

Perspective

Cite this article: Meregaglia M, Nicod E, Drummond M (2020). The estimation of health state utility values in rare diseases: overview of existing techniques. *International Journal of Technology Assessment in Health Care* **36**, 469–473. <https://doi.org/10.1017/S0266462320000665>

Received: 25 March 2020
Revised: 18 August 2020
Accepted: 20 August 2020
First published online: 25 September 2020

Key words:

Rare diseases; Utility; Direct techniques; Indirect techniques

There are several techniques for estimating health state utility values, each of which presents pros and cons in the context of rare diseases (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.

Rare diseases (RDs) are conditions affecting fewer than 1 in 2,000 people in the European Union, or less than 200,000 people in the United States (1). RDs are often severe; few have curative therapies, whereby most treatments aim to alleviate symptoms, enhance quality of life or delay the health status deterioration, with the ultimate goal of controlling or modifying the disease trajectory (2). Thus, patient-reported outcome measures (PROMs) are increasingly adopted in health technology assessment (HTA) to estimate the benefits of treatments in terms of quality of life (1;3), especially when their responses can be converted into health state utility values (HSUVs). HSUVs represent individual preferences for a given health state measured on a scale from zero (“death”) to one (“full health”), which, if combined with time spent in that state, generate quality-adjusted life years (QALYs) (4). QALYs are then incorporated into cost-utility models that inform reimbursement decisions in many HTA systems. Two groups of (direct and indirect) techniques exist to estimate HSUVs (4), the latter using a particular type of PROMs called multi-attribute utility instruments (MAUIs). Their application to RDs may be challenging because of their low incidence and high heterogeneity. This paper discusses the pros and cons of each technique in relation to RDs (Table 1).

Direct Techniques

The most common techniques for measuring HSUVs directly include standard gamble (SG) and time trade-off (TTO). SG involves trading a sub-optimal health state A with the risk of immediate death ($1-p$), and the HSUV is represented by p (i.e. the probability of returning to “full health”). TTO trades duration of life against quality of life, and the HSUV is the ratio between time in full health (X) and time (e.g. 10 yr) in state A (4). The person trade-off (PTO) is similar to TTO but focuses on persons instead of time as the trade-off unit. It requires indicating how many patients (X) in state A and (Y) in “full health,” respectively, are considered equal to saving one life year (5). The HSUV of A is calculated as Y/X (6), which is a “social” value as opposed to the “individual” one obtained from SG/TTO (5). The “rule of rescue” relies on the moral imperative people feel to rescue identifiable individuals facing an imminent risk of avoidable death, irrespective of cost-effectiveness considerations (6;7). It can provide HSUVs but raises measurement challenges, since it requires a two-stage procedure combining SG or TTO (evaluating individual utility) and PTO (evaluating social utility), and ethical issues, since prioritizing interventions based on “identifiability” is not morally justifiable, and contradicts with the impersonal logic underlying cost-utility analysis (6). Lastly, discrete choice experiments (DCEs) ask respondents to choose between hypothetical health states and derive HSUVs through regression techniques (8).

The measurement of HSUVs using direct techniques is conducted either with patients (or caregivers as proxy respondents), who value their own health state, or members of the public, who value hypothetical health states represented in “vignettes.” The use of “vignettes” is advantageous in RDs, since they can be designed to incorporate relevant health issues and limit the use of patient-level data. The risk is that the health states presented may not fully

Table 1. Overview of the available techniques and implications for RDs

Technique (group)	Technique	Brief description		PROS (P)/CONS (C) in RDs	Comments relating to RDs
Direct techniques	Standard gamble	Requires making a choice between two alternatives: (1) remaining in a given health state for the remaining lifespan or (2) undergoing a risky treatment with two possible outcomes: returning to full health with probability p , or dying with probability $(1-p)$; p is varied until the two alternatives are indifferent. The HSUV is equal to p .	C	Both SG and TTO are demanding tasks	Up to 75% of RDs affect children, and many result in cognitive impairment
	Time trade-off	Requires making a choice between two alternatives: (1) remaining in a given health state for the remaining lifespan (e.g. 10 yr) or (2) living in full health for a shorter period ($X < 10$). X is varied until the two alternatives are indifferent; the HSUV is equal to $X/10$.	P/C	Both SG and TTO tend to produce higher HSUVs, thus favoring allocation of resources to life-saving interventions	May favor treatments for life-threatening RDs (e.g. cystic fibrosis)
	Person trade-off	Estimates the “social value” of healthcare interventions by asking people how many individuals (X) in health state A they consider equivalent (to save) to individuals (Y) in full health. The utility of health state A is calculated as Y/X .	P	The public may assign greater value to treatments for severe diseases, thus reducing their ICER	Can lower the ICER of treatment for severe RDs
			C	Is quite demanding and requires a large number of respondents to minimize random measurement errors	Small-scale studies are usually conducted in RDs
	DCE	Requires making a binary choice between hypothetical health states; HSUVs are derived through regression models.	P	Allows to easily value health states considered worse than death	Useful for RDs with very poor quality of life (e.g. amyotrophic lateral sclerosis)
			P	Simpler task than traditional direct techniques (SG, TTO)	Up to 75% of RDs affect children, and many result in cognitive impairment
	Rule of rescue	Relies on the principle that people feel a moral imperative to rescue identifiable individuals facing avoidable death (or severe illness). HSUVs can be derived from a combination of SG/TTO and PTO.	P	Prioritizes the severity of the disease over treatment effectiveness and costs	Can lower the ICER of treatment for severe RDs
			P	Prioritizes life-saving interventions, but also interventions for highly visible physical deformities or disfigurements	Can favor treatments for life-threatening (e.g. cystic fibrosis) or visibly disabling RDs (e.g. Antley-Bixler syndrome)
			P/C	Gives absolute priority to identifiable patients and penalizes those unidentified, potentially causing ethical issues	Being few in number, RDs patients, and especially those with malformations and deformities, are more likely to gain visibility through media/fundraising campaign
			C	Presents measurement problems (two-stage procedure)	May require large studies that are unlikely to be performed in RDs
	Any	Use of “vignettes” (i.e. brief descriptions of hypothetical health states) to be valued by non-patient populations	P	All direct tasks can be performed also by the public using “vignettes”	Useful in case of vulnerable patients (e.g. children) and/or difficult to identify (e.g. due to geographical dispersion)
			C	The creation of realistic “vignettes” requires interviews with patients and clinical experts	A sufficiently large number of patients and clinical experts is often hard to find in RDs
			C	“Vignettes” present standard health states	Unable to capture the variability of health states in RDs

Indirect techniques	Generic MAUIs	Use of generic PROMs accompanied by a set of pre-calculated “tariffs” to generate HSUVs	P	Simple to administer, availability of children-specific tools (e.g. EQ-5D-Y)	Up to 75% of RDs affect children, and many result in cognitive impairment
			C	May not be sensitive enough to capture some relevant health issues	Most RDs present very heterogeneous symptoms not properly represented in generic PROMs
			P/C	Tend to produce lower HSUVs, thus they might favor the allocation of resources to non-fatal, chronic conditions	May favor treatments improving symptomatology and quality of life in chronic RDs (e.g. cutaneous lymphoma)
	Disease-specific MAUIs	Use of disease-specific PROMs accompanied by a set of pre-calculated “tariffs” to generate HSUVs	P	Include dimensions that are relevant for a specific condition	Can better capture the heterogeneity of individual RDs
			C	Do not allow comparability across different conditions	Less useful for the purpose of allocating resources among different RDs
			C	Specific MAUIs available only in few RDs	Very few RD-specific MAUIs identified; high costs in developing new, multilingual tools (to cope with geographical dispersion) in relation to the limited number of users
	Mapping	Use or development of an algorithm relating non-preference-based measures (e.g. RD-specific PROMs) to MAUIs	P	Allows to exploit a number of PROMs, although non-preference-based, and clinical measures into HTA processes	Clinical studies in RDs tend to use disease-specific PROMs, which can be converted onto HSUVs using mapping
			C	There is scarce “overlap” between disease-specific measures and generic MAUIs	RD-specific PROMs include very specific symptoms that are seldom captured in generic PROMs
			C	Requires large samples (possibly to be split into development and validation samples).	Only small sample sizes are achievable in most RDs
P/C			Allows to apply previously developed algorithms to derive HSUVs	There are several algorithms developed in common conditions (e.g. lung cancer) that can be applied to their rare variants (e.g. pleural mesothelioma); however, their application tends to overestimate HSUVs in RDs, which are usually more severe than similar common conditions	
ALL	ALL	Direct and indirect techniques involving patients	P/C	Require that patients fill in one (or more) questionnaires or perform direct choice tasks (unless using proxy-reporting or vignettes)	Patients are those who know best their own health status; however, sufficiently large and representative patient samples are difficult to recruit in RDs due to low patient numbers and geographical dispersion, and most patients are unable to reply questions or perform complex tasks (because of their young age or cognitive impairment).

DCE, discrete choice experiment; EQ-5D-Y, EuroQol Five-Dimensional Questionnaire, Youth Version; HSUV, health state utility values; HTA, Health Technology Assessment; ICER, incremental cost-effectiveness ratio; MAUI, multi-attribute utility instrument; PROM, patient-reported outcome measure; PTO, person trade-off; RD, rare disease; SG, standard gamble; TTO, time trade-off.

capture the experience of individual cases because of an extremely varied symptomatology. Moreover, the creation of realistic vignettes requires an extensive qualitative work (e.g. in-depth interviews and focus groups) involving patients that may be difficult to identify and recruit, and a sufficient amount of clinical expertise and literature that may be lacking in RDs. Lastly, there may be differences among valuation methods leading to inconsistent HSUV results (1;9–11).

SG and TTO may be challenging or unfeasible to administer in several RDs that affect children (around 75 percent) or are associated with cognitive and communication impairments, unless parent or caregiver proxy reporting is used (3;11). The PTO is usually performed by the public, who may assign a greater value to treatments for people with serious conditions, including RDs (7). However, the task requires large samples of participants to minimize measurement errors (5), while small-scale studies are usually conducted in RDs.

In the recent literature (7;12), the “rule of rescue” approach has been encouraged to value health states in RDs. This approach, by giving priority to identifiable people may favor RDs since patients are few in number, often children, or presenting visible deformities or disfigurements. Social media also play a role in increasing their recognizability and visibility in society compared to common conditions. Moreover, the estimated budget impact of rescue treatments for RD patients is perceived as negligible by society. In RDs, the “rule of rescue” has been discussed in relation to severe traumatic brain injury, where the decision to perform decompressive craniotomy is often taken irrespective of the patient’s subsequent quality of life, procedural costs, or trade-offs in using these resources to improve health in the wider community (13).

Lastly, DCEs may be promising in RDs, especially in those associated with very poor quality of life (e.g. amyotrophic lateral sclerosis, ALS), since health states can be valued as “worse than death” without altering the task, as is required with lead-time TTO. Moreover, DCEs are cognitively simpler than traditional direct techniques, since they require expressing a preference between state A and state B, without trading against risk of death or duration of life (8); thus, they are less affected by measurement errors when administered to vulnerable RD patients.

Indirect Techniques (MAUIs)

HSUVs can be estimated indirectly by using MAUIs, which are PROMs based on individual preferences, typically obtained in country-level surveys where members of the public value a sample of health states by using direct techniques or DCEs (8;14), and subsequently aggregated as mean scores. Therefore, MAUIs are provided with a value set of “tariffs” for every combination of the instruments’ domains/levels. The most popular generic MAUIs are the EuroQol 5-dimension (EQ-5D), the Health Utility Index (HUI), and the Short Form 6 Dimension (SF-6D) (14). Disease-specific MAUIs also exist, which are useful to provide HSUVs in conditions where generic ones are not appropriate, sensitive or responsive, or to compare HSUVs across different studies on a specific condition. However, these tools do not allow cross-disease comparisons and their role in HTA is often limited to providing additional supporting evidence of treatment benefits (beyond the cost-utility model) (15). In studies where MAUIs have not been used, “mapping” is an accepted alternative to generate HSUVs through *the development and use of a model or algorithm that uses data from other measures of health outcomes* (16), such as non-preference-based PROMs.

Indirect methods avoid asking patients the complex task of trading health states with different risks of death (SG) or years of remaining life (TTO). Such trade-offs need to be done only once by involving the public in the valuation exercise. The resulting “tariffs” are used to derive HSUVs by administering the corresponding MAUIs to patients. However, the difficulties encountered in the collection of PROMs in RDs also apply to MAUIs (3). First, the low prevalence of each RD results in small and heterogeneous samples affecting data collection and statistical analyses (1;3). Second, even though MAUIs are much easier to respond to compared to SG/TTO, they remain challenging for children and may need to rely on parent proxy reporting. Some simplified self-reported MAUIs, such as the EQ-5D-Y (Youth), are available for children and may facilitate the estimation of HSUVs in pediatric RDs. Third, the administration of MAUIs to RD patients may be challenging due to their geographical dispersion, which generally requires multi-site studies with related logistic and financial issues.

Fourth, generic MAUIs may not be sensitive enough to capture relevant health issues in RDs, particularly in the more heterogeneous conditions (1). In a recent survey, most RD patients reported that EQ-5D-5L did not capture important issues affecting their daily life, such as fatigue, relationship/social life, and co-morbidities (17). A systematic review of HSUVs in Duchenne Muscular Dystrophy (DMD) identified the use of EQ-5D or HUI3 in all studies deriving HSUVs, but that they did not capture relevant quality of life dimensions such as hope, fear, fatigue, social participation, and dignity (18). However, the level of sensitivity of MAUIs may vary according to the specific instrument adopted and individual RDs. For example, HUI3 compared to EQ-5D has a greater coverage of domains relevant for DMD patients such as ambulation and dexterity (18).

Using RD-specific MAUIs helps overcome the issue of poor sensitivity with generic ones, but only few instruments are available (e.g. ALS Utility Index, Short Bowel Syndrome-Quality of Life scale (15)), and the rarity of each condition can make the cost of new instruments development unsustainable (3). Lastly, “mapping” allows to exploit disease-specific, non-preference-based PROMs, which are preferred in clinical studies on RDs (11), but presents several pitfalls in RDs, such as lack of sufficiently large samples to develop and test algorithms, limited “overlap” between RD-specific and generic PROMs, or poor applicability of algorithms developed in similar non-RDs (19).

Implications for HTA

The impact of using different techniques to estimate HSUVs for HTA was assessed in a wide range of conditions, including some RDs (4). Overall, direct methods tend to produce consistently higher HSUVs than indirect methods. In ALS, the HSUVs derived from SG were significantly higher than those from EQ-5D for all severity levels (20). Similarly, in systemic sclerosis, the agreement between SF-6D and TTO/SG was poor, with SF-6D providing lower values than direct techniques (21). In esophageal cancer, TTO values were higher or lower than EQ-5D depending on tumor stage (22). Since the utility of death is fixed at zero (4), using direct techniques in RDs might favor new treatments for life-threatening diseases, including those with onset in early childhood (30 percent of children with RDs do survive to age 5 (3)), rare infectious diseases (e.g. tuberculosis), or rare cancers (e.g. pleural mesothelioma). Conversely, the use of MAUIs, giving more space for utility gain, may favor

treatments targeting symptoms relief and quality of life improvement in chronic RDs (e.g. cutaneous lymphoma). Lastly, the “rule of rescue” approach and PTO may advantage treatments for RDs in general, if people assign greater value to health gains in rare and severe conditions (12).

Conclusions

The estimation of HSUVs is crucial in RDs, given the growing use of PROMs to record quality of life gains from new treatments. However, there is no agreement on the most appropriate technique, and each may present pros and cons for individual RDs. Overall, the rarity of each condition allows the identification of only a few representative patients which affects the precision of the aggregate HSUVs resulting from the administration of MAUIs, or the evaluation of the individuals’ own health status in direct measurement tasks. In very heterogeneous RDs, different techniques can be used for patient subgroups to address their specific characteristics and increase the sample size (23). Moreover, there is a dearth of disease-specific MAUIs that could replace generic ones when these are not sensitive enough. The large number of RDs, the low prevalence for each, and patients’ geographical dispersion discourages the investment of resources in developing new multilingual tools (3), as well as performing ad hoc evaluation studies because of logistical issues and long timelines for recruitment and data collection. In most RDs affecting children (3), the use of children-specific MAUIs is encouraged. For its simplicity, the visual analogue scale (VAS) may be a further option, although it is a choice-less task and therefore less preferred than other direct techniques (9). Moreover, studies should take a family perspective to incorporate the HSUVs of parents (11;24). The use of less conventional approaches such as “vignettes,” PTO, “rule of rescue,” and DCEs requires further evidence on their usefulness in RDs and acceptability in HTA, given that some agencies already have special processes for the assessment of treatments for RDs (e.g. higher cost-effectiveness thresholds, reflecting the value of treating severe illnesses where no other treatment exists) (11;25). Overall, the establishment of a set of recommendations is required to inform the estimation of HSUVs across different RDs, and to address the HTA implications of using alternative techniques.

Acknowledgments. This research was funded by the European Commission’s Horizon 2020 research and innovation program and was undertaken 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. The authors are grateful to Dr Karen Facey and Dr Amanda Whittall for providing useful comments on this manuscript.

References

1. 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;13:228.
2. Institute of Medicine. *Rare diseases and orphan products: Accelerating research and development.* Washington, DC: The National Academies Press. 2010. Available from: <https://doi.org/10.17226/12953>.
3. Slade A, Isa F, Kyte D, Pankhurst T, Kerecuk L, Ferguson J et al. Patient reported outcome measures in rare diseases: A narrative review. *Orphanet J Rare Dis.* 2018;13:61.
4. Arnold D, Girling A, Stevens A, Lilford R. Comparison of direct and indirect methods of estimating health state utilities for resource allocation: Review and empirical analysis. *Br Med J.* 2009;339:b2688.
5. Nord E. The person-trade-off approach to valuing health care programs. *Med Decis Making.* 1995;15:201–08.
6. McKie J, Richardson J. The rule of rescue. *Soc Sci Med.* 2003;56:2407–19.
7. Silva EN, Sousa TR. Economic evaluation in the context of rare diseases: Is it possible? *Cad Saude Publica.* 2015;31:496–506.
8. Bansback N, Brazier J, Tsuchiya A, Anis A. Using a discrete choice experiment to estimate health state utility values. *J Health Econ.* 2012;31:306–18.
9. Brazier J, Rowen D. NICE DSU Technical Support Document 11: Alternatives to EQ-5D for generating health state utility values. 2011. Available from: <http://www.nicedsu.org.uk>.
10. Sireau N. *Rare diseases: Challenges and opportunities for social entrepreneurs.* Sheffield, UK: Greenleaf Publishing Limited; 2013.
11. Pearson I, Rothwell B, Olaye A, Knight C. Economic modeling considerations for rare diseases. *Value Health.* 2018;21:515–24.
12. Hughes DA, Tunnage B, Yeo ST. Drugs for exceptionally rare diseases: Do they deserve special status for funding? *QJM.* 2005;98:829–36.
13. Honeybul S, Gillett GR, Ho KM, Lind CR. Neurotrauma and the rule of rescue. *J Med Ethics.* 2011;37:707–10.
14. Brazier J, Ara R, Rowen D, Chevrou-Severac H. A review of generic preference-based measures for use in cost-effectiveness models. *Pharmacoeconomics.* 2017;35:21–31.
15. Rowen D, Brazier J, Ara R, Azzabi Zouraq I. The role of condition-specific preference-based measures in health technology assessment. *Pharmacoeconomics.* 2017;35:33–41.
16. Longworth L, Rowen D. Mapping to obtain EQ-5D utility values for use in NICE health technology assessments. *Value Health.* 2013;16:202–10.
17. Efthymiadou O, Mossman J, Kanavos P. Health related quality of life aspects not captured by EQ-5D-5L: Results from an international survey of patients. *Health Policy.* 2019;123:159–65.
18. Szabo SM, Audhya IF, Malone DC, Feeny D, Gooch KL. Characterizing health state utilities associated with Duchenne muscular dystrophy: A systematic review. *Qual Life Res.* 2020;29:593–605.
19. Meregaglia M, Whittall A, Nicod E, Drummond M. “Mapping” health state utility values from non-preference based measures: A systematic literature review in rare diseases. *Pharmacoeconomics.* 2020;38(6):557–574.
20. Green C, Kiebert G, Murphy C, Mitchell JD, O’Brien M, Burrell A et al. Patients’ health-related quality-of-life and health state values for motor neurone disease/amyotrophic lateral sclerosis. *Qual Life Res.* 2003;12:565–74.
21. Khanna D, Furst DE, Kee Wong W, Tsevat J, Clements PJ, Park GS et al. Reliability, validity, and minimally important differences of the SF-6D in systemic sclerosis. *Qual Life Res.* 2007;16:1083–92.
22. Wildi SM, Cox MH, Clark LL, Turner R, Hawes RH, Hoffman BJ et al. Assessment of health state utilities and quality of life in patients with malignant esophageal dysphagia. *Am J Gastroenterol.* 2004;99:1044–9.
23. Ara R, Brazier J, Young T. Recommended methods for the collection of health state utility value evidence in clinical studies. *Pharmacoeconomics.* 2017;35:67–75.
24. Simon NJ, Richardson J, Ahmad A, Rose A, Wittenberg E, D’Cruz B et al. Health utilities and parental quality of life effects for three rare conditions tested in newborns. *J Patient Rep Outcomes.* 2019;3:4.
25. Weinstein N, Martin M, Campbell R. Orphan drugs in the UK, do they meet the NICE highly specialised technology threshold? *Value Health.* 2017;20:660.