Improved methods and actionable tools for enhancing HTA

WP10 Guidance to Support Consistent HTA Appraisal for Orphan Medicinal Products

D10.2 Guidance on Use of Patient Reported Outcome Measures in HTA Appraisals of Rare Disease Treatments

E Nicod[1], M Meregaglia[1], A Whittal[1], S Upadhyaya[2], K Facey[3], M Drummond[4]

[1]SDA Bocconi – Italy, [2]National Institute for Health and Care Excellence – England, [3]University of Edinburgh – Scotland, [4]York University - England

March 2021

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 779312. The results presented reflect the authors' views and not those of the European Commission.

Contents

1. Overview of Work Package 10 and of PRO Guidance Development

IMPACT HTA Work Package 10 (WP10) has developed guidance on novel approaches to appraising rare disease treatments to support robust, accountable decision-making across Europe on high-cost products that have a limited evidence base with high uncertainty and intense stakeholder scrutiny. It explored how different forms of evidence obtained from a range of data sources and stakeholders can be integrated with economic modelling to inform robust, accountable decisions about OMPs and ultra-OMPs.

This was achieved through several work streams (WS) on (1) documentation of HTA appraisal/reimbursement processes for rare disease treatments (RDTs) in Europe to explore the need supplemental HTA appraisal processes for RDTs; (2) participant observation of appraisal processes to understand how decisions are being effectively made and how processes can be improved; (3) documentation of challenges and potential solutions of use of patient reported outcome (PRO) evidence and health state utility values (HSUVs) in HTA; (4) synthesis of challenges and potential solutions around design and implementation of outcome-based managed entry agreements in HTA [Deliverable 10.3]. The research outputs from these four workstreams have then feed into an overall HTA Appraisal Framework for RDTs [Deliverable 10.1].

The objective of WS3 was to create a guidance document on the use of PROMs in HTA for RDTs. PROMs of all types can be used in HTA, but in some jurisdictions they take the form of health state utility values (HSUVs), since these can be used to calculate the quality-adjusted life-years (QALYs) gained from treatments. We defined the nuances around use of PROMs/HSUVs in rare diseases, illustrated these with examples on use of PROMs/HSUVs in HTA in different countries, and provided guidance on when methodological approaches that differ from the standard may be suitable to measure impact on quality of life (QoL) in rare diseases. A mix of approaches was used including literature scoping/systematic reviews, documentation around use of PROMs in appraisal of RDTs and exploration of better use for evidence of clinical effectiveness and cost-effectiveness.

The following tasks were completed in this workstream (Figure 1.1):

- **Task 1: Review on use of PROMs in HTA for rare diseases [Chapter 3]**

Outlines advantages, challenges and potential solutions for using different types of PROMs in HTA, and the associated data collection and psychometric property considerations

- **Task 2: Use of HSUV techniques in rare diseases [Chapter 4]**

Discusses the advantages and challenges of using each available technique to obtain HSUVs in relation to the main features of RDs

- **Task 3: Systematic review of mapping in rare diseases [Chapter 5]**

Undertakes a systematic review of published studies using mapping to derive HSUVs from non-preference-based PROMs in RDs, and identifies critical issues in relation to the main features of RDs

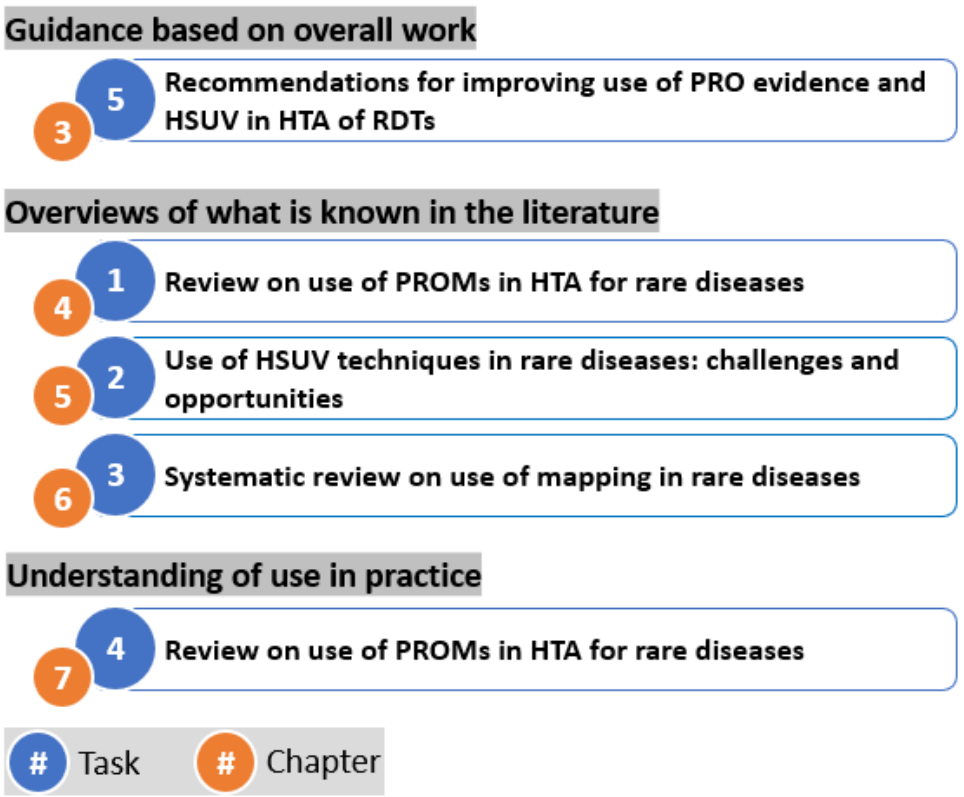
- **Task 4: Review of use of PRO data and HSUVs in HTA for rare disease treatments [Chapter 6]**

Reviews HTA reports for RDTs in four countries (England, Netherlands, France and Germany) and investigates how QoL considered through the use of PRO data and/or HSUVs was appraised and influenced the final decision. Differences across countries are contrasted and key learnings identified.

- **Task 5: Recommendations for improving use of PRO data and HSUV in HTA for rare disease treatments [Chapter 2]**

The emerging issues regarding the use of PROMs/HSUVs in HTA of rare and ultra-rare disease treatments were synthetized and critically analysed. These were then related to the overview of findings from the literature, and specifically on the nuances identified around use of PROMs and HSUVs in rare diseases. Findings from other workstreams in WP10 also fed into this work, namely around individual country appraisal processes for rare disease treatments from workstream 1 and participant observations and interview of appraisal committees from workstream 2. On this basis, five high-level recommendations were derived.

Figure 1.1 Overview of tasks conducted in this workstream



- **Follow-on work**

Task 6: Estimation of HSUVs in rare diseases: a scoping review comparing NICE technology appraisals with the existing literature

The work completed is being leveraged in ongoing follow-on work. A scoping review is being conducted to compare the methods used in the NICE appraisals with those reported in the literature. This will be used to generate a fuller understanding of what is available for these diseases in the literature, the types of techniques used to derive the HSUV values, and the types of challenges encountered and how they relate to the nature of rare diseases.

This deliverable report (D10.2) is a guidance document on how to better consider QoL in HTA for rare disease treatments, intended for HTA stakeholders. First, background information about PROMs, HSUVs and the nuances to consider when used in rare diseases is summarized in Chapter 2. Chapter 3 outlines the 5 key recommendations arising from this work. Chapters 4-7 include the results from the 4 respective tasks completed.

2. BACKGROUND

Background on PROMs

A PRO is defined as any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else. Accordingly, a PROM is a tool, such as a questionnaire or survey, that is used to measure and collect data on a PRO, usually pertaining to HRQoL, symptoms or treatment side effects, experience with care (adherence, satisfaction or health status). HRQoL is a multidimensional concept defined as the patient's subjective perception of the impact of disease and its treatment on daily life, physical, psychological and social functioning and well-being.

Figure 2.1 Types of PROMs

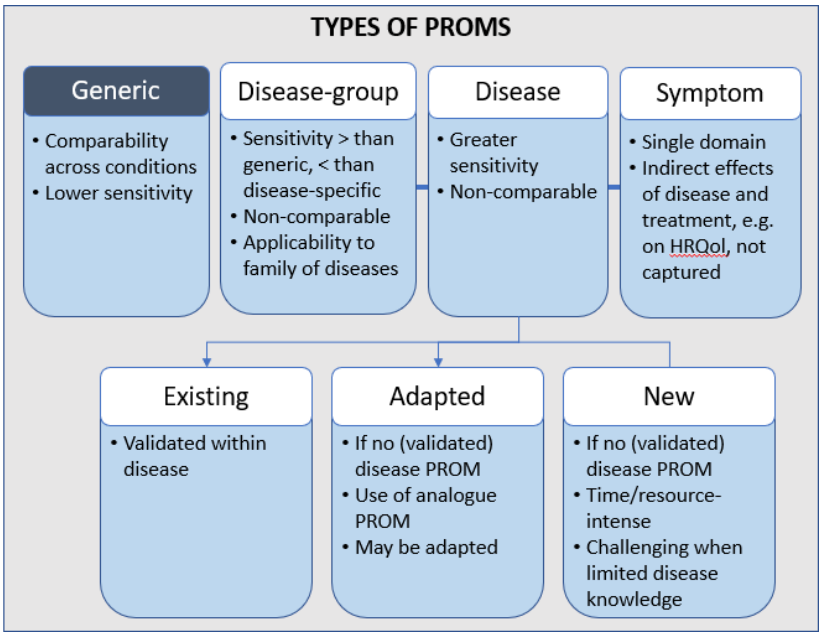


Figure 2.1 illustrates the different types of PROMs and their characteristics. Each have advantages and disadvantages. If a (validated) disease group-, disease- or symptom-specific PROMs is to be used, options include (1) use of existing validated PROM, (2) use of PROM for an analogue condition, possibly also adapting it, or (3) creation of new PROM.

Legend: PROMs: patient-reported outcome measures; PBM: preference-based measure

Overall, PROMs can be classified as generic, disease-group, disease-, or symptom-specific (Figure 2.1). ***Generic PROMs*** are condition non-specific, intended to fit a variety of a medical conditions, and can be used to compare patient groups with different conditions. They capture more general elements of HRQoL such as physical function, physical, mental, emotional health, social function, pain, etc. Examples of this type are the EuroQol-5 Dimension questionnaire classification (EQ-5D) or the 36-item Short-Form Survey (SF-36). Conversely, ***disease-specific PROMs*** are those created to record impacts that have been identified as important in specific conditions. An example is the ACTIVLIM for neuro-muscular disorders. These PROMs are more sensitive than generic PROMs to show symptoms associated with the condition and measure treatment effects, but are specific and thus not comparable across patient groups. ***Disease-group specific PROMs*** is a new category created within this work to distinguish a PROM developed for a specific group of similar conditions. Since it may be difficult to create a disease-specific PROM for each individual rare disease and particularly for the very rare ones, Morel et al (2011) highlighted the importance of PROMs for use across several diseases within a same family of diseases that share similarities, e.g. impact of chronic fatigue across diseases (Morel and Cano 2017). A common example used in oncology is the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30) (Bell et al. 2019). These PROMs are more sensitive than generic PROMs, but less sensitive than disease-specific PROMs. There are also ***symptom-specific PROMs,*** which focus on items affecting a single domain, here being a symptom, generally related to

the disease and/or treatment. Symptom-specific PROMs may fail to capture indirect effects of treatment on well-being, e.g. HRQoL, that may be important to patients (Deshpande et al. 2011). Examples include the Brief Fatigue Inventory or the M.D. Anderson Dysphagia Inventory.

Generic and disease-group/disease-specific PROMs can be preference-based or non-preference-based. **Non-preference-based measures (non-PBM) PROMs** are measured by summing answers within the different domains to provide scores that are then interpretable. **Preference-based measures (PBM) PROMs** calculate the impact of patients' ill health, and assign it a ''utility' score based on people's preferences for health states (Wolowacz et al. 2016; Neumann, Goldie, and Weinstein 2000). Specifically, patients' responses are used to generate a health state profile, which is converted into an index score (also called "health-state utility value, HSUV" or utility') based on societal preferences for the given health state, usually elicited through direct techniques (e.g. standard gamble, time trade-off).

HSUVs can be derived by converting patient (or public) responses using **indirect or direct** (Figure 2.2). **Indirect approaches** estimate HSUVs indirectly from generic or disease-specific PROMs. This can be done by using existing or by creating sets of preference 'weights' (or 'tariffs') for every combination of the PROM's domains/levels, e.g. EQ-5D. 'Mapping' represents a valuable alternative for utility assessment in cases where no preference-based PROMs are being collected. The mapping technique allows establishment of a statistical relationship between a 'source' measure (a generic or disease—specific non-PBM) and a 'target' one (usually a generic PBM, such as EQ-5D). A variety of regression methods are available for mapping, with the most common one being ordinary least squares (OLS), and the choice should be informed mainly by data distribution. Previously developed mapping algorithms can also be applied to disease-specific PRO data to obtain utility values for cost-effectiveness analyses (Longworth and Rowen 2013). In rare diseases, such mapping algorithms are rarely available, while being challenging to develop (Slade et al. 2018). This has been explored in Task 3 [Chapter 6].

Figure 2.2 Techniques to derive HSUVs

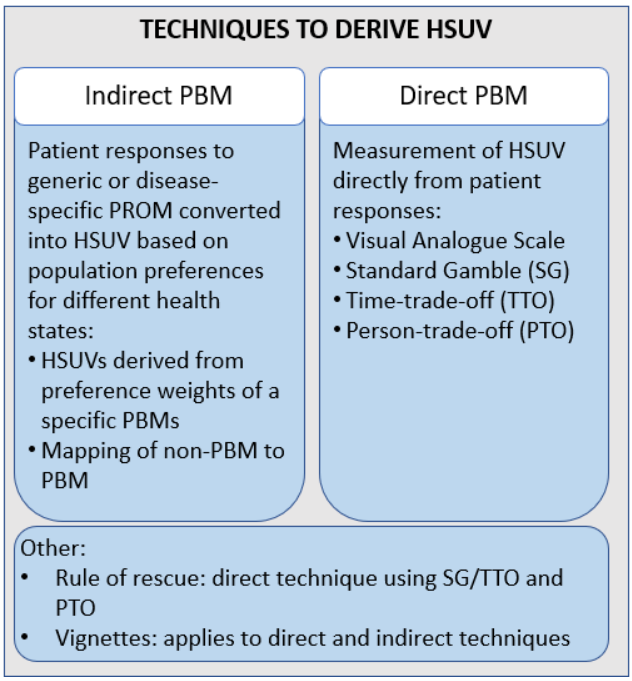


Figure 2.2 summarises the use of direct and indirect techniques to derive HSUVs. Indirect techniques are used to derive HSUVs from preference weights developed for disease-specific or generic PROMs (e.g. EQ-5D) using direct approaches, or through application of mapping algorithms relating non-PBMs to PBMs. Direct techniques measure HSUV directly from patient (or from the public using hypothetical descriptions) through various techniques, usually choice-based.

Legend: HSUV: health state utility value; PBM: preference-based measure; PROM: patient-reported outcome

Direct approaches exist to measure HSUVs of his/her own status directly from the patient or hypothetical health states ('vignettes' from the public). The most common ones are time-trade-

off (TTO) and standard gamble (SG). The SG involves trading alternative health states against each other with a risk of immediate death, whereas the TTO trades duration of life against quality of life. Their implementation may be too demanding for some patients with rare diseases (e.g. children), which may suggest using vignettes to be valued by non-patient populations (Bansback et al. 2008). The specific pros and cons of using the different direct and indirect techniques to derive HSUVs in rare diseases have been examined within Task 2 [Chapter 5].

HSUVs are usually measured on a scale of 0.0 to 1.0, where 1.0 represents full health, 0.0 death, and negative values represent health states worse than death. A common example of use of HSUVs is for the calculation of quality-adjusted life years (QALYs), which is a composite measure of a person's life weighed for their quality of life.

Approaches to using quality of life in HTA

PROMs are increasingly being used to derive information on a treatment's value, and are considered in HTA processes in two ways: (1) the added benefit assessment and (2) the cost-effectiveness assessment.

Each HTA body has its own requirements or preferences regarding the types of PROMs to consider in the appraisal of treatments (Figure 3.3). In added benefit assessments, HRQoL data captured through PROMs are used alongside consideration of morbidity/mortality for life-threatening conditions (e.g. cancer), and in less severe but disabling conditions (e.g. arthritis) to demonstrate their added benefit. In France and Germany, the Haute Autorité de Santé (HAS) and Institute for Quality and Efficiency in Health care (IQWiG) recommend use of both generic and disease-specific PROMs for their added benefit assessments (Haute Autorite de Sante 2018; Klakow-Franck 2014); other countries would accept disease-specific PROMs if the generic PROM was insufficiently sensitive to change, did not capture all dimensions of interest, or if its consideration allows for a more in-depth understanding (EUnetHTA 2015). Other countries do not recommend inclusion of disease-specific PROMs in general (EUnetHTA 2015). There is also a strong preference for PROMs measured with psychometrically appropriately validated instruments, collected in clinical trials (Klakow-Franck 2014; Haute Autorite de Sante 2018; EUnetHTA 2015); some also prefer PROMs representing a health outcome over a process measure (EUnetHTA 2015; Klakow-Franck 2014).

Figure 2.3. Considerations for use of PROMs in HTA

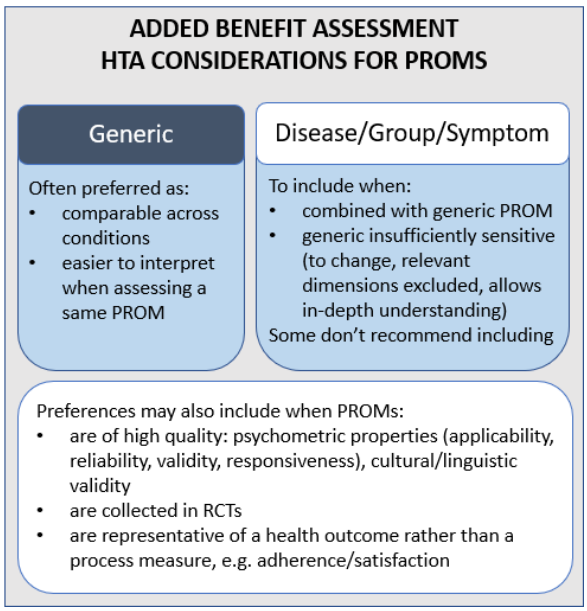


Figure 2.3 summarises considerations for selection and use of PROMs in HTA. Some countries prefer generic PROMs as it is easier for decision-makers to interpret and compare results across conditions. Other countries also require disease-group/disease-/symptom-specific PROMs combined with generic, or disease-group/disease-/symptom-specific PROMs when the generic PROM is insufficiently sensitive. Other countries do not require any disease-group/disease-/symptom-specific PROMs. Individual countries also have a number of preferences around psychometric properties, types of outcomes measures and data source for their collection.

Legend: PROMs: patient reported outcome measures; RCTs: randomised controlled trials

Cost-effectiveness analyses aim to capture a treatment's added benefit using a composite measure of length and HRQoL, the QALY being the most common measure of added benefit used in cost-effectiveness models. The HSUVs to collect should correspond to the different health states included in the economic model, for both the treatment and comparator (Brazier et al. 2019). On this basis, HSUVs identification is done via iterative searches by broadening the scope of the search when no suitable HSUVs were identified, until the appropriate estimates are selected (Brazier et al. 2019). Often, the "ideal" HSUV is not available, and a choice between the "best" or "most appropriate" HSUV estimate is to be made, based on patient characteristics, specific country evidence, consistency and preferences around use of measures within a disease/country (Brazier et al. 2019). Inclusion of additional HSUV estimates relative to, for example, adverse events, co-morbidities or impact on caregiver HRQoL would follow a similar process as above.

HTA body-specific preferences or requirements for use of HSUVs in HTA include indirect preference-based approaches, where some may additionally prefer the use of a specific PROM to derive the HSUVs (Figure 2.4). This would be the case of the EQ-5D, which is the most recommended PROM to be used in adults (National Institute for Health and Care Excellence. 2013; Haute Autorite de Sante 2018). WP5 of IMPACT-HTA is exploring its use in adolescents. Direct techniques are often preferred if indirect techniques are not suitable/possible, as they are more difficult to develop and implement. HTA bodies may also prefer that the HSUVs presented include measurement from the patient, use of validated PROMs, and/or, when possible, use of public preferences over patient preferences (Wolowacz et al. 2016; EUnetHTA 2015).

Figure 2.4. Considerations of use of HSUVs in HTA

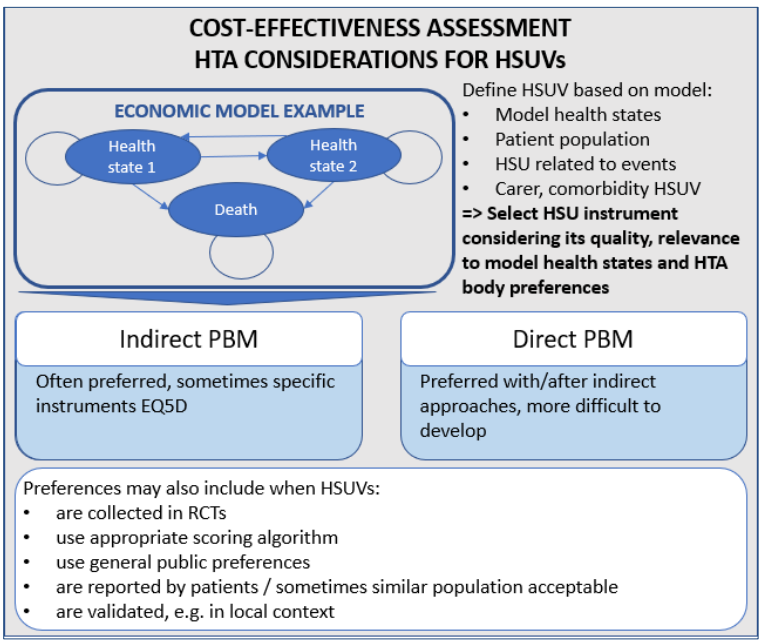


Figure 2.4 represents the different considerations when selecting HSUVs to include in an economic model for the purpose of HTA. The HSUVs to include should correspond to the model health states for the treatment and control arms, and be representative of the patient population. Other HSUVs relating to adverse events, co-morbidities or carers are to be included if there is correspondence to the model health states and if they are relevant in determination of a change with and without treatment. HTA bodies have different preferences or requirements in the types of PROMs to include, which are listed in the figure.

Legend: HTA: health technology assessment: HSU: health state utility; HSUV: health state utility value; PBM: preference-based measure; RCT: randomised controlled trial

Considerations for use of PROMs/HSUVs in HTA for rare diseases

Considering PRO data alongside clinical endpoints may improve the understanding of disease burden and treatment impacts, particularly in the absence of high-quality data, which is often the case in RDTs. The use of PRO evidence, however, is inconsistent across HTA bodies. There are a number of challenges in collecting high quality PROMs from patients living with a rare disease, which then has implications when they are used in HTA. Some of these

challenges are not specific to RDTs, but are exacerbated by the characteristics of rare conditions (e.g. small sample sizes, lack of disease natural history knowledge, etc.), and include:

- ability to collect data from small, hard to access, heterogeneous patient populations and related issues with statistical analyses (Slade et al. 2018; Benjamin et al. 2017);
- difficulty capturing meaningful and generalizable outcomes because of small and heterogeneous conditions and presentations (Bell et al. 2019; Benjamin et al. 2017; Pascoal et al. 2018; Slade et al. 2018; Morel and Cano 2017), e.g.:
    - challenge in demonstrating concept validity via concept saturation (Benjamin et al. 2017);
    - health outcomes can vary from patient to patient and as disease progresses (Morel and Cano 2017);
    - often complex, multi-system conditions, possibly confounded with co-morbidities, where it may be challenging to distinguish between symptoms from rare disease and those from co-morbidity (Morel and Cano 2017);
- frequent lack of information and understanding regarding disease progression and natural history, which makes it difficult to know which PROMs to use or how to develop new PROMs (Benjamin et al. 2017; Morel and Cano 2017; Tosi et al. 2019);
- frequent lack of validated PROMs for specific patient populations, particularly in children and adolescents (Slade et al. 2018);
- many issues that are important to patients are not captured with existing measures (Morel and Cano 2017; Swezey et al. 2019);
- multi-site/international studies maybe be used to gain a larger sample size, but psychometric and linguistic/cultural validation is then challenging (Benjamin et al. 2017);
- patients frequently cannot self-report because they are children or have cognitive impairments associated with the disease, putting a reliance on parent or other proxy measures (Benjamin et al. 2017).

Other observations around use of PROMs for rare diseases from the various workstreams in WP10 include:

- possible floor/ceiling effects due to lack of large sample sizes;
- health states in chronic conditions with early onset, e.g. infancy, may be given a higher value by patients as this is their “normal”;
- disease may affect multiple organs (potentially also in a heterogeneous manner), making it more challenging to select the appropriate PROM(s) to capture the relevant outcomes ensuring sufficient respondents to measure change;
- the use of unconventional approaches, such as alternative analyses to create proxy measures for utility scores, may be necessary for a RDT, but then pose difficulties for decision makers to interpret/accept;
- Capturing carer utility/disutility values is relevant but difficult and often results in high uncertainty.

The challenges identified in using PROMs in RDs are then also applicable to the HSUVs derived from these. A number of additional issues related to applying certain techniques to derive HSUVs for HTA in rare diseases include:

- the requirement for a large number of respondents to minimise random measurement errors (e.g. person-trade-off, development of mapping algorithms);
- challenges in identifying appropriate HSUV estimates that correspond to the health states in the model;
- lack of sensitivity of QALYs to baseline disease severity (does not account for how bad patients are before treatment (Towse and Garau 2018)), and to some changes in the patient’s condition (Pearson et al. 2018).

To better incorporate PROs into HTA decision making, a clearer understanding of the challenges in using PROMs/HSUVs in rare diseases and ways to tackle these is needed for all stakeholders. This was explored and discussed in Tasks 1-4 [Chapters 3-6]. Key considerations regarding use of PROMs and HSUVs in HTA for RDTs are explored in Tasks 1 and 2 [Chapters 3-4], respectively. Task 3 explores in more depth the use of mapping techniques in rare diseases [Chapter 5]. In Task 4, NICE HTA decisions for RDTs within their TA and HST programmes were analysed to understand the types of PROMs being considered, their impact on decision-making and how the issues highlighted by the Appraisal Committees relate to the identified common challenges encountered for RDTs [Chapter 6]. This was then contrasted with the PRO data appraised for the same treatments in Germany, France and The Netherlands. A better understanding of the nuances when dealing with PRO data and HSUV in rare diseases may be useful for HTA stakeholders when developing/selecting the appropriate PROMs, and when interpreting the PROM/HSUV evidence presented. The five recommendations derived from this work highlights the important nuances to understand and makes specific recommendations on how to deal with them [Capter 2], pointing to the relevant findings that arose across the different tasks and workstreams of WP10.

3. RECOMMENDATIONS FOR IMPROVING USE OF PATIENT REPORTED OUTCOME EVIDENCE AND HEALTH STATE UTILITY VALUES IN HEALTH TECHNOLOGY ASSESSMENT OF RARE DISEASE TREATMENTS

Elena Nicod, Michela Meregaglia, Amanda Whittal, Sheela Upadhyaya, Karen Facey, Michael Drummond, PhD

This chapter outlines the key recommendations arising throughout the different WS of WP10 around use of PRO data and HSUV in HTA of rare diseases. This version will be submitted to The Patient and may be revised based on the comments received.

3.1 Background

Diseases affecting a small number of patients were distinguished as rare more than 20 years ago within national or regional regulatory legislations (FDA Orphan Drug Act 1981; EMA Regulation (EC) No 141/2000). This was founded on the need to incentivise research and development given the small returns on investment to be expected in these conditions. More than 6,000 rare diseases exist, with the majority being extremely rare. They are often chronic, life-threatening, have early onset in infancy or children, and 80% are of genetic origin (Nguengang Wakap et al. 2020; Eurordis, Rare Diseases). To date, around 190 rare disease treatments ("orphan medicinal products") have been authorised by the European Medicines Agency (European Union, Orphan Medicinal Products).

Health Technology Assessment (HTA) is used to inform decisions about reimbursement or routine provision of a treatment to a specific population. It aims to assess how treatment effects translate into patient benefit through two main approaches: added benefit assessment and cost-effectiveness assessment. Added benefit is generally evaluated based on clinical and quality of life endpoints, the former being the main health outcome of interest given that it is more easily quantifiable and comparable across indications (e.g. survival gains) (Drummond et al. 2015). The latter provides a subjective self-report of a patient's health status without the interpretation of this report by a third person, e.g. clinician (FDA 2009). A patient-reported outcome measure (PROM) is the tool used to collect this information, which incorporates domains such as quality of life (QoL, e.g. daily life, physical, psychological and social functioning and well-being), symptoms, or treatment side effects. Two groups of PROMs exist. Generic PROMs are not condition-specific and are comparable across multiple indications. In contrast, disease-specific PROMs are developed for a specific condition, family of conditions (referred to as "disease-group"), or symptoms, and are thus more sensitive to change, but not comparable across indications.

In cost-effectiveness analyses, health effects are quantified using a composite measure of length of life adjusted for QoL called the quality-adjusted life year (QALY). QoL is weighed using health state utility values (HSUVs), which are numerical values ranging on a scale of 0.0 (representing death) to 1.0 (full health). HSUVs are derived by converting patient (or public) responses using indirect or direct approaches. The former converts generic (e.g. EQ-5D) or disease-specific PROMs into HSUVs through the use of pre-determined preference weights for every combination of the questionnaire's domains and levels, or through the creation or use of a statistical relationship between a source measure and a target one with existing value sets, referred to as "mapping". Direct techniques measure the patient's health status directly or hypothetical states ("vignettes") from non-patient populations through various techniques, which usually require participants to make trade-offs in a choice-based context. This approach is less preferred by HTA bodies (Brazier and Rowen 2011).

Therefore, QoL and cost-effectiveness results are examined in the context of HTA bodies' appraisal frameworks, which may also include other domains relevant to decision-making. These include, for example, the nature of the condition (England, Scotland), the impact of treatment beyond direct health benefits (England, Scotland), the implementation in clinical practice, whether it involves children (Germany), and ethical considerations (Bulgaria, Latvia). The nature of this additional evidence is generally qualitative and examined by the appraisal committee during the deliberative process.

Due to the nature of rare diseases in terms of, for example, small and heterogeneous populations, being severely life-threatening and debilitating, chronic and progressive, the evidence to demonstrate the added benefit of a treatment is often characterised by high levels of uncertainty (Nicod et al. 2019). This, in addition to their high prices (Luzzatto et al. 2018), has resulted in challenging reimbursement appraisals for many rare disease treatments undergoing HTA (Nicod, Maynou, et al. 2020). Considering that few treatments for rare diseases have a curative intent, it is essential to appropriately account for their impact on QoL when assessing their added benefit. However in rare diseases, this has shown to be challenging. In the added benefit approach, although PRO evidence can be considered in principle, they often do not receive sufficient consideration given they often lack validation or do not show significant effects due to the lack of sensitivity of the instrument or underpowered statistical tests. In the cost-effectiveness approach, PRO evidence can be considered, but a strong preference is for a HSUV, which may not reflect, or is not very sensitive to the changes in QoL that may matter most to the patient.

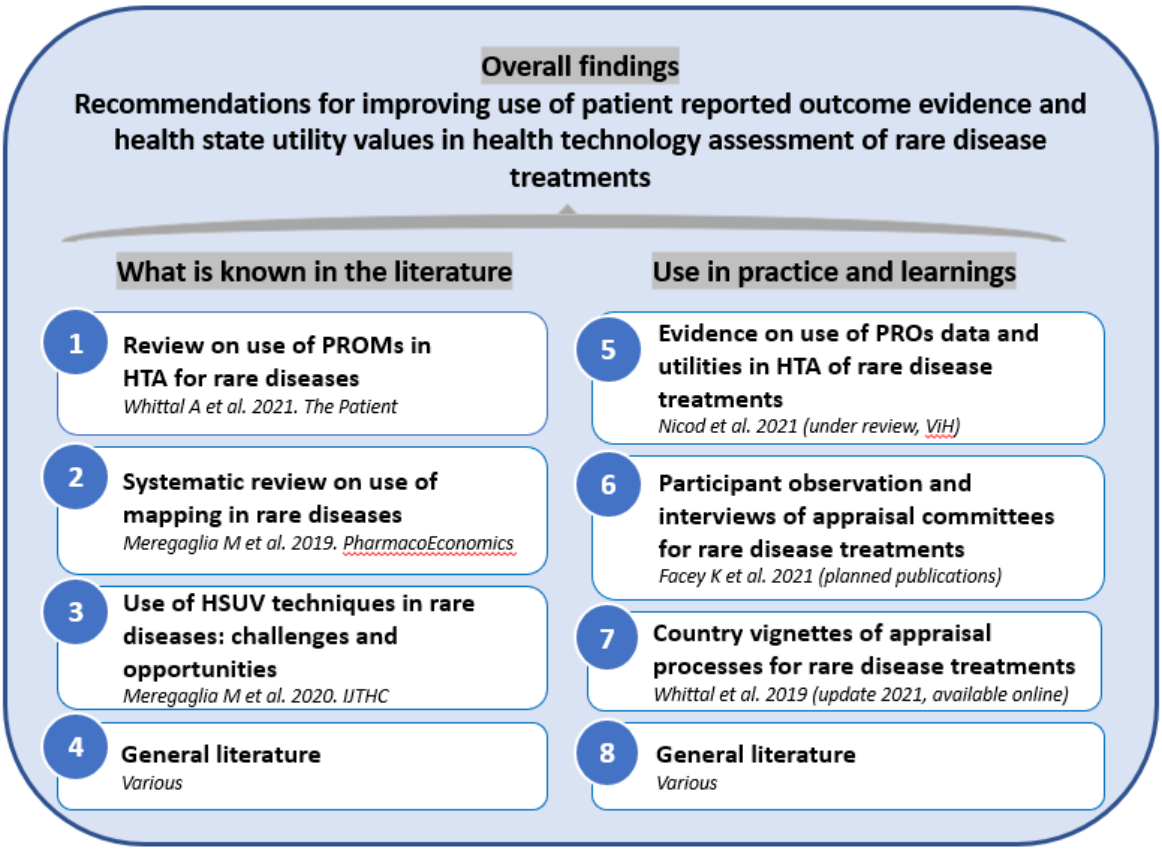
Work Package 10 (WP10) of the EU-funded project IMPACT-HTA on appraisal of rare disease treatments was undertaken to develop guidance on how to better use PRO evidence and HSUV estimates in HTA decision-making. Part of this work examined the nuances around use of PRO evidence and HSUV estimates in HTA of rare diseases (Whittal, Meragaglia, and Nicod 2020; M. Meregaglia, Nicod, and Drummond 2020; Michela Meregaglia et al. 2020). This paper outlines key recommendations for HTA bodies and stakeholders on how to better consider QoL through the use PRO and HSUV evidence in the appraisal of rare disease treatments.

3.2 Methods

Our recommendations for improved use of PRO evidence and HSUV estimates in HTA of rare disease treatments were derived from various workstreams examining (1) learnings from the literature, and (2) their use in practice (Figure 1). Learnings from the literature covered existing reviews on use of PROMs and PRO evidence (Whittal, Meregaglia, and Nicod 2021), techniques to derive HSUVs (M. Meregaglia, Nicod, and Drummond 2020), and use of mapping (Michela Meregaglia et al. 2020) in rare diseases. Evidence on their use in practice was derived from three different research activities. First, ethnographic participant observations of English-speaking Appraisal Committees in England, Scotland and Canada and interviews with the stakeholders involved in the appraisals. Second, a review of England's National Institute of Health and Clinical Excellence's Highly Specialised Appraisal (NICE HST) and Technology Appraisal (NICE TA) processes for 24 non-oncology rare disease treatments based on their published appraisal reports. This document analysis focused on how PRO evidence and HSUVs influenced the decision, and the extent to which other forms of evidence supported the interpretation (e.g., patient input and evidence). A cross-country comparison with the Netherlands (National Health Care Institute, ZIN), France (Haute Autorité de Santé, HAS), and Germany (Federal Joint Committee, G-BA) for the same treatments was also conducted (Nicod et al. 2021a). Third, country vignettes summarising individual appraisal processes for rare disease treatments, which include information about evidentiary requirements for PROMs and their respective appraisal frameworks (Nicod, Whittal, et al.

2020). These were then related to and augmented by existing literature to identify potential arguments or counterarguments around the recommendations being made.

Figure 3.1 Workstreams and evidence that fed into the development of the recommendations



3.3 Recommendations

The five recommendations are summarized in Table 3.1 and further elaborated on within the next sections.

Table 3.1 Recommendations for improving use of patient reported outcome evidence and health state utility values in health technology assessment of rare disease treatments

Recommendations for improving use of patient reported outcome evidence and health state utility values in health technology assessment of rare disease treatments
1. When appraising rare disease treatments, it is essential to understand impacts of the condition and treatment on patients' QoL

2. When critically assessing PRO evidence, challenges related to development and administration of PROMs should be taken into account
3. During the appraisal, interpretation of PRO evidence should recognise that lack of significant effect does not necessarily imply lack of benefit on QoL
4. Other forms of evidence such as qualitative evidence and expert input should be considered to enable a fuller appreciation of the impact of a medicine on QoL
5. It is important to consider family and carer perspectives to better capture the added benefit of a medicine

I. When appraising rare disease treatments, it is essential to understand impacts of the condition and treatment on patients' QoL

Burden of Condition

QoL for patients living with a rare disease is often very poor (Bogart and Irvin 2017). This is likely a consequence of the common issues encountered in, and often unique to, rare diseases around, for example, difficulties in diagnosis, lack of disease knowledge, information and expertise, demanding treatment pathways, uncertainty about the future, stigma, isolation, emotional stress, or the debilitating and progressive nature of these conditions (Hong, Villalonga-Olives, and Perfetto 2019; Bryson et al. 2019).

This was observed in the NICE appraisals analysed and observed, where the majority were debilitating and life-threatening (Nicod et al., draft). Particularly in the HST appraisals, it was also identified that the conditions were progressive, heterogeneous, multi-system, and with an early onset. Patients' QoL were considered to be impacted in all of the conditions appraised. From the documentation of appraisal processes in Europe, it's not clear that many countries specifically consider the burden of the disease in their appraisals.

Impact of treatment on QoL

When considering the intended treatment effects for all of these treatments, QoL was considered to be improved, whereas length of life improved for 75% of cases.

Comparing this to HTA evidentiary requirements for QoL across countries, the impact of treatment on QoL often appeared to be perceived as secondary or less important. Of 36 countries, only 47% included evidentiary requirements for PRO evidence. In the remaining cases, it was not required but accepted if presented for 25%, and not accepted in 27% of cases (Nicod, Whittal, et al. 2020). In those countries that do require QoL evidence, there was a general preference for trial data collected with a generic and validated PROM (EUnetHTA 2015).

When considering that most rare disease treatments are not curative, it is important to collect PRO data routinely and account for it when appraising rare disease treatments. The importance of QoL in drug development is increasingly recognised following the guidance on development and implementation of PRO measures published by the FDA in 2009, and the guidance on Patient-focused Drug Development published in 2020 (FDA 2009; Food and Drug Administration, Center for Drug Evaluation and Research, and Center for Biologics Evaluation and Research 2020). The latter focuses on ensuring that the evidence collected reflects what matters most to patients and the possible approaches to do so, including the use of qualitative methods. It is important for the HTA community to be aware of these important developments and ensure alignment with the type of evidence they are willing to accept. Early multi-stakeholder dialogues to discuss and align on PROs and QoL measurement could help ensure greater alignment and better use of more meaningful evidence.

II. When critically assessing PRO evidence, challenges related to development and administration of PROMs should be taken into account

In rare diseases, challenges measuring QoL arise in the selection, development, and administration of PROMs. Some of the challenges are not specific to rare diseases but are exacerbated by the nature of these conditions. It may be more difficult to develop or select appropriate PROMs in situations where disease progression and natural history are not well understood (Morel and Cano 2017; Benjamin et al. 2017; Tosi et al. 2019), or when disease affects multiple organs (potentially also in a heterogeneous manner) (Nicod et al., draft). It may also be the case that the selected PROM fails to capture a number of issues important to patients (Morel and Cano 2017; Swezey et al. 2019), or may not be validated in the population of interest, particularly when children and adolescents are affected (Slade et al. 2018). Further, there is a paucity of disease-specific PROMs developed and validated in rare diseases (M. Meregaglia, Nicod, and Drummond 2020).

There are also a number of challenges administering PROMs within small, hard to access, and heterogeneous patient populations (Bell et al. 2019; Benjamin et al. 2017; Pascoal et al. 2018; Slade et al. 2018; Morel and Cano 2017). These include difficulties demonstrating concept validity via concept saturation (Benjamin et al. 2017), capturing treatment effect when health outcomes vary from one patient to another or as disease progresses (Morel and Cano 2017), or distinguishing between symptoms and co-morbidities in often complex, multi-system conditions with possible confounding symptoms (Morel and Cano 2017). In addition, patients are frequently unable to self-report, as they may be children or cognitively impaired, and administration would need to rely on parent or other proxy measures (Benjamin et al. 2017). Multi-site/international studies may be used to gain a larger sample size, but psychometric and linguistic/cultural validation is then challenging (Benjamin et al. 2017). Other challenges around valuing health effects in rare diseases include possible floor/ceiling effects due to lack of large sample sizes, or a higher valuation by patients living with chronic conditions from early childhood, as a particular health state may be considered their “normal” (Nicod et al., draft).
Many of these issues were reflected in the analysis and observations of NICE appraisals (Facey, draft; Nicod et al. 2021a), and possibly explain the frequent inconclusive, or uncertain nature of the, PRO results. The most frequent issues related to the short trial duration, improvements being seen within only some of the measure’s dimensions, or minimally important differences (MID) not being demonstrated. Issues around ceiling effects were also common in chronic conditions. This is likely a consequence of patients adapting to their condition and ranking it as “their normal”, or of conditions presenting intermittent symptoms, where responses may change depending on the health state at that moment. There are also issues around face validity, when baseline values from patients of their own health status are higher than in the general population, and may be biased upwards because of, for example, the hope of participating in a trial.

In terms of the treatments intended for children, a wide range of approaches were used to derive HSUVs and very few children-specific PROMs, such as the PedsQL, were used, but even so were often inconclusive. Challenges in collecting PRO data from children were also repeatedly recognised (Facey, draft; Nicod et al. 2021b). This emphasises the need to better understand the limitations in applying existing approaches such as the EQ-5D in children, and in parallel to continue to develop methods for use in children (Hill et al. 2020). For example, a child-friendly EQ-5D version (EQ-5D-Y) is available for children and adolescents, and already applied in some rare diseases (e.g., cerebral palsy) (Burström et al. 2014).

Only three disease-specific PROMs were identified for five of 24 medicines appraised, and even when available, their results were mostly inconclusive. Considering these measures are expected to be more sensitive to change, their inconclusiveness may suggest that the nature of the certain rare disease populations are unsuitable for these more conventional approaches to measuring QoL or that the test was underpowered for this endpoint. Many of the PROMs considered focused on (groups of) symptoms, rather than on QoL. One possible explanation for this is that it may be easier to demonstrate added benefit through improved symptoms rather than through improved QoL in complex conditions (Hong, Villalonga-Olives, and Perfetto 2019). However, HTA tends to be more interested in QoL than specific symptoms and in such cases, this perspective may fail to have been captured.

The NICE appraisal committees also highlighted a number of aspects important to patients not captured by these measures. These included aspects around daily and social life, work, independence, dignity, mobility, adverse events, tolerability, dosing, fear, effect on family, and patient choice. Considering that aspects such as mobility or self-care are often included in QoL measures, such as the EQ-5D, this may indicate that these measures are not sufficiently sensitive to capture change.

The aim of HTA is to understand QoL with and without the treatment under review. It is therefore important to recognise that developing, validating, and using PROMs in rare diseases may be more challenging due to the small, heterogeneous and young nature of these patient populations. PRO evidence should be critically assessed in light of what is feasible in each specific context and based on what is known at the time of evidence generation.

III. During the appraisal, interpretation of PRO evidence should recognise that lack of significant effect does not necessarily imply lack of benefit on QoL

The consideration of PRO evidence in HTA adds an additional layer of complexity to the appraisal process. There are two main considerations here. First, in light of what is feasible in rare diseases, the interpretation of evidence may vary depending on individual evidentiary requirements of a given HTA body. Second, the QoL evidence may be different depending on the HTA approach used in that setting, with cost-effectiveness approaches often requiring an additional step in evidence translation to generate a HSUV. As a result, lack of significant effect of QoL evidence in HTA does not necessarily imply lack of benefit. The implications of these considerations in rare diseases are discussed here.

Different evidentiary requirements could lead to different interpretations of the same QoL evidence or to different evidence being appraised. This first scenario was seen in G-BA and HAS's appraisals of the Norfolk-DN data for inotersen, considered to be inconclusive for G-BA as it did not demonstrate clinical relevance despite being statistically significant, but modestly improved for HAS, since it was a co-primary endpoint (Nicod et al. 2021b). One must consider these requirements in light of what is feasible in rare diseases.

The most common HTA requirements for PRO evidence include the use of validated PROMs, PRO data from primary or secondary endpoints, and/or demonstration of MID. However, such

requirements are often more challenging to achieve in rare diseases. It is important to recognise that PROM validation requires large samples and is thus more challenging for rare diseases (Basch and Bennett 2014). Limited face validity is another common issue, also reflected in our analysis of NICE appraisals (Bell et al. 2019; Nicod et al., draft; Facey, draft). In such cases, it is essential to judge this evidence in light of the perspectives from patients and clinicians. Another common issue encountered is the failure to meet the MID criteria; however, this does not imply lack of improvement as, in a heterogeneous patient population, it may be met for some patients and not for others (Rüther et al. 2016)(Nicod et al., n.d.). In terms of the hierarchy of PRO endpoints, very few of those presented to support FDA labelling claims to date are sufficiently powered to assess impact on QoL because they are not primary endpoints in the trial (Hong, Villalonga-Olives, and Perfetto 2019). As a result, much of the QoL evidence produced in rare diseases may not meet HTA standards. Hopefully, this is likely to evolve with the trend towards patient-focused and patient-centred drug development (Food and Drug Administration, Center for Drug Evaluation and Research, and Center for Biologics Evaluation and Research 2020).

Nevertheless, there are implications for all stakeholders:

- For manufacturers, there is a need (1) for better evidence generation by, for example, ensuring that the PRO data being collected are high up in the evidence hierarchy as primary or tested secondary endpoints, and (2) to engage early on with decision-makers to understand what evidence they are willing to accept;
- For HTA decision-makers, there is a need (1) to recognise what is feasible in rare diseases, and (2) to better align their requirements with regulators and other HTA bodies, as there is only so much evidence that can be produced in small patient populations;
- For patients, there is a need for early involvement to ensure that the evidence collected and being appraised reflects what matters to them.
-

A second contrast is seen between the evidence for added benefit versus cost-effectiveness approaches. The former accounts for the most relevant PRO data, as discussed in the previous paragraph, while the latter require PRO data (source) to be translated into utility values. Ideally, the instrument to collect the source data would be a generic instrument, such as the EQ-5D. If not the PRO data collected would need to be mapped on to the EQ-5D, or used to derive another, disease-specific utility measure.

Deriving utility values from PRO data requires large datasets and is therefore prone to measurement error and high levels of uncertainty. This was reflected in the NICE appraisals analysed, where a large majority of the utility values derived from EQ-5D was uncertain. Challenges in identifying the appropriate utility estimates corresponding to the model's health states were also common (Towse and Garau 2018). These were explained by the limited trial data, data not being available for all health states, limited face validity, or issues around the data generating implausible health states (Nicod et al. 2021b), which are issues more common in rare diseases. Unconventional approaches (e.g. vignettes) were frequently used to derive HSUVs, but a number methodological issues made it more difficult for decision makers to interpret/accept such approaches (Towse and Garau 2018; Nicod et al. 2021b). The lack of sensitivity of QALYs to baseline disease severity and to changes in the patient's condition were also common (Towse and Garau 2018; Pearson et al. 2018). This may be due to possible ceiling effects, limited face validity or, in many of the cases, where no change was captured (Nicod et al. 2021b).

This additional layer of complexity when applying PRO evidence from rare diseases in HTA needs to be recognised. In light of this, committee members would benefit from discussing the feasibility and reasonableness of the evidence presented for each case during the deliberative process, and recognise benefit even when there is not conventional significant effect. This is

further discussed in the next recommendation. There is also a need for HTA bodies to work closely with patients and manufacturers to develop reasonable approaches that can be applied within the limitations of rare diseases.

IV. Other forms of evidence such as qualitative evidence and expert input should be considered to enable a fuller appreciation of the impact of a medicine on QoL

Conventional approaches for generating QoL evidence are not always suitable in rare diseases, in which case other approaches may be used, such as vignettes or published literature. Just over one third of HSUV estimates considered in the analysis and observations of NICE appraisals used such approaches, mostly vignettes. The reasons put forward to justify their use included issues around availability and appropriateness of the QoL data (Nicod et al. 2021b). This is aligned with NICE's Decision Support Unit view, whereby vignettes would be accepted when no other data based on validated PROMs are available (Brazier and Rowen 2011). However, these approaches were not always well received and a number of methodological issues were raised around health states being derived by a small number of clinicians, or QoL data not collected from patients (Nicod et al. 2021a). Further, the availability of only few specialists for a specific rare diseases may be a possible confounding factor. While vignettes can be an interesting alternative to measuring QoL for some rare conditions, their development needs to be robust and would benefit from methodological guidance on what would be acceptable evidence.

In the vignette approach, scenarios ("vignettes") based on hypothetical health states are built and respondents are asked to measure their QoL for each of the hypothetical health states being used. Development of "vignettes" requires rigorous qualitative research (e.g. in-depth interviews, focus groups) involving patients/carers and clinicians (Brazier and Rowen 2011; M. Meregaglia, Nicod, and Drummond 2020). It is essential that the face validity of these vignettes is then verified by clinical experts (e.g. recognition of the disease via the hypothetical health states described). QoL data collection from patients is preferred, but may not always be possible. It is important to involve patients when deciding on this and provide appropriate justification. Face validity of the HSUVs derived should also be checked in lay terms with patients to ensure they are accurate. A final distinction important to highlight is that HSUVs from preference-based measures such as the EQ-5D typically reflect societal preferences, whereas approaches such as vignettes represent patient (or other types of participants) preferences. This may be considered a major limitation for some decision-makers.

Published literature from another disease areas is another option when no QoL data is available. This was seen in NICE's appraisal of obeticholic acid used HSUVs derived for Hepatitis C given that the burden of both disease relates to liver failure and would therefore be comparable. However, it is important to seek clinical opinion on the appropriateness in another indication (Pearson et al. 2018).

There is a need for technical understanding of how HSUVs are derived in order to interpret them correctly. Often, different values derived when using different techniques. Patients and clinicians should be involved in interpreting the face validity of the data collected.

In the analysis of NICE appraisals, patient input was considered to complement the existing quantitative clinical evidence. This was mainly viewed as supporting the interpretation of the EQ-5D evidence, often considered uncertain. The type of perspective provided by patients served to support the interpretation of the uncertainty, inform on dimensions not captured by the PROM or provide patients' perspectives on other aspects such as administration preferences. The extent to which input is influential in decision-making depends on the HTA process and may be affected by the applicable decision-making framework. NICE has clear

processes for involvement of patient groups and patients and carers in terms of submission of information and attendance at the appraisal meeting. Their HTA reports carefully document the appraisal committee's deliberations and note patient inputs. Our NICE HST observations of volanesorsen, voretigene neparvovec and onasemnogene abeparvovec identified that the reports capture the patient input to committee well – showing a real impact in terms of describing the nature of the condition and commenting on the adequacy of PRO instruments and HSUV derivations. For example: for volanesorsen - EQ5D does not capture patient experiences (fear of eating, fear of pain); for voretigene nevoparvovec - lack of face validity of utility measures (losing vision would not lead to a state worse than death). It is also the case that a broader decision-making framework, such as SMC's ultra-orphan drug framework, allows for a more holistic consideration of aspects that may matter to patients beyond the QALY including aspects related to the nature of the condition and a joint statement from patients and clinicians that considers impacts beyond the QALY that are not captured in the economic model.

This emphasises the benefits of triangulation of evidence and inputs to better interpret QoL impact in decision-making, including inputs such as patient surveys, patient testimony, and clinical-expert input to support the interpretation of the clinical evidence. Furthermore, the evidence to inform an HTA can be expanded to include qualitative research such as patient-based evidence (from formal social sciences research), ethical analysis, etc. These elements can help explain the multi-faceted and complex nature of QoL, help with interpretation of clinical trial results and explore the face validity of assumptions in economic modelling. Patients would benefit from training and guidance on how to provide meaningful input.

V. It is important to consider family and carer perspectives to better capture the added benefit of a medicine

Rare diseases mostly affect children and are life-threatening, disabling and often complex. As such, they are burdensome for carers in terms of time commitment, reduced work, deteriorating mental health, difficulties in daily activities, and care and financial burden (Eurordis n.d.; Simpson et al. 2021). Some of the main nuances around rare diseases include the fact that these diseases are often less well understood, and therefore there is substantial uncertainty around when and how they may deteriorate, which increases stress levels for patients and family members/carers. Carers (parents) may also feel grief and desperation when caring for young children or infants who may be severely debilitated, or for whom there may be little hope of survival. The genetic nature of these conditions also means that siblings are often affected. The burden of care may also be increased due to the necessity of frequent travel to specialised centres and possibly uncoordinated care (Simpson et al. 2021).

In our analysis of NICE appraisals, impact on carer QoL was considered in only half of the cases where it was it was expected to be improved by the treatment (Nicod et al. 2021b). A mix of approaches were used to generate this evidence, including qualitative evidence, such as carer input or surveys, or different techniques to derive HSUVs based on factors like number of carers, report of challenges of living and caring for patients, study of people with limited mobility, or average between best-worse health points. Additionally during the observation of appraisal committees, carer HSUVs tended to be captured by simple applications of HSUV decrements with no clear basis. There was a lot of discussion about how this was captured in previous appraisals and whether circumstances of the treatment could be similar (voretigene and volanesorsen). Further, carer impact on QoL was mainly considered in the NICE HST appraisals, probably because this domain is included in their ultra-orphan drug framework (Nicod et al. 2021b).

Impact of disease and treatment on the QoL of carers is important to include when assessing the added benefit of treatment. However, evidence shows that carer QoL is inconsistently

accounted for and measured, and when it is, often results in high uncertainty (Nicod et al. 2021b; Pennington and Wong 2019). Guidance and greater alignment on approaches to develop robust carer QoL evidence across HTA bodies are needed.

3.4 Conclusions

This paper outlines recommendations on better use of PRO data and HSUV estimates in appraisal of rare disease treatments, derived from WP10 of the IMPACT-HTA project on appraisal of rare disease treatments. The crucial need for more optimal use of PRO and HSUV data stems from the fact that rare diseases have a substantial impact on patient and carer QoL. Their QoL therefore needs to be accounted for appropriately when assessing the benefit of a new treatment, considering that most treatments are not curative but improve outcomes and QoL. A better understanding is needed of the issues around generating PRO data and HSUV estimates from small, heterogeneous, and young populations. QoL evidence should be appraised in light of what is feasible and reasonable in each specific context, and whether it captures what matters most to patients. Better guidance and research is needed on alternative approaches for describing the impact on QoL of patients and carers with and without treatment. Triangulation of evidence, which would include patient and clinical input, may support the interpretation of the evidence and ensure that the decision is based on a more complete understanding of the impact of disease and treatment on QoL. Finally, education of stakeholders and decision-makers on the nuances and interpretation of PRO data and HSUV estimates in rare diseases may be beneficial to optimise their consideration in appraisal.

4. USE OF PROMS IN RARE DISEASES AND IMPLICATIONS FOR HTA

Amanda Whittal, Michela Meregaglia, Elena Nicod

A scoping review was conducted to identify the advantages, challenges and potential solutions for using different types of PROMs in HTA. This paper is now in The Patient (Whittal, Meregaglia, and Nicod 2021)w(Whittal et al. 2021). A full copy of the paper is available in Appendix A.

Abstract

Patient-reported outcome measures (PROMs) are used in Health Technology Assessment (HTA) for measuring patient experiences with disease and treatment, allowing a deeper understanding of the impact of treatment beyond clinical endpoints. Developing and administering PROMs for rare diseases poses unique challenges because of small patient populations, disease heterogeneity, lack of natural history knowledge, and short-term studies. This research aims to identify key factors to consider when using different types of PROMs in HTA for rare disease treatments (RDTs).

Methods: A scoping review of scientific and grey literature was conducted, with no date or publication type restrictions. Information on the advantages, challenges, and potential solutions in using different types of PROMs for RDTs, including psychometric properties, was extracted and synthesised.

Results: Twenty-seven of 66 records from PubMed were included, plus five records from the grey literature. PROMs for rare diseases face potential data collection and psychometric challenges resulting from small patient populations and disease heterogeneity. Generic PROMs are comparable across diseases, but are not sensitive to disease specificities. Disease-specific instruments are sensitive but do not exist for many rare diseases and rarely provide utility values required by some HTA bodies. Creating new PROMs is time and resource intensive. Potential solutions include pooling data (multi-site/international data collection), using computer assistant technology, or using generic and disease-specific PROMs in a complementary way.

Conclusions: PROMs are relevant in HTA for RDTs, yet pose a number of difficulties. A deeper understanding of the potential advantages, challenges, and solutions of each can help manage these difficulties.

5. THE ESTIMATION OF HEALTH STATES UTILITY VALUES IN RARE DISEASES: CHALLENGES AND OPPORTUNITIES OF EXISTING TECHNIQUES

Michela Meregaglia, Elena Nicod, Michael Drummond (2020)

This commentary paper discusses the pros and cons of using the different techniques to derive HSUVs in relation to the main features of RDs. This paper is published in the International Journal of Technology Assessment in Health Care (Michela Meregaglia, Nicod, and Drummond 2020). A full copy of the paper is available in Appendix B.

Abstract

There are several techniques for estimating health state utility values, each of which presents pros and cons in the context of RDs. Direct approaches (e.g. standard gamble and time trade-off) may be too demanding for patients with RDs, since most of them affect young children or cause cognitive impairment. The alternatives are using “vignettes” that describe hypothetical health states for the general public, which may not reflect the heterogeneous manifestations of RDs, or multi-attribute utility instruments (i.e. indirect techniques), such as EQ-5D, which may be less sensitive in capturing the specificities of RDs. The “rule of rescue” approach is a promising alternative in RDs, since it prioritizes identifiable patients with life-threatening or disabling conditions. However, it raises measurement challenges and ethical issues. Furthermore, the literature reports on relevant implications of choosing a technique over others for health technology assessment, which should be considered in relation to individual RDs.

6. "MAPPING" HEALTH STATE UTILITY VALUES FROM NON-PREFERENCE BASED MEASURES: A SYSTEMATIC LITERATURE REVIEW IN RARE DISEASES

Michela Meregaglia, Amanda Whittal, Elena Nicod, Michael Drummond

A systematic review identified all published studies using mapping to derive HSUVs from non-preference-based PROMs in RDs, and highlighted the critical issues in relation to the main features of RDs. This paper is published in PharmacoEconomics (Michela Meregaglia et al. 2020). A full copy is available in Appendix C.

Abstract

Background: The use of PROMs is increasing in RDs (i.e. affecting fewer than 1 in 2,000 people) to monitor the effects of disease and treatment on patient's symptomatology and daily life; however, these instruments seldom yield HSUVs for cost-utility analyses. In such a context, 'mapping' allows HSUVs to be obtained by establishing a statistical relationship between a 'source' (e.g. a disease-specific PROM) and a 'target' preference-based measure (e.g. the EuroQol-5 Dimension (EQ-5D) tool).

Objective: This study aimed to systematically review all published studies using 'mapping' to derive HSUVs from non-preference-based measures in RDs, and identify any critical issue related to the main features of RDs, which are characterized by small, heterogeneous, and geographically dispersed patient populations.

Methods: The following databases were searched during the first half of 2019 without time, study design or language restrictions: MEDLINE (via PubMed), the School of Health and Related Research Health Utility Database (ScHARRHUD) and the Health Economics Research Centre (HERC) database of mapping studies (version 7.0). The keywords combined terms related to 'mapping' with ORPHANET's list of RD indications (e.g. 'acromegaly'), in addition to 'rare' and 'orphan'; 'very rare' diseases (i.e. with less than 1,000 cases or families documented in the medical literature) were excluded from the searches. A predefined, pilot-tested extraction template (in Excel®) was used to collect structured information from the studies.

Results: Two groups of studies were identified in the review. The first group (n=19) developed novel mapping algorithms in thirteen different RDs. As a target measure, the majority used EQ-5D, and the others the Short-Form Six-Dimension (SF-6D) and 15D; most studies adopted Ordinary Least Squares (OLS) regression. The second group of studies (n=9) applied previously existing algorithms in non-RDs to comparable RDs, mainly in the field of cancer. The critical issues relating to 'mapping' in RDs included the availability of very few studies, the relatively high number of cancer studies, and the absence of research in paediatric RDs. Moreover, the reviewed studies recruited small samples, showed a limited overlap between RD-specific and generic PROMs, and highlighted the presence of cultural and linguistic factors influencing results in multi-country studies. Lastly, the application of existing algorithms developed in non-RDs tended to produce inaccuracies at the bottom of the EQ-5D scale, due to the greater severity of RDs.

Conclusions: More research is encouraged to develop algorithms for a broader spectrum of RDs (including those affecting young children), improve mapping study quality, test the generalizability of algorithms developed in non-RDs (e.g. HIV) to rare variants or evolutions of the same condition (e.g. AIDS wasting syndrome), and verify the robustness of results when mapped HSUVs are used in cost-utility models.

7. CONSIDERATION OF HEALTH-RELATED QUALITY OF LIFE IN THE APPRAISAL OF RARE DISEASE TREATMENTS: A CROSS-COUNTRY COMPARISON

Elena Nicod, Michela Meregaglia, Amanda Whittal, Sheela Upadhyaya, Karen Facey, Michael Drummond, PhD

This paper reviewed of England's National Institute of Health and Clinical Excellence's Highly Specialised Appraisal (NICE HST) and Technology Appraisal (NICE TA) processes for 24 non-oncology rare disease treatments based on their published appraisal reports. This document analysis focused on how PRO evidence and HSUVs influenced the decision, and the extent to which other forms of evidence supported the interpretation (e.g., patient input and evidence). A cross-country comparison with the Netherlands (National Health Care Institute, ZIN), France (Haute Autorité de Santé, HAS), and Germany (Federal Joint Committee, G-BA) for the same treatments was also conducted. This paper is to be submitted to Value in Health for publication, and may change following reviewer comments.

7.1 Introduction

Rare diseases are conditions affecting a small number of patients (e.g. less than 1 out of 2,000 people in Europe), which are most often life-threatening and/or severely debilitating, frequently genetic and with an early onset (Berdud, Drummond, and Towse 2018). Quality of life (QoL) of patients living with a rare disease is poorer than those living with more prevalent conditions (Bogart and Irvin 2017). This is partly explained by issues around diagnosis, and/or a lack of knowledge about the disease, its treatment pathways or treatment options (Bogart and Irvin 2017). 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 to provide that treatment 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 (Drummond et al. 2015). In the latter case, PRO evidence is collected directly from patients or proxies using patient-reported outcome measures (PROMs) (FDA 2009). PROMs are intended to capture aspects that matter most to patients about the impact of disease and treatment on symptoms, QoL, or health status (Kingsley and Patel 2017).

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). In added benefit assessments, QoL is considered when evaluating morbidity/mortality. In cost-effectiveness assessments, an economic evaluation models the 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. Techniques have been developed to translate PRO evidence into numerical values called health state utility values (HSUVs). HSUVs are cardinal values measured on a scale between 0, representing death, and 1, full health. 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 to the different model's health states, for both treatment and comparator arms (Brazier et al. 2019). Additional HSUVs relative to adverse events, co-morbidities or impact on carers are also frequently included.

More nuanced 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 (Whittal, Meragaglia, and

Nicod 2020; Benjamin et al. 2017). 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 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 that are 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, and further complicated by the nature of these diseases (Slade et al. 2018).

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 literature data, HSUVs being insufficiently sensitive to disease severity, or to changes that are important for patients (Pearson et al. 2018).

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 focus is often on demonstrating improvements in survival rather than in QoL. In addition, many rare disease treatments for cancers are for sub-populations, identified through a gene-expression test, 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. The objective of this research is to better understand how QoL of non-oncological rare disease treatments is appraised across different countries using different HTA approaches through review of HTA reports to identify the impact of conditions and their treatments on QoL and extent to which the QoL impact evidence presented influenced the decisions.

7.2 Methods

Study sample

Countries were selected to represent those that make decisions based on added benefit and those that focus on cost-effectiveness approaches in HTA with publicly available reports: England (NICE - National Institute for Health and Care Excellence), the Netherlands (ZIN – National Health Care Institute) for the cost-effectiveness approach, France (HAS – Haute Autorité de Santé), and Germany (G-BA - Federal Joint Committee) for the added benefit approach. In Germany, the Institute for Quality and Efficiency in Health Care (IQWiG) conducts the assessment and G-BA the appraisal. However, rare disease treatments are not required to undergo the assessment by IQWiG. Assessments by the Economic and Public Health Assessment Commission (CEEPS) at HAS, made separately from HAS's Transparency Committee's added benefit assessment, were not examined. Due to the greater availability of detailed reports, the analysis mainly focused on NICE appraisals, and reflections were made on how these contrast with the other countries' appraisals.

All treatments with a European Medicines Agency orphan designation and appraised by NICE within their Technology Appraisal (TA) and Highly Specialised Technology (HST) programmes before 1 June 2020 were selected (n=50). Cancer treatments were excluded given the greater likelihood of survival rather than QoL gains driving the assessments, and large availability of validated cancer-specific PROMs (Drummond et al. 2015; Kleijnen et al. 2017). Twenty-four treatments were selected (12 TA and 12 HST).

Data collection and analysis

Data collection consisted of extracting all relevant information from NICE's appraisal reports and, if needed, from the supporting documentation available on their website (e.g. committee papers, manufacturer submission). Two sets of information about (1) 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 included: (1) PRO evidence, HSUVs or patient-based evidence considered, (2) their source (e.g. trial, patient/clinical input), (3) their outcome (e.g. meaningful change), (4) the issues highlighted by the Committee, and (5) 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.), and HSUV on the technique used to derive them. Thematic analysis was undertaken to identify the latter two items (issues and their influence) based on the researcher's interpretation of the discussion reported in the published documents. The level of influence of PRO evidence on decisions was distinguished between "influence", when the Committee explicitly recognised and accounted for a change in QoL in their decisions, "possible influence", when PRO evidence was explicitly reported but considered limited by the Committee and it was therefore unclear whether it influenced the decision; "no influence", when PRO evidence was reported, most often inconclusive or not demonstrating any change in QoL and did not influence the Committee's decision. In the cross-country analysis, a much lower level of detail was available from the other countries' reports. The focus was therefore on the PRO evidence and HSUV estimates considered and their influence on the decision.

The burden of disease and impacts of treatments between the HST and TA were compared based on their relative likelihood of occurring to assess if there were some aspects of disease or treatment more likely to concern the types of treatments undergoing on programme over another. This was done by calculating the odds ratio of a specific aspect of disease or treatment occurring in the HST over the TA. The odds ratio was calculated by dividing the probability of the specific aspect occurring in the HST by the probability of the specific aspect occurring in the TA. 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 feasible in the different contexts.

7.3 Results

Impact of disease and treatment on quality of life

As Table 7.1 shows, most of the diseases undergoing the TA and HST are life-threatening and/or debilitating. There was a greater likelihood of the diseases undergoing the HST to affect children or infants (OR = 2.00), be heterogeneous (OR = 2.5) and multi-systemic (OR = 2.75), and (slightly) more likely to be progressive (OR = 1.71). With the exception of the prophylaxis treatment letermovir, all of the diseases targeted by the study sample affect patients' daily living and QoL due to symptoms.

In terms of the intended effects of the treatments, 67% of HST and 83% 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.

Table 7.1 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, and associated odds ratio (n=24)

		All n=24	HST n=12	TA n=12	odds ratio HST:TA
BURDEN OF DISEASE	Onset in infancy / childhood	0.63	0.83	0.42	**2.00**
	Progressive condition	0.79	1.00	0.58	1.71
	Heterogeneous condition	0.58	0.83	0.33	**2.50**
	Multi-systemic condition	0.63	0.92	0.33	**2.75**
	Debilitating condition	0.88	1.00	0.75	1.33
	Life-threatening condition	0.79	0.83	0.75	1.11
	Supportive care (no existing treatments)	0.38	0.58	0.17	**3.50**
	QoL impact from symptoms	0.96	1.00	0.92	1.09
	QoL impact on daily living	1.00	1.00	1.00	1.00
	QoL impact on families/carers	0.75	1.00	0.50	**2.00**
	QoL impact from current treatment	0.83	0.92	0.75	1.22
IMPACT TREATMENT	Length of life improved	0.75	0.67	0.83	0.80
	QoL improved from reduced symptoms	0.96	1.00	0.92	1.09
	QoL improved of daily living	1.00	1.00	1.00	1.00
	QoL improved of families/carers	0.71	1.00	0.42	**2.40**
	QoL improved compared to current treatment	0.92	1.00	0.83	1.20
	QoL improved due to administration mode	0.50	0.50	0.50	1.00

The odds ratio represents the greater likelihood of the various items relating to burden of disease and treatment impact to be relevant in the HST compared to the TA processes. For example, it is twice more likely that the diseases appraised in the HST programme have an onset in infancy or in children compared to those undergoing the TA programme

No previous treatments were available for 58% and 33% of those undergoing the HST and TA processes, respectively. Further, all of the HST and 83% of the TA treatments aim to improve patients' daily living and QoL over standard of care. In 50% of all cases, 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 improved 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 detail about patient numbers were provided in all other cases.

Influence of PRO evidence in NICE appraisals

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 (Figure 7.1). Several PROMs may have been considered for the same treatments. Examples of disease-group PROMs include the Paediatric Outcomes Data Collection Instrument (PDOCI) measuring functional outcomes in paediatric orthopedics(Lerman et al. 2005), and the St George's Respiratory Questionnaire (SGRQ) measuring overall health, daily life, and perceived well-being in patients with obstructive airways disease(Jones, Quirk, and Baveystock 1991). 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 7.2).

Figure 7.1 Types and influence of PRO evidence considered in NICE TA and HST appraisals of non-oncology rare disease treatments (n=24)

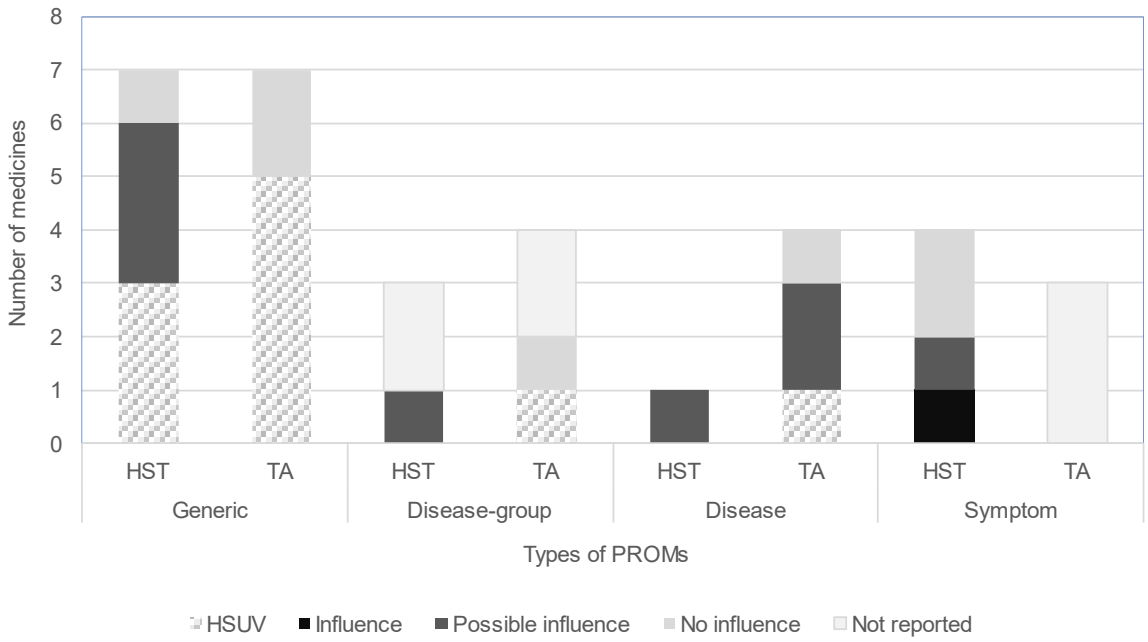


Table 7.2 Use and influence of PRO evidence in NICE TA and HST appraisals of non-oncology rare disease treatments (n=24)

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)	EQ5D -children 0.76 tx vs 0.43 no tx -adults 0.39 no tx -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	EQ5D 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)	EQ5D: mean improvement = 0.208 Survey: burden of disease and of current treatment on patients, carers/families	EQ5D used to derive HSUVs. Survey shows greatly impaired QoL of patients and carers from living with aHUS
Patisiran Hereditary transthyretin amyloidosis	**G-**EQ5D-5L (RCT, n=255) **D-**NIS, Norfolk-DN (RCT, n=255) **P-**Input	All PRO evidence significantly improved. NIS was trial's primary endpoint	EQ5D used to derive HSUVs. Effective based on significantly improved outcomes. PRO evidence captures most relevant treatment impacts, except: 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 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** - 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 Surveys: 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	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 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 -** patient and clinical input (oral administration)	QoL data: inconclusive (no change), except for change in GSRS	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 -** 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)	

MEDICINE Generic name - indication	PROM (by type) Instrument (source)	PRO EVIDENCE Description of results	APPRAISAL Influence of PRO evidence on decision
Burosumab for X-linked hypophosphataemia	None	No QoL evidence presented. Data being collected within trial (not reported)	
Inotersen Hereditary transthyretin amyloidosis	**D –** Norfolk QoL-DN (RCT)	Norfolk QoL-DN: no change in treatment arm, decrease in placebo arm	Not commented
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** no 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
Nintedanib Idiopathic pulmonary fibrosis	**G -** EQ-5D, PGI-C (RCT) **D -** SQRQ, SOBQ, CASA-Q **P -** patient input (tolerability)	PRO data: not reported Patient input: better tolerability profile impacting QoL, 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)
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
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

MEDICINE Generic name - indication	PROM (by type) Instrument (source)	PRO EVIDENCE Description of results	APPRAISAL Influence of PRO evidence on decision
Letermovir Cytomegalovirus (CMV)	**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, 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: 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: NA: no report available; MID: minimal important difference; EQ-5D: EuroQol-5 Dimension; 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-QoI: 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 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

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 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).

Further exploration of the influence of the PRO evidence on NICE decisions suggests that beyond those used to derive HSUVs, few of them had any influence on the decisions (Figure 7.2, Table 7.2).

Of the 14 appraisals considering generic PRO evidence, eight were used to derive the economic models' HSUVs and remaining six had unclear or no influence on the decisions (Table 7.2). 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 two dimensions, but considered uncertain due to the short trial duration. In all other cases (letermovir, asfotase alfa, voretigene and nintedanib), the disease-group PROMs, Functional Assessment of Cancer Therapy - Bone Marrow Transplantation (FACT-BMT), PODCI, Visual Function Questionnaire (VFQ), 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 uncertain and thus their influenced 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. Results from the AE-QoL data collected for lanadelumab were not commented on in the appraisal report.

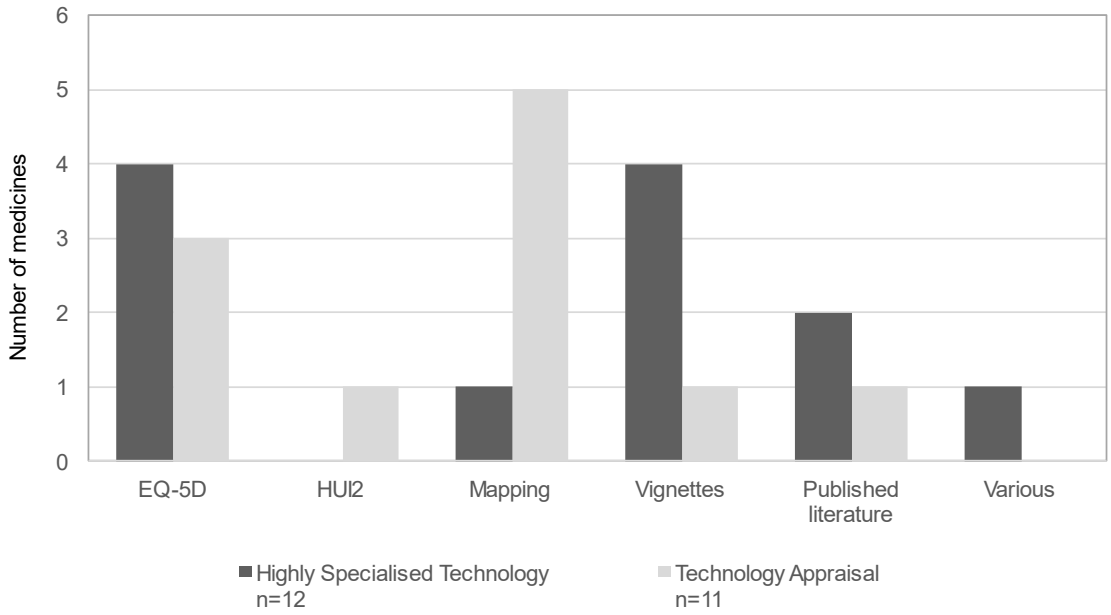
Of the seven treatments that considered symptom-specific PROMs, one of them influenced and another possibly influenced the decision. For patisiran, the Neuropathy Impairment Score (NIS) and Norfolk Quality of Life Questionnaire - Diabetic Neuropathy (Norfolk QoL-DN) data collected was statistically improved and contributed to recognizing treatment effectiveness. For eliglustat, no significant improvements were demonstrated for Fatigue Severity Scale (FSS) and Brief Pain Inventory (BPI) and it was unclear whether they were used to determine HSUVs for adverse events. 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. First, patient surveys provided information about impact of QoL on patients and carers (eculizumab), preferences for administration mode (eliglustat, colistimethate sodium and tobramycin DPI), or were used to derive HSUVs (elosulfase alfa). Respondents were patients and in one case also family members, two 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).

Influence of HSUV estimates in NICE appraisals

The most frequently used technique to derive HSUVs in NICE appraisals was the administration of EQ-5D (7/23), 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 7.2). 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. Mapping was more frequent in the TA and vignettes in the HST process. Additional HSUVs relating to adverse events (9/23), administration mode (4/23), carers (7/23) or other (7/23) were considered.

Figure 7.2 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 summarized in Table 7.3. Seven treatments used EQ-5D, two of which collected EQ-5D 3L in trials and 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). For all remaining cases, a number of issues were raised by the relevant committees, which included benefits not captured (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 given it is a preference-based PROM and would not require mapping to derive HSUVs, it was accepted by the relevant committee.

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 (nursinersen), 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 lack of correspondence between QoL data collected in the clinical trial and model health states (cerliponase), lack of negative values when deriving the PedsQLbeing considered unrealistic considering the condition's severity (cerliponase), 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; 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). There 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 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 recognized the challenges in collecting QoL data from children alongside the lack of validated PROMs.

Table 7.3 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	**List - Managed Access Agreement** ICER not specified, 3.0 QALY weight

MEDICINE Generic name, indication	**HSUV** Technique, appraisal	DECISION ICER, reasons
	alternative data **HSUV carers/siblings** => disutilities included, but 30 years considered to better reflect real life compared to life long	
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-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-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 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 **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. 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

MEDICINE Generic name, indication	**HSUV** Technique, appraisal	DECISION ICER, reasons
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 => HSUV of 0.840 as base case for visual acuity and HSUV decrement of 0.140 for disfigurement	**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
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

Legend: HSUV: health state utility values; QALY: quality-adjusted life years gained; ICER: incremental cost-effectiveness ratio; EQ-5D: EuroQol-5 Dimension; 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 2; DPI: tobramycin dry powders for inhalation; CFQ-R:cystic fibrosis questionnaire revised; FEV: Forced Expiratory Volume; SGRQ: St George Respiratory Questionnaire

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 (Tables 7.2, 7.3 and 7.4). First, many of the appraisal reports do 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, hardly any of the treatments appraised were considered to improve QoL, most being inconclusive. Only one treatment, inotersen, appraised by HAS was considered to provide a moderate improvement in QoL, as it was one of the trial's co-primary endpoint; whereas no meaningful clinically relevant change was recognized by NICE and G-BA. Two treatments, patisiran and lanadelumab, appraised by G-BA 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 7.4 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 (Gaucher disease severity score): clinically relevant and crucial, no significant differences
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
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 detail 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 o 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 included in report
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

Carer impact

Eighteen of the treatments were considered to have an impact on carers, whereas evidence on carer impact was considered for nine of these (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 (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 changes so as to better align with previous appraisals (patisiran), to include only children HSUVs (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.

7.4 Discussion

The vast majority of conditions investigated, particularly in NICE's HST, are life-threatening and/or debilitating. 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. Results also 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 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; overall its influence was fairly limited. Just over 1/3 of the HSUVs were accepted, even if in some cases they were recognized as not ideal. In the remaining cases, the HSUVs were highly uncertain and in most cases the relevant committee recognized 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.

In Germany and France, most of the PRO evidence was considered inconclusive due to the frequent lack of statistically significant improvement and/or as they did not meet the country-specific evidentiary requirements. In Germany, the PROM needs 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 (Rüther et al. 2016). This may be more frequent in heterogeneous and small patient populations (Schulz et al. 2019). 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 given the PRO evidence, 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.

Results therefore suggest that PRO evidence and HSUVs collected in rare diseases tend to be highly uncertain and often inconclusive. 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, and appear to be more flexible when interpreting the QoL evidence, e.g. in recognizing that all benefits are not captured by the measures used.

Overall a large amount of QoL data was collected, but these data were barely reported or referred to in the appraisal reports. However, the 28 different PROMs identified are most likely to cover concepts important for patients (FDA 2009). This lack of use suggests that either

there may be a loss of valuable information on the patient perspective, issues in capturing meaningful change using PROMs in rare diseases, or issues in accounting for these in the different HTA approaches. Consideration should be given on how to better use the available evidence for HTA.

Only three disease-specific PROMs were reported, but their consideration had a limited influence on the decision and included in only one case to derive HSUVs. This confirms the issue around a lack of validated disease-specific PROMs and their conversion into HSUVs through mapping (Whittal, Meragaglia, and Nicod 2020). 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 recognized that the QoL evidence did not capture the full range of dimensions of 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.

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 (Slade et al. 2018), but does not reflect the common reliance on proxy-reported data (Benjamin et al. 2017). In only one case was the PedsQL mapped to EQ-5D, but results were limited and the challenges in collecting data from children recognized (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 (Benjamin et al. 2017; Slade et al. 2018; Bell et al. 2019) was not entirely clear from the results. There were, however, cases where evidence on QoL was lacking to estimate the HSUV 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 (Benjamin et al. 2017). In one case (cerliponase), the HST Committee recognized 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 standard of care for these diseases require multi-disciplinary specialized 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 (Morel and Cano 2017).

Three quarters of the conditions appraised affect QoL of families and carers, with the treatments considered to improve their QoL (expect for the prophylaxis treatment letermovir).

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 8 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 (Goodrich, Kaambwa, and Al-Janabi 2012). There is a need for methodological guidance on when and how to include carer HSUVs in QALY and non-QALY approaches to HTA (Pennington and Wong 2019). 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 accurately 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 use of google translator for some of the countries. However, considering the analysis relied on a cross-country comparison, 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.

7.5 Conclusions

This study highlights some of the limitations and challenges in using PRO evidence and HSUVs in appraising rare disease treatments, based on a cross-country analysis. In most cases, PRO evidence and HSUVs failed to demonstrate meaningful change, were uncertain or inconclusive, and/or did not capture a number of domains important for patients. As a result, there is a need for improved development, testing, use and reporting of PRO evidence and of HSUVs. HTA bodies would also benefit from greater flexibility in allowing for use of other techniques to derive QoL evidence, such as vignettes or patient surveys, which today are often not accepted. However, there is a need to better understand how to develop methodologically sound vignettes and other less conventional evidence. Most importantly, patient and clinical input and patient evidence have shown to be crucial in providing information about the burden of illness, treatment benefits including those that matter most, and in helping in the interpretation of uncertain aspects 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.

Acknowledgments:

A particular thanks to Andrew Lloyd for his valuable time and input while pulling together these results.

8. REFERENCES

Bansback, Nick, Mark Harrison, John Brazier, Linda Davies, Jacek Kopec, Carlo Marra, Deborah Symmons, and Aslam Anis. 2008. “Health State Utility Values: A Description of Their Development and Application for Rheumatic Diseases.” *Arthritis Care and Research*. https://doi.org/10.1002/art.23813.

Basch, Ethan, and Antonia V Bennett. 2014. “Patient-Reported Outcomes in Clinical Trials of Rare Diseases.” *Journal of General Internal Medicine* 29 (3): 801–3. https://doi.org/10.1007/s11606-014-2892-z.

Bell, Jill A., Aaron Galaznik, Farrah Pompilus, Sara Strzok, Rafael Bejar, Fatima Scipione, Robert J. Fram, Douglas V. Faller, Stefan Cano, and Patrick Marquis. 2019. “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.” *Journal of Patient-Reported Outcomes*. https://doi.org/10.1186/s41687-019-0123-4.

Benjamin, Katy, Margaret K. Vernon, Donald L. Patrick, Eleanor Perfetto, Sandra Nestler-Parr, and Laurie Burke. 2017. “Patient-Reported Outcome and Observer-Reported Outcome Assessment in Rare Disease Clinical Trials: An ISPOR COA Emerging Good Practices Task Force Report.” *Value in Health*. https://doi.org/10.1016/j.jval.2017.05.015.

Berdud, M., M. Drummond, and A. Towse. 2018. “Establishing a Reasonable Price for an Orphan Drug.” *Research Papers*.

Bogart, Kathleen R., and Veronica L. Irvin. 2017. “Health-Related Quality of Life among Adults with Diverse Rare Disorders.” *Orphanet Journal of Rare Diseases*. https://doi.org/10.1186/s13023-017-0730-1.

Brazier, John, Roberta Ara, Ismail Azzabi, Jan Busschbach, Hélène Chevrou-Séverac, Bruce Crawford, Luciane Cruz, et al. 2019. “Identification, Review, and Use of Health State Utilities in Cost-Effectiveness Models: An ISPOR Good Practices for Outcomes Research Task Force Report.” *Value in Health*. https://doi.org/10.1016/j.jval.2019.01.004.

Brazier, John, and Donna Rowen. 2011. “NICE DSU Technical Support Document 11: Alternatives to EQ5D for Generating Health State Utility Values.”

Bryson, Brooke, Kathleen Bogart, Makenzie Atwood, Kelly Fraser, Taylor Locke, Kiley Pugh, and Mohamed Zerrouk. 2019. “Navigating the Unknown: A Content Analysis of the Unique Challenges Faced by Adults with Rare Diseases.” *Journal of Health Psychology*, February, 1359105319828150. https://doi.org/10.1177/1359105319828150.

Burström, Kristina, Å Bartonek, E W Broström, S Sun, and Ann-Charlotte Egmar. 2014. “EQ-5D-Y as a Health-Related Quality of Life Measure in Children and Adolescents with Functional Disability in Sweden : Testing Feasibility Andvalidity.” *Acta Paediatrica* 103 (4): 426–35. https://doi.org/10.1111/apa.12557.

Deshpande, PrasannaR, BLakshmi Sudeepthi, Surulivel Rajan, and CP Abdul Nazir. 2011. “Patient-Reported Outcomes: A New Era in Clinical Research.” *Perspectives in Clinical Research*. https://doi.org/10.4103/2229-3485.86879.

Drummond, Michael, Mark Sculpher, Karl Claxton, Greg Stoddart, and George Torrance. 2015. “Measuring and Valuing Health Effects.” In *Methods for Economic Evaluation in Health Care*, Fourth, 123–80. Oxford: Oxford University Press.

EUnetHTA. 2015. “Guidance - Endpoints Used for Relative Effectiveness Assessment: Health-Related Quality of Life and Utility Measures.” https://eunethta.eu/endpoints-used-for-relative-effectiveness-assessment-health-related-quality-of-life-and-utility-measures-amended-ja1-guideline-final-nov-2015/.

European Union. n.d. “Orphan Medicinal Products.” Accessed January 20, 2020.

https://ec.europa.eu/health/human-use/orphan-medicines_en.

Eurordis. “About Rare Diseases.” Accessed March 10, 2021a. https://www.eurordis.org/about-rare-diseases.

Eurordis. “Juggling Care and Daily Life.” Accessed March 16, 2021b. http://download2.eurordis.org/rbv/juggling_care_and_daily_life.infographic__final.pdf.

Facey, K. n.d. “Participant Observation of Appraisal Commitees for Orphan Medicines.”

FDA *Orphan Drug Act*. https://www.fda.gov/industry/designating-orphan-product-drugs-and-biological-products/e-cfr-21-part-316-orphan-drug-orphan-drug-act-1983.

FDA 2009. “Guidance for Industry - Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims.”

Food and Drug Administration, Center for Drug Evaluation and Research, and Center for Biologics Evaluation and Research. 2020. “Patient-Focused Drug Development: Collecting Comprehensive and Representative Input. Guidance for Industry, Food and Drug Administration Staff, and Other Stakeholders.” https://www.fda.gov/media/139088/download.

Goodrich, Kacey, Billingsley Kaambwa, and Hareth Al-Janabi. 2012. “The Inclusion of Informal Care in Applied Economic Evaluation: A Review.” *Value in Health*. https://doi.org/10.1016/j.jval.2012.05.009.

Haute Autorite de Sante. 2018. “Évaluation Des Technologies de Santé à La HAS: Place de La Qualité de Vie.” 2018. https://www.has-sante.fr/upload/docs/application/pdf/2018-10/note_de_synthese_place_de_la_qualite_de_vie.pdf.

Hill, Harry, Donna Rowen, Becky Pennington, Ruth Wong, and Allan Wailoo. 2020. “A Review of the Methods Used to Generate Utility Values in NICE Technology Assessments for Children and Adolescents.” *Value in Health* 23 (7): 907–17. https://doi.org/https://doi.org/10.1016/j.jval.2020.02.011.

Hong, Yoon Duk, Ester Villalonga-Olives, and Eleanor M Perfetto. 2019. “Patient-Reported Outcomes in Orphan Drug Labels Approved by the US Food and Drug Administration.” *Value in Health* 22 (8): 925–30. https://doi.org/https://doi.org/10.1016/j.jval.2019.03.010.

Jones, P. W., F. H. Quirk, and C. M. Baveystock. 1991. “The St George’s Respiratory Questionnaire.” *Respiratory Medicine*. https://doi.org/10.1016/S0954-6111(06)80166-6.

Kingsley, Charlotte, and Sanjiv Patel. 2017. “Patient-Reported Outcome Measures and Patient-Reported Experience Measures.” *BJA Education* 17 (4): 137–44. https://doi.org/10.1093/bjaed/mkw060.

Klakow-Franck, Regina. 2014. “The Relevance of Quality of Life for the Work of the Federal Joint Committee.” *Zeitschrift Fur Evidenz, Fortbildung Und Qualitat Im Gesundheitswesen*. https://doi.org/10.1016/j.zefq.2014.02.007.

Kleijnen, Sarah, Teresa Leonardo Alves, Kim Meijboom, Iga Lipska, Anthonius De Boer, Hubertus G. Leufkens, and Wim G. Goettsch. 2017. “The Impact of Quality-of-Life Data in Relative Effectiveness Assessments of New Anti-Cancer Drugs in European Countries.” *Quality of Life Research*. https://doi.org/10.1007/s11136-017-1574-9.

Lerman, Joel A., Elroy Sullivan, Douglas A. Barnes, and Richard J. Haynes. 2005. “The Pediatric Outcomes Data Collection Instrument (PODCI) and Functional Assessment of Patients with Unilateral Upper Extremity Deficiencies.” *Journal of Pediatric Orthopaedics*. https://doi.org/10.1097/01.bpo.0000149866.80894.70.

Longworth, Louise, and Donna Rowen. 2013. “Mapping to Obtain EQ-5D Utility Values for Use in Nice Health Technology Assessments.” *Value in Health*. https://doi.org/10.1016/j.jval.2012.10.010.

Luzzatto, L, H Hyry, A Schieppati, E Costa, S Simoens, and F Schaefer. 2018. “Outrageous Prices of Orphan Drugs: A Call for Collaboration.” *The Lancet* 392 (10149): 791–94.

Meregaglia, M., E. Nicod, and M. Drummond. 2020. “The Estimation of Health State Utility Values in Rare Diseases: Overview of Existing Techniques.” *International Journal of Technology Assessment in Health Care*. https://doi.org/10.1017/S0266462320000665.

Meregaglia, Michela, Elena Nicod, and Michael Drummond. 2020. “The Estimation of Health State Utility Values in Rare Diseases: Overview of Existing Techniques.” *International Journal of Technology Assessment in Health Care*. https://doi.org/10.1017/S0266462320000665.

Meregaglia, Michela, Amanda Whittal, Elena Nicod, and Michael Drummond. 2020. “‘Mapping’ Health State Utility Values from Non-Preference-Based Measures: A Systematic Literature Review in Rare Diseases.” *PharmacoEconomics*. https://doi.org/10.1007/s40273-020-00897-4.

Morel, Thomas, and Stefan J. Cano. 2017. “Measuring What Matters to Rare Disease Patients - Reflections on the Work by the IRDiRC Taskforce on Patient-Centered Outcome Measures.” *Orphanet Journal of Rare Diseases*. https://doi.org/10.1186/s13023-017-0718-x.

National Institute for Health and Care Excellence. 2013. “Guide to the Methods of Technology Appraisal 2013.” 2013. https://www.nice.org.uk/process/pmg9/resources/guide-to-the-methods-of-technology-appraisal-2013-pdf-2007975843781.

Neumann, Peter J., Sue J. Goldie, and Milton C. Weinstein. 2000. “Preference-Based Measures in Economic Evaluation in Health Care.” *Annual Review of Public Health*. https://doi.org/10.1146/annurev.publhealth.21.1.587.

Nguengang Wakap, Stéphanie, Deborah M. Lambert, Annie Olry, Charlotte Rodwell, Charlotte Gueydan, Valérie Lanneau, Daniel Murphy, Yann Le Cam, and Ana Rath. 2020. “Estimating Cumulative Point Prevalence of Rare Diseases: Analysis of the Orphanet Database.” *European Journal of Human Genetics*. https://doi.org/10.1038/s41431-019-0508-0.

Nicod, E., L. Annemans, A. Bucsics, A. Lee, S. Upadhyaya, and K. Facey. 2019. “HTA Programme Response to the Challenges of Dealing with Orphan Medicinal Products: Process Evaluation in Selected European Countries.” *Health Policy* 123 (2): 140–51. https://doi.org/10.1016/j.healthpol.2017.03.009.

Nicod, E., L. Maynou, L. Maynou, E. Visintin, and J. Cairns. 2020. “Why Do Health Technology Assessment Drug Reimbursement Recommendations Differ between Countries? A Parallel Convergent Mixed Methods Study.” *Health Economics, Policy and Law* 15 (3). https://doi.org/10.1017/S1744133119000239.

Nicod, E, M Meragaglia, A Whittal, S Upadhyaya, K Facey, and M Drummond. n.d. “Consideration of Health-Related Quality of Life in the HTA Appraisal of Orphan Medicinal Products: An Overview of Four Countries.” *Value in Health - under Review*.

Nicod, E, M Meregaglia, A Whittal, S Upadhyaya, K Facey, and M Drummond. 2021a. “Consideration of Health-Related Quality of Life in the HTA Appraisal of Orphan Medicinal Products: An Overview of Four Countries.” *Value in Health - to Be submitted Soon*.

Nicod, E, M Meregaglia, A Whittal, S Upadhyaya, K Facey, and M Drummond 2021b. “Consideration of Health-Related Quality of Life in the HTA Appraisal of Orphan Medicinal Products: An Overview of Four Countries.” *Value in Health - to be submitted soon*.

Nicod, E, A Whittal, M Drummond, and K Facey. 2020. “Country Vignettes on HTA Appraisal/Reimbursement Processes for Rare Disease Treatments.” 2020. https://www.impact-hta.eu/country-vignettes.

Pascoal, Carlota, Sandra Brasil, Rita Francisco, Dorinda Marques-Da-Silva, Agnes Rafalko, Jaak Jaeken, Paula A. Videira, Luísa Barros, and Vanessa Dos Reis Ferreira. 2018. “Patient and Observer Reported Outcome Measures to Evaluate Health-Related Quality of Life in Inherited Metabolic Diseases: A Scoping Review.” *Orphanet Journal of Rare Diseases*. https://doi.org/10.1186/s13023-018-0953-9.

Pearson, Isobel, Ben Rothwell, Andrew Olaye, and Christopher Knight. 2018. “Economic Modeling Considerations for Rare Diseases.” *Value in Health* 21 (5): 515–24. https://doi.org/10.1016/j.jval.2018.02.008.

Pennington, B, and R Wong. 2019. “Modelling Carer Health-Related Quality of Life in NICE Technology Appraisals and Highly Specialised Technologies.”

“Regulation (EC) No 141/2000 of the European Parliament and of the Council of 16 December 1999 on Orphan Medicinal Products.” 2000. *Official Journal of the European Union*.

Rüther, Alric, Deborah Elstein, Durhane Wong-Rieger, and Gordon Guyatt. 2016. “Aspects of Patient Reported Outcomes in Rare Diseases: A Discussion Paper.” *International Journal of Technology Assessment in Health Care* 32 (3): 126–30. https://doi.org/10.1017/S0266462316000271.

Schulz, Sandra, Anna Passon, Michael Kulig, Matthias Perleth, and Katja Matthias. 2019. “Orphan Drug Benefits Asssessments at the Federal Joint Committee in Germany.” In *HTAi Conference*. https://www.researchgate.net/publication/333632005_Orphan_Drug_Benefit_Assessments_At_The_Federal_Joint_Committee_In_Germany.

Simpson, Amy, Lara Bloom, Naomi J Fulop, Emma Hudson, Kerry Leeson-Beevers, Stephen Morris, Angus I G Ramsay, Alastair G Sutcliffe, Holly Walton, and Amy Hunter. 2021. “How Are Patients with Rare Diseases and Their Carers in the UK Impacted by the Way Care Is Coordinated? An Exploratory Qualitative Interview Study.” *Orphanet Journal of Rare Diseases*. Genetic Alliance UK, Third Floor, 86-90 Paul Street, London, EC2A 4NE, UK. amy.simpson@geneticalliance.org.uk. https://doi.org/10.1186/s13023-020-01664-6.

Slade, Anita, Fatima Isa, Derek Kyte, Tanya Pankhurst, Larissa Kerecuk, James Ferguson, Graham Lipkin, and Melanie Calvert. 2018. “Patient Reported Outcome Measures in Rare Diseases: A Narrative Review.” *Orphanet Journal of Rare Diseases*. https://doi.org/10.1186/s13023-018-0810-x.

Swezey, T., B. B. Reeve, T. S. Hart, M. K. Floor, C. M. Dollar, A. P. Gillies, and L. L. Tosi. 2019. “Incorporating the Patient Perspective in the Study of Rare Bone Disease: Insights from the Osteogenesis Imperfecta Community.” *Osteoporosis International*. https://doi.org/10.1007/s00198-018-4690-7.

Tosi, Laura L., Marianne K. Floor, Christina M. Dollar, Austin P. Gillies, Brendan Lee, Sandesh C.S. Nagamani, Frank Rauch, et al. 2019. “Assessing Disease Experience across the Life Span for Individuals with Osteogenesis Imperfecta: Challenges and Opportunities for Patient-Reported Outcomes (PROs) Measurement: A Pilot Study.” *Orphanet Journal of Rare Diseases*. https://doi.org/10.1186/s13023-019-1004-x.

Towse, Adrian, and Martina Garau. 2018. “Appraising Ultra-Orphan Drugs: Is Cost-per-QALY Appropriate? A Review of the Evidence.” https://www.ohe.org/publications/appraising-ultra-orphan-drugs-cost-qaly-appropriate-review-evidence.

Whittal, A., M. Meregaglia, and E. Nicod. 2021. “The Use of Patient-Reported Outcome Measures in Rare Diseases and Implications for Health Technology Assessment.” *Patient*. https://doi.org/10.1007/s40271-020-00493-w.

Whittal, A, M Meragaglia, and E Nicod. 2020. “The Use of Patient-Reported Outcome Measures (PROMs) in Rare Diseases and Implications for HTA.” *The Patient* under revi.

Wolowacz, Sorrel E., Andrew Briggs, Vasily Belozeroff, Philip Clarke, Lynda Doward, Ron Goeree, Andrew Lloyd, and Richard Norman. 2016. “Estimating Health-State Utility for Economic Models in Clinical Studies: An ISPOR Good Research Practices Task Force Report.” *Value in Health*. https://doi.org/10.1016/j.jval.2016.06.001.

Appendix A. The Use of Patient-Reported Outcome Measures in Rare Diseases and Implications for Health Technology Assessment

APPENDIX B. THE ESTIMATION OF HEALTH STATE UTILITY VALUES IN RARE DISEASES: OVERVIEW OF EXISTING TECHNIQUES

Appendix C. Mapping Health State Utility Values from Non-preference-based Measures: A Systematic Literature Review