Article:
Meregaglia, M., Whittal, A., Nicod, E. et al. (1 more author) (2020) 'Mapping' health state utility values from non-preference-based measures:a systematic literature review in rare diseases. Pharmacoeconomics. ISSN: 1179-2027

https://doi.org/10.1007/s40273-020-00897-4

‘Mapping’ health state utility values from non-preference based measures: a systematic literature review in rare diseases

A systematic review of mapping studies in rare diseases

Michela Meregaglia [1]*, Amanda Whittal [1], Elena Nicod [1], Michael Drummond [2]

[1] Research Centre on Health and Social Care Management (CERGAS), Bocconi University, Milan (Italy)
[2] Centre for Health Economics, University of York, York (United Kingdom)

*corresponding author: michela.meregaglia@unibocconi.it

Funding

This research was funded by the European Commission’s Horizon 2020 research and innovation programme and was undertaken under the auspices of IMPACT-HTA (Grant 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.

Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Authors’ contributions

All authors contributed to the study conception and design. MM and AW performed the literature searches. MM analysed the data and wrote the first draft of the manuscript. All authors commented on previous versions of the manuscript and approved the final manuscript.

Acknowledgments

The authors are thankful to Dr Karen Facey (University of Edinburgh) for her constructive suggestions and comments.

Conflict of interest

MM has no conflict of interest. AW has no conflict of interest. EN reports personal fees from Dolon Ltd outside the submitted work and has no conflict of interest. MD …..

ABSTRACT

Background: In rare disease (RD) studies, generic preference-based patient-reported outcome measures (PROMs) that yield health state utility values (HSUVs) are seldom collected, as they are considered not sensitive enough for these small and heterogeneous patient populations. In such cases, a HSUV can also be obtained by 'mapping' a more sensitive 'source' (e.g., disease-specific PROM) to a 'target' preference-based measure (e.g., EuroQol-5 Dimension (EQ-5D)) through a statistical relationship.

Objective: This study aimed to systematically review all published studies using 'mapping' to derive HSUVs from non-preference-based measures in RDs (i.e. affecting fewer than 1 in 2,000 people), and identify any critical issue related to the main features of RDs.

Methods: The following databases were searched during the first half of 2019 without time, study design or language restrictions: MEDLINE (via PubMed), the School of Health and Related Research Health Utility Database (ScHARRHUD) and the Health Economics Research Centre (HERC) database of mapping studies (version 7.0). The keywords combined terms related to 'mapping' with ORPHANET's list of RD indications (e.g., 'acromegaly'), in additional to 'rare' and 'orphan'. 'Very rare' diseases (i.e. with less than 1000 cases or families documented in the medical literature) were excluded from the searches. A predefined, pilot-tested extraction template (in Excel®) was used to collect structured information from the studies.

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

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

Key Points for Decision Makers

- In rare diseases (RDs), few studies (n=19) mapped non-preference-based measures onto generic preference-based ones (e.g. EQ-5D), while others (n=9) tested previous mapping algorithms developed in non-RDs on original patient-level data in RDs. More research is needed to provide a broader spectrum of RDs with mapping algorithms.
- Issues around mapping related to the small and heterogeneous nature of populations impacting on the ability to predict lower/higher HSUVs, to apply complex techniques or perform cross-validation, or on the risk to exclude certain items included in RD-specific PROMs due to lack of statistical power. Limitations in capturing varying levels of disease severity were also identified when applying existing mapping algorithms in non-RDs (e.g. lung cancer) to their more severe rare variants (e.g. pleural mesothelioma).
- Mapping-related limitations in RDs should be addressed, for example, by preliminarily assessing the degree of 'overlap' between RD-specific and generic patient-reported outcome measures, and performing extensive sensitivity analyses on mapped utility values included in economic models.

1. Background

In recent years, there has been an increased focus on placing patients at the centre of clinical research [1]. While traditional outcomes such as survival or biomarkers can demonstrate the physiological effects of treatment, the patient's voice may provide a more holistic assessment of its benefits [1]. A patient-reported outcome (PRO) is any report of a patient's experience with disease and treatment (e.g. symptoms, health status, health-related quality of life, access to care) that comes directly from the patient, without interpretation of the response by a clinician or anyone else [1-3]. Patient-reported outcome measures (PROMs) are the tools developed to measure PROs [4], usually in the form of self-completed questionnaires [1].

Despite their growing use in clinical research, the role of PROs is still unclear in health technology assessments (HTAs) used to inform reimbursement decisions and price negotiations. [5]. When cost per quality-adjusted life year (QALY) approaches are used, they would require a measure of quality of life in the form of health state utility values (HSUVs) derived from public preferences [6]. These can be obtained or derived from the two types of PROMs: generic or disease-specific [1]. The former applies a generic questionnaire applicable to all conditions, some of which have preference weights elicited from the general population and associated with the HSUVs described by different combinations of the tool's responses. The most frequently used tool is the EuroQol 5-Dimension (EQ-5D); other well-known preference-based PROMs are the Short-Form 6-Dimension (SF-6D), the 15D, the Health Utility Index (HUI), the Quality of Wellbeing (QWB) scale, and the Assessment of Quality of Life (AQoL) [7]. In contrast, disease-specific PROMs are designed to identify specific symptoms or change in functioning and quality of life from living with a specific disease. They have greater validity and sensitivity compared to generic PROMs, but are not comparable across conditions and do not have associated HSUVs [1]. In such cases, it is possible to develop or use an existing algorithm, through "mapping", to predict HSUVs from other measures of health outcomes, such as disease-specific PROMs [25].

In Europe, a disease is defined 'rare' when affecting less than 1 in 2,000 people; in the US, the definition applies to conditions affecting less than 200,000 people in total [8; 9]. The overall number of rare diseases (RDs) ranges between 6000 and 8000 [10]; of these, 75% affect children, who die before their fifth birthday in 30% of cases [9]. Patients with RDs usually experience a significant reduction in quality of life, including physical dysfunction and cognitive impairment [9]. Moreover, the rarity of each disease leads to a number of specific issues, such as a paucity of information, unavailability of treatments, delays in diagnosis, and social isolation [11]. In recent years, the European Organization for Rare Disorders (EURORDIS) listed patient's quality of life as a main priority in clinical research on RDs [8]. Thus, PROMs are increasingly adopted in clinical studies to monitor the natural history and progression of a rare condition, and the impact of diseases and treatments on patient's daily life. We are currently seeing a number of PROMs developed for specific RDs such as the phenylketonuria-specific Quality-of-Life questionnaire (PKU-QoL) [12], the Quality of Life

Assessment of Growth Hormone Deficiency in Adults (QoL-AGHDA) [13], and the Cushing's Quality-of-Life (CushingQOL) questionnaire [14].

There are known challenges in using PROMs in RDs [8-10; 15-17]. First, the very nature of RDs affecting small patient populations with heterogeneous clinical manifestation, progression and treatment response makes it difficult to recruit homogeneous samples, thus resulting in a wide range of responses. Second, in multi-centre international studies, which are required to cope with RD patients' geographical dispersion, researchers would need to translate PROMs in different languages and ensure their cross-cultural validity before pooling the data. Third, very few PROMs have been validated in RDs, especially in children and adolescents. Fourth, RDs are often associated with progressive disability and cognitive impairment, or affect vulnerable populations such as young children, which makes it difficult to collect PROMs unless it is done via proxy reporting by parents or caregivers.

The main issue with using generic PROMs in RDs concerns their lack of sensitivity in capturing specific health issues of heterogeneous patients for age, symptomatology, treatment response, and life contexts [6, 8, 10]. For example, in lysosomal storage disorders, even a measure that is specific to such disorders (in general) might not reflect the characteristics of each specific disease [6]. Moreover, generic PROMS might contain irrelevant items for some RDs (for example, in skin conditions including the rare cutaneous lupus erythematosus [18]), thus leading systematically to missing responses. The administration of multiple questionnaires within the same study can also be burdensome to patients, particularly to the more vulnerable populations [18-19].

Some preference-based, disease-specific PROMs have been developed in RDs, such as the Amyotrophic Lateral Sclerosis Utility Index (ALSUI) derived from five items of the Amyotrophic Lateral Sclerosis Functioning Rating Scale-Revised (ALSFRS-R) and based on the US population's preference scores [20-21]. The QLU-C10D is a preference-based PROM recently developed by the European Organisation for Research and Treatment of Cancer (EORTC) that might apply to rare cancers []. However, such measures do not allow comparability of the HSUVs across conditions, and therefore are not acceptable for most HTA agencies [22-23].

As mentioned above, HSUVs are required by several HTA agencies to populate cost-utility models assessing novel treatments. The use of HSUVs is relevant to understand the impact of drugs on quality of life and to compare this across conditions [24]. Since PROM studies in RDs are relatively scarce, and even more so those using preference-based PROMs, the potential for retrieving published HSUVs from the literature is extremely limited. Thus, where an HTA agency has recommended the use of a specific preference-based PROM, it is possible to obtain HSUVs for that PROM using the scores derived from any non-preference-based measure for which a mapping algorithm has been made available. For example, the National Institute for Health and Care Excellence (NICE) allows the use of mapping in the absence of EQ-5D data, which is

the recommended preference-based tool [25]. In other words, mapping 'bridges the gap' between the existing clinical evidence and that required by HTA [27].

In HTA when cost-utility models are used, patient experiences with disease and treatment can be captured through HSUVs. In rare diseases, these aspects are even more crucial to account for when measuring the impact of a treatment given the burdensome nature of these conditions. It is known that generic PROMs may not be sensitive enough for most rare conditions, whereby disease-specific PROMs may be more suitable. As such, there is a need to better understand how mapping from disease-specific PROMs is used in rare diseases. The objectives of this study are: (1) to systematically review all published studies using a mapping approach to derive HSUVs from non-preference-based measures in RDs; (2) to identify any critical issues around the use of mapping in RDs, and give suggestions for addressing them in future studies.

2. Methods

This study followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [28]. Three databases were searched without time, study design or language restrictions including MEDLINE (via PubMed), the School of Health and Related Research Health Utility Database (ScHARRHUD), and the Health Economics Research Centre (HERC) database of mapping studies [29]. The last date for conducting database searches was June 30th, 2019.

In the MEDLINE searches, the keywords combined terms related to 'mapping' with ORPHANET's list of RD indications* (e.g. 'acromegaly'), in addition to the terms 'rare' and 'orphan' (e.g. (((mapp*[Title/Abstract]) OR cross-walk*[Title/Abstract]) OR crosswalk*[Title/Abstract]) AND acromegaly [Title/Abstract]). Since ORPHANET reports 4183 different RDs, we excluded the 'very rare' diseases, e.g. those with less than 1000 cases or families documented in the medical literature worldwide [30]; thus leaving 1059 RD unique denominations for the online searches. All RD indications with an orphan designation from the European Medicines Agency (EMA) were also retrieved (169 in total) and used to search MEDLINE. Additionally, each medicine from the EMA list was searched on the NICE website, and corresponding HTA reports were downloaded (50 in total). The 'utility values' section of each report was reviewed to identify any mapping studies, which were used to cross-check whether these had been captured in the online searches. These additional searches aimed to minimise the risk of missing any mapping studies in RDs, as we assumed mapping would be done in relation to an existing drug and its HTA process. The two lists of disease terms (from ORPHANET and EMA) used for database searching are reported in the appendix (Supplementary Files 1-2). In ScHARRHUD, an ad hoc database on HSUVs, we used only mapping-related terms (i.e., mapp*, map*, cross-walk*, crosswalk*, deriv*, predict*, estimat*) in the abstract, and screened records to find studies on RDs. The HERC database (version 7.0, last update: 24-04-19) was screened by filtering the 'disease or patient group' column to identify any RD study not captured by the online searches. Additionally, we inspected the full-texts of studies whose 'disease category' was classified as 'various' to identify any RDs that could have been included in a mix of conditions. Lastly, we

manually searched the reference list of all eligible studies to avoid missing any relevant publication that may not have included the selected mapping-related terms used in our search.

Two reviewers (MM and AW) screened title and abstracts of the identified citations independently; any disagreement was resolved through further discussion or consultation with a senior author (MD). The same process was repeated with the full-text articles retrieved. The inclusion criteria were as follows: (1) full-text articles, (2) using a mapping approach to derive HSUVs, (3) from any non-preference based measure (e.g. PROM or other), (4) in any RDs reported by ORPHANET or within the EMA list [30]. Studies that recruited both patients with RD and non-RD were included. Editorials, commentaries, or conference abstracts were excluded.

We extracted the data from the included studies using Microsoft Excel®. After a pilot phase with a few studies, we defined the structure of information to extract as follows: study year, disease, country, study design, sample's characteristics, sample size, source and target PROMs, value set for EQ-5D (if applicable), regression techniques, goodness-of-fit measures, and explicit adherence to formal guidelines/recommendations. We also reported the prevalence (per 100,000) for each of the RDs included, as per ORPHANET's estimation.

3. Results

3.1 Literature search

The PRISMA flow diagram displays the process that lead to the selection of 30 mapping studies (Figure 1). The original database searches identified 46329 records; of these, 14960 were removed as duplicates. Four studies were manually identified from the reference lists of the eligible studies, and two from the HERC database, which came to a total of 31375 citations that were screened by title/abstract. 74 were selected for full-text inspection, of which 46 were eliminated because they were: (1) studies about HSUVs in RDs, but did not perform 'mapping'; (2) mapping studies in non-RDs; (3) conference abstracts. No additional studies were identified from the NICE reports. A total of 30 studies were eligible for inclusion.

3.2 Synthesis of included studies

The 28 studies included [15, 18, 20, 24, 31-45, 47-52] were split into two groups: 19 developing novel mapping algorithms in RDs (Table 1), and 9 applying existing algorithms to RD patient-level data (Table 2).

3.2.1 Novel mapping studies in RDs

The first group of studies (n=19) developed original mapping algorithms in 13 different RDs, one of which targeted two different blood cancers (i.e., multiple myeloma and non-Hodgkin's lymphoma) [43]. Four studies [18, 43, XX, XX] also recruited non-RD patient subgroups: one analysed a sample of patients with dermatological conditions including the rare variant lupus erythematosus [18], two others [XX, XX] used a

clinical sample including spinal cord injury and other non-RDs (i.e. heart disease, cancer, rheumatoid arthritis, osteoarthritis, psychiatric disorders, chronic obstructive pulmonary disease, and others), and the last one [43] addressed arthritis, multiple sclerosis and multiple myeloma/non-Hodgkin's lymphoma, but using different PROMs and thus developing separate algorithms for each condition. The range of study countries was quite broad and several studies had multiple locations.

As the source measure, 9 studies adopted RD-specific PROMs such as the Cystic Fibrosis Questionnaire-Revised (CFQ-R), which is a tool for monitoring psychosocial health in cystic fibrosis [15]. Three studies [18, 35, 43] used non-RD-specific PROMs, such as the Dermatology Life Quality Index (DLQI), which is applicable to different skin diseases [XX]. Four studies [20, 36, 40, 42] adopted a mix of RD-specific and non-RD-specific PROMs, including two [36, 40] on multiple myeloma that used both the general and the cancer-specific modules of the EORTC questionnaires assessing quality of life in cancer patients. Two studies [XX, XX] used a generic set of measures (Patient-Reported Outcomes Measurement Information System - PROMIS) evaluating health in the general population and in chronic conditions. Lastly, one study [44] only used a clinical measure (i.e. the Glasgow Outcome Scale, GOS). As target measure, EQ-5D was the most prevalent (n=15, of which only one used the 5-level version), while three studies [35, 39, 41] used the SF-6D, and one mapped to both EQ-5D and 15D. Among those using EQ-5D, 10 studies chose the English tariff, alone or in combination with the Dutch.

In terms of study design, eleven performed ad hoc cross-sectional surveys to collect PRO data for mapping [15, 33, 37, 38], or were secondary analyses of cross-sectional data collected for other purposes [18, 24, 31, 39, 41, XX, XX]. Five [20, 35, 40, 42, 44] used cross-sectional (usually baseline) data from cohort studies, two [36, 43] used longitudinal data collected within randomized controlled trials (RCTs), and one [32] identified three studies (two observational and one cross-sectional) from the literature, and accessed their data. The sample sizes ranged between 111 and 3542 (median value: 401); however, the study reporting the largest sample [18] recruited a mix of patients affected by various dermatological conditions, including rare lupus erythematosus, without specifying the numbers for each. In addition to condition-specific features, 13 studies reported adult age (e.g. above 12, 16 or 18 according to the study) among the inclusion criteria.

The majority of studies adopted the Ordinary Least Squares (OLS) regression, which is the most common approach in the mapping literature [25]. However, the most recent studies also explored more complex techniques by, such as the Bayesian model [36], and the Limited Dependent Variable Mixture Model (LDVMM) [44], which are more flexible in modelling EQ-5D data. Most studies provided summary measures of fit such as (adjusted) R-squared, and measures of predictive performance such as mean error (ME), mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE). Only three studies [15, 20, 44] reported Akaike/Bayesian Information Criteria (AIC/BIC). In general, high levels of error were found at the extremes of the EQ-5D utility scale, because of the tendency of mapping to over-/under-predict HSUVs in patients with very poor/good health [15, 20, 31, 32, 39, 41, 43, 44].

Only five studies explicitly embraced published recommendations in the field, including the MAPS Statement [46, XX] (n=2), ISPOR good practices [27] (n=1) and the study by Longworth [25] (n=2). However, five [33, 37, 38, 43, XX] were published before any guidelines for mapping were available.

3.2.2 Studies applying previous mapping to RDs data

The second group of studies (n=9) applied previous algorithms, usually retrieved from multiple studies, to original RD data. The studies addressed nine different RDs, all oncological except for Castleman's disease. One study [48] focused on two non-rare cancers (i.e. breast cancer and non-small cell lung cancer) as well as multiple myeloma/non-Hodgkin lymphoma, but used separate datasets for each condition. All of the included studies were RCTs, except for one prospective cohort study [47] and one review [XX]. Four studies [50-52, XX] were intercontinental, and all recruited small samples (<1000), except for one [XX] that used a dataset of mixed rare and non-rare cancers.

The studies applied previously published algorithms with four different purposes: (1) testing their external validity in a different database from the one used for mapping [47-48], (2) identifying the best available algorithms for a specific condition (e.g. ovarian cancer) [49-50], (3) deriving HSUVs for economic evaluation alongside RCTs [51-52] or for economic models [XX, XX], and (4) testing the comparability of mapped HSUVs with those derived from disease-specific preference-based PROMs (e.g. EORTC-8D) and generic PROMs (e.g. EQ-5D) [XX].

The six studies referred to twelve different original mapping studies (Table 2), all developed in oncology with the exception of one that recruited all patients that were referred to a large university hospital [19]. Since all twelve were conducted in non-RDs, no RD-specific PROMs were used as the source measure. Eight of the original mapping studies [53-60] were mapped from the EORTC QLQ-C30, which is a questionnaire widely used in oncology, two [61-62] used another popular cancer-specific tool (FACT-G), and one [63] used both; one study [19] used the SF-36, which is a generic non-preference-based tool. As a target measure, ten studies mapped onto the EQ-5D-3L, one to EQ-5D-3L, 6F-6D and 15D [57], and one [61] to time trade-off (TTO) utilities.

Difficulties in applying published algorithms to RD data were reported. First, most algorithms, and especially those using OLS, tend to under-predict HSUVs in poor health states (e.g. EQ-5D utility less than 0.5), or to predict values greater than 1 (e.g. outside the possible EQ-5D range) [47, 48, 51, XX]. The study testing the highest number of algorithms [47] showed that the response mapping approach performed best [63]. The application of more sophisticated techniques (e.g. mixture models) is encouraged to improve accuracy in mapping [48]. Second, the algorithms published were usually developed in populations that are very different from RD patients [52, XX]. Thus, some items that are relevant in RDs might not be significant in the algorithms applied. For example, dyspnoea is an important symptom in pleural mesothelioma; however, six (out of nine) of the algorithms tested by Arnold et al [47] did not include a dyspnoea score,

which would have led to an overestimation of HSUVs in patients with mesothelioma. Third, different 'source' PROM versions and/or EQ-5D value sets in the original mapping studies might affect mapped HSUVs [48, XX], although differences between questionnaire versions are generally small and EQ-5D tariffs have shown to be quite similar across European countries [48].

3.3 Critical issues around 'mapping' in RDs

The 25 reviewed studies highlighted a number of challenges around the existing mapping literature in RDs (Figure 2).

The first relates to the amount and context of existing literature on mapping applied to RDs. The searches identified only few studies (25 in 18 different RDs) compared to over 1000 RDs screened. Many of these were cancer studies (8/25), most of which came from the second group of studies applying published algorithms to original RD data. No mapping studies were performed in childhood diseases, and 68% of the studies retrieved comprised age (e.g. minimum 12 years old) as inclusion criteria, thus excluding paediatric patients.

The second issue relates to sample size, which were commonly small (<1000 patients in 13/19 novel mapping studies). This was because of the challenges to identify RDs patients [20, 31, 32, 41]. The limited size of recruited samples may affect the robustness of mapping coefficients. Moreover, the widely adopted OLS regression [] may not work for EQ-5D data in general [], and particularly in RDs, which are likely to present multimodality and peaks due to the high patients' heterogeneity in clinical manifestations, disease severity and quality of life impact even within the same condition [10, 15]. However, a small study's size hampers the applicability of techniques such as LDVMM dealing with multimodality, gaps, asymmetries and peaks, as recommended by the most recent guidelines for modelling EQ-5D data [27]. Lastly, small sample size makes it difficult to perform cross-validation tests in patient subgroups with peculiar characteristics such as disease subtypes or undertaking specific treatments [41]. The availability of external datasets to validate the developed algorithms may be lacking as well, due to the general paucity of experimental and observational studies in RDs.

Third, the accuracy of the mapping algorithms was reduced by the limited number of recruited patients in very poor health, which resulted in large errors at the lower bounds of the HSUVs distribution. For example, in the study by Acaster [15], only 3% of the recruited patients reported EQ-5D utilities below zero; the study by Kalaitzakis [35] on primary sclerosing cholangitis had a very small sub-sample of patients with end-stage liver disease, and encouraged future external algorithm validation in hospital-based samples. Similarly, another study [44] included six cases only with vegetative state in a sample of patients with traumatic brain injury. This results in challenges to accurately predict at the lowest scale's extremes, whereby existing algorithms tend to be more relevant for the 'average' RD patients, who are likely to be the target of clinical trials assessing new drugs, but less of use to predict individual patient-level HSUVs [33].

Fourth, generic preference-based PROMs such as EQ-5D might not capture relevant health issues in RDs. For example, one study [15] revealed that the CFQ-R respiratory domain, which is essential in monitoring cystic fibrosis's patients, was not a significant predictor of HSUVs. Similar findings were found in Cushing's syndrome, where some items of the CushingQol were excluded from the mapping models since they did not influence HSUVs [31]. This similarity was also seen in systemic lupus erythematosus (LupusQol dimensions omitted: planning, intimate relationships, burden to others, body image) [39], and motor neuron disease (ALSFRS-R domains omitted: communication, salivation, swallowing, hand use, respiratory function) [20]. Overall, it is rather frequent in the mapping literature that certain symptoms may not affect HSUV prediction [15]. This might be even more frequent in RDs where the symptomatology may be extremely specific.

Fifth, while the majority of the algorithms were mapped to EQ-5D, which is the mostly recommended preference-based PROM by HTA agencies worldwide, three studies [35, 39, 41] used the SF-6D. One study [42] used both EQ-5D and 15D and gave results slightly in favour of the latter, which appears to be more sensitive in capturing the perceived health of patients with chronic pain. However, agencies such as NICE in UK might not accept HSUVs derived from instruments other than EQ-5D.

Sixth, in RDs, trials are often multi-country and even multi-continental because of small patient numbers and patients' geographical dispersion. This may result in substantial inter-country heterogeneity in terms of socio-cultural attitudes, linguistic factors and wellbeing's perceptions that influence responses to PROMs [32, 33]. In particular, psychological domains are more likely to be affected by cultural differences and mentality compared to physical functioning [37]. For example, people living in some countries may be less willing to report anxiety/depression on the EQ-5D, thus the inclusion/exclusion of a country-specific sample would alter the coefficient of items related to the same concept in the source PROM [31]. This may represent a particular issue in RDs where the incidence of mental disorders is significantly higher than in other conditions due to misdiagnosis, social isolation, and financial distress [64]. Moreover, the presence of multiple countries within the same survey makes it difficult to select the best tariff for valuing EQ-5D, considering also that there may be no country-specific tariff available for several countries [31], especially in studies using the 5-level version.

Lastly, the generalizability of mapping algorithms developed in non-RDs to RDs is not straightforward. As anticipated before, some included studies [47, 48, XX] highlighted the tendency of the available algorithms, and especially those using OLS, to over-predict HSUVs in patients with poor health. Since rare variants of a condition are usually more severe than the condition itself (e.g. pleural mesothelioma vs. lung cancer), this type of error is particularly critical in RDs. Similar considerations apply to mapping studies developed in a range of conditions including RDs. This might help in increasing sample sizes for mapping, but raises concerns about multi-disease algorithms' ability to provide accurate estimates of HSUVs [18].

4. Discussion

This study reviewed all published studies mapping non-preference-based measures onto generic preference-based ones in RDs. Two systematic reviews by Dakin et al. [29, 65] provided a synthesis of all mapping studies from any clinical measure or PROMs in any condition, based on which the HERC database was established and is routinely updated. However, EQ-5D was the only 'target' measure considered, while we aimed to identify studies mapping to any preference-based measure in RDs. Our searches captured those studies identified in the HERC database, and three additional studies [35, 39, 41] using SF-6D. A previous review published in 2010 [66] identified 28 studies mapping non-preference-based PROMs onto any preference-based ones in any disease area; however, 23/25 of the studies included in our review were not captured as published after their search end date. Three more records [33, 37, 38] that did not use 'mapping' or 'cross-walking' (e.g. keywords adopted by Dakin et al. [29] and Brazier et al. [66]) were identified by reviewing the reference lists from included studies. Lastly, by focusing on single disease areas (e.g. individual RDs as search terms), we could expand the eligibility criteria to include also studies not developing original mapping but applying existing algorithms to original datasets, which gave insights onto the generalizability of mapping developed in similar non-RDs.

We also identified some published systematic reviews on quality of life and health utilities in individual RDs. The study by Forsythe [17] identified ten relevant studies reporting on quality of life and HSUVs in acute myeloid leukaemia, of which only one [XX] used a mapping approach (and was included in this review), and two [XX; XX] collected EORTC QLQ-C30 data that were converted into HSUVs by the review's authors using a published algorithm. Overall, this review recommended estimating HSUVs with larger sample sizes, in addition to routine monitoring (including PROMs administration) of patients with RDs and establishment of a set of recommendations to standardize HSUV elicitation across different RDs. Another systematic review on cystic fibrosis [67] identified only one study mapping from the CFQ-R onto EQ-5D-3L [15], which was consistent with our findings, and encouraged the development of future mapping studies to inform health economic modelling in this rare condition. Lastly, two systematic reviews [10, 68] were identified in Cushing's syndrome and included the same two mapping studies [31, 41] captured by our review. One review [68] stated that mapping HSUVs from the CushingQol questionnaire is possible, at least at group level, although further testing in independent patient samples is required; the other [10] highlighted the utility of mapping in deriving both EQ-5D and SF-6D utilities from the CushingQol.

This systematic review identified 30 studies using mapping in RDs (e.g. as per EMA's definition of RDs [8, 9]). Of these, 21 developed original algorithms with RD data, and 9 applied previous algorithms developed in different conditions to RD data, with the purpose of testing their external validity or obtaining HSUVs for cost-utility analyses of new treatments. In the second group, 8 out of 9 studies were performed in oncology likely due to the substantial amount of published algorithms in common neoplasms that may be suitable for rare cancers as well. Although it is not surprising to find very few studies in RDs, the proportion of RDs with a mapping algorithm is very low compared to the list of conditions screened in the database searches (e.g. around 1.4%). Also, there are no mapping studies in paediatric RDs (e.g. neuroblastoma), likely due to the

existence of more legal and ethical requirements to enrol children in clinical studies, and the lack of RD-specific PROMs intended for child self- (or parental proxy-) reporting.

Moreover, only five studies [15, 20, 39, 44, XX] explicitly embraced published recommendations in the mapping field, and only one [44] referred to the most recent guidelines (i.e., ISPOR good practices [27]). These, for example, recommend reporting AIC/BIC for model selection instead of summary measures of fit as the R-squared that is considered to provide limited information on the validity of mapping. In addition, they encourage considering alternatives to OLS for modelling HSUVs. Indeed, while most mappings studies still use OLS regression, the mapping literature has recently proposed alternative techniques such as beta-binomial regression [Khan 2014] and mixture models (e.g., ALDVMM [Alava 2012]). OLS, indeed, can accurately predict mean HSUVs, because its underestimation of high values is compensated by its overestimation of low values [69], but is not appropriate to estimate HSUVs at individual patient-level, which is particularly relevant in RDs due to the heterogeneity of patient populations. Moreover, ISPOR guidelines state that splitting a database for algorithms' internal validation may not be the right choice in case sample splitting further reduces sample size for estimation, and small sample is an issue that often affects mapping studies in RDs. However, previous recommendations [25] have suggested that the lower precision in coefficients of mapping algorithms from the reduction in sample size might be overcome by re-estimating the models using the full data set, once the preferred model has been selected by using the split-sample approach.

We also identified a number of critical issues around mapping in RDs that arose from the studies reviewed. Despite most of these challenges also apply to non-RDs, they might be particularly critical in RDs. For example, RCTs in RDs are more likely to involve multiple countries to deal with patients' dispersion and to increase study sample size, leading to inter-country variability in terms of language and culture that may affect HSUV estimates. Thus, it is recommended to use PROMs with validated translations and possibly showing consistent results across countries, and/or to include study's site as predictor in mapping models [32]. Moreover, although EQ-5D tariffs look quite similar across European countries [48], significant differences may arise in inter-continental studies. Thus, one could select the EQ-5D value set of the country with the biggest sample size [35], or use a weighted value set with weights derived from the relative country sample sizes, but the impact of using alternative sets on model coefficients should still be tested [18].

Moreover, clinical studies in RDs are obviously small because of the disease's rarity and consequent difficulties in recruiting representative patients. Small samples may affect the robustness of mapping coefficients and increase the risk of prediction errors, especially at the 'extremes' of EQ-5D distribution. However, the distance between actual and predicted HSUVs should be compared with the minimally important difference (MID), which is not yet established for all conditions [35]. Some studies reported MIDs between 0.033 and 0.082 in non-cancer conditions [50], which might be used as a reference for most RDs. One study [32] pooled data from multiple studies to increase the study's power; however, this might

introduce extra heterogeneity within the database, and attention should be given to inclusion criteria's comparability. Another study [43] pooled data across different time points of a RCT, although loss to follow-up raises the possibility of selection bias [20, 44].

A further issue relates to the selection of 'target' PROMs for mapping. Several items included in RD-specific PROMs turn out to be unrelated to generic PROMs. In a recent survey, 60% of a sample of patients with RDs revealed that important health issues such as 'fatigue', 'social life' and 'comorbidities' were not captured by EQ-5D-5L [70]. Thus, the degree of 'overlap' between 'source' and 'target' measures should be assessed in advance using proper correlation tests (e.g., Pearson) in order to foresee the mapping's predictive ability [18, 43]. Moreover, the lack of interchangeability among different preference-based measures should be considered when using algorithms mapping to SF-6D or 15D to derive HSUVs in HTA processes requiring EQ-5D [39, 48].

Lastly, the generalizability of mapping algorithms developed in similar non-RDs to RD datasets needs careful attention. Overall, there is a tendency for mapping to overestimate HSUVs in poor health states. This should be considered when applying existing algorithms to RDs, since the rare variant or evolution of a condition (e.g., AIDS wasting syndrome) is usually more severe than the condition itself (e.g., HIV), where it is more likely that a mapping study is available [71]; for example, five-year survival in rare cancers is much lower than in common cancers (47% vs. 65%, source RARECAREnet). As a rule, the estimation population must be as similar as possible to the RD population in order to minimize prediction errors. For example, in the study by Arnold [47] on pleural mesothelioma, the algorithm by Jang [53] presented a good performance likely due to being developed in a comparable disease (i.e., non-small cell lung cancer). Similarly, in the study by Rowen [XX], the worst performing algorithm [54] was that developed in a breast cancer population, which little resembled the multiple myeloma dataset adopted. Moreover, identical 'source' PROM versions should be used in mapping algorithms and RD studies using them [48]. A justification for choosing a specific algorithm, such as its external validation or wide adoption in HTA, should be given when many competing algorithms exist, as it is for the numerous mapping studies in oncology that are potentially applicable to rare cancers. However, the preferred algorithm might not generate the most plausible HSUVs, thus it is essential to evaluate the impact of using different algorithms on predicted HSUVs and related cost-utility results [50-51]. Overall, careful sensitivity analyses should be performed when cost-utility models are populated with mapped HSUVs [31, 32, 49, 50].

The study also presents some limitations. First, the list from ORPHANET was screened until a prevalence/incidence of around 0.01 per 100,000, thus excluding the 'very rare' diseases for which less than 1000 cases (or families) are documented in the literature. This resulted in searching around one fourth of the RDs reported by ORPHANET. However, terms related to very RDs for which an orphan drug is available (resulting from the EMA website search) were used for database searching. Overall, very RDs are extremely unlikely to be provided with mapping algorithms because of the difficulty in recruiting sufficiently large

samples, the lack of disease-specific PROMs, and the lack of treatments that incentivize the estimation of HSUVs for HTA. Moreover, by searching title/abstract in PubMed, we are likely to have missed some studies applying existing mapping algorithms, but mentioning them only in the full-text. We also did not include other databases (e.g., EMBASE) that might have provided more eligible studies. Second, a quality assessment of the included studies was not performed due to the absence of checklists developed for this purpose, and we just reported the number of studies explicitly referring to any guidelines in the mapping field (e.g., the MAPS Statement [46]). Third, since mapping is considered as a 'second best' solution [25, XX], there might be studies in RDs that collected HSUVs directly from patients using generic preference-based PROMs or direct elicitation techniques (e.g., TTO). There might also be studies that applied HSUVs estimated for other similar non-RDs to the RDs in question. Although such studies were beyond the scope of this review, we used those retrieved by the online searches to identify any additional records and to inform the interpretation of the review results.

5. Conclusion

This systematic review synthetized the available mapping literature on RDs to derive HSUVs. New mapping algorithms should be developed to cover a much broader spectrum of RDs, including childhood diseases such as Duchenne muscular dystrophy, Gaucher disease or neuroblastoma, to inform the reimbursement decisions for new drugs. This would allow exploiting the usability of RD-specific PROMs for HTA purposes beyond clinical assessment. In addition, the quality of future studies should be enhanced by following up-to-date recommendations in the field. An incentive to perform high-quality mapping studies, especially in the UK, could come from NICE recommending undertaking mapping in the absence of EQ-5D data collection in clinical trials. However, the limitations of using such approach to derive HSUVs should be acknowledged in HTA processes, and possibly overcome by future research.

References

1. Weldring T, Smith SM. Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs). Health Serv Insights. 2013; 6:61-8.
2. U.S Department of Health and Human Services Food and Drug Administration. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labelling Claims (Guidance for Industry). U.S. FDA, Clinical/Medical 2009. Available at: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf.
3. Deshpande PR, Rajan S, Sudeepthi BL, Abdul Nazir CP. Patient-reported outcomes: A new era in clinical research. Perspect Clin Res. 2011; 2(4):137-44.
4. Kingsley C, Patel, S. Patient-reported outcome measures and patient-reported experience measures. BJA Education 2017; 17 (4); 137–144.
5. Brogan A, DeMuro C, Barrett AM, et al. Payer Perspectives on Patient-Reported Outcomes in Health Care Decision Making: Oncology Examples. J Manag Care Spec Pharm, 2017; 23(2):125-134.
6. Rüther A, Elstein D, Wong-Rieger D, Guyatt G. Aspects of patient reported outcomes in rare diseases: a discussion paper. Int J Technol Assess Health Care. 2016; 32(3):126-30.
7. Richardson J, Khan MA, Iezzi A, Maxwell A. Comparing and explaining differences in the magnitude, content, and sensitivity of utilities predicted by the EQ-5D, SF-6D, HUI3, 15D, QWB, and AQoL-8D multi-attribute utility instruments. Med Decis Mak. 2015; 35:276–291.
8. Benjamin K, Vernon MK, Patrick DL, et al. 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; 20(7):838-855.
9. Slade A, Isa F, Kyte D, et al. Patient reported outcome measures in rare diseases: a narrative review. Orphanet J Rare Dis. 2018; 13(1):61.
10. Knoble N, Nayroles G, Cheng C, Arnould B. llustration of patient-reported outcome challenges and solutions in rare diseases: a systematic review in Cushing's syndrome. Orphanet J Rare Dis. 2018; 13(1):228.
11. Heuyer T, Pavan S, Vicard C. The health and life path of rare disease patients: results of the 2015 French barometer. Patient Relat Outcome Meas. 2017; 8:97-110.
12. Regnault A, Burlina A, Cunningham A, et al. Development and psychometric validation of measures to assess the impact of phenylketonuria and its dietary treatment on patients' and parents' quality of life: the phenylketonuria - quality of life (PKU-QOL) questionnaires. Orphanet J Rare Dis. 2015; 10:59. doi: 10.1186/s13023-015-0261-6.
13. McKenna SP, Doward LC, Alonso J, et al. The QoL-AGHDA: An instrument for the assessment of quality of life in adults with growth hormone deficiency. Qual Life Res. 1999; 8(4):373-83.
14. Nelson LM, Forsythe A, McLeod L, et al. Psychometric evaluation of the Cushing's Quality-of-Life questionnaire. Patient. 2013; 6(2):113-24.
15. Acaster S, Pinder B, Mukuria C, Copans A. Mapping the EQ-5D index from the cystic fibrosis questionnaire-revised using multiple modelling approaches. Health Qual Life Outcomes. 2015; 13: 33.
16. Basch E, Bennett AV. Patient-reported outcomes in clinical trials of rare diseases. J Gen Intern Med. 2014; 29 Suppl 3:S801-3.
17. Forsythe A, Brandt PS, Dolph M, et al. Systematic review of health state utility values for acute myeloid leukemia. Clinicoecon Outcomes Res. 2018; 10:83-92.
18. Ali FM, Kay R, Finlay AY, et al. Mapping of the DLQI scores to EQ-5D utility values using ordinal logistic regression. Qual Life Res. 2017; 26:3025–3034.
19. Rowen D, Brazier J, Roberts J. Mapping SF-36 onto the EQ-5D index: how reliable is the relationship? Health Qual Life Outcomes 2009; 7:27.
20. Moore A, Young CA, Hughes DA. Mapping ALSFRS-R and ALSUI to EQ-5D in Patients with Motor Neuron Disease. Value Health. 2018; 21(11):1322-1329.

21. Beusterien K, Leigh N, Jackson C, et al. Integrating preferences into health status assessment for amyotrophic lateral sclerosis: the ALS Utility Index. Amyotroph Lateral Scler Other Motor Neuron Disord. 2005; 6(3):169-76.
22. Rowen D, Brazier J, Ara R, Azzabi Zouraq I. The Role of Condition-Specific Preference-Based Measures in Health Technology Assessment. Pharmacoeconomics. 2017; 35 (Suppl 1):33-41.
23. Versteegh MM, Leunis A, Uyl-de Groot CA, Stolk EA. Condition-specific preference-based measures: benefit or burden? Value Health. 2012; 15(3):504-13.
24. Aygören-Pürsün E, Bygum A, Beusterien K, et al. Estimation of EuroQol 5-Dimensions health status utility values in hereditary angioedema. Patient Prefer Adherence. 2016; 10: 1699–1707.
25. Longworth L, Rowen D. Mapping to obtain EQ-5D utility values for use in NICE health technology assessments. Value Health. 2013; 16(1):202-10.
26. Longworth L, Rowen D. NICE DSU Technical Support Document 10: The Use of Mapping Methods to Estimate Health State Utility Values [Internet]. NICE Decision Support Unit Technical Support Documents. 2011.
27. Wailoo AJ, Hernandez-Alava M, Manca A, et al. Mapping to Estimate Health-State Utility from Non-Preference-Based Outcome Measures: An ISPOR Good Practices for Outcomes Research Task Force Report. Value Health. 2017; 20(1):18-27.
28. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 2009; 6(7):e1000097.
29. Dakin H, Abel L, Burns R, Yang Y. Review and critical appraisal of studies mapping from quality of life or clinical measures to EQ-5D: an online database and application of the MAPS statement. Health Qual Life Outcomes. 2018; 16(1):31.
30. Prevalence of rare diseases: bibliographic data. Orphanet Report Series, Rare Diseases Collection, June 2018, Number 2: Diseases list by decreasing prevalence, incidence, or number of published cases. Available at: http://www.orpha.net/orphacom/cahiers/docs/GB/Prevalence_of_rare_diseases_by_decreasing_prevalence_or_cases.pdf.
31. Badia X, Roset M, Valassi E, et al. Mapping CushingQOL scores to EQ-5D utility values using data from the European Registry on Cushing's syndrome (ERCUSYN). Qual Life Res. 2013; 22(10): 2941–2950.
32. Badia X, Trainer P, Biermasz NR, et al. Mapping AcroQoL scores to EQ-5D to obtain utility values for patients with acromegaly. J Med Econ. 2018; 21(4):382-389.
33. Busschbach JJ, Wolffenbuttel BH, Annemans L, et al. Deriving reference values and utilities for the QoL-AGHDA in adult GHD. Eur J Health Econ. 2011; 12(3):243-52.
34. Gu NY, Bell C, Botteman MF, et al. Estimating preference-based EQ-5D health state utilities or item responses. Patient 2012; 5(3): 185-197.
35. Kalaitzakis E, de Valle MB, Rahman M, et al. Mapping chronic liver disease questionnaire scores onto SF-6D utility values in patients with primary sclerosing cholangitis. Qual Life Res. 2016; 25: 947-957.
36. Kharroubi SA, Edlin R, Meads D, et al. Use of Bayesian Markov chain Monte Carlo methods to estimate EQ-5D utility scores from EORTC QLQ data in myeloma for use in cost-effectiveness analysis. Med Decis Making. 2015; 35(3):351-60.
37. Kołtowska-Häggström M, Jonsson B, Isacson D, Bingefors K. Using EQ-5D to derive general population-based utilities for the quality of life assessment of growth hormone deficiency in adults (QoL-AGHDA). Value Health. 2007; 10(1):73-81.
38. Kołtowska-Häggström M, Kind P, Monson JP, Jonsson B. Growth hormone (GH) replacement in hypopituitary adults with GH deficiency evaluated by a utility-weighted quality of life index: a precursor to cost-utility analysis. Clin Endocrinol (Oxf). 2008; 68(1):122-9.
39. Meacock R, Harrison M, McElhone K, et al. Mapping the disease-specific LupusQoL to the SF-6D. Qual Life Res. 2015; 24(7):1749-58.
40. Proskorovsky I, Lewis P, Williams CD, Jordan K, et al. Mapping EORTC QLQ-C30 and QLQ-MY20 to EQ-5D in patients with multiple myeloma. Health Qual Life Outcomes. 2014; 12:35.
41. Roset M, Badia X, Forsythe A, Webb SM. Mapping CushingQoL scores onto SF-6D utility values in patients with Cushing's syndrome. Patient. 2013; 6(2):103-11.
42. Vartiainen P, Mäntyselkä P, Heiskanen T, et al. Validation of EQ-5D and 15D in the assessment of health-related quality of life in chronic pain. Pain. 2017; 158(8):1577-1585.

43. Versteegh MM, Leunis A, Luime JJ, et al. Mapping QLQ-C30, HAQ, and MSIS-29 on EQ-5D. Med Decis Making. 2012; 32(4):554-68.
44. Ward Fuller G, Hernandez M, Pallot D, et al. Health State Preference Weights for the Glasgow Outcome Scale Following Traumatic Brain Injury: A Systematic Review and Mapping Study. Value Health. 2017; 20(1):141-151.
45. Wijnen BFM, Mosweu I, Majoie MHJM, et al. A comparison of the responsiveness of EQ-5D-5L and the QOLIE-31P and mapping of QOLIE-31P to EQ-5D-5L in epilepsy. Eur J Health Econ. 2018; 19(6):861-870.
46. Petrou S, Rivero-Arias O, Dakin H, et al. Preferred reporting items for studies mapping onto preference-based outcome measures: The MAPS statement. J Med Econ. 2015; 18(11):851-7.
47. Arnold DT, Rowen D, Versteegh MM, et al. Testing mapping algorithms of the cancer-specific EORTC QLQ-C30 onto EQ-5D in malignant mesothelioma. Health Qual Life Outcomes. 2015; 13:6.
48. Crott R, Versteegh M, Uyl-de-Groot C. An assessment of the external validity of mapping QLQ-C30 to EQ-5D preferences. Qual Life Res. 2013; 22(5):1045-54.
49. Hess LM, Brady WE, Havrilesky LJ, et al. Comparison of methods to estimate health state utilities for ovarian cancer using quality of life data: a Gynecologic Oncology Group study. Gynecol Oncol. 2013; 128(2):175-80.
50. Hettle R, Borrill J, Suri G, Wulff J. Estimating health-state utility values for patients with recurrent ovarian cancer using Functional Assessment of Cancer Therapy - General mapping algorithms. Clinicoecon Outcomes Res. 2015; 7:615-27.
51. Meng Y, McCarthy G, Berthon A, Dinet J. Patient-reported health state utilities in metastatic gastroenteropancreatic neuroendocrine tumours - an analysis based on the CLARINET study. Health Qual Life Outcomes. 2017; 15(1):131.
52. Vernon M, Robinson D Jr, Trundell D, et al. Deriving health utility values from a randomized, double-blind, placebo-controlled trial of siltuximab in subjects with multicentric Castleman's disease. Curr Med Res Opin. 2016; 32(7):1193-200.
53. Jang RW, Isogai PK, Mittmann N, et al. Derivation of utility values from European Organization for Research and Treatment of Cancer Quality of Life-Core 30 questionnaire values in lung cancer. J Thorac Oncol. 2010; 5(12):1953–7.
54. Crott R, Briggs A. Mapping the QLQ-C30 quality of life cancer questionnaire to EQ-5D patient preferences. Eur J Health Econ. 2010; 11(4):427–34.
55. Kim EJ, Ko SK, Kang HY. Mapping the cancer-specific EORTC QLQ-C30 and EORTC QLQ-BR23 to the generic EQ-5D in metastatic breast cancer patients. Qual Life Res. 2012; 21(7):1193–203.
56. Kim SH, Jo MW, Kim HJ, Ahn JH. Mapping EORTC QLQ-C30 onto EQ-5D for the assessment of cancer patients. Health Qual Life Outcomes. 2012; 10:151.
57. Kontodimopoulos N, Aletras VH, Paliouras D, Niakas D. Mapping the cancer-specific EORTC QLQ-C30 to the preference-based EQ-5D, SF-6D, and 15D instruments. Value Health. 2009; 12(8):1151–7.
58. McKenzie L, van der Pol M. Mapping the EORTC QLQ C-30 onto the EQ-5D instrument: the potential to estimate QALYs without generic preference data. Value Health. 2009; 12(1):167–71.
59. Proskorovsky I, Lewis P, Williams CD, et al. Mapping EORTC QLQ-C30 and QLQ-MY20 to EQ-5D in patients with multiple myeloma. Health Qual Life Outcomes. 2014; 12:35.
60. Versteegh MM, Leunis A, Luime JJ, et al. Mapping QLQ-C30, HAQ, and MSIS-29 on EQ-5D. Med Decis Making. 2012; 32(4):554–68.
61. Dobrez D, Cella D, Pickard AS, et al. Estimation of patient preference-based utility weights from the functional assessment of cancer therapy - general. Value Health. 2007; 10(4):266-72.
62. Cheung YB, Thumboo J, Gao F, et al. Mapping the English and Chinese versions of the Functional Assessment of Cancer Therapy-General to the EQ-5D utility index. Value Health. 2009; 12(2):371-6.
63. Longworth L, Yang Y, Young T, et al. Use of generic and condition-specific measures of health-related quality of life in NICE decision-making: a systematic review, statistical modelling and survey. Health Technol Assess 2014; 18:1–224.
64. Nunn R. “It’s not all in my head!” - The complex relationship between rare diseases and mental health problems. Orphanet J Rare Dis. 2017; 12: 29. doi: 10.1186/s13023-017-0591-7.

65. Dakin H. Review of studies mapping from quality of life or clinical measures to EQ-5D: an online database. Health Qual Life Outcomes. 2013; 11:151.
66. Brazier JE, Yang Y, Tsuchiya A, Rowen DL. A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. Eur J Health Econ. 2010; 11(2):215-25.
67. Mohindru B, Turner D, Sach T, et al. Health State Utility Data in Cystic Fibrosis: A Systematic Review. Pharmacoecon Open. 2019. doi: 10.1007/s41669-019-0144-1.
68. Santos A, Crespo, Aulinas A, et al. Quality of life in Cushing's syndrome. Pituitary. 2015; 18(2):195-200.
69. Crott R. Direct Mapping of the QLQ-C30 to EQ-5D Preferences: A Comparison of Regression Methods. Pharmacoecon Open. 2018; 2(2): 165–177.
70. 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(2):159-165.
71. Joyce VR, Sun H, Barnett PG, et al. Mapping MOS-HIV to HUI3 and EQ-5D-3L in Patients With HIV. MDM Policy Pract. 2017; 2(2):2381468317716440.
72. 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; 12(1):171.