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

SYSTEMATIC REVIEW

'Mapping' Health State Utility Values from Non-preference-Based Measures: A Systematic Literature Review in Rare Diseases

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

Abstract

Background The use of patient-reported outcome measures (PROMs) to monitor the effects of disease and treatment on patient symptomatology and daily life is increasing in rare diseases (RDs) (i.e. those affecting less than one in 2000 people); however, these instruments seldom yield health state utility values (HSUVs) for cost-utility analyses. In such a context, 'mapping' allows HSUVs to be obtained by establishing a statistical relationship between a 'source' (e.g. a disease-specific PROM) and a 'target' preference-based measure [e.g. the EuroQol-5 Dimension (EQ-5D) tool].

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

Methods The following databases were searched during the first half of 2019 without time, study design, or language restrictions: MEDLINE (via PubMed), the School of Health and Related Research Health Utility Database (ScHARRHUD), and the Health Economics Research Centre (HERC) database of mapping studies (version 7.0). The keywords combined terms related to 'mapping' with Orphanet's list of RD indications (e.g. 'acromegaly') in addition to 'rare' and 'orphan'. 'Very rare' diseases (i.e. those with fewer 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 13 different RDs. As a target measure, the majority used EQ-5D, and the others used 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 published algorithms in non-RDs to comparable RDs, mainly in the field of cancer. The critical issues relating to 'mapping' in RDs included the availability of very few studies, the relatively high number of cancer studies, and the absence of research in paediatric RDs. Moreover, the reviewed studies recruited small samples, showed a limited overlap between RD-specific and generic PROMs, and highlighted the presence of cultural and linguistic factors influencing results in multi-country studies. Lastly, the application of existing algorithms developed in non-RDs tended to produce inaccuracies at the bottom of the EQ-5D scale, due to the greater severity of RDs.

Conclusions More research is encouraged to develop algorithms for a broader spectrum of RDs (including those affecting young children), improve mapping study quality, test the generalisability 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.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s40273-020-00897-4) contains supplementary material, which is available to authorized users.

Extended author information available on the last page of the article

Published online: 10 March 2020

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. EuroQol-5 Dimension (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.

A number of critical issues emerged from these studies in relation to the main features of RDs, which typically affect vulnerable, heterogeneous, and isolated patient populations. In particular, the applicability of existing mapping algorithms in non-RDs (e.g. lung cancer) is hampered by the greater severity of their 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]. Indeed, 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]. According to the US Food and Drug Administration (FDA), 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 importance in clinical research, the use of PROMs in health technology assessment (HTA) to inform drug pricing and reimbursement is still inconsistent [5]. In jurisdictions using the cost per quality-adjusted life year (QALY) approach, most PROMs are unsuitable for HTA since they do not provide health state utility values (HSUVs) based on public preferences [6]. PROMs, indeed, can be either generic or disease specific [1]. The former are usually composed of a generic questionnaire, and some are provided with a set of preference weights elicited from the general population and associated with the health states 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 Six-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 and their impact on the patient's functioning and quality of life, thus presenting greater validity and sensitivity compared to generic PROMs; however, they do not allow cost-utility comparisons across a variety of conditions [1].

In Europe, a disease is defined 'rare' when it affects less than one in 2000 people; in the USA, the definition applies to conditions affecting fewer 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 Organisation for Rare Disorders (EURORDIS) listed patients' 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 patients' daily life. So far, a number of PROMs have been developed specifically for 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 the specific health issues of heterogeneous patients regarding 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 individual 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 frequently leading to missing responses. The administration of multiple questionnaires within the same study can also be burdensome to patients, particularly to the more vulnerable ones [18, 19].

Some preference-based, disease-specific PROMs have been developed in RDs, such as the Amyotrophic Lateral Sclerosis Utility Index (ALSUI), which is 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 [22]. However, such measures do not allow comparability of the HSUVs across conditions, and therefore are not acceptable for most HTA agencies [23, 24].

As mentioned above, HSUVs are required by several HTA agencies to populate cost-utility models assessing novel treatments that, in RDs, are usually characterised by high costs. Moreover, the availability of HSUVs for a range of RDs is relevant to raise awareness of their quality-of-life impact, provide a common basis for comparison with other conditions, and incentivise the development of new therapeutic options [25]. Since studies on RDs are relatively rare overall, and even more so those using preference-based PROMs, the potential for retrieving published HSUVs from the literature is extremely limited. In this context, 'mapping' has been defined as the development and use of an algorithm to predict HSUVs by using data from other measures of health outcomes, such as disease-specific PROMs [26]. 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 [26]. In other words, mapping 'bridges the gap' between the existing clinical evidence and that required by HTA [27].

The objectives of this study are to (1) systematically review all published studies using a mapping approach to derive HSUVs from non–preference-based measures in RDs and (2) 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 30, 2019.

In the MEDLINE searches, the keywords combined terms related to 'mapping' with Orphanet's list of RD indications (e.g. 'acromegaly'), besides '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, i.e. those with fewer than 1000 cases or families documented in the medical literature worldwide [30], thus leaving 1059 RD unique denominations for the online searches. Moreover, we retrieved all drugs with an orphan designation from the European Medicines Agency (EMA) website (https://www.ema.europa.eu/en, accessed in February 2019) and used the corresponding 149 unique RD indications to search MEDLINE. We also searched the same drugs on the NICE website (https://www.nice.org.uk/) and reviewed the related HTA reports (50 in total). These additional searches aimed to minimise the risk of missing any studies, as we assumed mapping would be mainly done in relation to an existing drug and its HTA process. The two lists of disease terms (from Orphanet and EMA) used for MEDLINE searching are reported in the appendix (see the electronic supplementary material, Supplementary Files 1 and 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 might not have included the selected mapping-related terms used in our searches.

Two reviewers (MM and AW) screened the titles 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) use of a mapping approach to derive HSUVs, (3) from any non-preference based measure, and (4) in any RDs reported by Orphanet (or within the EMA list). Studies that recruited both patients with RDs and non-RDs 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 the following: study year, disease, country, study design, sample's characteristics, sample size, source and target measures, value set for EQ-5D (if applicable), regression techniques, goodness-of-fit/predictive accuracy 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 led to the selection of 28 studies (Fig. 1). The original database searches identified 46,329 records; of these, 14,960 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 31,375 citations, which were screened by title/abstract. Seventy-four 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; and (3) conference abstracts. No additional studies were identified from the NICE reports. A total of 28 studies were eligible for inclusion.

3.2 Synthesis of Included Studies

The 28 studies included [15, 17, 18, 20, 25, 31–45, 48–55] were split into two groups: 19 developing novel mapping algorithms in RDs (Table 1) and nine 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, since one targeted two different blood cancers (i.e. multiple myeloma and non-Hodgkin's lymphoma) [44]. Four studies [18, 40, 42, 44] also recruited non-RD patient subgroups: one analysed a sample of patients with dermatological conditions including the rare lupus erythematosus [18]; two others [40, 42] used a clinical sample including an RD (i.e. 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 [44] addressed two non-RDs (i.e. arthritis and multiple sclerosis) and the rare 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, nine 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, 34, 44] used non-RD-specific PROMs, such as the Dermatology Life Quality Index (DLQI), which is applicable to different skin diseases [46]. Four studies [20, 35, 39, 43] adopted a mix of RD-specific and non-RD-specific PROMs, including two [35, 39] on multiple myeloma that used both the general [i.e. Quality of Life Core Questionnaire 30 (QLQ-C30)] and the cancer-specific module [i.e. Quality of Life Questionnaire–Multiple Myeloma Module (QLQ-MY20)] of the EORTC questionnaires assessing quality of life in cancer patients. Two studies [40, 42] 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 [45] only used a clinical measure [i.e. the Glasgow Outcome Scale (GOS)]. As a target measure, EQ-5D was the most prevalent ($n = 15$, of which only one study used the 5-level version), while three studies [34, 38, 41] used the SF-6D, and one [43] mapped to both EQ-5D and 15D. Among those using EQ-5D, ten studies chose the English tariff, alone or in combination with the Dutch.

In terms of study design, 11 studies performed ad hoc cross-sectional surveys to collect PRO data for mapping [15, 33, 36, 37] or were secondary analyses of cross-sectional data collected for other purposes [18, 25, 31, 38, 40–42]. Five [20, 34, 39, 43, 45] used cross-sectional (usually baseline) data from cohort studies, two [35, 44] used longitudinal data collected within randomised controlled trials (RCTs), and one [32] identified three studies (two observational

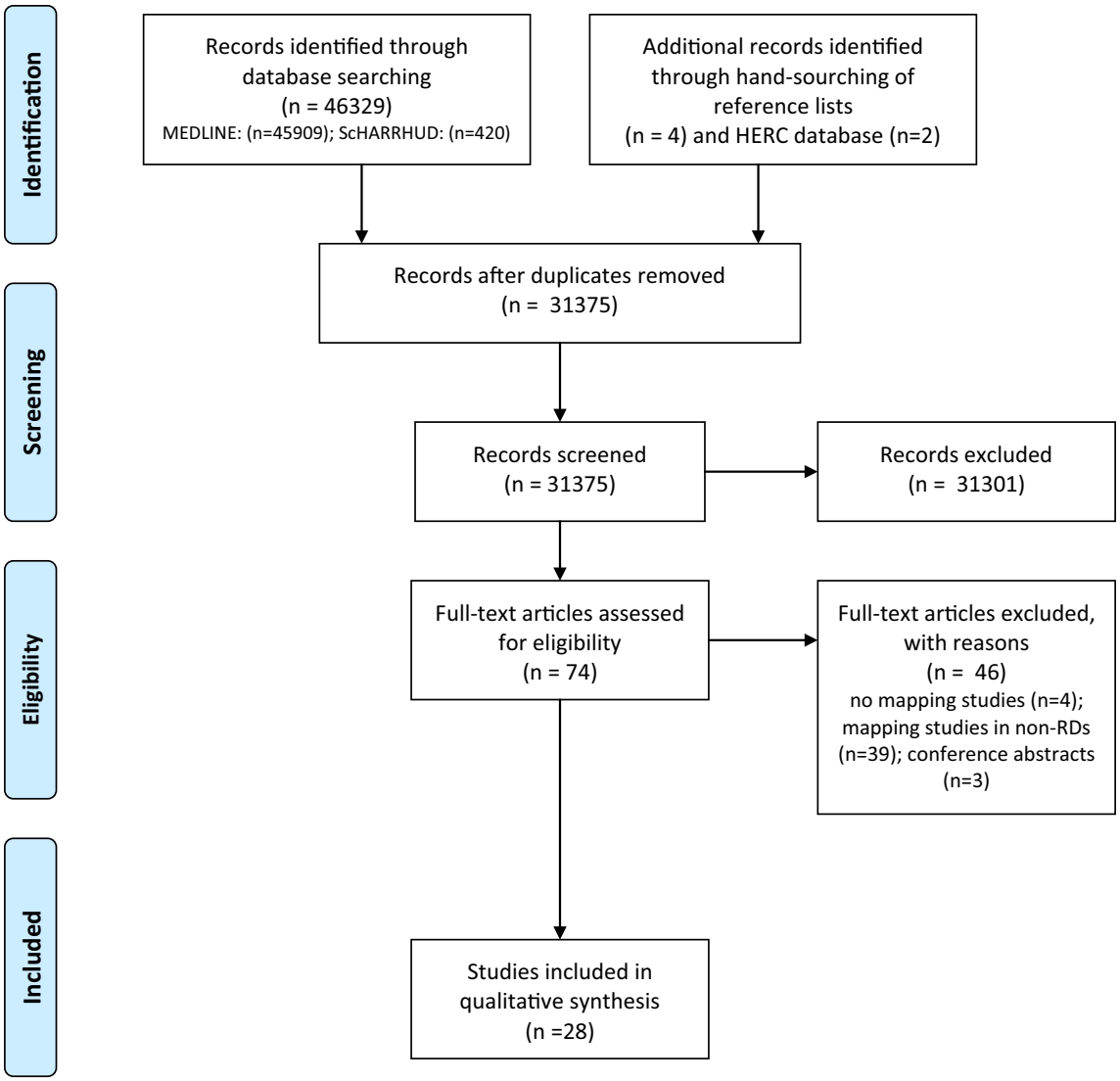

Fig. 1 Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2009 flow diagram showing study selection. *ScHARRHUD* School of Health and Related Research Health Utility Database, *HERC* Health Economics Research Centre

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 [26]. However, the most recent studies also explored more complex techniques, such as the Bayesian model [35] and the limited dependent variable mixture model (LDVMM) [45], 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, 45] 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, 38, 41, 44, 45].

Only five studies [15, 20, 38, 42, 45] explicitly embraced published recommendations in the field, including the MApping onto Preference-based measures reporting Standards (MAPS) statement [47] ($n = 2$), International Society of Pharmacoeconomics and Outcome Research (ISPOR) good

Table 1 Overview of studies ($n = 19$) developing novel mapping algorithms in RDs

First author, year [ref.]	RD (cases/100,000)	Cancer	Country	Study design	Sample's description (n)	Source	Target (tariff)	Regression technique(s)	Goodness-of-fit Predictive accuracy	Guidelines cited
Acaster, 2015 [15]	Cystic fibrosis (7.4)	No	UK	Cross-sectional survey	Adults aged 18+ ($n = 401$)	CFQ-R	EQ-5D-3L (UK)	OLS; Tobit; TPM	Adjusted R^2; AIC/BIC MSE; RMSE; ANOVA (subgroups analysis); ICC	Longworth 2013 [26]
Ali, 2017 [18]	Dermatological conditions including lupus erythematosus (50.0)	No	13 European countries	Cross-sectional survey	Any patients ($n = 3542$[a])	DLQI	EQ-5D-3L (UK)	OLS	None MAE; MSE	None
Aygören-Pürsün, 2016 [25]	Hereditary angio-oedema (1.5)	No	Spain, Germany and Denmark	Cross-sectional survey (HAE-BOIS-Europe)	At least 12 years; HAE type 1 or type 2 ($n = 111$)	HAE-BOIS-Europe survey items	EQ-5D-3L (UK)	Manual cross-walking	None None	None
Badia, 2018 [32]	Acromegaly (5.5)	No	Spain, NL and UK	Three previous studies	Adults aged 18+ ($n = 245$)	AcroQol	EQ-5D-3L (UK)	TPM; Tobit; GAM	Adjusted R^2 ME; MAE; RMSE; ICC; t test; Spearman r	None
Badia, 2013 [31]	Cushing's syndrome (5.9)	No	23 European countries	Cross-sectional study (ERCUSYN)	Pituitary-dependent and adrenal-dependent syndrome ($n = 129$)	CushingQoL	EQ-5D-3L (UK)	GLM; Tobit	Adjusted R^2; ME; MAE; RMSE	None
Busschbach, 2011 [33]	Growth hormone deficiency (0.39)	No	Belgium and NL	Cross-sectional study (ERCUSYN)	1st cohort (general population): 6875 (Belgium) + 1400 (NL). 2nd cohort (patients): 370 (Belgium) + 286 (NL). 3rd cohort (patients): 64 (NL)	QoL-AGHDA	EQ-5D-3L (NL, Belgium)	Multiple regression	R^2 ICC; t test	None

Table 1 (continued)

First author, year [ref.]	RD (cases/100,000)	Cancer	Country	Study design	Sample's description (n)	Source	Target (tariff)	Regression technique(s)	Goodness-of-fit Predictive accuracy	Guidelines cited
Kalaitzakis, 2016 [34]	Primary sclerosing cholangitis (8.1)	No	Sweden and England	Cohort study	Any patients ($n=163$)	CLDQ	SF-6D	OLS; generalised linear, median, and kernel regression	Adjusted R^2 RMSE; MAE; Pearson r	None
Kharroubi, 2015 [35]	MM (11.9)	Yes	UK, New Zealand and South Africa	Trial (MYELOMA-IX)	Any patients ($n=1839$)	QLQ-C30/QLQ-MY20	EQ-5D-3L (NS)	OLS; Bayesian models	R^2; adjusted R^2 MAE; RMSE	None
Kołtowska-Häggström, 2008 [37]	Growth hormone deficiency (0.39)	No	England and Wales	Cross-sectional survey	Adults from the general population ($n=921$)	QoL-AGHDA	EQ-5D-3L (UK)	Not specified	Adjusted R^2 None	None
Kołtowska-Häggström, 2007 [36]	Growth hormone deficiency (0.39)	No	Sweden	Cross-sectional survey	Adults (18–85 years) from the general population ($n=1945$)	QoL-AGHDA	EQ-5D-3L (6 European countries)	OLS	R^2 Correlation (r)	None
Meacock, 2015 [38]	Systemic lupus erythematosus (50.0)	No	UK	Cross-sectional survey	Adults aged 18+, literate in English, without major psychiatric disease ($n=320$)	LupusQoL	SF-6D	OLS	R^2 MAE; RSME	Longworth 2013 [26]
Moore, 2018 [20]	Motor neuron disease (amyotrophic lateral sclerosis) (3.9)	No	UK	Baseline data from a cohort study (TONiC)	Any patients ($n=595$)	ALSFRS-R; NPS; MND-HADS	EQ-5D-5L (UK)	OLS; Tobit; multinomial logistic regression	AIC MSE; MAE	MAPS statement [47]
Proskorovsky, 2014 [39]	MM (11.9)	Yes	UK and Germany	Cross-sectional data from a cohort study	Adults aged 18+ ($n=154$)	QLQ-C30/QLQ-MY20	EQ-5D-3L (UK)	Multivariate regression model	Adjusted R^2 RMSE	None
Revicki, 2009 [40]	Various conditions including spinal cord injury (32.0)	No	US	Cross-sectional survey (PROMIS)	Adults aged 18+ ($n=531$; full clinical sample: $n=7883$)	PROMIS global items ($n=10$) and selected five domain item banks	EQ-5D-3L (US)	OLS	Adjusted R^2 Bland-Altman plot; ICC	None

Table 1 (continued)

First author, year [ref.]	RD (cases/100,000)	Cancer	Country	Study design	Sample's description (*n*)	Source	Target (tariff)	Regression technique(s)	Goodness-of-fit Predictive accuracy	Guidelines cited
Roset, 2013 [41]	Cushing's syndrome (5.9)	No	Spain, NL, France, Italy and Germany	Cross-sectional survey (CushingQoL original validation study)	Adults (aged 18+); CS of pituitary or adrenal origin ($n=116$)	CushingQoL	SF-6D	GLM	Adjusted R^2 ME; MAE; RMSE	None
Thompson, 2017 [42]	Same as Revicki, 2009 [40]					PROMIS global items ($n=10$)	EQ-5D-3L (US)	Linear and equipercentile equating	R^2 PMCC; ICC; MAE; MSE	MAPS statement [47]
Vartiainen, 2017 [43]	Chronic pain (referred to tertiary pain clinics) (12.0)	No	Finland	Prospective cohort study (KROKIETA)	Adults (age 18–75) with chronic non-cancer pain ($n=391$)	BDI; BNSQ; BPI Intensity/Interference; CPAQ; PASS	EQ-5D-3L/15D (NS)	Multiple OLS	R^2 None	None
Versteegh, 2012 [44]	MM/NHL (11.9/11.6)	Yes	NL	Two randomised controlled trials (HOVON 24 and HOVON 25)	Untreated patients; NHL: Ann Arbor stages II–IV, or intermediate or high-grade malignancy (MM: $n=137$; NHL: $n=108$)	QLQ-C30	EQ-5D-3L (NL, UK)	OLS	Adjusted R^2 MAE; RSME; normalised RMSE	None
Ward Fuller, 2017 [45]	Traumatic brain injury (37.8)	No	Australia	Cross-sectional data from a retrospective cohort study (Victorian State Trauma Registry)	Adults (≥ 16 years) with Injury Scale severity score ≥ 3 ($n=3437$)	GOS	EQ-5D-3L (UK)	LDVMM	AIC/BIC ME; MAE; RMSE	ISPOR Good Practices [27]

AIC Akaike information criteria, *ALSFRS-R* Amyotrophic Lateral Sclerosis Functional Rating Scale–Revised, *BDI* Beck Depression Inventory, *BIC* Bayesian information criteria, *BNSQ* Basic Nordic Sleep Questionnaire, *BPI* Brief Pain Inventory, *CFQ-R* Cystic Fibrosis Questionnaire–Revised, *CLDQ* Chronic Liver Disease Questionnaire, *CPAQ* Chronic Pain Acceptance Questionnaire, *CS* Cushing's syndrome, *DLQI* Dermatology Life Quality Index, *EQ-5D-3L/-5L* EuroQol 5-Dimension 3-Level/5-Level, *GAM* generalised additive model, *GLM* generalised linear model, *GOS* Glasgow Outcome Scale, *HAE* hereditary angioedema, *HAE-BOIS* Hereditary Angioedema Burden of Illness Study, *ICC* interclass correlation coefficient, *ISPOR* International Society of Pharmacoeconomics and Outcome Research, *LDVMM* limited dependent variable mixture model, *MAE* mean absolute error, *MAPS* MApping onto Preference-based measures reporting Standards, *ME* mean error, *MM* multiple myeloma, *MND-HADS* Hospital Anxiety and Depression Scale for motor neuron disease, *MSE* mean squared error, *NHL* non-Hodgkin lymphoma, *NL* Netherlands, *NPS* Neuropathic Pain Scale, *NS* not specified, *OLS* ordinary least squares, *PASS* Pain Anxiety Symptoms Scale, *PMCC* product-moment correlation coefficient, *PROMIS* Patient-Reported Outcomes Measurement Information System, *QLQ-C30* Quality of Life Core Questionnaire 30, *QLQ-MY20* Quality of Life Questionnaire–Multiple Myeloma Module, *QoL-AGHDA* Quality of Life Assessment of Growth Hormone Deficiency in Adults, *RD* rare disease, *RMSE* root mean squared error, *SF-6D* Short-Form Six-Dimension, *TONiC* Trajectories of Outcomes in Neurological Conditions, *TPM* two-part model, *UK* United Kingdom, *US* United States

[a] *n* refers to the overall sample; the *n* of patients with lupus erythematosus is not reported in the original mapping study

Table 2 Overview of studies ($n=9$) applying existing mapping algorithms to original RD data

RD study						Mapping algorithms					
First author, year	Rare disease (cases/100,000)	C	Country	Study design	Sample's description (n)	First author, year	Disease(s)	Country	Source	Target	Regression technique(s)
Arnold, 2015 [48]	Pleural mesothelioma (3.1)	Y	UK	Prospective cohort study (SWAMP trial)	Unsuitable for surgery; WHO performance score 0, 1, or 2 ($n=73$)	Crott, 2010 [57]	Breast cancer	Various	QLQ-C30	EQ-5D	OLS
						Jang, 2010 [56]	NSCLC	Canada	QLQ-C30	EQ-5D	Linear
						Kim EJ, 2012 [58]	Breast cancer	Korea	QLQ-C30	EQ-5D	OLS
						Kim SH, 2012 [59]	Multiple cancers	Korea	QLQ-C30	EQ-5D	OLS
						Kontodimopoulos, 2009 [60]	Gastric cancer	Greece	QLQ-C30	EQ-5D; SF-6D; 15D	OLS
						Longworth, 2014 [64]	MM; breast cancer; lung cancer; multiple cancers	QLQ-C30 (MM): several countries worldwide; QLQ-C30 (breast/lung cancer): Canada. FACT-G: US	QLQ-C30; FACT-G	EQ-5D	OLS; Tobit; two-part model; splining; response mapping; LDVMM
						McKenzie, 2009 [61]	Oesophageal cancer	UK	QLQ-C30	EQ-5D	OLS; ordered probit
						Proskorovsky, 2014 [39]	MM	UK and Germany	QLQ-C30/ QLQ-MY20	EQ-5D	Multivariate regression model
						Versteegh, 2012 [44]	MM/NHL; arthritis; multiple sclerosis	NL and UK	QLQ-C30; HAQ; MSIS-29	EQ-5D	OLS
Crott, 2013 [49]	MM/NHL (11.9/11.6)	Y	NL	RCTs (HOVON 24/25)	Previously untreated patients (MM: $n=132$; NHL: $n=108$)	Crott, 2010 [57]	As above				

Table 2 (continued)

RD study						Mapping algorithms					
First author, year	Rare disease (cases/100,000)	C	Country	Study design	Sample's description (*n*)	First author, year	Disease(s)	Country	Source	Target	Regression technique(s)
Forsythe, 2018 [17] using data from Grulke, 2012 [68] and Perić, 2016 [69]	Acute myeloid leukaemia/haematopoietic stem cell transplantation (2.5/0.65)	Y	Grulke 2012 [68]: countries worldwide; Perić 2016 [69]: Croatia	Review; Grulke 2012 [68]: review; Perić 2016 [69]: cross-sectional	Grulke 2012 [68]: various cancers, including non-rare ($n=2800$); Perić 2016 [69]: chronic graft-versus-host disease ($n=38$)	Crott, 2010 [57]	As above				
Hess, 2013 [50]	Ovarian cancer (30.0)	Y	Canada; US	RCTs (GOG-0152 and GOG-0172)	GOG-0152: advanced ovarian cancer; GOG-0172: stage III ovarian cancer (baseline: $n=746$; time 4: $n=569$)	Dobrez, 2007 [62]	Multiple cancers	US	FACT-G	TTO	OLS
						Cheung, 2009 [63]	Multiple cancers	Singapore	FACT-G	EQ-5D	OLS; CLAD
Hettle, 2015 [51]	Ovarian cancer (30.0)	Y	European countries; US	RCT (NCT00753545)	Patients with PSROC who have received two or more courses of chemotherapy ($n=247$)	Dobrez, 2007 [62]	As above				
						Cheung, 2009 [63]	As above				
						Longworth, 2014 [64]	As above				
Meng, 2017 [52]	Gastroenteropancreatic NETs (0.21)	Y	European countries; US; India	RCT (CLARINET study)	Patients with advanced, well- or moderately differentiated, non-functioning, somatostatin receptor-positive NETs of grade 1 or 2 ($n=204$)	McKenzie, 2009 [61]	As above				
						Longworth, 2014 [64]	As above				

Table 2 (continued)

RD study						Mapping algorithms					
First author, year	Rare disease (cases/100,000)	C	Country	Study design	Sample's description (*n*)	First author, year	Disease(s)	Country	Source	Target	Regression technique(s)
Pan, 2010 [53]	Myelodysplastic syndromes (1.5)	Y	US	RCT	Adults (18+); IPSS score ≥ 0.5 (*n* = 170)	Kontodimopoulos, 2009 [60]					
Rowen, 2012 [54]	MM (11.9)	Y	US	RCT (VISTA - NCT00111319)	Adults (18+) (*n* = 674; 5650 observations)	Crott, 2010 [57]	As above				
						Kontodimopoulos, 2009 [60]	As above				
						McKenzie, 2009 [61]	As above				
Vernon, 2016 [55]	Castleman's disease (1.0)	N	Countries worldwide	RCT (CNTO328MCD2001)	Adults (aged 18+); HIV negative; HHV-8 negative (*n* = 79)	Rowen, 2009 [19]	Any	UK	SF-36	EQ-5D	GLS models

C cancer, *CLAD* censored least absolute deviation, *EQ-5D* EuroQol 5-Dimension (3-Level), *FACT-G* Functional Assessment of Cancer Therapy–General, *GLS* generalised linear squares, *HAQ* Health Assessment Questionnaire, *HHV* human herpes virus, *HIV* human immunodeficiency virus, *IPSS* International Prognostic Scoring System, *LDVMM* limited dependent variable mixture model, *MM* multiple myeloma, *MSIS-29* Multiple Sclerosis Impact Scale, *N* no, *NET* neuroendocrine tumour, *NHL* non-Hodgkin lymphoma, *NL* Netherlands, *NSCLC* non-small cell lung cancer, *OLS* ordinary least squares, *PSROC* platinum-sensitive relapsed serous ovarian cancer, *QLQ-C30* Quality of Life Core Questionnaire 30, *QLQ-MY20* Quality of Life Questionnaire–Multiple Myeloma Module, *RCT* randomised controlled trial, *RD* rare disease, *SF-36* Short Form (36-item), *SF-6D* Short-Form Six-Dimension, *SWAMP* South West Area Mesothelioma and Pemetrexed, *TTO* time trade-off, *UK* United Kingdom, *US* United States, *WHO* World Health Organization, *Y* yes

practices [27] ($n=1$), and the study by Longworth and Rowen [26] ($n=2$). However, five [33, 36, 37, 40, 44] were published before any guidelines for mapping were available.

3.2.2 Studies Applying Previous Mapping to RD 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 [49] 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 nine studies were RCTs, except for one prospective cohort study [48] and one review [17]. Four studies [17, 51, 52, 55] were intercontinental, and all recruited small samples (< 1000), except for one [17] 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 [48, 49], (2) identifying the best available algorithms for a specific condition (i.e. ovarian cancer) [50, 51], (3) deriving HSUVs for economic evaluation alongside RCTs [52, 55] or for economic models [17, 53], and (4) testing the comparability of mapped HSUVs with those derived from disease-specific preference-based PROMs (i.e. EORTC-8D) and generic PROMs (i.e. EQ-5D) [54].

The nine studies referred to 12 different original mapping studies (Table 2), all developed in oncology, with the exception of one that recruited all patients referred to a large university hospital [19]. Since all 12 were conducted in non-RDs, no RD-specific PROMs were used as the source measure. In detail, eight of the original mapping studies [39, 44, 56–61] were mapped from the EORTC QLQ-C30, which is a questionnaire widely used in oncology, two [62, 63] used another popular cancer-specific tool [Functional Assessment of Cancer Therapy–General (FACT-G)], and one [64] used both; one study [19] used the 36-item Short-Form (SF-36), which is a generic non–preference-based tool. As a target measure, ten studies mapped onto the EQ-5D-3L, one onto EQ-5D-3L, SF-6D, and 15D [60], and one [62] onto time trade-off (TTO) utilities.

Overall, the nine studies reported several difficulties in applying published algorithms to RD data. First, most algorithms, and especially those using OLS, tend to over-predict HSUVs in poor health states (e.g. EQ-5D utility less than 0.5) or to predict values greater than 1 (i.e. outside the possible EQ-5D range) [48, 49, 52, 54]; indeed, the study testing the highest number of algorithms [48] showed that response mapping performed best [64]. Second, the available algorithms were usually developed in populations that are very different from RD patients [54, 55]. 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. [48] did not include a dyspnoea score, with a consequent 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 [49, 54].

3.3 Critical Issues Around 'Mapping' in RDs

The 28 reviewed studies highlighted a number of challenges around the existing mapping literature in RDs (Fig. 2).

The first relates to the amount and context of existing literature on mapping applied to RDs. The searches identified only a few studies (28 in 20 different RDs) compared to over 1000 RDs screened. In addition, the number of cancer studies is relatively high (11/28), most of which came from the second group of studies applying published algorithms to original RD data. Moreover, no mapping studies were performed in childhood diseases, and 13 (out of 19) of the original mapping studies included age (e.g. minimum 12 years old) as an inclusion criterion, thus excluding paediatric patients.

The second issue relates to sample sizes, which were commonly small (< 1000 RDs patients in 16 out of 19 novel mapping studies). This was because of the challenges in identifying RDs patients, as also recognised by some of the included studies [20, 31, 32, 41]. The limited size of recruited samples may affect the robustness of

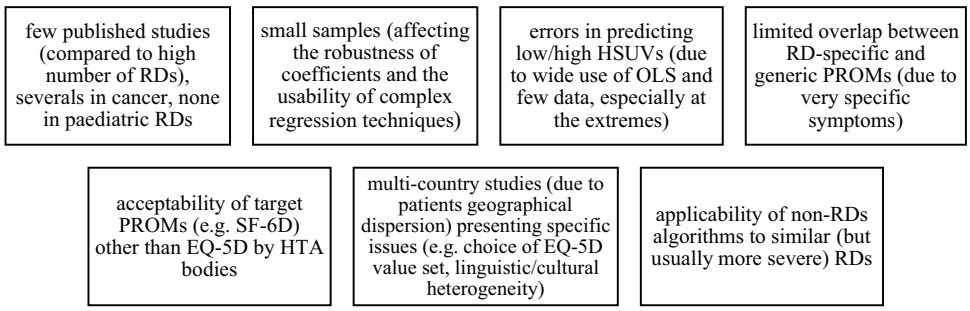

Fig. 2 Critical issues emerged from mapping studies in RDs. *EQ-5D* EuroQol-5 Dimension, *HSUV* health state utility value, *HTA* health technology assessment, *OLS* ordinary least squares, *PROM* patient-reported outcome measure, *RD* rare disease, *SF-6D* Short-Form Six-Dimension

mapping coefficients. Moreover, the widely adopted OLS regression [65] may not work for EQ-5D data in general [27], and particularly in RDs, which are likely to present multimodality and peaks due to the high patient heterogeneity in clinical manifestations, disease severity, and quality-of-life impact even within the same condition [10, 15]. However, a small sample size hampers the applicability of more complex techniques such as LDVMM, which is recommended by the most recent guidelines for modelling EQ-5D data [27]. Lastly, a small sample size makes it difficult to perform cross-validation tests in patient subgroups with peculiar characteristics such as disease subtypes or those 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 et al. [15], only 3% of the recruited patients reported EQ-5D utilities below zero; the study by Kalaitzakis et al. [34] 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 [45] 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 'average' RD patients, who are likely to be the target of clinical trials assessing new drugs, but are less use in predicting 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 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], as well as in systemic lupus erythematosus (LupusQoL dimensions omitted: planning, intimate relationships, burden to others, and body image) [38] and motor neuron disease (ALSFRS-R domains omitted: communication, salivation, swallowing, hand use, and 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 [34, 38, 41] used the SF-6D. One study [43] 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 the 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 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 [36]. 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 [66]. 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 generalisability of mapping algorithms developed in non-RDs to RDs is not straightforward. As anticipated before, some included studies [48, 49, 54] 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, 67] 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

12 studies identified in the HERC database and three additional studies [34, 38, 41] using SF-6D. A previous review published in 2010 [65] identified 28 studies mapping non-preference-based PROMs onto any preference-based ones in any disease area; however, it did not capture any of the studies included in our review, since 17 (out of 19) were published after 2009, and the more dated two [36, 37] used different mapping-related terms. Indeed, three records [33, 36, 37] that did not use 'mapping' or 'cross-walking' (e.g. keywords adopted by Dakin et al. [29] and Brazier et al. [65]) were identified by reviewing the reference lists of the included studies. Lastly, by focusing on a single disease area (i.e. RDs), 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 generalisability of mapping developed in similar non-RDs.

We also identified some relevant systematic reviews in individual RDs. The study by Forsythe et al. [17] identified ten studies reporting on quality of life and HSUVs in acute myeloid leukaemia, of which only one [53] used a mapping approach (and was included in our review) and two [68, 69] collected EORTC QLQ-C30 data that were converted into HSUVs by the review's authors using a published algorithm [57]. Overall, this review recommended estimating HSUVs with larger samples, in addition to routine monitoring (including PROMs administration) of patients with RDs and establishment of a set of recommendations to standardise HSUV elicitation across different RDs. A systematic review on cystic fibrosis [70] 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, 71] were identified in Cushing's syndrome and included the same two mapping studies [31, 41] captured by our review. One review [71] stated that mapping HSUVs from the CushingQoL questionnaire is possible, at least at the 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 28 studies using mapping in RDs, i.e. conditions affecting no more than one person in 2000, according to the European definition [8, 9]. Of these, 19 developed original algorithms with RD data, and nine applied previous algorithms developed in different conditions to RD data; among the latter, eight studies were performed in oncology, likely because of the substantial number 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 (less than 2%) compared to the list of conditions screened in the database searches. Also, no mapping studies were found in paediatric RDs, likely because of the greater legal and ethical requirements around enrolling children in clinical studies and the lack of RD-specific PROMs intended for child self-reporting (or parental proxy-reporting).

Moreover, only five studies [15, 20, 38, 42, 45] explicitly embraced published recommendations in the mapping field, and only one [45] referred to the most recent guidelines (i.e. ISPOR good practices [27]), which, for example, recommend reporting AIC/BIC for model selection instead of summary measures of fit (e.g. R-squared). In addition, ISPOR encourages considering alternatives to OLS for modelling HSUVs, such as beta-binomial regression [72] and mixture models (e.g. adjusted LDVMM [73]). Indeed, OLS can accurately predict mean HSUVs, because its underestimation of high values is compensated by its overestimation of low values [74], but is not appropriate to estimate HSUVs at the 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 this further reduces sample size for estimation, and small sample is an issue that often affects mapping studies in RDs. However, previous recommendations from Longworth and Rowen [26] 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 applying 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 site as a predictor in mapping models [32]. Moreover, although EQ-5D tariffs look quite similar across European countries [49], 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 [34], 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 [34]. Some studies reported MIDs between 0.033 and 0.082 in non-cancer conditions [51], 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 [44] pooled data across different time points of an RCT, although loss to follow-up raises the possibility of selection bias [20, 45].

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 [75]. 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, 44]. 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 [38, 49].

Lastly, the generalisability of mapping algorithms developed in similar non-RDs to RD datasets needs careful attention. Overall, the mapping tendency to overestimate HSUVs in poor health states is critical, 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 [76]; for example, 5-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 minimise prediction errors. For example, in the study by Arnold et al. [48] on pleural mesothelioma, the algorithm by Jang et al. [56] 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 et al. [54], the worst performing algorithm [57] was that developed in a breast cancer population, which little resembled the multiple myeloma dataset adopted. In principle, the choice of a specific algorithm should be justified when many competing algorithms exist, as it is for several mapping studies in oncology that are potentially applicable to rare cancers, although the preferred algorithm might not generate the most plausible HSUVs [51, 52]. Moreover, identical 'source' PROM versions should be used in mapping algorithms and studies using them [49]. Overall, careful sensitivity analyses (e.g. evaluating the impact of using different algorithms) should be performed when cost-utility models are populated with mapped HSUVs [31, 32, 50, 51].

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 RDs. This resulted in searching around one fourth of the RDs reported by Orphanet. Overall, very RDs are extremely unlikely to be provided with mapping algorithms, because of insufficiently large samples, a lack of disease-specific PROMs, and a lack of treatments that incentivise the estimation of HSUVs for HTA. However, 29 terms related to very RDs for which an orphan drug is available (as resulting from the EMA website) were used for database searching. Second, by searching only title/abstract in PubMed, we are likely to have missed some studies applying existing mapping algorithms, but that mention them only in the full text. We also did not include other databases (e.g. EMBASE) that might have provided more eligible studies. Third, 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. Fourth, since mapping is considered as a 'second best' solution [26, 54], there might be studies in RDs that collected HSUVs directly from patients using generic preference-based PROMs or direct elicitation techniques (e.g. TTO), or that applied the HSUVs estimated for other similar non-RDs to the RD 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 synthetised the available mapping literature on RDs to derive HSUVs in the absence of preference-based measures. Overall, 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) and inform the reimbursement decisions for new drugs. This would allow exploitation of 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, which recommends undertaking mapping in the absence of EQ-5D data collection in clinical trials. However, the limitations of using such an approach to derive HSUVs should be acknowledged in HTA processes, and possibly overcome by future research.

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

Author 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.

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

Compliance with Ethical Standards

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.

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 has no conflict of interest.

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 Educ. 2017;17(4):137–44.
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):125–34.
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;35:276–91.
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):838–55.
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. Illustration 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. https://doi.org/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;26:3025–34.
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):1322–9.
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. McTaggart-Cowan H, King MT, Norman R, et al. The EORTC QLU-C10D: the canadian valuation study and algorithm to derive cancer-specific utilities from the EORTC QLQ-C30. MDM Policy Pract. 2019;4(1):2381468319842532.
23. 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.
24. 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.
25. 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;10:1699–707.
26. 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.
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):2941–50.
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):382–9.
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. 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;25:947–57.
35. 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 Mak. 2015;35(3):351–60.
36. 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.
37. 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.
38. 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.
39. 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.
40. Revicki DA, Kawata AK, Harnam N, et al. Predicting EuroQol (EQ-5D) scores from the patient-reported outcomes measurement information system (PROMIS) global items and domain item banks in a United States sample. Qual Life Res. 2009;18(6):783–91.
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. Thompson NR, Lapin BR, Katzan IL. Mapping PROMIS Global Health Items to EuroQol (EQ-5D) utility scores using linear and equipercentile equating. Pharmacoeconomics. 2017;35(11):1167–76.
43. 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):1577–85.
44. Versteegh MM, Leunis A, Luime JJ, et al. Mapping QLQ-C30, HAQ, and MSIS-29 on EQ-5D. Med Decis Mak. 2012;32(4):554–68.
45. 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):141–51.
46. Finlay AY, Khan GK. Dermatology Life Quality Index (DLQI)—a simple practical measure for routine clinical use. Clin Exp Dermatol. 1994;19(3):210–6.
47. 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.
48. 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.
49. 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.
50. 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.
51. 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.
52. 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.
53. Pan F, Peng S, Fleurence R, et al. Economic analysis of decitabine versus best supportive care in the treatment of intermediate- and high-risk myelodysplastic syndromes from a US payer perspective. Clin Ther. 2010;32(14):2444–56.
54. Rowen D, Young T, Brazier J, Gaugris S. Comparison of generic, condition-specific, and mapped health state utility values for multiple myeloma cancer. Value Health. 2012;15(8):1059–68.
55. 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.
56. 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.
57. 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.
58. 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.
59. 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.
60. 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.
61. 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.
62. 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.
63. 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.
64. 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.
65. 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.
66. 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. https://doi.org/10.1186/s13023-017-0591-7.
67. 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.
68. Grulke N, Albani C, Bailer H. Quality of life in patients before and after haematopoietic stem cell transplantation measured with the European Organization for Research and Treatment of Cancer (EORTC) Quality of Life Core Questionnaire QLQ-C30. Bone Marrow Transplant. 2012;47(4):473–82.
69. Perić Z, Desnica L, Duraković N, et al. Which questionnaires should we use to evaluate quality of life in patients with chronic graft-vs-host disease? Croat Med J. 2016;57(1):6–15.
70. Mohindru B, Turner D, Sach T, et al. Health state utility data in cystic fibrosis: a systematic review. Pharmacoecon Open. 2019. https://doi.org/10.1007/s41669-019-0144-1.
71. Santos A, Crespo, Aulinas A, et al. Quality of life in Cushing's syndrome. Pituitary. 2015;18(2):195–200.
72. Khan I, Morris S. A non-linear beta-binomial regression model for mapping EORTC QLQ- C30 to the EQ-5D-3L in lung cancer patients: a comparison with existing approaches. Health Qual Life Outcomes. 2014;12:163.
73. Hernández Alava M, Wailoo AJ, Ara R. Tails from the peak district: adjusted limited dependent variable mixture models of EQ-5D questionnaire health state utility values. Value Health. 2012;15(3):550–61.
74. Crott R. Direct mapping of the QLQ-C30 to EQ-5D preferences: a comparison of regression methods. Pharmacoecon Open. 2018;2(2):165–77.
75. 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):159–65.
76. 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.

Affiliations

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

✉ Michela Meregaglia
michela.meregaglia@unibocconi.it

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

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