



Evidence on the relationship between PROMIS-29 and EQ-5D: a literature review

Tianxin Pan^{1,2} · Brendan Mulhern³ · Rosalie Viney³ · Richard Norman² · An Tran-Duy¹ · Janel Hanmer⁴ · Nancy Devlin¹

Accepted: 10 June 2021 / Published online: 28 June 2021
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2021

Abstract

Purpose EQ-5D and PROMIS-29 are both concise, generic measures of patient-reported outcomes accompanied by preference weights that allow the estimation of quality-adjusted life years (QALYs). Both instruments are candidates for use in economic evaluation. However, they have different features in terms of the domains selected to measure respondents' self-perceived health and the characteristics of (and methods used to obtain) the preference weights. It is important to understand the relationship between the instruments and the implications of choosing either for the evidence used in decision-making. This literature review aimed to synthesise existing evidence on the relationship between PROMIS-29 (and measures based on it, such as PROMIS-29+2) and EQ-5D (both EQ-5D-3L and EQ-5D-5L).

Methods A literature review was conducted in PubMed and Web of Science to identify studies investigating the relationship between PROMIS-29 and EQ-5D-based instruments.

Results The literature search identified 95 unique studies, of which nine studies met the inclusion criteria, i.e. compared both instruments. Six studies examined the relationship between PROMIS-29 and EQ-5D-5L. Three main types of relationship have been examined in the nine studies: (a) comparing PROMIS-29 and EQ-5D as descriptive systems; (b) mapping PROMIS-29 domains to EQ-5D utilities; and (c) comparing and transforming PROMIS-29 utilities to EQ-5D utilities.

Conclusion This review has highlighted the lack of evidence regarding the relationship between PROMIS-29 and EQ-5D. The impact of choosing either instrument on the evidence used in cost-effectiveness analysis is currently unclear. Further research is needed to understand the relationship between the two instruments.

Keywords Patient-reported outcome measures · Preference-accompanied measures · PROMIS-29 · EQ-5D · Descriptive systems · Utility scores · Comparison

Introduction

There is a proliferation of patient-reported outcome measures (PROMs) designed for different purposes. One such set of measures are accompanied by value sets (also referred to as utilities) which cover core dimensions of health and generate preference weights typically obtained from the general population or patients for calculating quality-adjusted life years (QALYs) in cost-effectiveness analysis (CEA). There are different types of generic preference-accompanied PROMs. Traditional PROMs include those which comprise a concise descriptive system using a fixed number of question items, accompanied by a set of values (e.g. EQ-5D [1], HUI [2, 3]), and those which obtain patients' health states on a longer generic questionnaire (e.g. SF-36 [4]) and then reduce it from multiple items to one attribute per dimension

✉ Tianxin Pan
tianxin.pan1@unimelb.edu.au

¹ Health Economics Unit, Centre for Health Policy, Melbourne School of Population and Global Health, University of Melbourne, 207 Bouverie Street, Melbourne, VIC 3010, Australia

² School of Population Health, Curtin University, Perth, WA, Australia

³ Centre for Health Economics Research and Evaluation, University of Technology Sydney, Sydney, NSW, Australia

⁴ Department of General Internal Medicine, University of Pittsburgh, Pittsburgh, PA, USA

to generate utilities (e.g. SF-6D [5]). More recently, PROMs based on item response theory (IRT)—most notably, the Patient-Reported Outcomes Measurement Information System (PROMIS), have been developed [6]. PROMIS currently has over 100 health domains with a calibrated “item bank” of questions that can be administered by standard short forms, custom short forms, or computer adaptive testing [7]. The PROMIS Profile measures (i.e. PROMIS-29, PROMIS-43 and PROMIS-57) have been developed to meet the demand for a fixed short-form version of an instrument that is relevant for most health contexts. Utilities for PROMIS-29 are available, but (to date) only for the US [8, 9].

Before introducing the key instruments, it is important to define terms to allow comparability. We define both PROMIS-29 and EQ-5D to be preference-accompanied PROMs, i.e. instruments which seek to describe a set of health states that are mutually exclusive and exhaustive, and which are accompanied by preference weights. We use the term preference-accompanied rather than preference-based because both these PROMs can be considered as measurement systems that are independent of the preferences that accompany them, although a central part of their use is in the application of a preference algorithm to them, for example, in the estimation of QALYs. Elsewhere, preference-accompanied PROMs are known as utility instruments, or multi-attribute utility instruments (MAUIs) [3, 10]. Each preference-accompanied PROM consists of a set of dimensions, that are used to measure respondents’ self-perceived health. Elsewhere, these dimensions are also known as attributes or domains. In this paper, we allow the use of the terminology and descriptions used by the instrument developers in each case.¹

PROMIS-29 is the brief generic instrument within the PROMIS system that consists of 29 fixed items (questions) covering seven health domains. It has been quickly taken up in many research and clinical settings [9, 14]. From the three PROMIS profile measures reported in Cella et al. (2019) [14], most users have opted to use PROMIS-29 (personal correspondence, Health Measures, 23rd July 2020). The EQ-5D is a very concise generic instrument which has five dimensions with one item in each dimension and has been widely used in many countries. Both PROMIS-29 and EQ-5D are self-report measures of core health domains and accompanied by preference weightings. Therefore, both are potentially important candidates for use in economic evaluation. However, they differ in the way they measure and value health, which has the potential to yield differences in health

states described and the estimates of QALYs to be used in CEA in the decision-making process [15].

The EQ-5D and PROMIS-29 differ in the health domains covered. The EQ-5D includes five dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression; the five dimensions used in seeking self-reported health from patients are fixed. Each dimension has one question. The initial version of the EQ-5D has three levels of problems within a particular EQ-5D dimension (EQ-5D-3L); a more recent version retains the same five dimensions but increases the levels of problems on each to five (EQ-5D-5L) [16]. For EQ-5D-5L, there are five response levels for each dimension: no problems = 1; slight problems = 2; moderate problems = 3; severe problems = 4; and unable/extreme problems = 5. EQ-5D-5L health states can be described as a five-number string, representing the level of each dimension in the order in which they appear in the questionnaire (as described above) [17].

The PROMIS-29 comprises 4-item short forms of seven domains: Physical Function, Depression, Anxiety, Fatigue, Sleep Disturbance, Ability to Participate in Social Roles and Activities (Social Roles), Pain Interference, plus a 0–10 Pain Intensity numeric rating item. As one of the PROMIS Profile measures, PROMIS-29 is derived from the PROMIS item banks and has shown evidence of strong psychometric properties such as reliability and validity [14]. In PROMIS-29, each item (or question) has five response levels ranging in value from one to five (raw score), and the domain sum raw score ranges from 4 (1×4) to 20 (5×4). Health state for each domain in PROMIS-29 (and in other PROMIS measures is expressed in T-scores, which have a mean of 50 and a standard deviation of 10 compared with the general US population.

The stated preference methods used to obtain utilities for each instrument are also very different. The development of the EQ-5D was motivated in part the aim to facilitate QALY estimates for use in economic evaluation [1]; the five dimensions in the descriptive system are identical to the five domains used in the stated preference exercises that generate the utilities for it. Utilities for the EQ-5D-3L are available for many countries and were typically developed using either Time Trade-Off (TTO) or, for some older value sets, Visual Analogue Scale (VAS) methods [18]. Utilities for the EQ-5D-5L have been produced in many countries and are typically obtained using a protocol which includes both Discrete Choice Experiments (DCE) and the composite Time Trade-Off (cTTO), with resulting value sets based either on cTTO or a hybrid of cTTO and DCE [19]. Alongside this, DCE methods including duration have also been used as a standalone approach for the estimation of value sets that are used in decision-making [20].

Two valuation studies have produced utilities that can be applied to PROMIS-29. An ‘official’ preference-based scoring system (PROPr) has been developed to attach utilities to

¹ For example, EQ-5D uses “dimension” to describe key components of self-perceived health, whereas the corresponding term in PROMIS measures is “domain”. For more detail on the differences in terms and descriptions, see Pan et al. [11–13].

health states described by PROMIS domains based on multi-attribute utility theory using Standard Gamble [8]. The PROPr system was not designed for PROMIS-29 specifically but for any PROMIS measures that collect measurements on seven PROMIS domains: Cognitive Function, Depression, Fatigue, Pain Interference, Physical Function, Sleep Disturbance and Social Roles. Among them, six domains are included in PROMIS-29 (the exception being Cognitive Function). To facilitate the calculation of the PROPr summary score for PROMIS-29, the PROMIS team developed a function to predict the missing Cognitive Function score based on Profile measures [21]. A further, ‘unofficial’ valuation study is available, which focussed on PROMIS-29 domains with a 10-year duration (i.e. respondents were asked to consider decrements in health lasting 10 years) using DCE, has produced preference weights for the PROMIS-29 [9].

The choice of a particular instrument for measuring health outcomes can have a non-trivial effect on estimates of QALYs and QALY gains. The use of either EQ-5D-3L, EQ-5D-5L or PROMIS-29 will potentially yield different results and evidence for decision-making, deriving both from differences in what is measured in each case and how it is valued. Previous work has used a literature review to compare the psychometric properties of different preference-accompanied measures and has proven it as a valid and useful method [22, 23]. This literature review aims to synthesise existing evidence on the relationship between EQ-5D and PROMIS-29-based instruments² to establish what is currently known about the comparative properties of these measures and their utilities and to identify where there are gaps in knowledge.

Methods

We employed a literature search in PubMed and Web of Science, from inception to 28 Feb 2021, to identify studies reporting on both instruments and examining their relationships. We used combinations of the following keywords: “PROMIS-29”, “PROMIS Profile”, “PROMIS”, “Patient-Reported Outcomes Measurement Information System” and “EQ-5D”, “EQ-5D-5L”, “EQ-5D-3L”, “EuroQoL” and “Euro-QoL”. There were 77 records yielded from PubMed, and 101 records from Web of Science. After removal of duplicates, a total of 95 studies remained.

A two-stage study selection process was employed. The first stage involved the identification of papers reported

original research, and there was an indication in the title or abstract that it contained both instruments, i.e. PROMIS-29 and EQ-5D (EQ-5D-3L or EQ-5D-5L). In the second stage, full text articles were assessed. Studies that included only one or two domains in PROMIS-29 were excluded because we were interested in the two instruments as generic, multi-dimensional measures of self-perceived health. Papers were included if they investigated the following:

- Comparing PROMIS-29 and EQ-5D as descriptive systems,
- Response mapping between PROMIS-29 domains and EQ-5D dimensions,
- Mapping of PROMIS-29 domain scores to EQ-5D utility scores (refers to functions for which an instrument without utility weights is “mapped” onto one with weights),
- Comparing the characteristics of the value sets available for PROMIS-29 and EQ-5D,
- Comparing PROMIS-29 utility scores and EQ-5D utility scores using population or patient data,
- Scale transformation (refers to functions which seek to establish the link between instruments which both have existing utility weights), and
- Validating an existing mapping or scale transformation algorithm.

Data extracted included the instruments, patient or population group, country of origin, sample size, types of relationship evaluated, statistical methods, psychometric properties (reliability, validity and responsiveness), model specification, performance (within and external validation) and authors’ comments on the study.

Results

Search results

A total of 95 unique studies were identified from the literature search, resulting in 73 initial inclusions after title and abstract screening. Among the remaining studies, only nine studies investigated the relationship between PROMIS-29 and EQ-5D, including six using the EQ-5D-5L measure.³

² PROMIS-29-based instruments refer to PROMIS-29 and measures based on it, such as PROMIS-29 v1.0, PROMIS-29 v2.0, PROMIS-29+2 v2.1. PROMIS-29+2 v2.1 is the recommended instrument to calculate PROPr scores, which consist of the PROMIS-29 and two items from Cognitive Function-Abilities v2.0 [24]. EQ-5D instruments include EQ-5D-3L and EQ-5D-5L.

³ It is worth noting that there is one ongoing study uses item response theory to create a crosswalk to the EQ-5D-5L from PROMIS-29 domains [25]. After confirming with the authors, only abstract was available at this stage. Also, we identified two further studies which compared EQ-5D-5L utilities with PROPr utilities. However, the author did not use PROMIS-29 to calculate PROPr scores but administered 14 questions (2 for each PROPr domains) to respondents. Detailed information on the questions was not reported. Thus the two studies were not included in our literature review. For more information, please refer to [26, 27].

Fig. 1 Flow chart of the literature review

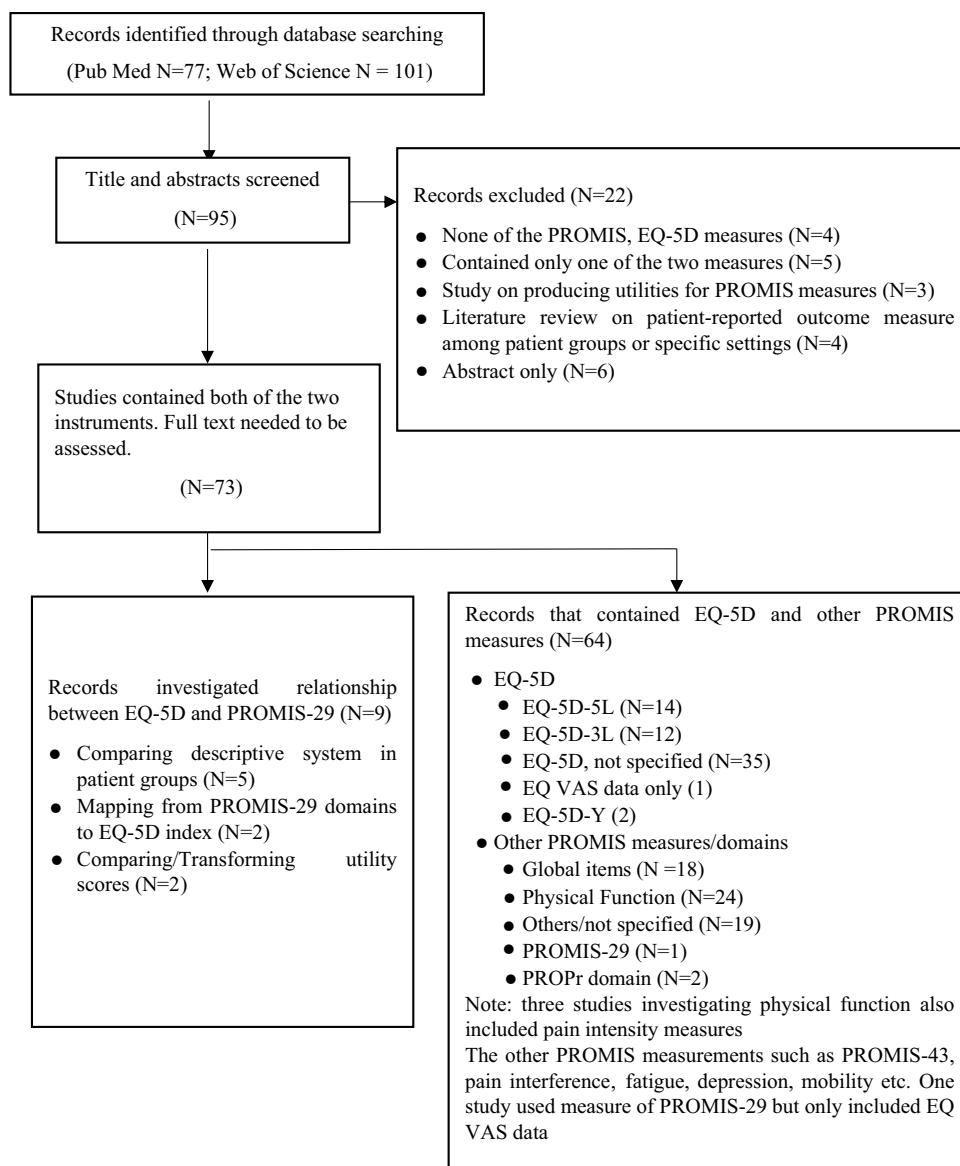


Figure 1 presents the flow chart of the inclusion process of the literature review.

An overview of the evidence

Of the nine included studies, seven were from the United States and two were from Europe and UK. Three of the nine sampled from the general population only, five used patient data and one used one sample from both the general population and patients. The sample size ranged from 20 to 2623. Table 1 provides a summary of the included studies.

Among the six categories of comparison or associations we wished to explore (as listed in our inclusion criteria), we found four have been addressed in the literature on

PROMIS-29 and EQ-5D-3L, and three on PROMIS-29 and EQ-5D-5L (as shown in Table 2). Specifically, five studies collecting patient data using both instruments provided information on known-group validity and or responsiveness [33, 34, 36, 37, 40]. Two studies predicted EQ-5D-3L and EQ-5D-5L utilities from PROMIS-29 domain scores, respectively [28, 38]. There were two studies focussing on utility scores. One compared and transformed the PROMIS-29 utilities (obtained using DCE) to EQ-5D-3L utilities [30]. The other examined the convergent and known-group validity between PROPr utilities and EQ-5D-5L values (using the EQ-5D-3L crosswalk link function that maps the descriptive system of EQ-5D-5L to that of 3L and then using the US TTO value set) [31].

Table 1 A summary of existing evidence

First author	Journal	Year	Preference-weighted patient-reported outcomes		Relationship investigated	Disease or patient group	Country	Sample size
			PROMIS-29	EQ-5D				
Revicki, D. [28]	Quality of life research	2009	PROMIS-29 (version not stated)	EQ-5D-3L	Mapping of PROMIS-29 domain scores to predict EQ-5D utilities	General population and various diseases	US	1658
Hartman, J.D [30]	Quality of life research	2018	PROMIS-29 v1.0	EQ-5D-3L	Comparing descriptive systems; comparing utility scores; scale transformation from PROMIS-29 utilities to EQ-5D utilities	General population	US	2623
Hammer, J. [31]	Plos one	2018	PROMIS-29 and cognitive function 4-item short-form	EQ-5D-5L	Comparing utility scores;	General population	US	983
Khanna, D [33]	Arthritis care & research	2019	PROMIS-29 v2.0	EQ-5D-5L ^a	Comparing descriptive systems	Patients with systemic sclerosis	US	267
Murphy, L [34]	Journal of scleroderma and related disorders	2020	PROMIS-29 v2.0	EQ-5D-5L	Comparing descriptive systems	Patients with systemic sclerosis	US	267
Wyngaert, K. [36]	BMC nephrology	2020	PROMIS-29 v2.0	EQ-5D-3L	Comparing descriptive systems	Patients with end-stage kidney disease	Belgium	113
Dunn, S. [37]	Journal of cardiovascular nursing	2020	PROMIS-29 (version not stated)	EQ-5D-5L	Comparing descriptive systems	Patients with heart disease and moderate to severe hopelessness	US	20
Klapproth, C. P. [38]	Health and quality of life outcomes	2020	PROMIS-29 v2.0	EQ-5D-5L	Mapping of PROMIS-29 domain scores to predict EQ-5D utilities	General population	UK, France, Germany	1509 (UK) 1501 (France) 1502 (Germany)
Kempton, C [40]	Haemophilia	2021	PROMIS-29 v2.0	EQ-5D-5L	Comparing descriptive systems	Patients with haemophilia	US	126

^aIn Khanna et al. [33], the authors did not specify which EQ-5D version was used in the study. However, in their related study and publication by Murphy et al. [34], the author stated that EQ-5D-5L was used. Some studies did not specify which version of PROMIS-29 instrument was used

Key findings

Comparing PROMIS-29 and EQ-5D as descriptive systems

Six studies compared the health states described by PROMIS-29 and EQ-5D. One US study provided a descriptive analysis of PROMIS-29 and EQ-5D-3L responses from the general population (2623 adults). Hartman and Craig [30] found that, in their sample, 74 out of 243 possible health states for EQ-5D-3L were observed, and the two most prevalent health states accounted for more than 55% of the sample (43% of the sample reported no problems on five dimensions); there were 1670 health states out of the nearly 0.4 sextillion possible health states ($5^{28} \times 11$) for PROMIS-29 and the ten most prevalent ones accounted for less than 10% (4% of the sample reported no problems on the 29 items). It indicates that EQ-5D-3L has a large ceiling effect and may not be sensitive to mild health problems, particularly among those in good health.

We found five published studies that did not directly examine the relationship between the two instruments but reported PROMIS-29 domain scores, EQ-5D level of problems and/or EQ-5D utilities among different patient groups. Wyngaer et al. examined the association between measures of physical function, risk of falls and quality of life using PROMIS-29 (T-scores) and EQ-5D-3L (index score and level of problems) among patients on maintenance haemodialysis in Belgium. There was a statistically significant association between physical health measures and EQ-5D-3L index score. Patients with severely impaired functional performance showed statistically significant lower scores on all dimensions of EQ-5D-3L except for anxiety/depression, but only on one PROMIS-29 domain (physical function) [36].

The remaining four studies used EQ-5D-5L. Khanna and colleagues reported PROMIS-29 domain scores and EQ-5D-5L health states and utilities between treatment and control groups before and after an intervention using patient data [33]. Among patients with longer disease duration, there was a statistically significant difference between the treatment and control groups in the physical function domain scores for PROMIS-29 but not in the levels of mobility on the EQ-5D-5L; and there was a statistically significant difference in levels of pain and discomfort dimension for EQ-5D-5L but was not found in pain-related domains for PROMIS-29. In a more recent study, using the same data, Khanna and colleagues examined the association between baseline PROMIS-29 fatigue and follow-up social participation, physical function and overall quality of life measured by EQ-5D-5L index scores. They found baseline fatigue only predicted social participation scores 16 weeks later, and there was negative associations [34]. Dunn et al. tested the reliability and validity of the State-Trait Hopelessness Scale in patients with heart disease and moderate to severe

hopelessness using PROMIS-29 and EQ-5D-5L. The authors reported the EQ-5D-5L level sum scores and found that State Hopelessness Subscale was positively associated with the summary score, and some of the PROMIS-29 domains (depression, fatigue and social roles) [37]. Kempton et al. developed a new instrument (HRDq) for measuring disease-related distress among haemophilia patients, and EQ-5D-5L and PROMIS-29 were used to assess the validity of HRQq. The author reported correlation between EQ-5D-5L index score and HRDq total scores and between PROMIS-29 domain and HRDq. There were moderate correlations in the expected directions [40].

Mapping PROMIS-29 domains to EQ-5D utility score

Revicki et al. [28] selected five PROMIS domains (physical function, fatigue, pain impact, anxiety, and depression) and mapped the T-scores for the five domains to EQ-5D-3L utility based on a sample of 1658 respondents in the US, using a linear ordinary least square regression model. The model explained 57% of the variance in the EQ-5D utility score. Regression coefficients for physical function ($b = 0.0077$), fatigue ($b = -0.0021$), pain impact ($b = -0.0040$), anxiety ($b = -0.0023$) and depression ($b = -0.0022$) were all statistically significant at 1% level in the model.⁴ These results indicate that there are conceptual overlaps in the health domains covered in the two instruments. Fatigue is important in EQ-5D utility scores but is not explicitly measured in the descriptive system. It was unclear that what the remaining 43% of the variance in EQ-5D utility scores was explained by. The authors examined the performance of the model by testing the agreement between predicted and actual EQ-5D utility scores and assessing known-group validity. The results suggested good fit of the model but there was some overestimation for EQ-5D scores under 0.40.

More recently, Klapproth et al. collected EQ-5D-5L and PROMIS-29 profiles from general populations in the UK, France and Germany to find the best models to predict the EQ-5D-5L crosswalk from all seven PROMIS-29 domains [38]. The authors applied stepwise regression with backward selection, starting with full models that incorporated linear, quadratic and cubic effects for all seven PROMIS-29 domains and age and sex, for each included country separately. In addition, the authors compared Revicki's model and their optimal models by applying all models to the European data. Besides the country-specific coefficients, the

⁴ It is worth noting that the coefficient for Physical Function is positive, whereas the other coefficients are negative. This is because in PROMIS system, for positively worded concepts like Physical Function, higher domain raw scores and T-scores represent better health; negatively worded concepts like Anxiety and Pain, lower scores indicate better health.

Table 2 Types of the relationships that have been investigated in literature

Type of relationship	PROMIS-29 & EQ-5D-3L	PROMIS-29 & EQ-5D-5L
Comparing descriptive systems	Yes	Yes
Response mapping	No	No
Mapping of PROMIS-29 domain to EQ-5D utility scores	Yes (model development and validation)	Yes (model development and in-sample cross-validation; using crosswalk value set)
Comparing characteristics of the value sets	No	No
Comparing utility scores using patient/population data	Yes	Yes
Scale transformation	Yes (model development)	No

main differences between Revicki's model and Klapproth's model is the latter incorporated two additional PROMIS-29 domains (social roles and sleep disturbance) and higher-order effects for health state utilities prediction. It is reported that the application of Revicki's model to the European data would systematically underestimate the EQ-5D-5L crosswalk for the UK and France but not for Germany, while Revicki's model performs well in upper ranges of health. The authors highlighted that due to the country-specific nature of health state utility, mapping algorithms for health state utility should not be generalised across countries.

Comparing PROMIS-29 utility score to EQ-5D utility score

Characteristics and correlation of the utility scores Two studies from the US provided descriptive statistics on the range and distribution of the utility scores using data from the general population, both examined convergent validity [30, 31] and one also assessed known-group validity [31]. Although the two studies used different value sets for PROMIS-29 and different EQ-5D measures,⁵ both showed that PROMIS-29 utility score had a wider range and was more granular than the EQ-5D utility score. There were discontinuities in distribution and ceiling effects for EQ-5D utility score, while PROMIS-29 utility score did not have a ceiling effect. It indicates that there is a higher chance to report mild health problems rather than full health in PROMIS-29 compared to EQ-5D utility scores. Although the PROMIS-29 utility score was much lower than that of EQ-5D, PROMIS-29 and its value sets might be more sensitive to capture changes in utilities, particularly for mild health problem. The correlation coefficients between the

actual utility scores ranged from 0.70 to 0.77 [30, 31]. Hanmer et al. (2018) examined known-group validity showed that both EQ-5D-5L value sets and PROPr can distinguish between socio-demographic groups and the presence of self-reported chronic conditions [31]. However, EQ-5D-5L and PROPr provided different utility estimates across different groups.

Scale transformation Hartman and Craig [30] developed a utility scale transformation of PROMIS-29 utility to EQ-5D-3L utility, but we have not found any study validating this model. Using a sample of 2623 adults from the US, they transformed PROMIS-29 utility loss (i.e. 1 minus the utility value) to EQ-5D utility loss. The best fit model suggested that EQ-5D-3L utility loss equals $0.1784 \times (\text{PROMIS-29 loss})^{0.7286}$.⁶ The authors also assessed external validity between EQ-5D-3L and rescaled PROMIS-29 utility by comparing to general health responses (EQ VAS scores and 5-level self-rated health), and they found that with the worsening of self-reported health, the mean utility loss of both instruments increased more rapidly (which means utility score decreased more rapidly) as might be expected.

Discussion

This review found only nine studies that investigated the relationship between PROMIS-29 and EQ-5D, five of which have been published very recently (in 2020 or 2021). There are three main findings of note from the literature. First, PROMIS-29 and EQ-5D instruments have clearly different descriptive systems. There are overlaps but also divergence

⁵ Hartman et al. (2018) used a DCE value set for PROMIS-29 [9] and a TTO value set for EQ-5D-3L [29]; Hanmer et al. (2018) used PROPr and the EQ-5D crosswalk value set [32]. As we mentioned earlier, there are two studies that used 14 questions (i.e. not PROMIS-29) to calculate PROPr [26, 27], the authors compared PROPr with EQ-5D-5L utilities using the US value set for the latter [35].

⁶ In Hartman and Craig (2018), the authors tested four models: linear, non-linear power model, polynomial (quadratic and cubic), and two-part piecewise linear models and tested the goodness-to-fit for each model based on root mean square error (RMSE). Based on this criterion, the best fit was the non-linear power function as it had the smallest RMSE (0.0984).

in the health domains measured.⁷ PROMIS-29 captures a broader range of health states and does not have ceiling effects compared to EQ-5D-3L, indicating that PROMIS-29 may have more descriptive power. However, there is little evidence on the relative merit of PROMIS-29 over the EQ-5D-5L, given the reduced ceiling effects in the latter compared with its EQ-5D-3L predecessor. Second, mapping from PROMIS-29 domains to predict EQ-5D-3L and EQ-5D-5L utilities is feasible. However, the evidence on the performance of the prediction model is mixed, and this has implications for using the predicted utility score. Third, PROMIS-29 utilities and EQ-5D (3L and 5L) utilities are highly correlated, but PROMIS-29 utilities are much lower than EQ-5D utilities. The implication of this for their use in cost-effectiveness analysis is not obvious, as the focus is on the incremental *changes* in QALYs, so some of the differences in utilities between the instruments may cancel out. It is possible to rescale PROMIS-29 utilities to EQ-5D-3L utilities, but we have not found external studies to assess the model performance.

The body of evidence is surprisingly small, given the widespread use of EQ-5D in the health utility field and beyond, and the growing importance of PROMIS-29 as a concise multidimensional preference-accompanied measure [14, 41]. The number of studies is also small compared to the number of studies exploring the relationship among alternative widely used preference-accompanied PROMs, such as SF-6D, Quality of Well-Being Scale, the Health Utilities Index (HUI) [e.g. 42, 43–46]. For the comparison between PROMIS-29 and EQ-5D-3L, we did not find any study examining the correlations between the domain responses of the two measures and no study conducted response mapping. There is an increasing number of studies including both PROMIS-29 and EQ-5D-5L; however, there is currently a dearth of literature regarding the value set characteristics and scale transformation. This is particularly important, given the growing evidence on the superiority of the EQ-5D-5L over EQ-5D-3L as a descriptive system, reflected in increasing numbers of clinical trials and other studies opting to use the EQ-5D-5L rather than the three-level version [47–52].

There are several methodological considerations arising from the existing literature. First, evidence on the performance of the prediction model of EQ-5D-3L utilities from PROMIS-29 domain scores is mixed. Recent studies pointed out that the US model developed by Revicki et al. [28] was

regression-based, which was not suitable for mapping due to regression to the mean (which is a common problem across all mapping studies), and the predicted EQ-5D utilities were inaccurate particularly when applying the algorithm to populations from another country [53, 54]. There is no external validation for the recent developed European prediction model of EQ-5D-5L utilities from PROMIS-29 profile. The validation and the choice of model may have an impact on the calculation of QALYs. Second, methodological considerations regarding the choice of value sets for instruments have not been fully addressed in scale transformation. Scale transformation of preference-accompanied PROMs can help to bridge the gap between instruments in economic evaluations. However, the discordance in scale results not only from the difference in instruments used, but also from the value sets chosen for the instruments. Hanmer et al. used the US EQ-5D-3L crosswalk value set for their EQ-5D-5L measure [31] but we know that there are important differences between the values produced by crosswalking between 5 and 3L and EQ-5D-5L value sets [35, 55]. The two available preference weights for PROMIS-29 also differ in several ways, including the domains included and elicitation method. It is important to understand the characteristics of the value set before comparing utility scores using data from the general population or patient group. Particularly, with the availability of US preference weights for the EQ-5D-5L [35], an updated analysis of scale transformation is possible and needed. Third, the recruited participants in the existing literature may not be representative, either of the general population or of specific disease groups. The respondents in one of the US studies using general population samples were more likely to be older and white [31]. Seven of the nine studies were from the US, and these results may not be generalisable to other settings where PROMIS-29 (and other PROMIS measures) have been translated and used [14, 38, 56].

Due to the lack of evidence on the relationship between the two instruments, the implications of instrument choice for the evidence to be considered in the decision-making process is not clear. PROMIS-29 does not have ceiling and floor effect when describing health states and is more sensitive to differences in health states compared to EQ-5D. However, it takes more time to complete the former instrument. That PROMIS-29 and EQ-5D cover some similar but also some different domains which may influence the choice of which to use, e.g. in clinical settings. The fact that PROMIS-29 generates lower utility values compared to EQ-5D might make PROMIS-29 a less attractive instrument to use (for more detail on the differences in utilities see Pan et al. (2020) [11]). However, as noted earlier, the effect of this on estimates of QALY gains is ambiguous. Perhaps the more important limiting factor in use of PROMIS-29 for QALY estimation to inform health technology appraisal is the fact that utilities

⁷ The overlaps in the health domains in PROMIS-29 and EQ-5D are: Anxiety and Anxiety/Depression (the former being domains in PROMIS-29 and the latter being dimensions in EQ-5D); Depression and Anxiety/Depression; Pain Interference and Pain/Discomfort; Physical Function and Mobility; Social Roles and Usual Activities. There is no explicitly corresponding PROMIS-29 domains of Fatigue and Sleep Disturbance in EQ-5D, and no corresponding EQ-5D dimension of Self-care in PROMIS-29.

for it are thus far only available in the US. Researchers collecting PROMIS-29 can estimate utility in multiple ways: generated from PROMIS-29, rescaled EQ-5D utility from PROMIS-29 utility, predicted EQ-5D utility or HUI utility [57] from PROMIS-29 domains with different models; similarly, EQ-5D-5L utilities can be obtained either by crosswalking or by use of EQ-5D-5L value sets. This complicates the assessment of how instruments and utilities compare; it also risks gaming and the choices of utilities being made on the basis of what will yield the most favourable cost-effectiveness results.

Given the gaps in research evidence and limitations of the existing literature, this review has identified several future research areas. First, a direct comparison of PROMIS-29 and EQ-5D (3L and 5L) as the descriptive system should be the starting point to understand the associations and differences between the two measures. The descriptive systems provide information on aspects of health measured by the instruments and the feasibility of mapping approach, which depends on the assumption that there is conceptual overlaps between the descriptive systems [58]. Second, more research work is needed around exploring relationships between PROMIS-29 and EQ-5D-5L, given its use is superseding EQ-5D-3L. Such research includes a comparison of the two instruments as descriptive systems, comparing the characteristics of the value set, validation of the EQ-5D-5L mapping algorithm from PROMIS-29 domains particularly in patient groups, and scale transforming of utility scores using data from the general population and patient populations. In ongoing work, we compare the characteristics of all the theoretical values from both value sets [11], and assess the impact of differences in value sets using patient data [59]. Third, evidence from countries outside the US is also needed. International network and collaboration such as the PROMIS Health Organization International Committee could help coordinate and promote the development, validation and use of PROMIS instruments across countries. Fourth, with the availability of preference weights for PROMIS-29, additional analysis on comparing the utility scores generated from two instruments can be conducted among studies that have already collected data from general populations and patient groups. Lastly, further work is needed to collect and report patient outcomes from both instruments in a variety of clinical contexts in order to improve understanding of the differences between them and the impact of these differences in longitudinal settings.

Conclusion

This review has highlighted the dearth in studies examining the relationship between PROMIS-29 and EQ-5D. Given the widespread use of EQ-5D in the health utility

field and beyond, and the growing use of PROMIS-29, this is an important gap in evidence. Future work is needed to better understand the relationship between the two descriptive systems and their utility scores, so that the implications for the use of evidence from these instruments in reimbursement decisions is clear.

Acknowledgements We acknowledge the EuroQol Research Foundation for the support of this study.

Author contributions ND conceived and designed the study. TP performed the literature search and data analysis. TP and ND wrote the first draft of the paper. BM, RV, RN, AT-D and JH provided advice on the first draft and revised the article critically for important intellectual content. All authors read and approved the final manuscript.

Funding Funding for this study was provided by the EuroQol Research Foundation.

Declarations

Conflict of interest Nancy Devlin, Brendan Mulhern, Rosalie Viney and Richard Norman are members of the EuroQol Group. Janel Hanmer contributed to development of PROPr. Views expressed in this paper are those of the author and not necessarily those of the EuroQol Research Foundation.

References

1. Devlin, N. J., & Brooks, R. (2017). EQ-5D and the EuroQol Group: Past, present and future. *Applied Health Economics and Health Policy*, 15(2), 127–137.
2. Feeny, D., Torrance, G., & Furlong, W. (1996). *Health utilities index*. Lippincott-Raven Press.
3. Horsman, J., Furlong, W., Feeny, D., & Torrance, G. (2003). The health utilities index (HUI): Concepts, measurement properties and applications. *Health and Quality of Life Outcomes*, 1, 54.
4. Ware, J. E., Jr., & Sherbourne, C. D. (1992). The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Medical Care*, 30(6), 473–483.
5. Brazier, J., Roberts, J., & Deverill, M. (2002). The estimation of a preference-based measure of health from the SF-36. *Journal of Health Economics*, 21(2), 271–292.
6. Cella, D., Riley, W., Stone, A., Rothrock, N., Reeve, B., Yount, S., Amtmann, D., Bode, R., Buysse, D., Choi, S., Cook, K., Devellis, R., DeWalt, D., Fries, J. F., Gershon, R., Hahn, E. A., Lai, J. S., Pilkonis, P., Revicki, D., ... Group P. C. (2010). The patient-reported outcomes measurement information system (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *Journal of Clinical Epidemiology*, 63(11), 1179–1194.
7. Health Measures. (2020). List of Adult Measures. Retrieved 27 April, 2020, from <http://www.healthmeasures.net/explore-measurement-systems/promis/intro-to-promis/list-of-adult-measures>
8. Dewitt, B., Feeny, D., Fischhoff, B., Cella, D., Hays, R. D., Hess, R., Pilkonis, P. A., Revicki, D. A., Roberts, M. S., Tsevat, J., Yu, L., & Hanmer, J. (2018). Estimation of a preference-based summary score for the patient-reported outcomes measurement

- information system: The PROMIS((R))-Preference (PROPr) scoring system. *Medical Decision Making*, 38(6), 683–698.
9. Craig, B. M., Reeve, B. B., Brown, P. M., Cella, D., Hays, R. D., Lipscomb, J., Simon Pickard, A., & Revicki, D. A. (2014). US valuation of health outcomes measured using the PROMIS-29. *Value Health*, 17(8), 846–853.
 10. Kennedy-Martin, M., Slaap, B., Herdman, M., van Reenen, M., Kennedy-Martin, T., Greiner, W., Busschbach, J., & Boyle, K. S. (2020). Which multi-attribute utility instruments are recommended for use in cost-utility analysis? A review of national health technology assessment (HTA) guidelines. *The European Journal of Health Economics*, 21(8), 1245–1257.
 11. Pan, T., Mulhern, B., Viney, R., Norman, R., Hanmer, J., & Devlin, N. (2020). How do the utilities for PROMIS-29 and EQ-5D-5L compare? A comparison of PROPr and EQ-5D-5L value sets, and implications for users. In: *37th EuroQol scientific plenary meeting*.
 12. EQ-5D. (2020). Terminology—EQ-5D. Retrieved 19 May, 2020, from <https://euroqol.org/support/terminology/>
 13. Health Measures. (2020). Glossary. Retrieved 19 May, 2002, from <http://www.healthmeasures.net/resource-center/measurement-science/glossary>
 14. Cella, D., Choi, S. W., Condon, D. M., Schalet, B., Hays, R. D., Rothrock, N. E., Yount, S., Cook, K. F., Gershon, R. C., Amtmann, D., DeWalt, D. A., Pilkonis, P. A., Stone, A. A., Weinfurt, K., & Reeve, B. B. (2019). PROMIS((R)) adult health profiles: efficient short-form measures of seven health domains. *Value Health*, 22(5), 537–544.
 15. Tsuchiya, A., Brazier, J., & Roberts, J. (2006). Comparison of valuation methods used to generate the EQ-5D and the SF-6D value sets. *Journal of Health Economics*, 25, 334–346.
 16. Herdman, M., Gudex, C., Lloyd, A., Janssen, M., Kind, P., Parkin, D., Bonsel, G., & Badiu, X. (2011). Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of Life Research*, 20(10), 1727–1736.
 17. Devlin, N., Parkin, D., & Janssen, B. (2020). *Methods for analysing and reporting EQ-5D data*. Springer.
 18. Szende, A., Oppe, M., & Devlin, N. (2007). *EQ-5D value sets: Inventory, comparative review and user guide*. Springer.
 19. Oppe, M., Devlin, N. J., Hout, B., Krabbe, P. F. M., & Charro, F. (2014). A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. *Value in Health*, 17(4), 445–453.
 20. Norman, R., Cronin, P., & Viney, R. (2013). A pilot discrete choice experiment to explore preferences for EQ-5D-5L health states. *Applied Health Economics and Health Policy*, 11(3), 287–298.
 21. Dewitt, B., Jalal, H., & Hanmer, J. (2020). Computing PROPr utility scores for PROMIS® profile instruments. *Value in Health*, 23(3), 370–378.
 22. Brazier, J., Connell, J., Papaioannou, D., Mukuria, C., Mulhern, B., Peasgood, T., Jones, M. L., Paisley, S., O’Cathain, A., Barkham, M., Knapp, M., Byford, S., Gilbody, S., & Parry, G. (2014). A systematic review, psychometric analysis and qualitative assessment of generic preference-based measures of health in mental health populations and the estimation of mapping functions from widely used specific measures. *Health Technology Assessment*, 18(34), vii–viii, xiii–xxv 1–188.
 23. Longworth, L., Yang, Y., Young, T., Mulhern, B., Hernandez-Alava, M., Mukuria, C., Rowen, D., Tosh, J., Tsuchiya, A., & Evans, P. (2014). Use of generic and condition specific measures of health related quality of life in NICE decision making: Systematic review, statistical modelling and survey. *Health Technology Assessment*, 18, 9.
 24. Health Measures. (2019). Preference-based Scores. Retrieved 10 March, 2020, from www.healthmeasures.net/score-and-interpret/calculate-scores/preference-based-scores
 25. Dewitt, B., Yu, L., & Hanmer, J. (2020). Using item response theory to estimate EQ-5D-5L utilities from PROMIS, Society for Medical Decision Making 42nd Annual Meeting. <https://smdm.confex.com/smdm/2020/meetingapp.cgi/Paper/13540>
 26. Hanmer, J., DeWalt, D. A., & Berkowitz, S. A. (2021). Association between food insecurity and health-related quality of life: A nationally representative survey. *Journal of General Internal Medicine*. <https://doi.org/10.1007/s11606-020-06492-9>
 27. Hanmer, J. (2020). Cross-sectional validation of the PROMIS-preference scoring system by its association with social determinants of health. *Quality of Life Research*. <https://doi.org/10.1007/s11136-020-02691-3>
 28. Revicki, D. A., Kawata, A. K., Harnam, N., Chen, W. H., Hays, R. D., & Cella, D. (2009). 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. *Quality of Life Research*, 18(6), 783–791.
 29. Shaw, J. W., Johnson, J. A., & Coons, S. J. (2005). US valuation of the EQ-5D health states: Development and testing of the D1 valuation model. *Medical Care*, 43(3), 203–220.
 30. Hartman, J. D., & Craig, B. M. (2018). Comparing and transforming PROMIS utility values to the EQ-5D. *Quality of Life Research*, 27(3), 725–733.
 31. Hanmer, J., Dewitt, B., Yu, L., Tsevat, J., Roberts, M., Revicki, D., Pilkonis, P. A., Hess, R., Hays, R. D., Fischhoff, B., Feeny, D., Condon, D., & Cella, D. (2018). Cross-sectional validation of the PROMIS-preference scoring system. *PLoS One*, 13(7), e0201093.
 32. The Economics Network. Available from: https://www.economicnetwork.ac.uk/health/EQ_5D_index_calculator.xls. Retrieved 29 May, 2020.
 33. Khanna, D., Serrano, J., Berrocal, V. J., Silver, R. M., Cuencas, P., Newbill, S. L., Battyany, J., Maxwell, C., Alore, M., Dyas, L., Riggs, R., Connolly, K., Kellner, S., Fisher, J. J., Bush, E., Sachdeva, A., Evnin, L., Raisch, D. W., & Poole, J. L. (2019). Randomized controlled trial to evaluate an internet-based self-management program in systemic sclerosis. *Arthritis Care & Research*, 71(3), 435–447.
 34. Murphy, S. L., Whibley, D., Kratz, A. L., Poole, J. L., & Khanna, D. (2020). Fatigue predicts future reduced social participation, not reduced physical function or quality of life in people with systemic sclerosis. *Journal of Scleroderma and Related Disorders*. <https://doi.org/10.1177/2397198320965383>
 35. Pickard, A. S., Law, E. H., Jiang, R., Pullenayegum, E., Shaw, J. W., Xie, F., Oppe, M., Boyle, K. S., Chapman, R. H., Gong, C. L., Balch, A., & Busschbach, J. J. V. (2019). United States valuation of EQ-5D-5L health states using an international protocol. *Value Health*, 22(8), 931–941.
 36. VandenWyngaert, K., Van Craenenbroeck, A. H., Eloit, S., Calders, P., Celie, B., Holvoet, E., & Van Biesen, W. (2020). Associations between the measures of physical function, risk of falls and the quality of life in haemodialysis patients: A cross-sectional study. *BMC Nephrology*, 21(1), 7.
 37. Dunn, S. L., DeVon, H. A., Buursma, M. P., Boven, E., & Tintle, N. L. (2020). Reliability and validity of the state-trait hopelessness scale in patients with heart disease and moderate to severe hopelessness. *Journal of Cardiovascular Nursing*, 35(2), 126–130.
 38. Klapproth, C. P., van Bebber, J., Sidey-Gibbons, C. J., Valderas, J. M., Leplege, A., Rose, M., & Fischer, F. (2020). Predicting EQ-5D-5L crosswalk from the PROMIS-29 profile for the United Kingdom, France, and Germany. *Health and Quality of Life Outcomes*, 18(1), 389.
 39. van Hout, B., Janssen, M. F., Feng, Y.-S., Kohlmann, T., Busschbach, J., Golicki, D., Lloyd, A., Scalone, L., Kind, P., & Pickard, A. S. (2012). Interim scoring for the EQ-5D-5L: Mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value in Health*, 15(5), 708–715.

40. Kempton, C. L., Michaels Stout, M., Barry, V., Figueroa, J., Buckner, T. W., Gillespie, S., Ellen Lynch, M., Mattis, S., Whitten, S., & McCracken, C. (2021). Validation of a new instrument to measure disease-related distress among patients with haemophilia. *Haemophilia*, 27(1), 60–68.
41. Yount, S. E., Cella, D., & Blozis, S. (2019). PROMIS(R): Standardizing the patient voice in health psychology research and practice. *Health Psychology*, 38(5), 343–346.
42. Richardson, J., Iezzi, A., Khan, M. A., & Maxwell, A. (2014). Validity and reliability of the assessment of quality of life (AQoL)-8D multi-attribute utility instrument. *Patient*, 7(1), 85–96.
43. Rowen, D., Brazier, J., Tsuchiya, A., & Alava, M. H. (2012). Valuing states from multiple measures on the same visual analogue scale: A feasibility study. *Health Economics*, 21(6), 715–729.
44. Whitehurst, D. G., Bryan, S., & Lewis, M. (2011). Systematic review and empirical comparison of contemporaneous EQ-5D and SF-6D group mean scores. *Medical Decision Making*, 31(6), E34–44.
45. Dakin, H., Abel, L., Burns, R., & Yang, Y. (2018). 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 and Quality of Life Outcomes*, 16(1), 31.
46. Whitehurst, D. G. T., Stirling, B., & Lewis, M. (2011). Systematic review and empirical comparison of contemporaneous EQ-5D and SF-6D group mean scores. *Medical Decision Making*, 31(6), E34–44.
47. Thompson, A. J., & Turner, A. J. (2020). A Comparison of the EQ-5D-3L and EQ-5D-5L. *Pharmacoeconomics*. <https://doi.org/10.1007/s40273-018-0623-8>
48. Feng, Y., Devlin, N., & Herdman, M. (2015). Assessing the health of the general population in England: How do the three- and five-level versions of EQ-5D compare? *Health and Quality of Life Outcomes*, 13, 171.
49. Buchholz, I., Janssen, M. F., Kohlmann, T., & Feng, Y. S. (2018). A systematic review of studies comparing the measurement properties of the three-level and five-level versions of the EQ-5D. *Pharmacoeconomics*, 36(6), 645–661.
50. Janssen, M. F., Pickard, A. S., Golicki, D., Gudex, C., Niewada, M., Scalone, L., Swinburn, P., & Busschbach, J. (2013). Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: A multi-country study. *Quality of Life Research*, 22(7), 1717–1727.
51. Janssen, M., Bonsel, G., & Luo, N. (2018). Is EQ-5D-5L better than EQ-5D-3L? A head-to-head comparison of descriptive systems and value sets from seven countries. *Pharmacoeconomics*. <https://doi.org/10.1007/s40273-018-0623-8>
52. Agborsangaya, C. B., Lahtinen, M., Cooke, T., & Johnson, J. A. (2014). Comparing the EQ-5D 3L and 5L: Measurement properties and association with chronic conditions and multimorbidity in the general population. *Health and Quality of Life Outcomes*, 12, 74.
53. Klapproth, C. P., Leplege, A., Gibbons, C., Coste, J., Valderas, J. M., Rose, M., & Fischer, F. (2017). Predicting EQ-5D index scores from PROMIS Profile 29 in the United Kingdom, France, and Germany. *Quality of Life Research*, 26(1), 49–49.
54. Thompson, N. R., Lapin, B. R., & Katzan, I. L. (2017). Mapping PROMIS global health items to EuroQol (EQ-5D) utility scores using linear and equipercentile equating. *Pharmacoeconomics*, 35(11), 1167–1176.
55. Mulhern, B., Feng, Y., Shah, K., Janssen, M. F., Herdman, M., van Hout, B., & Devlin, N. (2018). Comparing the UK EQ-5D-3L and English EQ-5D-5L value sets. *Pharmacoeconomics*, 36(6), 699–713.
56. Health Measures. (2020). Available translations: PROMIS® measures are available in languages other than English. Retrieved 29 May, 2020, from <http://www.healthmeasures.net/explore-measurement-systems/promis/intro-to-promis/available-translations>
57. Hays, R. D., Revicki, D. A., Feeny, D., Fayers, P., Spritzer, K. L., & Cella, D. (2016). Using linear equating to map PROMIS global health items and the PROMIS-29 V. 2 profile measure to the health utilities index—mark 3. *Pharmacoeconomics*, 34(10), 1015–1022.
58. Brazier, J. E., Yang, Y., Tsuchiya, A., & Rowen, D. L. (2010). A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. *The European Journal of Health Economics*, 11(2), 215–225.
59. Mulhern, B., Pan, T., Norman, R., Tran-Duy, A., Hanmer, J., Viney, R., & Devlin, N. (2021). Are generic measures of health related quality of life comparable? Understanding the measurement relationship between EQ-5D-5L, PROMIS-29 and PROPr. In: *EuroQol Academy Virtual Meeting*, 2021.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.