-QoL: angioedema quality of life questionnaire; WPAI:GH: work productivity and activity impairment questionnaire - general health; HADS: hospital anxiety and depression scale; EQ-5D-5L: EuroQol-5 Dimension-5 Level

Sources: [16-39]

Table 3. Types and influence of HSUVs and patient evidence considered in NICE TA and HST appraisals of non-oncology rare disease treatments (n=24)

		Techniques to derive HSUVs considered in NICE appraisals by HSUV technique and level of influence					**Patient evidence**
		Generic PROMs	Mapping	Vignettes	Published literature	Various	
NICE Highly Specialised Technology	Asfotase alfa			≈			
	Eculizumab	≈					+++
	Patisiran	≈					+++
	Voretigene			≈			
	Cerliponase			≈			
	Elosulfase alfa					+++	+++
	Ataluren				++		
	Migalastat	≈					+++
	Eliglustat		++				+++
	Strimvelis				++		
	Burosumab			≈			
NICE Technology Appraisal	Inotersen	≈					
	Mannitol	+++					
	Colistimethate sodium and tobramycin dry powders for inhalation (DPI) [=antibiotics]		≈				+++
	Nintedanib	≈					++
	Lumacaftor–ivacaftor		≈				
	Mepolizumab		+++				
	Obeticholic acid				++		
	Holoclar		++				
	Pirfenidone*						
	Darvadstrocel			≈			
	Nusinersen		≈				
	Letermovir	≈					+++
	Lanadelumab	+++					

Legend: HSUV: Health State Utility Value; NICE: National Institute for Health and Care Excellent; PRO: Patient Reported Outcome; PROM: Patient Reported Outcome Measure

*No HSUV results were considered as this is a re-assessment and no new QoL evidence was provided

+++	**Accepted** – HSUV results / patient evidence were influential in the final decision
++	**Not commented** - HSUV results / patient evidence were reported and suggested some type of benefit, but no discussion about these results was reported – assumption is that they are likely to have influenced the decision
≈	**Uncertain** - HSUV results were reported, but were considered uncertain

Table 4

Table 42. Use and influence of HSUVs in NICE TA and HST appraisals of non-oncology rare disease treatments (n=23)*

MEDICINE Generic name, indication	**HSUV** Technique, appraisal	DECISION ICER, reasons
Eculizumab Atypical haemolytic uraemic syndrome	**EQ-5D** - all benefits not captured due to lack of data - ERG's HSUV lower than manufacturers (10 versus 25 QALYs) => in both cases, substantial increase in QoL recognised	**Restrict - monitoring and stopping rules** Cost-consequence model ~10-25 QALYs => QoL underestimated due to lack of data => magnitude of benefit substantial despite uncertainty
Patisiran Hereditary transthyretin amyloidosis	**EQ-5D: 5L mapped to 3L** => uncertain assumptions around HSUV, duration of treatment benefit **HSUV after stopping treatment** **-** uncertain evolution after stopping => little effect on ICER **HSUV carer** - estimates revised to align with inotersen => considered acceptable **HSUV adverse events (gastro-intestinal, GI)** - possible overlap with impact captured in EQ-5D => value between manufacturer's estimate and no disutility => scenario analysis using pessimistic GI disutilities ~£125k/QALY **Benefits not captured**: ability to work, carry out daily activities, more active family and social life, maintain independence and dignity	**List - commercial agreement** ~£80-125k/QALY => no QALY weighing (~9.16 QALYs) => ICER acceptable due to additional factors (severity, rarity, size of health benefits, benefits not captured, innovativeness, impact on carers)

MEDICINE Generic name, indication	**HSUV** Technique, appraisal	DECISION ICER, reasons
Voretigene Inherited retinal dystrophies (caused by RPE65-mediated IRD)	**Vignettes** - implausible lowest health state [worse than death (-0.04)] given patients confirmed adapting to disease - few clinicians involved in development - focus of clinicians focus on vision loss rather than QoL => possible underestimation of QoL => EQ-5D more appropriate due to focus on QoL (and not vision loss) **TTO (published literature)** - not robust, good complement to vignettes => HSUV to fall between vignettes (company) and TTO (ERG) **HSUV adverse events** => suitable, small effect on ICER **HSUVs carers (published literature)** => only children included (adults excluded)	**List - commercial agreement** ICER range £114,956 (company)--–£155,750 (ERG) => 1.2 QALY weight (QALY gains 12.1-17.7)
Cerliponase Nuronal ceroid lipofuscinosis type 2	**PedsQL** - Trial QoL data not used as HSUVs unavailable for all model health states => preference for trial data, but recognition that possibility of negative values excluded, unrealistic given the severity of disability **Vignettes/EQ-5D (5L mapped to 3L)** - validation of vignettes and completion of EQ-5D 5L by clinical experts. 5L mapped to 3L - issues with robustness: additional elements such as pain and frequency of seizures included, but their association to motor and language scales defining health states unclear => neither source of data sufficiently robust, suggesting lack of correspondence between vignette and model health states => EQ-5D 3L mapped to HSUVs using vignettes considered, given no alternative data **HSUV carers/siblings** => disutilities included, but 30 years considered to better reflect real life compared to life long	**List - Managed Access Agreement** ICER not specified, 3.0 QALY weight

MEDICINE Generic name, indication	**HSUV** Technique, appraisal	DECISION ICER, reasons
Elosulfase alfa Mucopolysaccharidosis type IVa	**Various approaches and sources** Issues around capturing QoL: - QOL rarely collected in trials, as challenging particularly for children (e.g. recollection of how they felt before treatment) - potential issue around questions: it's not about the activities they can do post-treatment, but about how they feel => EQ-5D not collected in trial, limited evidence on QOL => lack of developed/validated methods => impact of adverse effects on QoL not included => treatment improves QOL and HSUV increment considered appropriate => uncertainty remains in HSUV modelled	**List - Managed Access Agreement + commercial agreement** Cost-consequence model: limited impact on incremental QALYs QoL not appropriately captured due to challenges in measuring relevant effects and collecting data from children. No QoL measures collected in trials
Ataluren Duchenne muscular dystrophy	**HSUV scoliosis and carers (published literature)** - uncertainty around scoliosis not occurring after puberty (model assumption), or applying different HSUVs after loss of walking. Company's assumption: QOL linked to ability to walk greater since loss of walking would occur later. Clinical experts commented plausibility if loss is in upper limb muscle strength when ability to walk is lost, for which no evidence was presented => unreasonable to assume different HSUVs across treatment group once ability to walk is lost given no evidence	**List - Managed Access Agreement** Managed Access Agreement to capture carer HSUV using EQ-5D and Child Health Utility 9D Cost-consequence model ~2.389-8.562 QALY gains Wider benefits: indirect costs/benefits (ability to work of carers, decrease in out-of-pocket costs)

MEDICINE Generic name, indication	**HSUV** Technique, appraisal	DECISION ICER, reasons
Migalastat Fabry disease	**EQ-5D (questionnaire, Dutch cohort study with UK tariff) - enzyme replacement therapy and complications (comparator)** - to measure disutility of patients undergoing enzyme replacement therapy - similar HSUV as for end-stage renal disease, stroke, heart complications - patients/clinicians emphasised major impact on QoL => uncertain disutility values **HSUV infusion (DCE)** - 506 people from UK general population - HSUV infusion > HSUV complications => not comparable since different methods used (uncertain face validity) => patient input: recognition of added benefit of migalastat over ERT infusion (convenience from oral administration) => decreasing infusion-disutility by 50% decreased QALY gains (from 0.98 to 0.34 incremental QALYs)	**Restrict - if ERT + patient access scheme** Confidential cost-consequence model Miglastat considered to have similar benefits compared to ERT, with the main advantage of oral administration (patient input). Main concern about adherence with oral administration. Main driver of model infusion disutility
Eliglustat **Gaucher disease**	**SF36 mapped to EQ-5D (published algorithm)** **HSUV adverse events** - HSUV decrements applied **HSUV oral administration** - HSUV increment (0.12) based on preference for oral administration (vignette commissions by manufacturer) => too high, ERG's estimate of 0.05 more plausible	**List - Patient Access Scheme** Cost-consequence model Model driven by QoL (mode of administration)

MEDICINE Generic name, indication	**HSUV** Technique, appraisal	DECISION ICER, reasons
Asfotase alfa Paediatric-onset hypophosphatasia	**Vignettes** **-** 9 clinical experts completing EQ-5D for each level of severity (6MWT) => reasonable face validity (suitability of measure in capturing concept of interest) => not collected in trials - health states in the Markov model defined based on severity levels of 6MWT that, however, may not capture all the relevant symptoms => measure accepted due to lack of available evidence => HSUV for most severe health state very low (0.23), potentially overestimating benefits (more space for HSUV gain) => lack of correspondence between vignettes and model health states **EQ-5D (European patient survey)** => aligned with values in vignette study	**List - Managed Access Agreement + commercial agreement** Cost-consequence model ~14-25 QALYs HSUV considered to reasonably capture impact on QoL, risk of underestimation compensated by carer disutility not included in model
Strimvelis Adenosine deaminase deficiency–severe combined immunodeficiency	Trial QoL data not included in model because limited **HSUV QoL (published literature - no detail)** - Full health HSUV from general population => since no data on long term effect, these were explored within sensitivity and scenario analyses. The committee agreed lower values should be used **HSUV intravenous immunoglobulin (IVIG) or severe infections** - ERG: 0.75 HSUV included - plausibility confirmed by clinical experts **HSUV carer** **-** improved fast after treatment - no approach to measure - to be considered qualitatively during deliberations	**List** £12-120K/QALY (14.0-19.6 QALY gained) Impact of changes of QoL on model not reported

MEDICINE Generic name, indication	**HSUV** Technique, appraisal	DECISION ICER, reasons
Burosumab X-linked hypophosphataemia	**Vignettes** - 6 clinicians value QoL of patients with XLH aged 18, 40 and 60 years using EQ-5D 5L - some missing data, company inferred 1 for healed health states - scored by clinicians not patients, not from trials => approach deemed appropriate (in absence of alternatives), but highly uncertain **HSUV carer (literature)** - published literature on people with limited mobility => acceptable, not robust	**List - Managed Access Agreement + commercial agreement** £113-£150K/QALY (5.52-15.99 QALYs gained) Most/less conservative assumptions included/excluded carer disutility (and different stopping ages) resulting in ICERS ranging from £112-149k/QALY. Unclear to what extent variation due to inclusion/exclusion of carer disutility
Inotersen Hereditary transthyretin amyloidosis	**EQ-5D (Brazilian registry converted with UK tariffs, source model HSUVs)** - modelling of values from dataset with a number of assumptions, e.g. cap to ensure HSUVs do not exceed the general population => model could generate implausible health state classifications => not ideal, but acceptable, considered uncertain **HSUV carer** - 1 in stages 1-2, 2 in stage 3	**List - commercial agreement** £96,697-£150,636/QALY (no QALY weighing) HSUV values did have some effect on model, but generally uncertain => unclear if driving the model => time-dependent HSUVs used within each health state
Mannitol Cystic fibrosis	**HUI2 (trial)** - mean disutility at baseline (0.988), average change at each timepoint added to baseline to calculation HSUV for each health state => HUI2 baseline considered high given multiple comorbidities => EQ-5D measure preferred => difficulty to value health states in chronic conditions. Standard method of using general population's valuation of QoL descriptions to generate HSUVs appropriate **HSUV lung transplant and pulmonary exacerbations (literature)**	**Restrict - clinical parameters, 2nd line** ICER<£30K Model changes with extension of life, little with changes in QoL - patients confirmed treatment improved QoL, considered important => HSUVs values very uncertain

MEDICINE Generic name, indication	**HSUV** Technique, appraisal	DECISION ICER, reasons
Colistimethate sodium and tobramycin dry powders for inhalation (DPI) [=antibiotics] Pseudomas lung infection in cystic fibrosis	**HSUV Colistimethate sodium DPI** **CFQ-R mapped to EQ-5D** => no preference-based model considered a methodological limitation **Health utility study linking EQ-5D responses to FEV% health states** => issue around establishing relationship, but considered more appropriate compared to manufacturer's model (mapping) **HSUV Tobramycin DPI (patient input)** => DPI to improve QoL in terms of speed and adherence compared to nebuliser	**List - Patient Access Scheme** Drivers of cost-effectiveness model: cost of interventions and their comparators, QALY gains/losses Colistimethate sodium DPI: small QALY loss (based on HSUV/QoL evidence) but substantial cost savings over nebuliser Tobramycin DPI: dominant - small QALY gain (no HSUV/QoL evidence, based on patient input) and cost saving (DPI dominated nebuliser)
Nintedanib Idiopathic pulmonary fibrosis	**EQ-5D (trial)** - model based on predicted FVC changes and rate of exacerbations **HSUV adverse events** - serious gastro-intestinal events, rash related events => model did not include ~~diarrhea~~diarrhoea-adverse events as not severe and affected a small proportion of patients => committee did not agree, and considered it to affect QoL **HSUV exacerbations** => possible gains in QOL not captured in QALY (tolerability profile, reduced dosing frequency) => lack of sensitivity to change	**Restrict - clinical parameters + Patient Access Scheme** Dominant over pirfenidone (survival equal, differences in QALYs) Committee recognised that additional impact on QoL not captured in model

MEDICINE Generic name, indication	**HSUV** Technique, appraisal	DECISION ICER, reasons
Lumacaftor–ivacaftor Cystic fibrosis	**HSUV QoL (multivariate mixed model)** - repeated regression analysis to model relationship between EQ-5D, lung function and pulmonary exacerbations in trials - no change in EQ-5D + little opportunity to demonstrate improved QoL due to ceiling effect - clinical experts state that EQ-5D capture most important effects in cystic fibrosis - committee tested model with values from another study (Lancaster) that better captured changes in QOL using EQ-5D in patients with similar levels of severity, resulting in increased ICER by ~65K/QALY => HSUV not captured adequately, uncertainty in model => however, trial data used, which is the biggest trials conducted in cystic fibrosis to date **HSUV lung transplant (literature)**	**Reject** ~-218-349K/QALY Model mostly driven by changes in life years gained When HSUVs from other study were used (Lancaster), ICER increased by 65K
Mepolizumab Severe refractory eosinophilic asthma	**SGRQ (MENSA trial) mapped to EQ-5D** - mapping algorithm based on population with chronic obstructive pulmonary disease - used as baseline value, adjusted due to differences between treatment arms and ages => considered acceptable **EQ-5D (DREAM trial)** - values adjusted for differences in baseline utilities values => baseline adjustment considered appropriate **HSUV exacerbation** - mid-point between trial data and published value => little change when using different disutility values, approach acceptable	**List - Patient Access Scheme** ~£29k/QALY Little effect of QoL on ICER. EQ-5D mapped from SGRQ considered in model, EQ-5D values from DREAM trial accounted for in the interpretation of QoL impact. Drivers included exacerbation rates, age-related mortality estimates and attrition rates

MEDICINE Generic name, indication	**HSUV** Technique, appraisal	DECISION ICER, reasons
Obeticholic acid Primary biliary cholangitis	No HSUVs data collected in trials **Published literature and expert assessment** - Chronic Hepatitis C and previous Technology Appraisal reports => some issues raised, but accepted	**List - Patient Access Scheme** ~£33K/QALY, additional factors considered: ICER underestimated in trial due to lack of adjustment up to recommended dose in some patients + innovative nature + potential to return to normal life + opportunity cost of liver transplant on other patients needing it
Holoclar Limbal stem cell deficiency after eye burns	**Mapping (HSUV visual acuity)** - combination of visual acuity in both best and worst seeing eyes - published mapping algorithm => model did not capture: negative effect on donor eye => if donor disutility captured, ICER likely to decrease **HSUV from pain, burning, photophobia** - base case value attached to presence of moderate or severe pain/burning/ photophobia derived from EQ-5D 3L tariff and uses the level 2 and 3 decrements of -0.123 and -0.386 respectively. Alternative values of no decrement and that derived from the general population SG method of -0.291 for both moderate and severe were used **Disfigurement HSUV** - Bespoke standard gamble exercise performed by 520 UK participants who were presented with various clinical scenarios describing moderate to severe limbal stem cell deficiency, including an image of a patient's eye with this condition showing the extent of the disfigurement typically present - estimated at 0.308 => applied from non-reference case methods and likely to be exaggerated => patients with one eye may prioritise impact of disfigurement over visual acuity, and those with two eyes affected may prioritise visual acuity over disfigurement => cataract disutilities considered more appropriate estimate of impact on QoL	**Restrict - subgroups and 1 eye + Patient Access Scheme** £6,948-£30,415-£42,139/QALY (lower values with 1 eye) Best plausible ICER was above £20K/QALY (includes ERG's estimate of disfigurement decrement). The committee accepted that if the model had considered a negative impact on donors, it would most likely be cost-effective => accepted for reimbursement

MEDICINE Generic name, indication	**HSUV** Technique, appraisal	DECISION ICER, reasons
	=> HSUV of 0.840 as base case for visual acuity and HSUV decrement of 0.140 for disfigurement	
Pirfenidone Idiopathic pulmonary fibrosis	no info (resubmission, no new data)	**Restrict - clinical parameters + Patient Access Scheme** £32,643-£38,687/QALY
Darvadstrocel Crohn's disease	**Vignette** => considered robust given significant number of participants (n=835 general public and n=162 patients with Crohn's disease) => reliable estimates of HSUVs => vignettes used considered appropriate (even if EQ-5D not collected in trial), also aligns with values in literature => HSUVs in some heath states might be too low, and that correctly derived HSUVs for these 3 health states could result in higher ICERs	**Reject** £143,131/QALY Very uncertain model. HSUVs may have some influence on ICER levels

MEDICINE Generic name, indication	**HSUV** Technique, appraisal	DECISION ICER, reasons
Nusinersen Spinal muscular atrophy	**HSUV expert elicitation** => not based on formal elicitation methods (may differ if other clinicians were to redo exercise) => questions asked to clinicians not available, making it difficult to interpret => health states based on motor function may not have captured QOL impact, differences in HSUVs between health states small **HSUV PedsQL mapped to EQ-5D** - published algorithm for later onset, and HSUVs adapted for the early onset model based on assumed correspondence of health states (values confidential) => limited face validity, not considered appropriate => challenge in babies and children => HSUV techniques not ideal, results highly uncertain **HSUV carer** - best health state based on general population HSUV, worse health state based on cross-sectional study of SMA patients, adjusted for each health state - equal transitions between these 2 points (values confidential) => based on assumptions and not on evidence => key driver in ICER (better ICER for later onset, worse for early onset due to carer disutility "saved" from early death - seen as "perverse" effect) => to be included, but highly uncertain **Disutility due to bereavement** **-** applied as -0.04	**Restrict - Types 1,2,3 + Managed Access Agreements** ICER not specified Key driver in models - may impact differently early and late onset models: carer disutility (highly uncertain, difficult to quantify), resource costs

MEDICINE Generic name, indication	**HSUV** Technique, appraisal	DECISION ICER, reasons
Letermovir Cytomegalovirus	**EQ-5D-3L (published literature)** - Long term disutility associated with haematopoietic stem cell transplant derived from a mix of EQ-5D 5L and 3L values from two published studies - ERG proposed alterative approach based on difference between mean HSUVs of patients in trial (PN001, 48 weeks) and the general population from another study => ERG approach preferable	**List - commercial agreement** <£24,269/QALY likely <£20,000/QALY ICER likely to decrease due to QoL not captured in evidence (when considering PROM data)
Lanadelumab Hereditary angioedema	**Published literature** Committee accepted alternative approach to EQ-5D-5L (recognised as insufficiently sensitive). Published study used to derive HSUVs, which collected EQ-5D-5L about health state today and health state during last attack	**Restrict - indication + commercial agreement** <£20,000/QALY QALY gains small relatively to costs, ICER could change with different clinical scenarios

*No HSUVs were reported for one treatment (pirfenidone) given it was a re-assessment; therefore, it was excluded from this analysis, which focuses on the 23 remaining treatments.

Legend: HSUV: health state utility values; QALY: quality-adjusted life years gained; ICER: incremental cost-effectiveness ratio; EQ-5D: EuroQol-5 Dimension; EQ-5D-3L: EuroQol-5 Dimension-3 Level; EQ-5D-5L: EuroQol-5 Dimension-5 Level; GI: gastro-intestinal; QoL: quality of life; ERG: Evidence Review Group; IRD: Inherited retinal dystrophies ; TTO: time-trade off; PedsQL: Paediatric Quality of Life Inventory - Parent Report for Toddlers; DCE: discreet choice experiment; ERT: enzyme replacement therapy; 6MWT: 6-minute walk test; IVIG: intravenous immunoglobulin; HUI2: Health Utility Index Mark 2; DPI: tobramycin dry powders for inhalation; CFQ-R:cystic fibrosis questionnaire revised; FEV: Forced Expiratory Volume; SGRQ: St George Respiratory Questionnaire

Sources: [16-39]

Table 5. Use and influence of PRO evidence in HAS, G-BA and ZIN appraisals of non-oncology rare disease treatments

MEDICINE Generic name - indication	PRO EVIDENCE AND APPRAISAL		
	HAS (France)	G-BA (Germany)	ZIN (Netherlands)
Asfotase alfa Paediatric-onset hypophosphatasia	no details provided	no trial QoL data, no conclusion	no details provided
Eculizumab Atypical haemolytic uraemic syndrome (aHUS)	no details provided	NA	no details provided
Patisiran Hereditary transthyretin amyloidosis	no details provided	Norfolk QoL-DN: statistically improved; validity and reliability confirmed; possible bias from higher missing values after 18 months in control group; no MID, effect size's hedges calculated for dossier; clinically relevant difference	NA
Voretigene Inherited retinal dystrophies	VFQ: not demonstrated. Secondary judgment criterion, no hierarchical test	VFQ: unsuitable. Transferability and MID from NEI VFQ-25 to new VFQ inappropriate	Vignettes, EQ-5D-5L, HUI3: not adequately collected
Cerliponase Nuronal ceroid lipofuscinosis type 2	PedsQL, CLN2-QoL, EQ-5D 5L: exploratory consideration of QoL, stabilisation in treatment group versus degradation in natural history data	PedsQL: no benefit in QoL recognised due to lack of comparative data and clinical relevance of change CLN2-QoL: not considered as no data on its development (by company) and validation provided	NA
Elosulfase alfa Mucopolysaccharidosis type IVa	no details provided	no details provided	no PROM data collected in trial
Ataluren Duchenne muscular dystrophy	no details provided	PODCI: not statistically significant. Quality and patient relevance not demonstrated due to lack of information	PedsQL, PODCI: not statistically significant
Migalastat Fabry disease	no details provided	SF-36: inconclusive	SF-36: inconclusive
Eliglustat Gaucher disease	no details provided	BPI, FSS, SF-36: no significant differences	SF-36, BPI, FSS, DS3: clinically relevant and crucial, no significant differences

MEDICINE Generic name - indication	PRO EVIDENCE AND APPRAISAL		
	HAS (France)	**G-BA (Germany)**	**ZIN (Netherlands)**
Strimvelis Adenosine deaminase deficiency–severe combined immunodeficiency	NA	NA	NA
Burosumab for X-linked hypophosphataemia	SF-36, PROMIS: exploratory, not usable	SF-10: lack of information on questionnaire development, restrictions in content validity, reliability and validity. Results not accounted for	NA
Inotersen Hereditary transthyretin amyloidosis	Norfolk QoL-DN: modest improvement as co-primary endpoint SF-36: not discussed	SF-36: biased due to missing values Norfolk-DN: no valid MID based on hedge's g, effects not clinically relevant. Statistically significant improvement, but not clinically relevant C-SSRS: not discussed	NA
Mannitol for cystic fibrosis	no details provided	NA	CFQ-R: no significant improvements. Overall effect around improving QoL and reducing pulmonary exacerbations
Colistimethate sodium and tobramycin dry powders for inhalation (DPI) [=antibiotics] Pseudomas lung infection in cystic fibrosis	no details provided	NA	no details provided
Nintedanib Idiopathic pulmonary fibrosis	EQ-5D, EORTC QLQ-30, QLQ-LC13: no expected improvement	G-BA/IQWiG EQ-5D VAS: statistically improved, benefit not proven given hedge's g SGRQ: not discussed	SGRQ: not clinically relevant
Lumacaftor–ivacaftor Cystic fibrosis	no details provided	not collected in trial	EQ-5D: used to derive HSUVs

MEDICINE Generic name - indication	PRO EVIDENCE AND APPRAISAL		
	HAS (France)	**G-BA (Germany)**	**ZIN (Netherlands)**
Mepolizumab Severe refractory eosinophilic asthma	NA	no details provided	no details provided
Obeticholic acid for primary biliary cholangitis	no details provided	PBC-40: validated measure. Responsiveness and MID not examined. Marginal change, but clinical relevance not determined	PBC-40: no improvement
Holoclar for limbal stem cell deficiency after eye burns	NA	no details provided	NA
Pirfenidone Idiopathic pulmonary fibrosis	no details provided	SGRQ, WHO QoL: no proof of added benefit	EQ-5D, SGRQ: from published paper. Unclear benefit as baseline values and validation difficult to verify
Darvadstrocel Crohn's disease	Van Assche Score, IBDQ: exploratory secondary endpoints, no change captured	IBDQ: not designed or validated for target population, no information on MID. Inconclusive QoL benefit	NA
Nusinersen Spinal muscular atrophy	PedsQL: not possible to quantify QoL benefit due to low response rates	PedsQL: QoL not demonstrated. Caregiver experience included	PedsQL mapped to EQ-5D
Letermovir Cytomegalovirus	EQ-5D-3L: unsuitable, used to derive HSUVs for different health states rather than change in QoL associated with an illness	EQ-5D, FACT-BMT FACT-BMT considered validated in patient population	not details provided
Lanadelumab for hereditary angioedema	AE-QoL: unusable as exploratory endpoint	AE-QoL: statistically improved, considered clinically relevant based on Hedge's g	NA

Legend: NA: no report available; MID: minimal important difference; Norfolk-DN: Norfolk quality of life-diabetic neuropathy; VFQ: visual function questionnaire; PedsQL: Paediatric Quality of Life Inventory - Parent Report for Toddlers; CLN2-Qol: CLN2 quality of life instrument; PODCI: paediatric outcomes data collection instrument; FSS: fatigue severity scale; BPI: brief pain inventory; CFQ-R: cystic fibrosis questionnaire revised; HUI3: health utility index 3; SGRQ: St George Respiratory Questionnaire; FACT-BMT: functional assessment of cancer therapy; AE-QoL: angioedema quality of life questionnaire; SF-36: short form 36; DS3: Gaucher disease severity score: PROMIS:

patient reported outcome measurement information system; SF-10: short form 10; EORTC-QLQ 30: European Organisation for Research and Treatment of Cancer - Quality of life 30; QLQ-LC13: modular supplement to EORTC QLQ 30 for use in lung cancer; NEI: National Eye Institute; PBC-40: quality of life for primary biliary cirrhosis; WHO QOL: Quality of life - WHO; IBDQ: Inflammatory Bowel Disease Questionnaire