



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

Amanda Whittal<sup>1</sup> · Michela Meregaglia<sup>1</sup> · Elena Nicod<sup>1</sup>

Accepted: 12 December 2020 / Published online: 19 January 2021  
© The Author(s) 2021

## Abstract

**Background** Patient-reported outcome measures (PROMs) are used in health technology assessment (HTA) to measure patient experiences with disease and treatment, allowing a deeper understanding of treatment impact 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.

**Objective** 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 of and the challenges and potential solutions when using different types of PROMs for RDTs, including psychometric properties, was extracted and synthesized.

**Results** Of 79 records from PubMed, 32 were included, plus 12 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 not sensitive to disease specificities. Disease-specific instruments are sensitive but do not exist for many rare diseases and rarely provide the 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-assisted technology, or using generic and disease-specific PROMs in a complementary way.

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

## 1 Introduction

A patient-reported outcome (PRO) is a report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by anyone. Accordingly, a patient-reported outcome measure (PROM) is a tool, such as a questionnaire or a survey, used to measure and collect data on a PRO, usually related to health-related quality of life (HRQoL), symptoms or

### Key Points for Decision Makers

Patient-reported outcome measures (PROMs) for rare diseases face potential challenges resulting from small patient populations and disease heterogeneity.

Data collection, psychometric properties and each specific type of PROM face unique challenges.

Each of the challenges have potential solutions that can be considered and selected to fit specific contexts.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s40271-020-00493-w>.

✉ Amanda Whittal  
amanda.whittal@unibocconi.it

<sup>1</sup> Centre for Research on Health and Social Care Management (CERGAS), SDA Bocconi School of Management, Via Sarfatti 10, 20136 Milan, Italy

treatment side effects or experience with care (adherence, satisfaction or health status) [1].

Various types of PROMs exist to capture PROs, with the main distinction being between generic measures and disease/condition/treatment-specific measures [2]. Generic PROMs are not specific to a disease, condition or treatment, but can be used across different populations. They more generally capture

such aspects as quality of life (QoL); HRQoL; physical function; physical, mental and emotional health; social function; pain, etc. [2]. Examples include the Short Form-36 (SF-36) or the World Health Organization Quality of Life (WHOQOL) questionnaire. Disease-group-specific PROMs relate to a specific group of conditions or diseases, or similar diseases. These PROMs tend to be more sensitive than generic measures, but less sensitive than PROMs tailored to a specific rare disease (RD). A common example used in oncology is the European Organization for Research and Treatment of Cancer Quality of Life (EORTC QLQ-C30) questionnaire [3]. Disease/condition/treatment-specific PROMS (hereafter referred to as ‘disease-specific’) are tailored to measure symptoms, effects of treatment or other aspects related to a specific condition or disease [2].

Generic and disease-specific PROMs can be further divided into preference or non-preference based. Non-preference-based PROMs are presented as profiles or by summing answers to provide a total score that is interpretable on its own [2]. Preference-based PROMS are measured in a way in which health state utility values (HSUVs) can be derived. Instead of answers being summed, they are used to create an index score (based on societal preferences for a particular health state), which allows calculation of quality-adjusted life-years (QALYs) [24]. The most common of these types of PROMs are the EuroQol 5-Dimensions (EQ-5D), the Health Utility Index (HUI3), and the Short Form 6 Dimension (SF-6D) [4].

PROMs are increasingly being used to derive information on a treatment’s value and are often accounted for during health technology assessment (HTA) processes when making decisions on whether to provide a treatment for routine use [5]. Patient perspectives provide crucial information for decision makers in these contexts [6], particularly in RDs [7]. The high unmet need, severe and disabling nature of the condition and scarcity of adequate data for RDs means clinical trials need creative and pragmatic supplements to conventional measures, to capture treatment effects from patient perspectives [7] and help ensure the measurement of meaningful outcomes. Well-designed PROMs can support clinical endpoints, which are often challenging in RDs and may rely on surrogate endpoints [8]. Some countries (e.g. the UK) use QALYs to quantify health outcomes for HTA and use preference-based generic PROMs to derive HSUVs for calculating QALYs to be included within economic models [9]. Other non-QALY-based HTA systems (e.g. Germany) use PROMs as sources of additional evidence for the deliberative process [10, 11]. Currently, the focus of HTA bodies is largely on generic PROMs, and the use of PROM evidence in decision making is inconsistent [10].

The use of PROMs in HTA for rare disease treatment (RDTs) also poses a number of challenges, some of which are not specific to RDs but may be exacerbated by the inherent characteristics of such conditions:

- Data are usually collected from small patient populations [12], which may result in inaccurate aggregate results.
- Conditions and presentations can be heterogeneous, which make it difficult to capture meaningful and generalizable outcomes [3, 12–15].
- Information and understanding regarding disease progression and natural history is lacking, which makes it difficult to know which PROMs to use or how to develop new PROMs [12, 13, 16, 17].
- The number of studies is insufficient, which makes it difficult to obtain representative samples in literature reviews [18].
- Many issues that are important to patients are not captured with existing measures/methods [19].
- Existing value frameworks largely fall short of consistently measuring outcomes that matter to patients [16].
- Psychometric and linguistic validation of newly developed PROMs is challenging to attain [12].
- Patients are often children or have cognitive impairments associated with the disease, which makes it hard or impossible for patients to self-report and often places a reliance on proxy measures such as parent proxies [13].

These challenges have important implications for the use of PROMs in HTA for RDTs; a thorough understanding of challenges (and potential solutions) can be beneficial for all stakeholders involved in these processes.

The aim of this research was to review the current literature on the use of PROMs in RDs and identify key factors to consider when using PROMs for HTA of RDTs. These identified factors are then interpreted and discussed, with the goal of providing useful, evidence-based insights that can support HTA stakeholders when considering PROM results during RDT appraisal. This work is not about the details of selecting, adapting or developing PROMs for a particular RD, as this is a complex process for which an entirely separate piece of research is needed.

This study was conducted within IMPACT HTA, an EU Horizon 2020 project examining new and improved methods in costs, health outcomes and economic evaluation in the context of HTA and health system performance measurement (<https://www.impact-hta.eu/>). This work package (WP10) focuses specifically on HTA appraisal of medicinal products for RDs. Results will feed into a guidance document intended for HTA stakeholders on the use of PROMs in HTA for rare diseases.

## 2 Methods

### 2.1 Study Design

A scoping review of scientific (PubMed) and grey (Google) literature was conducted, following the PRISMA-ScR

(Preferred Reporting Items for Systematic Reviews and Meta-Analyses—extension for scoping reviews) checklist [20] to ensure accurate and comprehensive information reporting. Only PubMed was used as an inclusive database; other database test searches (Science Direct, Springer Link) provided overlapping hits. All hits from each search string were exported to Excel. One reviewer screened the titles (eliminating unrelated articles), read the abstracts (to eliminate unsuitable articles not detected by title screening), read the full text for promising and included articles and extracted the information. A second senior researcher read and reviewed the full-text articles selected for inclusion and the extracted information. To ensure all relevant information was captured, literature that was already available for the project was also included, and references of all selected articles were checked to retrieve any relevant literature that was not captured by the search strings. Searches were conducted until May 2020, with no date or study design limitations.

In PubMed, search terms were open and included patient reported outcome measure\*, patient reported outcome\*, prom\*, rare disease\*, RDT\*, orphan medicinal product\*, OMP\*, challenge\*, recommend\*, healthy technology assessment, HTA, appraisal\*. For the grey literature, search terms included patient reported outcome measure, rare disease, health technology assessment. The full search string combinations are listed in Appendix 1 in the electronic supplementary material (ESM).

To encompass a wide perspective from both scientific and real-world practice viewpoints, the search was broad and incorporated various types of literature. This included original research, reviews, commentaries, discussion papers, policy papers, conference/webinar/symposium presentations and position papers.

## 2.2 Article Selection

Articles were included if they were in English and provided any insight into PROMs for RDs in terms of the advantages, challenges and potential solutions, both in general and specifically related to HTA.

Articles were excluded if they only described the application or development of a PROM without a description of the advantages, challenges or potential solutions relevant for use with RDs or if they referred to aspects of PROMs not relevant for the purposes of this research.

The PubMed search identified 103 scientific articles, resulting in 44 records included for analysis (see Fig. 1). This included 23 original research, nine reviews, four commentaries/editorials/short communications, three conference/webinar/symposium presentations, two discussion/perspective papers, two reports and one position statement. The remaining 59 articles were excluded because they were related to QoL or experience, but not PROMs; they were

about effectiveness, not QoL; or they were generally about evidence.

## 2.3 Information Extraction

Relevant information from all included articles identified as being related to the advantages, challenges and potential solutions of using PROMs for RDTs was extracted and summarized in an Excel template. Extracted information included authors, date of publication, journal, title, country, type of research, research objective(s)/research questions(s) and key advantages/challenges/solutions mentioned in the text. This information was used to identify which aspect of PROMs in HTA the article was most applicable for, or the article 'focus'. An overview of the characteristics of selected articles is displayed in Table 1, and the detailed information that was extracted is in Appendix 2 in the ESM.

Key findings were derived from each article included in the analysis and grouped into pre-defined categories related to PROMs for RDTs. Categories were identified as those areas requiring understanding of all stakeholders to better ensure successful use of PROMs for RDTs in HTA: (1) psychometric properties, (2) existing generic PROMs, (3) existing disease-group-specific PROMs, (4) existing disease-specific PROMs and (5) creating new disease-specific PROMs. Consideration was also given to whether the PROMs were preference or non-preference based. The potential solutions of included articles were discussed among authors regarding factors that may hinder or facilitate solution implementation. This is explored further in the discussion section to indicate which solutions may be more or less feasible to implement and why, and what HTA bodies could do to facilitate the success of such solutions.

## 3 Results

Our results, summarized in Table 2, outline the potential challenges and their solutions when using PROMs for RDTs, and implications for HTA.

### 3.1 General Considerations for the Use of Patient-Reported Outcome Measures (PROMs) in Rare Diseases (RDs)

#### 3.1.1 Potential Data Collection/Masurement Challenges and Solutions

*Diversity of Use of PROMs in RDs* Researchers are using a wide variety of PROM types for the same condition or group of conditions, making the comparison of results across populations more challenging [15, 21]. Recommended core outcome measures based on existing guidelines could be

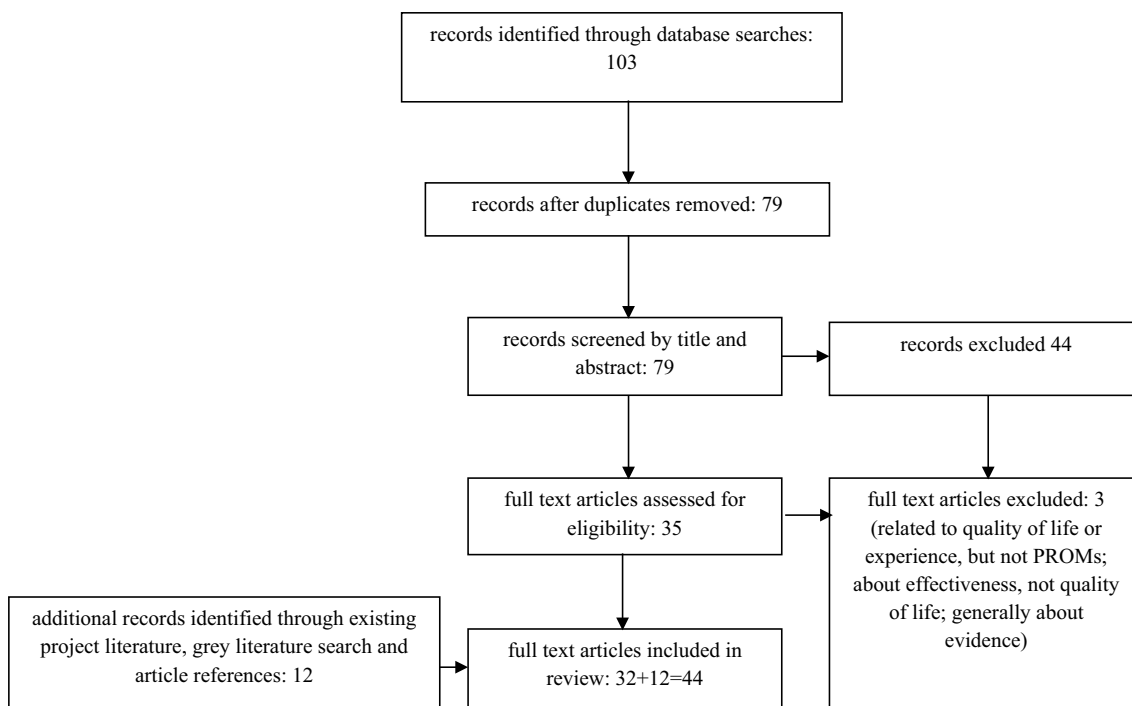


Fig. 1 Article selection flow chart

developed to provide a standard set of PROMs for specific RDs or groups of RDs to ensure improved consistency and comparability across populations [5, 15, 21]. Disease and treatment characteristics from the perspective of all stakeholders, especially patients and carers, should be considered when developing these core outcome measures. Concept elicitation interviews, for example, could be conducted with as many patient and carer groups as possible to evaluate differences in patient experiences across disease subtypes. These studies should take into account the most important features that cause variation in disease experience, such as disease group, age, ethnicity, or disease severity. They should further aim to identify the most important symptoms within various subtypes and focus on core signs and symptoms that apply to most or all patients. Working with patient and clinical experts at an early stage is essential for capturing the meaning and importance of all potential endpoints [5, 13].

*Small, Heterogeneous Populations* The small sample sizes and heterogeneous populations inherent to RDs result in sampling, data collection and statistical analysis issues, which often mean that conventional methods of selecting, developing or adapting PROMs are not effective [15, 22].

Small population sizes result in sampling and data collection issues when trying to recruit enough patients for clinical trials or PROM development/validation [23]. Collaborating with patient advocacy groups and clinical care networks may help to maximize patient recruitment [23]. Some software,

such as that based on Bayesian item response theory, can offer statistical methods to overcome the small sample size challenge while maintaining adequate psychometric qualities [24].

Multicentre or international data collection can increase sample sizes and allow pooling of data from different locations. This may include, for example, using a research network to collect data; all sites can identify eligible participants via electronic health records and use various recruitment methods. This enables efficient identification of eligible participants, an available sample and a standardized approach that allows for pooling of information [25]. However, challenges remain when collecting multi-site data. First, this may entail linguistic and cultural validation of these PROMs. Neglecting cultural specificities may lead to a lack of cultural validity, which makes comparison of results difficult and may result in dropouts and missing data [15]. Moreover, it may be difficult to engage patients over a wide area [26], and cross-cultural variations in research protocols can exist between centres [14]. It is essential that extra efforts are made to engage participants, and data collection should be standardized across locations as much as possible. Collecting and pooling international data can be an effective solution to overcome the small sample size issue only with the presence of high-quality study design and methods and psychometric, linguistic and cultural validity [14].

In establishing cultural validity, a statistical Rasch measurement theory calculation can first determine whether

**Table 1** Characteristics of included studies

Characteristics	N (%) / 44
Country	
USA	18 (41)
International	8 (18)
UK	7 (16)
France	2 (5)
Canada	2 (5)
Germany	2 (5)
Belgium	1 (2)
Netherlands	1 (2)
Ireland	1 (2)
Switzerland	1 (2)
Portugal	1 (2)
Type of research	
Original research	23 (52)
Review	9 (20)
Presentation (symposium, conference, webinar)	3 (7)
Report	2 (5)
Discussion/perspective paper	2 (5)
Short communication	2 (5)
Position statement	1 (2)
Editorial	1 (2)
Commentary	1 (2)
Focus	
Current issues and/or suggestions for using PROMs in RD	12 (27)
Method of developing PROMs for a specific disease	8 (19)
Examining psychometric properties of a PROM used for an RD	6 (14)
Methods to incorporate patient perspectives into PROM development and use	4 (9)
Applying existing PROMs to a specific RD	3 (7)
Methods for creating a general disease-specific PROM	2 (5)
Challenges capturing clinical outcomes in RD trials	2 (5)
Methods of using existing PROMs for a specific RD	1 (2)
Outcomes measures use for trials in a specific disease	1 (2)
Assessing data collection and/or psychometric properties of existing PROMs	1 (2)
Identifying and selecting existing disease-specific PROMs	1 (2)
Adding items to a PROM for a specific disease	1 (2)
Challenges in mapping for PROMs in RD	1 (2)
Examining trends of PROMs over time	1 (2)
Identifying existing PROMs and application	1 (2)

PROM patient-reported outcome measure, RD rare disease

significant country or language differences exist [15]. To better ensure cultural validity, it has been recommended to consider the following six types of cross-cultural equivalence: conceptual, semantic, operational, item, measurement and functional equivalence [27, 28], with the first three being particularly important [28]. One approach to achieving conceptual equivalence is the simultaneous development of instruments in different cultural settings. To ensure semantic equivalence, forward and back translation and cognitive debriefing in a small sample of the target population

is recommended [28–30]. Finally, an understanding of response styles in different settings and the use of different measurement approaches may help to address operational equivalence [28].

In terms of heterogeneity, there may be substantial variability even within one RD, so measures in which patients answer the same questions may not capture each manifestation of the disease [31]. However, collecting information on every PROM for every domain of a disease can be too demanding on patients, especially with a small sample,

**Table 2** Challenges and solutions in using patient-reported outcome measures to inform the appraisal of rare disease treatments

Topic	Potential challenges	Potential solutions	Interpreted implications for HTA: feasibility of solution implementation and what is needed from HTA bodies
Data collection/measurement	<p>Diversity of use of PROMs for RDTs</p> <p>There is a diversity of use of existing PROMs by researchers, making comparison of results difficult because different outcome measures are used [13, 20]</p>	<p>Develop recommended core outcome measures for disease area [13, 20]</p> <p>To help ensure accuracy of core outcome measures, the perspective of all stakeholders should be considered insofar as possible, especially patients and carers, e.g. concept-elicitation interviews can evaluate differences in patient experience across disease subtype [5, 11] To help ensure accuracy of core outcome measures, the perspective of all stakeholders should be considered insofar as possible, especially patients and carers, e.g. concept-elicitation interviews can evaluate differences in patient experience across disease subtype [5, 11]</p>	<p>Feasible: Core outcome measures would require initial upfront agreement and development, but, once developed, this is a tool that can be used sustainably with minor adjustments</p> <p>Need for recognition and buy-in of core outcome measures developed for specific disease areas, and time commitment to gather and include stakeholder perspectives</p>
	<p>Small, heterogeneous populations</p>	<p>With small population sizes, it can be difficult to recruit enough patients for trials or PROM development/validation [13, 21, 22]</p>	<p>Feasible: Requires collaboration stakeholder willingness for planning and time commitment but has the potential to save a substantial amount of time later in the process</p>
		<p>Consider collaborating with patient advocacy groups and/or clinical care networks to maximize recruitment [22]</p> <p>Consider using specialized statistical software that can work with small sample sizes while maintaining adequate psychometric properties [23]</p>	<p>Feasible: Using statistical software depends on resources and knowledge, but using available tools to overcome as many data collection challenges as possible does not require substantial time or structural changes</p>
		<p>Multi-site/international data collection to pool samples and gain larger sizes [12, 13, 24–28, 45]</p>	<p>Feasible with additional challenges: Multi-site/international data collection is a good way to overcome the small sample size issue but poses challenges with regard to obtaining cross-cultural validity, and may thus require more consolidated and adhered to guidance to produce data of sufficient quality</p>

Table 2 (continued)

Topic	Potential challenges	Potential solutions	Interpreted implications for HTA: feasibility of solution implementation and what is needed from HTA bodies
	<p>Measures in which each patient answers the same questions may not accurately capture each particular manifestation of the disease; PROMs are needed that can capture heterogeneity as much as possible without being too taxing on patients [13, 22, 29]</p>	<p>Tailor PROMs to condition/therapy while maintaining a set of standard core outcome measures [5, 29]</p>	<p>Feasible: Core outcome measures would require initial upfront agreement and development, but, once developed, this is a tool that can be used sustainably with minor adjustments Probably feasible: Tailoring PROMs to each condition would require significant time and resources Need for recognition of PROM development approaches that better manage heterogeneity Need for flexibility in accounting for HRQoL impact from different PROMs that would allow a fuller picture to be captured from more heterogeneous conditions</p>
	<p>Difficulty with self-reporting May not accurately capture patient experience: patients are often children/have cognitive impairment and are not able to self-report, different people may interact differently with instruments differently, measures may reflect disease and treatment as well as environmental or contextual factors [11, 26, 29, 30]</p>	<p>Use of parent or clinician proxy measures [13, 33] Use of children-specific PROMs [33]</p>	<p>Feasible: Proxy measures can and have been used, but care must be taken to ensure they are capturing the perspective of the patient as best as possible</p>
Psychometric properties	<p>Instruments are often not fit for purpose, and HTA evaluators are often not convinced a PROM is measuring what it is claimed to be measuring [29, 31]  Conventional methods are not always suitable for psychometric analysis in RDs because they require large samples and high-quality data [13, 33, 35]</p>	<p>Use of observer-reported outcomes [32] Help of an interviewer when inability to self-report is due to physical impairments [22, 24, 30, 32]  Prior discussion with the relevant evaluating agencies can help to ensure a PROM is compatible with their standards [5, 22, 32]  RD populations can be combined with populations with similar disease presentations to increase sample size [14]</p>	<p>Need for recognition of challenges in collecting PROM data from certain patient populations and ensuring alternatives are accepted  Probably feasible: This requires stakeholder willingness for planning and time commitment but has the potential to save a substantial amount of time later in the process Need for early and, if possible, iterative engagement between RDT developers and HTA evaluators Feasible: Combining populations with similar disease characteristics would require guidelines and best practices, but, if done properly, provides a promising solution to overcoming the limited RD sample size</p>

Table 2 (continued)

Topic	Potential challenges	Potential solutions	Interpreted implications for HTA: feasibility of solution implementation and what is needed from HTA bodies
<p>Validated generic PROMs are often preferred by HTA agencies</p> <p>Preference-based generic measures provide HSUV data</p> <p>Generic PROMs allow for comparability across conditions and populations [3, 9, 29]</p> <p>Use of disease-group-specific PROMs Advantages:</p> <p>More sensitive than generic and more widely applicable than disease-specific PROMs</p>	<p>Practical limitations exist for current PROMs for RDs in terms of feasibility and response rates, and they often have poor content validity and poor face validity due to data quality [3]</p> <p>Can be unresponsive and miss important information for specific RDs [12, 13, 16, 29]</p>	<p>Mixed-methods psychometric research is the best fit in RDs to best maximize clinical interpretability, increase conceptual understanding and avoid potential measurement problems [14]</p>	<p>Probably feasible: Mixed-methods research is a good approach to minimize potential problems but also requires time and resource investment</p> <p>Need for recognition of why it may be more challenging to demonstrate measurement properties in RDs</p> <p>Need for recognition that PROM data may be more uncertain for RDs and acceptance of innovative approaches to better deal with small samples/lower quality data</p>
<p>Use of generic PROMs</p>	<p>Practical limitations exist for current PROMs for RDs in terms of feasibility and response rates, and they often have poor content validity and poor face validity due to data quality [3]</p> <p>Can be unresponsive and miss important information for specific RDs [12, 13, 16, 29]</p>	<p>Use of expert panel review to determine face validity/generalizability; hybrid concept-elicitation or cognitive interviews or linking items to international classification systems to determine content validity [11, 36]</p> <p>Use both a generic and a disease-specific instrument for RDs in a complementary way [12]</p>	<p>Probably feasible: Qualitative data can be a good approach to ensuring validity without relying on large sample sizes but would require time and resource investment</p> <p>Need for recognition of the importance of, and willingness to consider, other forms of evidence in informing HRQoL impact</p> <p>Feasible with additional challenges: This could be a very valuable solution, but it depends on the specific HTA body and they type of data they are willing to accept; if an HTA agency only wants preference-based generic PROMs, the impact of adding disease-specific measures will likely be minimal</p> <p>Need for broadening of the willingness of HTA agencies to accept different forms of HRQoL data, including that from both generic and disease-specific measures</p>
<p>Advantages:</p> <p>Validated generic PROMs are often preferred by HTA agencies</p> <p>Preference-based generic measures provide HSUV data</p> <p>Generic PROMs allow for comparability across conditions and populations [3, 9, 29]</p> <p>Use of disease-group-specific PROMs Advantages:</p> <p>More sensitive than generic and more widely applicable than disease-specific PROMs</p>	<p>Often do not correspond specifically enough to the disease; may include a mix of conceptually different items, some of which may be entirely irrelevant and thus insufficient to grasp RD specificity [14, 16, 35]</p>	<p>Existing item banks can be used to find the best match between the concept of interest and the instrument [11]</p> <p>A systematic review to identify the most relevant PROMs may be needed. Existing tools to aid in selection include COSMIN, ePROVIDE™ and PROMIS [13]</p>	<p>Probably feasible: Disease-group PROMs are a promising solution, but the definition of 'disease group' needs to be clearly defined as to whether it refers to disease families, symptom- or function-specific PROMs or PROMs similar to those for common diseases</p>



Table 2 (continued)

Topic	Potential challenges	Potential solutions	Interpreted implications for HTA: feasibility of solution implementation and what is needed from HTA bodies
<p>Use of disease-specific PROMs</p> <p>Advantages:</p> <p>More sensitive and responsive than generic and disease-family PROMs; more likely to capture meaningful outcomes</p>	<p>Substantial heterogeneity in the manifestation of the RD in question may mean it is not possible to measure distinct outcomes across the population, making application of these general disease-specific PROMs difficult [22]</p>	<p>It may be important to limit the scope of applicability, so that concept-specific instruments are created that could be applicable across a family of closely related RDs and not just any similar disease [14]</p> <p>A multi-attribute questionnaire that poses questions most relevant for patients based on previous answers may help make the PROM more applicable across heterogeneous RD manifestations [5]</p> <p>Mixed-methods frameworks using qualitative and quantitative data may help maximize the applicability of the PROM to a different condition [5, 7]</p>	<p>Need for sufficient consideration of alternative sources of QoL evidence during the HTA deliberative process</p>
<p>Use of disease-specific PROMs</p> <p>Advantages:</p> <p>More sensitive and responsive than generic and disease-family PROMs; more likely to capture meaningful outcomes</p>	<p>Cannot make comparisons across patient groups [12]</p>	<p>Use both a generic and a disease-specific instrument for RDs in a complementary way [12]</p>	<p>Feasible with additional challenges: This could be a very valuable solution, but it depends on the specific HTA body and the type of data they are willing to accept; if an HTA agency only wants preference-based generic PROMs, the impact of adding disease-specific measures will likely be minimal</p> <p>HTA bodies would need to account for both generic and disease-specific instruments equally</p>
<p>Too much outcome measure heterogeneity from disease-specific measures hinders the ability to reliably/reproducibly capture significant change in disease [19]</p>	<p>Consider the following general approach for RDs: develop a variety of measures with the same basic presentation that include some features of generic measures, and add appropriate disease-specific aspects [13]</p>	<p>Probably not feasible: Tailoring PROMs to each condition would require significant time and resources</p> <p>HTA bodies would need to account for both generic and disease-specific instruments equally</p>	<p>Probably not feasible: Tailoring PROMs to each condition would require significant time and resources</p> <p>HTA bodies would need to account for both generic and disease-specific instruments equally</p>

Table 2 (continued)

Topic	Potential challenges	Potential solutions	Interpreted implications for HTA: feasibility of solution implementation and what is needed from HTA bodies
Validated, disease-specific PROMs for RDs are lacking [13, 15, 19, 38–40]	Concordance between generic and disease-specific QoL data is often lacking, making it complicated to conduct a mapping exercise that allows derivation of HSUVs from disease-specific PROMs [29]	If no (validated) disease-specific PROMs exist for a target condition, validated disease-group PROMs could be considered, or a new PROM could be created if resources permit [29, 38]	Probably feasible: Disease-group PROMs are a promising solution, but the definition of ‘disease group’ needs to be clearly defined as to whether it refers to disease families, symptom- or function-specific PROMs or PROMs similar to those for common diseases Feasible with additional challenges: Creating new PROMs requires significant time and resources, particularly for RD populations Need for recognition of limited number of RD-specific PROMs and risk that generic PROMs may not be sufficiently sensitive Need for provision of guidance on acceptable HRQoL measures in situations where existing PROMs or PROM development are unsuitable
The literature reports only four preference-based disease-specific PROMs yielding HSUVs in RDs (i.e. ALSUI, ABC-UI, MF-8D, and an algorithm for SBS-QoL) [42]	The degree of ‘overlap’ between generic and disease-specific PROMs should be assessed using proper correlation tests before conducting a mapping exercise [41]	Feasible with additional challenges: This approach in and of itself is good and specifically relevant for QALY-based HTA agencies looking for HSUVs, but the frequent lack of concordance between generic and disease-specific data means that mapping will often not be a viable solution Need for recognition that it is not always possible to map disease-specific PROMs onto generic ones, in which case an alternative should be used to generate HSUVs, such as referring to published literature or conducting ad hoc valuation studies Need for research focused on developing new approaches that would better make mapping a more viable solution for RDs	Probably not feasible: This approach is valuable, but developing such preference-based algorithms for additional PROMs would require significant time and resources Need for recognition of the limited number of preference-based disease-specific PROMs, which makes it very difficult to derive HSUVs in RDs

**Table 2** (continued)

Topic	Potential challenges	Potential solutions	Interpreted implications for HTA: feasibility of solution implementation and what is needed from HTA bodies
<p>Creating new disease-specific PROMs Advantages: Can be well-tailored to disease; high possibility of capturing meaningful outcomes</p>	<p>PROMs are time and resource intensive to create; it cannot realistically be done for every RD and manifestation [5, 23, 36, 44]  The natural history of most RDs is poorly understood, making it hard to identify concepts of interest [14]</p>	<p>It is important to use innovative and flexible PROM strategies for RDs, for instance, computer-assisted technology can ease the process by streamlining responses, reducing the burden on patients and allowing multi-site data collection [11, 13, 15, 24]  All available sources of information should be used to understand the natural history of an RD [11]  Incorporate patient voice early and throughout process of PROM development [5, 7, 46, 47]  Focus on the most common symptoms and impact that seem to be most important to patients [49]</p>	<p>Feasible with additional challenges: Creating new PROMs requires significant time and resources, particularly for RD populations Need for recognition that innovative and flexible PROM strategies is required  Feasible: The effort required to gather and use all sources of possible information is valuable, as the lack of understanding of the natural history is a key challenge Probably feasible: Incorporation of the patient voice requires willingness, participation and time commitments of stakeholders Need acceptance of a variety of sources of information</p>
<p>Effective approaches to developing PROMs are not always clear [38]</p>	<p>Take into account existing development guidance (e.g. FDA) and examples of PROM development [38, 50]</p>	<p>Feasible: Referring to and following any available high-quality guidance is only a matter of taking the time to do the research HTA agencies could set requirements for PROMs to be developed in accordance with existing guidance</p>	

*ABC-UI* Aberrant Behaviour Checklist Utility Index, *ALSUJ* Amyotrophic Lateral Sclerosis Utility Index, *FDA* US Food and Drug Administration, *HRQoL* health-related quality of life, *HSUV* health state utility value, *HTA* health technology assessment, *MF-8D* Myelofibrosis 8 Dimensions, *PROM* patient-reported outcome measure, *QoL* quality of life, *RD* rare disease, *RDT* rare disease treatment, *SBS-QoL* Short Bowel Syndrome health-related Quality of Life

which can result in fatigue and missing data [15]. Therefore, a primary challenge is to identify a PROM that has the most appropriate content possible, as well as a method of data collection in which patients can realistically participate [23]. PROMs need to be tailored to the patient, condition and therapy, but should at the same time contain some core comparable outcome measures, as mentioned previously [31].

*Difficulty with Self-Reporting* Obtaining information from patients with RDs can often be challenging. Patients are not always able to self-report [13, 32], as they are often children and/or may have cognitive impairment or functional limitations from their illness. Moreover, individual and cultural differences may influence how people interact with instruments. Furthermore, patients' responses may not only reflect disease and treatment experience, but also other environmental or contextual factors [31]. Therefore, self-report measures might not accurately capture the patient experience [28].

Several possible ways of dealing with these challenges exist. PROM information can be obtained through proxy PROMs or PROM instruments designed for children, although the reliability of parent-proxy responses still needs further investigation since child and adult preferences can differ [33]. PROMs can be completed with the help of an interviewer when patients are able to report but have physical impairments and cannot complete paper, computer or phone measures [23]. When no self-report or parent proxy is possible, it may be more suitable to use other outcome assessment measures, such as clinician- or observer-reported measures, performance outcomes or survival-based outcome measures [32], although it is also important to minimize clinician burden in terms of recruitment and data entry [26].

HTA bodies usually prefer established PROMs that are easy to interpret and are often critical of poor-quality PROM data [10]. While poor-quality data can understandably not be accepted, it is important that decision makers recognize innovative approaches to PROM use and development in light of the challenges posed by the paucity of PROMs for RDs and the small, heterogeneous populations.

### 3.1.2 Potential Challenges and Solutions with Psychometric Properties

PROMs are often not fit for purpose; the evaluators are often not convinced that a PROM is measuring what it claims, or supporting evidence may be insufficient [34]. As such, it is difficult for HTA bodies and payers to accept the results that can be realistically expected from RD PROMs [31]. Thus, if a drug is being developed for approval, then discussion and collaboration with the relevant agencies is essential to both ensure the PROM is attuned to their standards and to come to an agreement about generating evidence to reduce uncertainty [5, 23, 33].

PROMs need to be as valid and responsive as possible [9, 13] and allow for accurate interpretation of results [31], yet PROMs for RDs are often not validated for the population in which they are being used [5, 14]. Evaluating the psychometric properties of PROMs in RDs is challenging, as small population sizes and lower-quality data mean conventional methods are not always appropriate [15, 35, 36]. To deal with this, RD populations can be combined with populations with similar disease presentations to increase the sample size. Rasch measurement theory in particular is a potential solution that can be used for small sample sizes when combining RD populations with similar disease presentations; the use of differential item functioning can then be used to determine whether the responses of the combined groups are equivalent and, if they are, both samples can be used as a larger sample to validate the PROM [16]. Conventional methods are still used for measuring psychometric properties of RD PROMs [17, 37], but these require larger sample sizes and may not always be appropriate for such PROMs. Mixed-methods psychometric research is the best fit in RDs, as it can help to maximize clinical interpretability, improve conceptual understanding and avoid potential measurement problems [16].

Practical limitations exist for current PROMs for RDs in terms of feasibility and response rates, and they often have poor content validity and poor face validity due to issues with data quality [10]. Validating a PROM can be challenging [18], yet it is perhaps the most important psychometric property to address. Content validity (measuring concepts of importance to patients) is of utmost importance and should always be checked [38]. Vinik et al. [39] described the validation of a condition-specific questionnaire for an RD using different approaches (e.g. correlation, linear regression) to assess elements such as floor or ceiling effect and invariance, but these are conventional approaches and not ideal for such small sample sizes. However, some approaches do not require large sample sizes. For example, face validity/generalizability can be checked by expert panel review, and content validity can be checked by linking items to international classification systems [38]. Hybrid concept-elicitation/cognitive interviews can also be used to test content validity in new populations [13].

### 3.2 Generic PROMs

HTA bodies often prefer validated generic PROMs [10], as the standardized questions used allow for comparability across diseases and populations [31]. Preference-based generic PROMs (e.g. EQ-5D, HUI, SF-6D, 15D, Assessment of Quality of Life, Quality of Well-Being scale) from which HSUVs can be derived are preferred when economic analyses are conducted: "... having utility data is of course critical for accurate cost effectiveness analysis and there

aren't that many instruments out there that do have the utility information" [9].

However, generic PROMs can pose the challenge of being unresponsive and missing important disease- and population-specific data [14, 15, 18, 31]. This is a particular issue for PROMs in RDs because the small, heterogeneous samples and variation in treatment impact increase the possibility of generic PROMs being insufficiently applicable.

Since disease-specific PROMs tend to be more sensitive for distinguishing changes in health within a specific population or disease [15], it has been suggested to use both a generic and a disease-specific instrument for RDs in a complementary way. This enables comparability across populations and sufficient data for economic analysis, as well as the ability to detect small but important changes specific to the condition [14]. An objective systematic approach for RDs might be to develop a variety of measures that include some constant features of generic measures as well as measures related to the specific personal and societal factors appropriate for patients and disease-specific aspects. This would include, for example, basic QoL questions, with added disease-specific questions. For instance, a HTA-specific approach similar to that of the European Network for HTA could use a disease-specific PROM for effect assessment and a generic PROM for utility analysis [31]. The combination of generic and disease-specific instruments requires the willingness of HTA bodies to accept such evidence. Since many prefer generic PROMs, this would necessitate a change in requirements, or at least discussion to come to an agreement with decision makers in a particular country regarding what they would accept for a given RDT.

### 3.3 Disease-Group-Specific PROMs

The main advantage of disease-group-specific PROMs is that they are more widely applicable to various conditions than disease-specific PROMs, and—while they are not as sensitive as disease-specific PROMs—they are more sensitive than generic measures. It is nearly impossible to create disease-specific PROMs for every RD, making disease-group-specific PROMs across similar conditions a practical alternative.

The challenges with using disease-group-specific PROMs primarily revolve around their applicability and responsiveness. First, many of these instruments are not specifically compatible enough with the target disease and may include some items that are not applicable for the target disease/population [16]. Thus, the responsiveness of disease-group-specific PROMs to the RD or manifestation to which they are applied may not be well-established [36] or sufficient to grasp the RD's specificity [18]. Using such existing instruments from one context of use to another is valuable and needs to be carried out in a way that increases

applicability as much as possible. To facilitate the use of existing instruments, previously created item banks can be used to select and match instrument to the concept of interest (COI). Instruments closest to the COI that can be disaggregated should be selected, if possible, to only include relevant subscales [13]. A systematic review to identify the most relevant PROM may be needed. Existing supportive tools can facilitate the selection process [5], such as the COSMIN (CONsensus-based standards for the selection of health Measurement INSTRUMENTS) guidelines [40], the ePROVIDE™ database (contains a range of information on PROMs, including critical review on the measurement properties) [40, 41] or PROMIS (a cooperative group programme of research aiming to develop, validate and standardise item banks to capture PROM data across a wide range of conditions and domains) [15, 42]. To increase applicability for these instruments in general, the scope of applicability may need to be limited so that concept-specific instruments are created that could be applicable across a closely related group of RDs and not just any similar disease [16].

Similarly, if there is substantial heterogeneity in manifestation of the RD in question, which is often the case, it may not be possible to measure distinct outcomes across the population [23], making the application of an existing disease-group-specific PROM difficult. A multi-attribute questionnaire may therefore be useful when working with disease-group PROMs, to make the PROM more applicable across heterogeneous manifestations of an RD. Mixed-methods frameworks may also be a practical approach to optimize the applicability of a PROM in a new context of use [3]. The US FDA suggests using mixed methods in clinical trials to capture patient experience qualitatively and quantitatively and gives recommendations for identifying what is important to patients [7, 43–45].

### 3.4 Disease-Specific PROMs

The advantage of disease-specific PROMs is that they are more sensitive and responsive than generic PROMs and disease-group PROMs, making them more likely to capture meaningful outcomes of specific conditions.

Disease-specific measures pose the challenge that they can only make comparisons within the same patient group [14]. As disease-specific and generic instruments assess different aspects of QoL, the use of both instruments in a complementary way has been suggested [14].

Moreover, if multiple disease-specific measures for different conditions (and manifestations of a condition) exist and are used, this can lead to outcome measure heterogeneity. Outcome measure heterogeneity hinders the reliable and reproducible capture of a significant change in disease or health status and the synthesis and meta-analysis needed for evidence-base generation [46]. To manage this challenge,

recommended core outcome measures could be developed for disease-specific instruments to ensure a level of comparability [5, 15].

Disease-specific PROMs for RDs are generally lacking [15, 17, 46–48]. Those that have been developed may have been validated for a specific population that is not the target population, and clarity on them within the expert community is often lacking [49]. If no (validated) PROMs exist for an RDT, validated PROMs from other, similar diseases could be considered, or a new PROM can be created if resources permit. In addition, in QALY-based systems, generic preference-based PROMs yielding HSUVs (e.g. EQ-5D) are often preferred by HTA bodies. The ‘mapping’ technique can potentially allow the conversion of disease-specific PROM responses onto HSUVs derived from generic PROMs, but the lack of concordance between disease-specific and generic PROMs means it is complicated to conduct a mapping exercise in practice, and their degree of ‘overlap’ should be assessed in advance using proper correlation tests [31, 50].

Furthermore, the amount of preference-based disease-specific PROMs available to inform cost-utility analyses is limited, and only four have been identified in RDs: the Amyotrophic Lateral Sclerosis Utility Index (ALSUI), the Aberrant Behaviour Checklist Utility Index (ABC-UI) for fragile X syndrome, the Myelofibrosis 8 Dimensions, and a preference-based scoring algorithm for the Short Bowel Syndrome health-related Quality of Life (SBS-QoL) scale [51]. Utility data are often lacking for conditions affecting infants and young children because most instruments are not designed for such young age groups, yet about 80% of RDs affect children [52]. Additionally, the benchmarking of HSUVs estimated for similar diseases is limited by the resemblance of health states being compared [53]. Thus, the usage of preference-based disease-specific PROMs in HTA is generally limited to interventions where it is inappropriate to use a generic PROM.

The development of new algorithms to derive HSUVs from disease-specific PROMs is encouraged to evaluate RDTs where the use of generic PROMs is not appropriate. The range of HSUVs in diseases with similar characteristics can be used as a benchmark to validate results of such new preference-based instruments in RDs. For example, HTA bodies could use utilities benchmarked from similar diseases to define reasonable intervals for the incremental cost-utility ratio produced by preference-based RD-specific PROMs [53].

### 3.5 Creating New Disease-Specific PROMs

The advantage of developing new disease-specific PROMs is that they have the potential to be well-tailored to a specific disease, thus making them highly likely to capture

meaningful outcomes. A new PROM is extremely time and resource intensive to create well, often requiring several steps, patient and clinician engagement and qualitative and perhaps quantitative analysis [3, 24, 54]. This problem is amplified for RDs because of the heterogeneous disease presentation and small populations, which make it difficult to access and recruit (enough) patients to collect data for PROM development [38].

Several approaches can be used to optimize the PROM development process. For instance, computer-assisted technology (CAT) can streamline instrument development by helping reduce response burden on patients and increase completion rates, and multi-attribute questionnaires using skip patterns and computer adaptive testing can be customized to the individual. Such technologies enable a small but specific number of questions to be presented, selected based on a person’s answers to previous questions. This also allows disease-specific items highlighted by patients to be incorporated into the questionnaire, thus tailoring the PROMs to a patient’s specific symptoms without having to develop a completely new instrument [17]. Additionally, web-based approaches such as electronic PROMs allow data to be collected internationally, locally and in real time. This gives patients the freedom and flexibility to complete PROMs when it is convenient for them, which can improve dropout and missing rates and allows data to be collected from multiple sources and locations [13, 15, 26].

Additionally, the natural history of most RDs is poorly understood. Without sufficient information about the disease, it can be difficult to identify concepts of interest for meaningful treatment benefit and to clearly determine what outcome(s) should be measured [16]. To maximize knowledge as much as possible, all available sources of information should be used to understand the natural history of an RD. Engaging with patient advocacy groups and the RD community can help provide the full picture, from disease symptom onset to correct diagnosis and treatment [13]. It has also been recommended that a PRO consortium that incorporates the patient voice throughout all stages could be beneficial for developing PROMs that capture disease-specific patient experience and challenges [55]. It is crucial to partner with and listen to patients and caregivers early and systematically to identify meaningful treatment outcomes that resonate with their experience, preferences, expectations and values and compensate for a lack of natural history knowledge [5, 7, 56]. Patient or patient representative involvement via, for example, discussions or interviews can be used to explore and prioritize patients’ health concerns for a given RD [19, 57] and can help develop an understanding of natural history as much as possible. This approach can be used to identify the most common symptoms among patients and what they consider most important [58].

Effective approaches to developing PROMs are not always clear. For instance, the current FDA guidance for reviewing and evaluating existing PROMs does not address disease-specific issues in the development of PROMs, which is especially important for RDs [18]. When developing a disease-specific PROM, it may therefore be helpful to refer to any existing guidelines and examples of stages that may be useful. The FDA has documented guidelines [46], and an example of PROM development stages has been published [59].

## 4 Discussion

RDs pose unique challenges and require PROM strategies that are flexible and innovative [3, 16]. This scoping review synthesized the details of challenges and potential solutions in the literature for the use of PROMs for RDTs in HTA. To our knowledge, this is the first study to thoroughly review the literature and comprehensively identify the key challenges and existing potential solutions for the use of PROMs for RDTs in HTA. This work can be useful in helping HTA stakeholders understand the specificities of using and developing PROMs (and associated HSUVs) in RDTs for HTA.

An overarching takeaway is that it is essential for HTA stakeholders to be aware of the potential challenges that may arise when using PROMs in RDTs, which are similar in rare and non-rare diseases, but are exacerbated in RDs (e.g. heterogeneity of disease presentation and diversity of outcomes, data collection/psychometric property challenges due to small sample sizes, lack of sensitivity of generic PROMs, lack of disease-specific proms). In HTA, the added benefit of treatment that PROMs aim to demonstrate may not be accurately captured or interpretable because of these challenges. This requires HTA stakeholders to recognize the need for potential innovative solutions. Some potential solutions have been identified with this research, such as the use of core outcome measures, stakeholder communication to agree on acceptable and feasible PROM data and combining populations with similar diseases for PROM development or validation. However, many reported solutions are still conventional and not necessarily appropriate for RDTs; there remains a substantial need for more effective, innovative solutions. The solutions that were identified in this search were reviewed and discussed among the research team for appropriateness and feasibility.

Solutions that were agreed to be both appropriate and feasible to implement for RDTs were as follows:

- *Development of core outcome measure set across disease, disease subtype or similar disease* This requires initial upfront agreement and development; however, once developed, it is a tool that can be used sustainably with minor adjustments.
  - *Proactive stakeholder collaboration and discussion to agree on acceptable and feasible PROM data* This requires stakeholder willingness for planning and time commitment but has the potential to save a substantial amount of time later in the process.
  - *Combining populations with similar disease characteristics to increase sample size* Guidelines and best practices are needed for this but, if done properly, provides a promising solution to overcoming the limited RD sample size.
  - *CAT to streamline the PROM development process* Although this depends on resources, using available tools to overcome as many data collection challenges as possible does not require substantial time or structural changes.
  - *Take existing guidance into account (e.g. FDA)* Referring to and following any available high-quality guidance is only a matter of taking the time to do the research.
  - *Use of disease-group PROMs when no disease-specific PROM exists* This is a very promising solution, but requires further research regarding what ‘disease group’ actually entails. In this paper, we used the term ‘disease group’ to refer to any methods using PROMs across similar diseases, but definitions vary in the literature: some refer to disease families [16], others relate to symptom- or function-specific PROMs or PROMs that capture similar symptoms in analogous conditions [47], and still others refer to using PROMs similar to those for common diseases [60]. Thus, the parameters of such PROMs still need to be better defined.
- A solution that was considered appropriate and probably feasible to implement was mixed-methods research, which can serve to avoid potential measurement issues and maximize the applicability of disease-group-specific PROMs, but it does require time and resource investment.
- Solutions that were agreed to be appropriate and feasible but that entailed more potential challenges were as follows:
- *Use of generic and disease-specific measures* This could be a very valuable solution, but it depends on the specific HTA body and the type of data they are willing to accept. For example, if HTA agencies only want preference-based generic PROMs, the impact of adding disease-specific measures will likely be minimal. This solution would thus require a broadening of the willingness of HTA agencies to accept different forms of QoL data. In QALY systems, this would require clarity around how data not included in the economic model weighs in the decision.
  - *Mapping disease-specific measures to generic PROMs to enable HSUVs to be derived for QALY-based systems*

This approach in and of itself is good, and many HTA agencies are willing to use it, but it relies heavily on the degree of overlap between generic and disease-specific measures, making the mapping exercise particularly difficult for RD PROMs, and very few mapping algorithms are available in RDs [31, 50, 51]. Therefore, it may be necessary for HTA bodies to recognize that it is often not possible to map disease-specific PROMs onto generic ones, in which case an alternative should be used to generate HSUVs, such as referring to published literature or conducting ad hoc valuation studies.

- *Multi-site or international data collection* This is a good way to overcome the small sample size issues for PROM development and validation but poses challenges with regard to obtaining cross-cultural validity and may thus require more consolidated and adhered to guidance to produce data of sufficient quality.
- *Developing new disease-specific PROMs when none exist* While this solution can lead to PROMs that are well-tailored to the disease and can be created to be preference based, this is very resource and time intensive relative to the number of users and may often be beyond stakeholders' resources.

In terms of putting some of these solutions into practice and working towards further innovative solutions, HTA decision makers require a willingness to accept other forms of QoL evidence than are currently expected. HTA PROM requirements and preferences differ across jurisdictions, so one solution cannot be recommended across all; however, the general challenges are relevant for all stakeholders, and solutions can be specified to particular requirements with proactive collaboration between key stakeholders. This is in line with suggestions in the literature that, in order for PROMs to be integrated into HTA in a more standardised and sustainable way that contributes added value to the assessment, there is a need for international agreement on the evidentiary requirements that is accepted by all stakeholders [5, 16]. It has further been suggested that patient-relevant outcomes and endpoints should be discussed in advance with HTA bodies and other stakeholders via joint scientific advice meetings or qualification procedures, so that optimal evidence-generation plans can be designed and agreed on. When patient evidence suggests that novel PROMs or the adaptation of existing outcome measures to make them more relevant to patients are needed, the prospect of innovative measures or methodologies (such as individualized outcome measures) to capture patient benefit should be accepted in the HTA process [5, 16].

Additionally, the deliberative HTA process needs to allow for sufficient consideration of evidence around QoL. For countries with a QALY system, HSUVs from a generic PROM are used in the economic model, and the

disease-specific PROM (if considered) would be deliberated. The former is likely to have more weight in the decision, but this approach needs to be re-evaluated considering the frequent inability of generic PROMs to accurately capture PROM evidence [14, 15, 18, 31]. If disease-specific PROMs were considered and given equal weight, the solution of using generic and disease-specific PROMs together would be a more feasible solution. In countries with non-QALY systems, both generic and disease-specific PROMs would be deliberated in parallel, but the interpretation of the generic is often easier since committee members are more familiar with such measures. A proposed solution to this that could be developed would be to provide a benchmark that supports the interpretation and comparison of the results from the different PROMs.

This study has some limitations that should be acknowledged. First, the searches were only conducted in PubMed. While this is a comprehensive database, others may still have provided additional articles. Furthermore, we cannot clearly recommend one concrete approach for selecting and using PROMs for RDTs in HTA, as this is a complex process with additional factors that must be taken into account but which were beyond the scope of this research. Some points of relevance were not captured by the searches but are relevant for HTA bodies, including the change in disease course over time, slowing disease progression or maintaining function, which are significant for patients but difficult to capture and factor into clinical benefit; impact on family is similarly highly important and relevant for decision makers to factor into disease severity and treatment benefit.

Despite these limitations, this study contributes a situational analysis of where we are today and points to areas where further PROM research is needed, along with constructive discussions around what may or may not be acceptable for improving the development and use of PROMs for HTA in RDTs.

## 5 Conclusion

The usefulness of PROMs in HTA for RDTs may be undermined by practical challenges. A better understanding of the potential advantages, challenges and solutions when using PROMs for RDTs can help improve their use in HTA. This review provides an overview of the critical issues and some potential solutions for the use of PROMs for RDs in HTA. Some solutions can be taken forward, but solutions are often conventional ones that may have limitations in RDs. There is a pressing need for HTA stakeholders to acknowledge these limitations and discuss innovative approaches and non-standard solutions.



## Declarations

**Funding** This research was funded by the European Commission's Horizon 2020 research and innovation programme under the auspices of IMPACT-HTA (grant number 779312). The results presented here reflect the authors' views and not the views of the European Commission. The European Commission is not liable for any use of the information communicated.

**Conflicts of interest** EN is employed part-time and AW is contracted on a project basis by Dolon Ltd. Amanda Whittal, Michela Meregaglia and Elena Nicod have no conflicts of interest that are directly relevant to the content of this article.

**Availability of data and material** The information extraction document is available upon request.

**Consent for publication** All authors consented to the submission of this work for publication.

**Author contributions** AW, EN and MM conceptualized the design of this work. AW performed the searches, extracted the information and analysed the data. EN supported the analysis with comments and input throughout the process to shape the clarity of the results. AW, EN and MM drafted the work and/or substantially revised it.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

## References

1. US FDA. Patient-reported outcomes (PROs) in medical device decision-making. [Internet]. 2019. <https://www.fda.gov/about-fda/cdrh-patient-engagement/patient-reported-outcomes-pros-medical-device-decision-making#whatare>. Accessed 25 June 2019.
2. Smith A. What are PROMs. York Health Econ Consort. 2014;44:1–2.
3. Bell JA, et al. A pragmatic patient-reported outcome strategy for rare disease clinical trials: application of the EORTC item library to myelodysplastic syndromes, chronic myelomonocytic leukemia, and acute myeloid leukemia. *J Patient-Rep Outcomes*. 2019. <https://doi.org/10.1186/s41687-019-0123-4>.
4. Brazier J, Ara R, Rowen D, Chevrou-Severac H. A review of generic preference-based measures for use in cost-effectiveness models. *Pharmacoeconomics*. 2017;35(s1):21–31. <https://doi.org/10.1007/s40273-017-0545-x>.
5. Sabino G, Mills A, Jonker AH, et al. Patient-Centered Outcome Measures in the Field of Rare Diseases: International Rare Diseases Research Consortium (IRDiRC), 2016. [http://www.irdirc.org/wp-content/uploads/2016/03/PCOM\\_Post-Workshop\\_Report\\_Final.pdf](http://www.irdirc.org/wp-content/uploads/2016/03/PCOM_Post-Workshop_Report_Final.pdf). Accessed 5 Nov 2019.
6. Whitty JA, et al. Patient preferences in the medical product life-cycle. *Patient*. 2020;13(1):7–10. <https://doi.org/10.1007/s40271-019-00400-y>.
7. Contesse MG, Valentine JE, Wall TE, Leffler MG. The Case for the use of patient and caregiver perception of change assessments in rare disease clinical trials: a methodologic overview. *Adv Ther*. 2019;36(5):997–1010. <https://doi.org/10.1007/s12325-019-00920-x>.
8. Bai JP, Barrett JS, Burckart GJ, Meibohm B, Sachs HC, Yao L. Strategic biomarkers for drug development in treating rare diseases and diseases in neonates and infants. *AAPS J*. 2013;15(2):447–54.
9. Miller PA, Mulla SM, Adams-Webber T, Sivji Y, Guyatt GH, Johnston BC. Patient-reported outcomes in rare lysosomal storage diseases: key informant interviews and a systematic review protocol. *Int J Technol Assess Health Care*. 2016;32(6):393–9. <https://doi.org/10.1017/S0266462316000568>.
10. Morel T. How PROs can change classical clinical trials and endpoints ; how relevant regulations and patient involvement can influence PROs. In: 9th Symposium of the Swiss Clinical Trial Organisation, 2019.
11. Skoupá J, Annemans L, Hájek P. Health economic data requirements and availability in the European Union: results of a survey among 10 European countries. *Value Health Reg issues*. 2014;4:53–7.
12. Acaster S. Patient-reported outcome and observer-reported outcome assessment in rare disease trials. *Value Heal*. 2017;20(7):856–7. <https://doi.org/10.1016/j.jval.2017.05.017>.
13. Benjamin K, Vernon MK, Patrick DL, Perfetto E, Nestler-Parr S, Burke L. Patient-reported outcome and observer-reported outcome assessment in rare disease clinical trials: an ISPOR COA Emerging Good Practices Task Force Report. *Value Health*. 2017. <https://doi.org/10.1016/j.jval.2017.05.015>.
14. Pascoal C, et al. Patient and observer reported outcome measures to evaluate health-related quality of life in inherited metabolic diseases: A SCOPING review. *Orphanet J Rare Dis*. 2018. <https://doi.org/10.1186/s13023-018-0953-9>.
15. Slade A, et al. Patient reported outcome measures in rare diseases: A narrative review. *Orphanet J Rare Dis*. 2018. <https://doi.org/10.1186/s13023-018-0810-x>.
16. Morel T, Cano SJ. Measuring what matters to rare disease patients - Reflections on the work by the IRDiRC taskforce on patient-centered outcome measures. *Orphanet J Rare Dis*. 2017. <https://doi.org/10.1186/s13023-017-0718-x>.
17. Tosi LL, et al. 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 J Rare Dis*. 2019. <https://doi.org/10.1186/s13023-019-1004-x>.
18. Knoble N, Nayroles G, Cheng C, Arnould B. Illustration of patient-reported outcome challenges and solutions in rare diseases: A systematic review in Cushing's syndrome. *Orphanet J Rare Dis*. 2018. <https://doi.org/10.1186/s13023-018-0958-4>.
19. Swezey T, et al. Incorporating the patient perspective in the study of rare bone disease: insights from the osteogenesis imperfecta community. *Osteoporos Int*. 2019. <https://doi.org/10.1007/s00198-018-4690-7>.
20. Tricco AC, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med*. 2018;169(7):467–73. <https://doi.org/10.7326/M18-0850>.
21. Report O. Patient-Centred Outcome Measures in Research & Healthcare October 2018, no. October, pp. 1–14, 2018
22. Davio K. Patient-reported outcome measures are key to understanding rare disease. *Am J Managed Care*, [Internet]. 2018, <https://www.ajmc.com/view/patientreported-outcome-measures-are-key-to-understanding-rare-diseases> Accessed 11 Nov 2019.

23. Basch E, Bennett AV. Patient-reported outcomes in clinical trials of rare diseases. *J Gen Intern Med.* 2014;29(SUPPL. 3):801–3. <https://doi.org/10.1007/s11606-014-2892-z>.
24. Garrard L, Price LR, Bott MJ, Gajewski BJ. A novel method for expediting the development of patient-reported outcome measures and an evaluation of its performance via simulation. *BMC Med Res Methodol.* 2015;15(1):1–14. <https://doi.org/10.1186/s12874-015-0071-5>.
25. Hollin IL, Dimmock AEF, Bridges JFP, Danoff SK, Bascom R. Collecting patient preference information using a clinical data research network: demonstrating feasibility with idiopathic pulmonary fibrosis. *Patient Prefer Adhere.* 2019;13:795–804. <https://doi.org/10.2147/PPA.S201632>.
26. Javaid MK, et al. The RUDY study platform - a novel approach to patient driven research in rare musculoskeletal diseases. *Orphanet J Rare Dis.* 2016;11(1):1–9. <https://doi.org/10.1186/s13023-016-0528-6>.
27. Herdman M, Fox-Rushby J, Badia X. A model of equivalence in the cultural adaptation of HRQoL instruments: the universalist approach. *Qual Life Res.* 1998;7:323–35.
28. Devlin N, Lorgelly P, Herdman M. Can We Really Compare and Aggregate PRO Data Between People and Settings? Implications for Multi-Country Clinical Trials and HTA. *OHE Res. Pap.*, no. April, 2019.
29. Koller M, et al. Translation procedures for standardised quality of life questionnaires: The European Organisation for Research and Treatment of Cancer (EORTC) approach. *Eur J Cancer.* 2007;43(12):1810–20. <https://doi.org/10.1016/j.ejca.2007.05.029>.
30. Wild D, et al. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: Report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value Health.* 2005;8(2):94–104. <https://doi.org/10.1111/j.1524-4733.2005.04054.x>.
31. Rütther A, Elstein D, Wong-Rieger D, Guyatt G. Aspects of patient reported outcomes in rare diseases: A discussion paper. *Int J Technol Assess Health Care.* 2016;32(3):126–30. <https://doi.org/10.1017/S0266462316000271>.
32. Hong YD, Villalonga-Olives E, Peretto EM. Patient-reported outcomes in orphan drug labels approved by the us food and drug administration. *Value Health.* 2019;22(8):925–30. <https://doi.org/10.1016/j.jval.2019.03.010>.
33. Sampson C, Garau M. How should we measure quality of life impact in rare disease? recent learnings in spinal muscular atrophy. 2019, *Office of Health Economics*, no. 002146.
34. IQVIA. “PROving is worth: how to develop a PRO strategy to distinguish your product with regulators and payers webinar (Europe),” [Internet]. 2018, <https://www.iqvia.com/event/s/2018/09/proving-its-worth-how-to-develop-a-pro-strategy-to-distinguish-your-product> Accessed 18 Sep 2018.
35. Noël E, et al. Treatment needs and expectations for Fabry disease in France: Development of a new Patient Needs Questionnaire. *Orphanet J Rare Dis.* 2019;14(1):1–10. <https://doi.org/10.1186/s13023-019-1254-7>.
36. Johnston BC, et al. Limited responsiveness related to the minimal important difference of patient-reported outcomes in rare diseases. *J Clin Epidemiol.* 2016;79:10–21. <https://doi.org/10.1016/j.jclinepi.2016.06.010>.
37. Witt S, et al. Psychometric properties and minimal important differences of SF-36 in Idiopathic Pulmonary Fibrosis. *Respir Res.* 2019;20(1):1–11. <https://doi.org/10.1186/s12931-019-1010-5>.
38. Gelhorn HL, et al. The measurement of physical functioning among patients with Tenosynovial Giant Cell Tumor (TGCT) using the Patient-Reported Outcomes Measurement Information System (PROMIS). *J Patient-Rep Outcomes.* 2019;3(1):1–10. <https://doi.org/10.1186/s41687-019-0099-0>.
39. Vinik EJ, et al. Norfolk QOL-DN: Validation of a patient reported outcome measure in transthyretin familial amyloid polyneuropathy. *J Peripher Nerv Syst.* 2014;19(2):104–14. <https://doi.org/10.1111/jns5.12059>.
40. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res.* 2012;21(4):651–7.
41. Mapi Research Trust. ePROVIDE Online Support in Clinical Outcome Assessments. [https://eprovide.mapi-trust.org/search?form\[searchText\]=&form\[ezxform\\_token\]=vm5mY6mwXBf4tdf8IIZbV5vT3P8YndbDBnIiXAAa3gLI](https://eprovide.mapi-trust.org/search?form[searchText]=&form[ezxform_token]=vm5mY6mwXBf4tdf8IIZbV5vT3P8YndbDBnIiXAAa3gLI). Accessed 20 Nov 2019.
42. Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, Ader D, Fries JF, Bruce B, Rose M. The patient-reported outcomes measurement information system (PROMIS): progress of an NIH roadmap cooperative group during its first two years. *Med Care.* 2007;45:S3–11.
43. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research. In: Patient-focused drug development: collecting comprehensive and representative input. Silver Spring, MD; 2018.
44. Chalasani M, Vaidya P, Mullin T. Enhancing the incorporation of the patient’s voice in drug development and evaluation. *Res Invol Engag.* 2018;4:10.
45. FDA. Patient-focused drug development guidance public workshop: methods to identify what is important to patients and select, develop or modify fit-for-purpose clinical outcomes assessments [Internet]. 2018. <https://www.fda.gov/downloads/Drugs/NewsEvents/UCM620707.pdf>. Accessed 25 Sep 2018.
46. Klingels K, et al. Development of a patient-reported outcome measure for upper limb function in Duchenne muscular dystrophy: DMD Upper Limb PROM. 2016, <https://doi.org/10.1111/dmcn.13277>.
47. Salek MS, Ionova T, Johns JR, Oliva EN. Appraisal of patient-reported outcome measures in analogous diseases and recommendations for use in phase II and III clinical trials of pyruvate kinase deficiency. *Qual Life Res.* 2019;28(2):399–410. <https://doi.org/10.1007/s11136-018-2025-y>.
48. Behan L, Leigh MW, Dell SD, Quittner AL, Hogg C, Lucas JS. Validation of pediatric health-related quality of life instruments for primary ciliary dyskinesia (QOL-PCD). *Pediatr Pulmonol.* 2019;54(12):2011–20. <https://doi.org/10.1002/ppul.24507>.
49. Solebo AL, Barry RJ, Keane PA, Rahi JS, Denniston AK. Underutilisation of reproducible, child appropriate or patient reported outcome measures in childhood uveitis interventional research. *Orphanet J Rare Dis.* 2019;14(1):1–11. <https://doi.org/10.1186/s13023-019-1108-3>.
50. Meregaglia M, Whittal A, Nicod E, Drummond M. Mapping’ health state utility values from non-preference-based measures: a systematic literature review in rare diseases. *Pharmacoeconomics.* 2020. <https://doi.org/10.1007/s40273-020-00897-4>.
51. Rowen D, Brazier J, Ara R, Azzabi I, Rowen D. The role of condition-specific preference-based measures in health technology assessment. *Pharmacoeconomics.* 2017;35(s1):33–41. <https://doi.org/10.1007/s40273-017-0546-9>.
52. Wakap SN, Lambert DM, Olry A, Rodwell C, Gueydan C, Lanneau V, Murphy D, Le Cam Y, Rath A. Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. *Eur J Hum Genet.* 2020;28(2):165–73.
53. Teynor M, et al. Benchmarking utility estimates in rare diseases may guide the selection of data inputs in health economic modeling. In: 2018, European Conference on Rare Diseases and Orphan Products.
54. Gounder MM, Maddux LA, Paty J, Atkinson TM. Prospective development of a patient-reported outcomes

- instrument for desmoid tumors or aggressive fibromatosis. *Cancer*. 2020;126(3):531–9. <https://doi.org/10.1002/cncr.32555>.
55. Gipson DS, et al. NephCure accelerating cures institute: a multi-disciplinary consortium to improve care for nephrotic syndrome. *Kidney Int Rep*. 2018;3(2):439–46. <https://doi.org/10.1016/j.ekir.2017.11.016>.
  56. Morel T, Aymé S, Cassiman D, Simoens S, Morgan M, Vandebroek M. Quantifying benefit-risk preferences for new medicines in rare disease patients and caregivers. *Orphanet J Rare Dis*. 2016;11(1):1–12. <https://doi.org/10.1186/s13023-016-0444-9>.
  57. Dell SD, et al. Primary ciliary dyskinesia: First health-related quality-of-life measures for pediatric patients. *Ann Am Thorac Soc*. 2016;13(10):1726–35. <https://doi.org/10.1513/AnnalsATS.201603-198OC>.
  58. Propp R, et al. Development and content validation of the Muscular Dystrophy Child Health Index of Life with Disabilities questionnaire for children with Duchenne muscular dystrophy. *Dev Med Child Neurol*. 2018;61(1):75–81. <https://doi.org/10.1111/dmcn.13977>.
  59. Swigris JJ, Wilson SR, Green KE, Sprunger DB, Brown KK, Wamboldt FS. Development of the ATAQ-IPF: A tool to assess quality of life in IPF. *Health Qual Life Outcomes*. 2010;8:1–9. <https://doi.org/10.1186/1477-7525-8-77>.
  60. Lanar S, Acquadro C, Seaton J, Savre I, Arnould B. To what degree are orphan drugs patient-centered? A review of the current state of clinical research in rare diseases. *Orphanet J Rare Dis*. 2020;15:1–18.