





REVIEW

Open Access



# The psychometric properties of the Quality of Life in Neurological Disorders (Neuro-QoL) measurement system in neurorehabilitation populations: a systematic review

Rebecca Ataman<sup>1,2,3</sup> , Rehab Alhasani<sup>4</sup>, Line Auneau-Enjalbert<sup>1,2,5</sup>, Adria Quigley<sup>6,7</sup> ,  
Henry Ukachukwu Michael<sup>8,9</sup>  and Sara Ahmed<sup>1,2,5,10\*</sup> 

## Abstract

**Objective** To systematically review the literature of existing evidence on the measurement properties of the Quality of Life in Neurological Disorders (Neuro-QoL) measurement system among neurorehabilitation populations.

**Data sources** The Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) guided this systematic review in which we searched nine electronic databases and registries, and hand-searched reference lists of included articles.

**Study selection** Two independent reviewers screened selected articles and extracted data from 28 included studies.

**Data extraction** COSMIN's approach guided extraction and synthesizing measurement properties evidence (insufficient, sufficient), and the modified GRADE approach guided synthesizing evidence quality (very-low, low, moderate, high) by diagnosis.

**Data synthesis** Neuro-QoL has sufficient measurement properties when used by individuals with Huntington's disease, Multiple Sclerosis, Parkinson's disease, stroke, lupus, cognitive decline, and amyotrophic lateral sclerosis. The strongest evidence is for the first four conditions, where test-retest reliability, construct validity, and responsiveness are nearly always sufficient (GRADE: moderate-high). Structural validity is assessed only in multiple sclerosis and stroke but is often insufficient (GRADE: moderate-high). Criterion validity is sufficient in some stroke and Huntington's disease domains (GRADE: high). Item response theory analyses were reported for some stroke domains only. There is limited, mixed evidence for responsiveness and measurement error (GRADE: moderate-high), and no cross-cultural validity evidence.

**Conclusions** Neuro-QoL domains can describe and evaluate patients with Huntington's disease, multiple sclerosis, Parkinson's disease, and stroke, but predictive validity evidence would be beneficial. In the other conditions captured in this review, a limited number of Neuro-QoL domains have evidence for descriptive use only. For these conditions,

\*Correspondence:  
Sara Ahmed  
sara.ahmed@mcgill.ca

Full list of author information is available at the end of the article

further evidence of structural validity, measurement error, cross-cultural validity and predictive validity would enhance the use and interpretation of Neuro-QoL.

**Keywords** Patient reported outcome measures, Nervous system diseases, Psychometrics

## Introduction

The need for adult rehabilitation is expected to increase as people live longer, and as the prevalence of chronic disease and disability rises [1]. In 2019, about 255 million people globally required neurological rehabilitation, making individuals with neurological conditions one of the largest patient groups requiring rehabilitation services. Furthermore, individuals with neurological conditions have one of the highest ‘disability weights’ or calculated disease severity [2]. Therefore, it is imperative that these individuals receive the best possible rehabilitation care. Targeted and robust patient-reported outcome measures can positively impact patient care by improving patient-provider communication and offering standardized assessments to identify areas to target in rehabilitation and monitor outcomes [3, 4].

The Quality of Life in Neurological Disorders (Neuro-QoL) measurement system is a patient-reported measurement system that assesses health-related quality of life in populations with neurological conditions, especially stroke, multiple sclerosis, amyotrophic lateral sclerosis, Parkinson’s disease, epilepsy, and muscular dystrophy [5, 6]. Neuro-QoL can be used to optimize patient care by obtaining a patient’s experience of their symptoms, treatment side effects, functioning, and well-being. In 2011, Neuro-QoL-version 1.0 was first published, encompassing 17 domains within physical, mental, and social health [5, 6]. Version 2.0 followed in 2015, with changes to some domains and statistics [7]. In each version, domains are structured as either scales or item banks, allowing for efficient measurement through fixed-length short forms and/or computerized adaptive testing (CATs) [8].

Since its initial publication over a decade ago, there has been considerable uptake of Neuro-QoL by researchers and clinicians. There have been at least 200 publications in which authors used Neuro-QoL [9], including those recommending the use of Neuro-QoL as part of a core outcome set [10], standardized survey [11] or within clinical trials [12]. Despite the increasing use of Neuro-QoL, there is no synthesis of its available psychometric evidence to inform evidence-based use in adults receiving rehabilitation care. Existing syntheses of Neuro-QoL are limited, focusing only on specific aspects like a single domain such as fatigue [13], or a particular diagnosis such as traumatic brain injury [14] or Parkinson’s disease [15]. There is a need for a comprehensive synthesis of the current evidence of Neuro-QoL within all neurological rehabilitation populations to highlight psychometric

strengths, weaknesses, and gaps, and inform the use of Neuro-QoL in clinical practice.

Thus, the objective of this review was to systematically review the literature of existing evidence on the measurement properties of Neuro-QoL measures among neuro-rehabilitation populations.

## Methods

This study was part of a larger systematic review encompassing Neuro-QoL and related measurement systems: Patient Reported Outcomes Measurement Information System (PROMIS), Traumatic Brain Injury Quality of Life (TBI-QoL) and Spinal Cord Injury Quality of Life (SCI-QoL). While we originally intended to report our findings from all these measurement systems in one consolidated manuscript, the volume of information precluded this option. Thus, here we focus on the measurement properties of the Neuro-QoL according to the CONsensus-based Standards for the selection of health Measurement Instruments (COSMIN) 2018 guidelines [16]. COSMIN has not developed comprehensive guidelines for extracting and rating of Rasch analyses. Therefore, the research team applied criteria previously developed and used in an earlier systematic review to inform Rasch synthesis and interpretation [17]. We report this manuscript in accordance with the PRISMA guidelines for systematic reviews [18].

## Literature search and eligibility

Two reviewers independently searched electronic databases (MEDLINE, EMBASE, PsycINFO and HaPI (Ovid), CINAHL (EBSCO), Cochrane Library and Web of Science), and clinical trials registries (ISRCTN Registry and ClinicalTrials.gov) from inception to March 23<sup>rd</sup> 2024. The search strategy (Psychometric properties AND (Neuro-QoL OR PROMIS OR TBI-QoL OR SCI-QoL) AND Rehabilitation Conditions; MEDLINE search strategy in Appendix A) was developed using a measurement properties search filter validated by COSMIN [19] and drawing from other search strategies for COSMIN reviews by the same research team (e.g., [17]).

After deduplication in EndNote X9 [20], two reviewers independently screened titles and abstracts, followed by the full texts of the manuscripts. We included peer-reviewed articles in English or French providing original data on Neuro-QoL, PROMIS, TBI-QoL, or SCI-QoL measurement properties, feasibility, or interpretability among any rehabilitation population—we did not impose restrictions on diagnosis (Table 1). We excluded articles

**Table 1** COSMIN guidelines and extended Rasch criteria for evaluated measurement properties

	Measurement property	Definition	Data management and interpretation <sup>1</sup>
COSMIN definitions of psychometric properties	Content validity	The degree to which the content of a measure is an adequate reflection of the construct to be measured	<i>COSMIN synthesis:</i> Adequate if the development paper reported clear descriptions of the measurement aim, target population, dimensions measured, and item selection process Measure should be comprehensive, comprehensible and relevant according to clinicians/researchers, caregivers and patients
	Structural validity	The degree to which the scores of a measure are an adequate reflection of the dimensionality of the construct to be measured	<i>COSMIN synthesis:</i> Exploratory or confirmatory factor analysis with adequate model fit (e.g., Kaiser-Meyer-Olkin test 0.8–1.0, Bartlett's test significant)
	Internal consistency	The degree of interrelatedness among the items	<i>COSMIN synthesis:</i> Cronbach's alpha(s) $\geq 0.70$ for each unidimensional scale or subscale AND at least low evidence for sufficient structural validity <i>Meta-analysis:</i> Weighted mean and range of results calculated for Cronbach's alpha where possible
	Cross-cultural validity	The degree to which the performance of the items on a translated or culturally adapted measure are an adequate reflection of the performance of the items of the original version	<i>COSMIN synthesis:</i> No important differences found between group factors (such as age, gender, language) in multiple group factor analysis OR no important differential item functioning for group factors (McFadden's R2 < 0.02)
	Reliability	The proportion of the total variance in the measurements which is due to 'true' differences between patients	<i>COSMIN synthesis:</i> Intra-class correlation coefficient or spearman's correlation $\geq 0.70$ OR Rater reliability: > 0.8 AND Rater separation: < 0.2 <i>Meta-analysis:</i> Weighted mean and range of results calculated for ICC where possible
	Measurement error	The systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured	<i>COSMIN synthesis:</i> Smallest detectable change < minimal important change
	Criterion validity	The degree to which the scores of a measure are an adequate reflection of a 'gold standard'	<i>COSMIN synthesis:</i> For predictive validity, prediction should be clinically meaningful
	Construct validity	The degree to which the scores of a measure are consistent with hypotheses based on the assumption that the measure validly assesses the construct to be measured	<i>COSMIN synthesis:</i> The result is in accordance with the hypothesis 1. Correlations with (changes in) instruments measuring similar constructs should be $\geq 0.50$ . 2. Correlations with (changes in) instruments measuring related, but dissimilar constructs should be lower, i.e., 0.30–0.50. 3. Correlations with (changes in) instruments measuring unrelated constructs should be < 0.30. 4. Correlations defined under 1, 2, and 3 should differ by a minimum of 0.10. <i>Meta-analysis:</i> Weighted mean and range of results calculated for correlations to a measure.
	Responsiveness	The degree to which a measure can detect change over time in the construct to be measured	<i>COSMIN synthesis:</i> The result is in accordance with the hypothesis OR area under the curve $\geq 0.70$ . 1. Meaningful changes between relevant (sub) groups (e.g., patients with expected high vs low levels of the construct of interest)
Rasch analysis	Structural validity	Defined by COSMIN as above	<i>COSMIN synthesis:</i> Unidimensional, locally independent, monotonic with adequate model fit
	Person and item reliability and separation	The reproducibility of the person or item in a relative location on the measure (reliability) The spread of high and low performers or easy and hard items within a measure (separation)	<i>Narrative synthesis</i> Person or item separation > 2.0 <sup>2</sup> Person or item reliability > 0.8 <sup>2</sup>
IRT	Structural validity	Defined by COSMIN as above	<i>COSMIN synthesis:</i> Unidimensional, locally independent, monotonic with adequate model fit

*COSMIN* Consensus-based standards for the selection of health measurement instruments, *CTT* classical test theory, *SD* standard deviation, *ICC* intra-class correlation, *TCC* test characteristic curve

<sup>1</sup>See COSMIN manual and a priori hypotheses (Appendix B) for full details

<sup>2</sup>Boone WJ, Noltemeyer A. Rasch analysis: A primer for school psychology researchers and practitioners *Cogent Educ.* 2017;4(1):1416898

that: (1) did not investigate the measurement properties of these measurement systems (e.g., used as an outcome measure only); (2) used these measurement systems to validate another measure [16]; (3) were published before 2004 (this being the year of the first PROMIS publication); (4) were posters or abstracts or (5) pediatric or (6) non-rehabilitation populations (e.g., mental health, focus on surgical modality such as for orthopedic injuries, etc). We resolved disagreements between reviewers by consensus or with another research team member when necessary. We hand-searched reference lists of all included articles for possible inclusion.

#### Data extraction

After training sessions and 2–3 article pilots, two reviewers independently extracted data concerning the methods and results of the estimated measurement properties, study characteristics, and study population using the structured forms from COSMIN [16]. We consulted a third reviewer in the case of disagreement.

#### Data analysis

Two reviewers independently assessed the measurement properties in each study. They rated content validity against COSMIN criteria [16, 21]. All other measurement properties were rated using Terwee and colleagues' standards [22] as "sufficient" (+), "insufficient" (–), or "indeterminate" (?). When the COSMIN criteria for good measurement properties did not include the statistical test being used in the included study, we summarized and reported the measurement properties narratively.

The research team decided a priori that there is no gold standard measure that could be used to assess the criterion validity of Neuro-QoL. We set a priori hypotheses based on recommendations by de Vet and colleagues [23] for testing construct validity and responsiveness (Appendix B).

#### Data synthesis

Based on their clinical experience, and the study and patient characteristics, the research team grouped the studies by measurement system (Neuro-QoL, PROMIS, TBI-QoL, SCI-QoL), domain (e.g., fatigue, stigma, etc.) and diagnosis (i.e., amyotrophic lateral sclerosis, cognitive decline or mild cognitive impairment, Huntington's disease, lupus, stroke, multiple sclerosis, Parkinson's disease, and mixed neurological conditions). In this manuscript, we report the synthesis of Neuro-QoL's measurement properties.

Two reviewers independently summarized the results for each measurement property across studies (i.e., +/-/±/?). We gave an overall "sufficient" (+) or "insufficient" (–) rating if >75% of measurement property results across studies were concurrent. We have an

"inconsistent" (±) rating if no rating exceeded 75% and no appropriate explanation for inconsistency could be given. We gave an "indeterminate" (?) rating if the results did not achieve a sufficient or insufficient score (i.e., greater than 25% but less than 75% sufficient ratings).

#### Quality assessment

Two independent reviewers assessed the methodological quality of individual studies using the COSMIN risk of bias checklist [16, 24]. Each checklist item is rated as "very good", "adequate", "doubtful" or "inadequate". The overall rating of the methodological quality for a measurement property was based on the worst item rating [16, 24].

Two independent reviewers then graded the quality of evidence for each property per subgroup using the COSMIN modified Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach [16, 25]. The quality of evidence was rated "high", "moderate", "low", or "very low" after considering risk of bias, inconsistency, imprecision, and indirectness.

#### Results

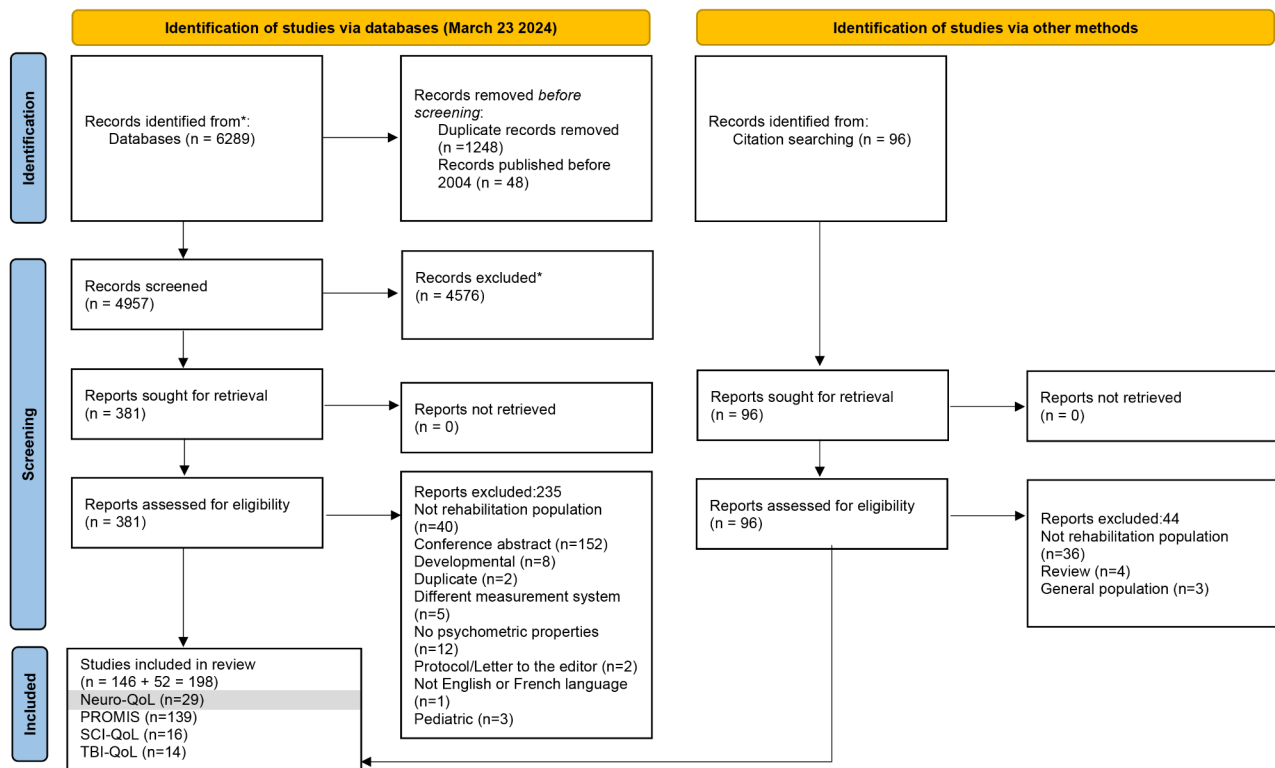
We retrieved a total of 6289 articles and 4957 articles remained following deduplication. Title and abstract screening resulted in 381 included articles. The full-text screen resulted in 146 included articles and reference checks resulted in an additional 52 included articles for a total of 198 included articles. Twenty-nine of these were Neuro-QoL articles and of these, three were identified in the reference checks (Fig. 1).

Authors from all but two studies [26, 27] collected data from the United States only. In alphabetical order, patient populations included amyotrophic lateral sclerosis (n=3), cognitive decline or mild cognitive impairment (n=2), Huntington's disease (n=8), lupus (n=1), mixed neurological conditions (n=3), multiple sclerosis (n=8), Parkinson's disease (n=4) and stroke (n=8). Some studies included a population with mixed diagnoses. The full study and population characteristics can be found within the extraction table (Appendix C). We divided the results in each Neuro-QoL domain according to diagnosis (Table 2, Appendix D).

#### Content Validity of Neuro-QoL

The original development of Neuro-QoL included Alzheimer's disease, multiple sclerosis, amyotrophic lateral sclerosis, Parkinson's disease, stroke, and adult and pediatric epilepsy conditions, and has been described extensively [5, 28–32]. The results from these studies indicate that Neuro-QoL domains possess sufficient content validity, encompassing comprehensibility, relevance, and comprehensiveness, based on high-quality evidence across these diagnoses. Since the initial Neuro-QoL

PRISMA 2020 flow diagram for new systematic reviews which included searches of databases, registers and other sources



PRISMA 2020 flow diagram for new systematic reviews which included searches of databases, registers and other sources

## \*Exclusion Reasons:

- Not psychometric properties (n=2404)
- Not Rehabilitation population (n=820)
- Protocol (n=90)
- Paediatric (n=497)
- Review (n=138)
- Conference abstract (n=82)
- Developmental (n=36)
- Translation cultural validation (n=24)
- Animal study (n=7)
- Consensus statement (n=2)
- Guidelines (n=1)
- Not English or French language (n=5)
- Not PROMIS (n=451)
- Letter to the editor (n=2)

Fig. 1 PRISMA 2020 flow diagram

development, only one article added to the evidence for the content validity of Neuro-QoL in rehabilitation. The results of this article indicate sufficient comprehensibility of the Spanish version of all Neuro-QoL domains across neurological conditions based on high-quality evidence from cognitive debriefing with patients and caregivers [27].

### Amyotrophic lateral sclerosis

Evidence from the single article reporting measurement properties for amyotrophic lateral sclerosis indicates that the Neuro-QoL fatigue domain has sufficient construct validity (2 out of 2 tested hypotheses met) based on high-quality evidence [33].

### Cognitive decline or mild cognitive impairment

In individuals at risk for cognitive decline, construct validity for the cognitive function domain (v2.0) is currently indeterminant (2/3 hypotheses met) and

responsiveness sufficient (4/5 hypotheses met) based on moderate quality evidence due to a low sample size (n=76) [34]. The other evidence available for this subgroup is for the executive function domain (v1.0), which was combined with general concerns (v1.0) to form the cognitive function domain in version 2.0. For individuals already assessed with mild cognitive impairment, there is sufficient evidence for construct validity (1/1 hypotheses met) and responsiveness (1/1 hypotheses met) based on high-quality evidence. However, test-retest reliability is currently assessed as insufficient ( $r=0.35$ ) based on low-quality evidence since the authors provided inadequate information on patient stability between tests and they calculated a correlation coefficient as opposed to an intraclass correlation coefficient [35].

### Huntington's disease

For the ability to participate in social roles and activities domain, both the short form and CAT have sufficient









**Table 2** (continued)

Domain	Type	COSMIN				IRT or Rasch							
		# articles	Structural validity	Internal consistency	Reliability	Measurement error	Cross-cultural validity	Criterion validity	Construct validity	Responsiveness	Structural validity	Person, item reliability, separation	Cross-cultural validity
Ability to participate in social roles and activities	Item bank	1									?, M		
Anxiety	SF	1			±; H								
Cognitive function	SF	3	±; H	±; H	±; H		±; H						
Cognitive function, executive function (v1.0)	SF	1			±; H								
Cognitive function, general cognitive concerns (v1.0)	SF	1			±; H								
Depression	SF	1			±; H								
Emotional and behavioural dyscontrol	SF	1			±; H								
Fatigue	SF	1			±; H								
Lower extremity function—mobility	SF	1			±; H								
Positive affect and well-being	SF	1			±; H								
Positive psychological function	Item bank	1											
Satisfaction with social roles and activities	SF	2			±; H								
Satisfaction with social roles and activities	Item bank	1											
Sleep disturbance	SF	2	±; H	±; H	±; H								
Stigma	SF	1			±; H								
Upper extremity	SF	1			±; H								

CAT computerized adaptive test, SF short form

GRADE approach: H high, M moderate, L low; V/L very low

Measurement property rating: + Sufficient; – Insufficient; ? Indeterminant

Person/Item/Rel. person/item reliability, IRT item response theory

\*All domains are the current version (v2.0), which have not changed from the original version (v1.0). The exceptions are the Cognitive Function, General Cognitive Concerns (v1.0) and Cognitive Function, Executive Function domains (v1.0). These were combined into Cognitive Function (v2.0) in the version 2 update

evidence for test-retest reliability, construct validity (4/4 and 8/11 hypotheses met, respectively), and responsiveness (5/6 and 6/6 hypotheses met, respectively) [36, 37]. Currently, the internal consistency of the short form receives an indeterminate rating due to the lack of structural validity evidence, despite having a Cronbach's alpha greater than 0.7 (specifically 0.94) [37].

For the anxiety CAT, the emotional and behavioral dyscontrol short form and CAT and positive affect and well-being short form and CAT, there is sufficient evidence for test-retest reliability and responsiveness (all 2/2 hypotheses met) based on moderate-high quality evidence (low sample sizes). While the emotional and behavioural dyscontrol CAT has evidence of sufficient measurement error, measurement error for the emotional and behavioural dyscontrol short form, anxiety CAT and positive affect and wellbeing short form and CAT is rated as insufficient because the standard error of measurement (SEM) (3.11, 1.83, 1.72, 2.41 respectively) is greater than the minimal important change (MIC) ( $2.65 \pm 8.22$ ,  $0.62 \pm 6.08$ ,  $-0.91 \pm 5.28$ ,  $-1.25 \pm 6.67$ ) [38].

For the cognitive function (v2.0) domain, both the short form and CAT demonstrate sufficient test-retest reliability and responsiveness (1/1 hypothesis met) [39]. The executive function v1.0 short form demonstrates sufficient criterion and construct validity (3/3 hypotheses met) while the general cognitive concerns short form currently has insufficient (AUC=0.68) and indeterminate evidence (2/3 hypotheses met), respectively [40]. Both of these v1.0 short forms currently lack structural validity evidence, meaning that internal consistency currently receives an indeterminate rating despite both forms having a Cronbach's Alpha greater than 0.7 [40].

The depression CAT has evidence for sufficient test-retest reliability and construct validity (2/2 hypotheses met) based on moderate quality evidence. However, since the SEM (1.85) is greater than the MIC ( $1.68 \pm 8.90$ ), measurement error is currently rated as insufficient based on moderate quality evidence [38].

The lower extremity function—mobility short form and CAT have sufficient evidence of test-retest reliability, measurement error, and responsiveness (9/12 and 10/12 hypotheses met, respectively) based on moderate or high-quality evidence [41, 42]. Currently, the short form is assigned an indeterminate rating for internal consistency due to the absence of evidence for structural validity, even though it has a Cronbach's Alpha greater than 0.7, specifically 0.93 [40]. The CAT has sufficient evidence for criterion validity (3/3 equations with the area under the curve (AUC) greater than 0.7) but indeterminate construct validity (8/11 hypotheses met) based on high-quality evidence [41].

The satisfaction with social roles and activities and stigma short forms and CATs have sufficient evidence

of test-retest reliability and responsiveness (2/2, 9/11 hypotheses met and 2/2, 9/11 hypotheses met, respectively) based on moderate quality evidence due to small sample sizes [36–38]. The satisfaction with social roles and activities short form also has sufficient construct validity based on high-quality evidence (4/4 hypotheses met) [37]. While the stigma CAT has sufficient measurement error, the SEM for the short form is insufficient (SEM=2.54, MIC= $1.43 \pm 4.62$ ) [38].

For the upper extremity domain, both the short form and CAT have sufficient evidence of test-retest reliability based on moderate quality evidence due to small sample sizes [41, 42]. The CAT also has sufficient evidence for measurement error, criterion validity (3/3 AUC>0.7) and construct validity (9/11) based on high-quality evidence. In contrast, the short form currently has insufficient evidence for measurement error (SEM: 3.82, MIC decline:  $-2.87$  (6.89), MIC improvement:  $-0.25$  (7.16)). Furthermore, there is not yet consistent evidence to make a definitive rating for internal consistency (Cronbach's alpha 0.94 but no structural validity evidence) or responsiveness (8/11 hypotheses met).

#### Lupus

In the executive function and general cognitive concerns (v1.0) domains, the short form has sufficient evidence for test-retest reliability [43].

#### Mixed neurological conditions

For a mixed neurological group including individuals with epilepsy, stroke, amyotrophic lateral sclerosis, multiple sclerosis, and Parkinson's disease, the short forms for the ability to participate in social roles and activities, executive function (v1.0), depression, and lower extremity domains have evidence for sufficient construct validity (1/1 hypothesis met) and responsiveness (1/1 hypothesis met) [44]. For the cognitive function (v2.0) domain, there is sufficient evidence for construct validity only (1/1 hypothesis met) [45]. For all of these domains, internal consistency is indeterminate due to a current lack of evidence for structural validity [44, 45].

#### Multiple sclerosis

The total Neuro-QoL scale (i.e., all domains together) has sufficient construct validity based on high-quality evidence (1/1 hypothesis met) [26].

The ability to participate in social roles and activities and anxiety short forms have sufficient evidence of test-retest reliability and construct validity [46, 47]. However, despite Cronbach's alpha being greater than 0.7 (0.89 [46], 0.96 [47] and 0.93 [46], 0.94 [47], respectively), evidence for insufficient structural validity due to both the comparative fit index and Tucker-Lewis index being less than 0.95 and the root mean square error of approximation

greater than 0.06 [47] currently results in an insufficient rating for internal consistency.

The two cognitive function short forms from v1.0 of Neuro-QoL (i.e. executive function and general cognitive concerns) have evidence for sufficient test-retest reliability [46] and construct validity based on high-quality evidence (15/17 hypotheses met) [46, 47]. However, executive function currently has evidence of insufficient structural validity (CFI and TFI < 0.95) [47], thus resulting in an insufficient rating for internal consistency as well [46, 47].

For the communication scale and depression short form, there is sufficient evidence for structural validity [47], internal consistency [46, 47] and construct validity (15/17 hypotheses met) [46, 47]. The depression short form also has sufficient evidence of test-retest reliability based on high-quality evidence [46].

The fatigue short form has sufficient evidence of test-retest reliability and construct validity (13/17 hypothesis met) [47]. However, despite Cronbach's alpha being greater than 0.7 (0.93 [47]), evidence for insufficient structural validity due to both CFI and TLI being less than 0.95 [47] currently results in an insufficient rating for internal consistency too. Measurement error for the fatigue short form is currently indeterminant because the MIC has not been calculated [48].

The emotional and behavioral dyscontrol, fatigue, lower extremity—mobility, positive affect and well-being, sleep disturbance, stigma, and upper extremity domains all have sufficient evidence for test-retest reliability and construct validity (13–16/17 hypotheses met). Current evidence of insufficient structural validity (CFI and TLI < 0.95) or a lack of evidence (stigma) currently results in insufficient evidence for internal consistency despite Cronbach's alpha being > 0.7 in all domains [46, 47].

#### **Parkinson's disease**

The anxiety, depression, cognitive function v1.0 forms (executive function and general cognitive concerns), emotional and behavioural dyscontrol, fatigue, lower extremity—mobility, positive affect and wellbeing, satisfaction with social roles and activities, sleep disturbance, stigma, and upper extremity short forms have sufficient evidence for test-retest reliability, measurement error and responsiveness (1/1 hypotheses met) based on moderate to high-quality evidence. For all domains, Cronbach's alpha is > 0.7 but without evidence of structural validity, internal consistency is currently rated as insufficient. All domains but anxiety, depression, and emotional and behavioral dyscontrol have sufficient evidence for construct validity (ranging from 8–10/11 hypotheses met). These three domains currently have indeterminant evidence for construct validity (6–7/11 hypotheses met) [49].

#### **Stroke**

The ability to participate in social roles and activities short form has insufficient evidence of inter-rater reliability between patients and proxy raters (ICC = 0.55) based on high-quality evidence [50]. The ability to participate in social roles and activities item bank has indeterminant structural validity based on moderate quality evidence because information on monotonicity was not reported [51].

The cognitive function (v2.0) [52–54] domain has sufficient evidence for construct validity (3/3 hypotheses met). There is not yet enough evidence for sufficient structural validity (RMSEA > 0.06) [52], inter-rater reliability between patients and proxies (ICC = 0.54) [53] internal consistency (Cronbach's alpha > 0.7 but insufficient evidence for structural validity) [52] or criterion validity (AUC = 0.691) [50]. The executive function (v1.0) and general cognitive concerns (v1.0) domains currently have insufficient evidence for inter-rater reliability between patients and proxies (ICC = 0.56, 0.59, respectively) [50].

The short forms for anxiety, depression, emotional and behavioural dyscontrol, fatigue, mobility, positive affect and wellbeing, satisfaction with social roles and activities, sleep disturbance, stigma and upper extremity have evidence for insufficient inter-rater reliability between patients and proxies (ICC = 0.32–0.53) [50]. The sleep disturbance short form has insufficient structural validity due to RMSEA not meeting cutoffs (RMSEA = 0.12). Due to insufficient structural validity, internal consistency is also insufficient [52].

The positive psychological function and satisfaction with social roles and activities item banks have indeterminant evidence of structural validity because monotonicity was not reported. Positive psychological function person and item reliability meet Rasch criteria (> 0.8) [51].

#### **Discussion**

We performed a systematic review to assess the strength and quality of the measurement properties of Neuro-QoL. We reported the measurement properties of Neuro-QoL according to diagnosis (amyotrophic lateral sclerosis, cognitive decline or mild impairment, Huntington's disease, lupus, mixed neurological conditions, multiple sclerosis, Parkinson's disease and stroke) and domain (e.g., fatigue, stigma, etc.). Consequently, this manuscript serves as a comprehensive reference for researchers and clinicians, offering diagnosis-specific recommendations for the application of Neuro-QoL. All results and associated recommendations are based on COSMIN's rating system according to currently available evidence. COSMIN's requirements are more extensive and stringent than others in the field (e.g., International

Society for Quality of Life Research (ISOQOL) [55], International Society for Pharmacoeconomics and Outcomes Research (ISPOR) [56, 57], HealthMeasures reporting standards [58]). For example, ISOQOL provides guidance concerning the minimal measurement properties and associated statistical tests recommended for the use of patient-reported outcome measures in patient-centered outcome research [55]. COSMIN does the same but goes a step further by providing cut-off scores to rate evidence sufficiency and criteria to assess evidence quality. The ratings and recommendations we made in this review based on COSMIN provide guidance regarding gaps in the available evidence. As such, the recommendations in this review are subject to change as new evidence becomes available.

In most cases, there is only one article providing measurement property evidence for each domain in each diagnosis. However, this limited evidence is strong as most studies' measurement properties were rated sufficient based on high-quality evidence. As a result, our general recommendation is that Neuro-QoL is appropriate for use across various neurological conditions, per HealthMeasures guidelines [59, 60]. Specific recommendations for each diagnosis are detailed in Table 3.

The least amount of evidence is available for the diagnoses of amyotrophic lateral sclerosis, cognitive decline or mild impairment, lupus, and mixed neurological conditions, both in terms of the evidence for different measurement properties and domains. For example, despite amyotrophic lateral sclerosis being a core neurological condition for which Neuro-QoL was developed, the only measurement property evidence available in rehabilitation, aside from content validity, pertains to the construct validity of the fatigue short form [33]. Rehabilitation researchers and clinicians are currently limited in the domains and interpretations they can make based on research evidence when using Neuro-QoL within these populations.

In contrast, Huntington's disease, multiple sclerosis, Parkinson's disease and to a lesser extent, stroke have a range of measurement evidence available across most (though not all,) Neuro-QoL domains. Of these, only Huntington's disease has evidence for CATs [37, 38]. Although evidence for these diagnoses is typically confined to 1–2 articles per domain, these articles often offer high-quality evidence for different measurement properties. Thus, rehabilitation researchers and clinicians have more information available to justify their use and interpretation of Neuro-QoL in these domains. However, a notable gap is the lack of predictive use, such as employing admission scores to forecast discharge or follow-up outcomes. Further research in this area may be particularly fruitful.

There were some patterns in measurement properties ratings across domains or diagnosis within this review which suggest avenues for future research. For example, criterion and construct validity and responsiveness tend to be sufficient in Huntington's disease, multiple sclerosis, Parkinson's disease and stroke. Additionally, across various conditions, it's common to find that Cronbach's alpha exceeds 0.7. However, it is often rated as insufficient, typically due to either a lack of adequate structural validity evidence or the presence of evidence indicating insufficient structural validity [46, 47]. The only instances where there is sufficient structural validity evidence was for general cognitive concerns and depression short forms, and the communication scale when used with individuals with multiple sclerosis [47]. The structural validity of Neuro-QoL domains in rehabilitation populations is a key area for future research.

There is either test-retest (Huntington's disease, multiple sclerosis, Parkinson's disease) or inter-rater (stroke) reliability evidence available. In all populations test-retest reliability was sufficient ( $ICC > 0.7$ ) and inter-rater reliability between patients and proxies was insufficient ( $ICC < 0.7$ ). Inter-rater reliability for all 12 short forms used in stroke may be negatively affected due to stroke's effects on patients' self-awareness, and vision, hearing, and cognitive status, all of which may contribute to patients responding differently from caregivers' [61]. Future research should consider these patient variables so that Neuro-QoL can be interpreted more holistically.

Measurement error evidence was available for Neuro-QoL domains when used in Huntington's disease and Parkinson's disease. While Parkinson's disease has evidence for sufficient construct validity across all domains [49], the evidence is mixed for Huntington's disease because the SEM is often greater than the MIC [38]. Further estimates of both SEM and MIC may clarify this measurement property, allowing researchers and clinicians to interpret change scores more confidently.

There are some Neuro-QoL domains (e.g., bowel function, sexual function, urinary/bladder function [7, 8]) that have no evidence across any of the rehabilitation populations studied. In this case, these are only item pools and no Neuro-QoL measure has yet been developed for these domains. Similarly, there are some instances where an article was included but the evidence was uninterpretable, resulting in no evidence for that diagnosis and domain. For example, an article containing evidence of the construct validity of the Neuro-QoL CAT for stroke patients only reported p values rather than correlations or other statistics providing construct validity evidence [62]. Researchers should consider generating evidence for these Neuro-QoL diagnoses and domains for which there is currently no evidence available in rehabilitation.

**Table 3** Specific recommendations for the use of Neuro-QoL according to the results of this review

Diagnoses	Current evidence indicates can be used to	More evidence required to
<i>Amyotrophic lateral sclerosis (ALS)</i>	Appropriately gather information from this population Describe patient fatigue at admission	Evaluate or predict patient fatigue Describe, evaluate or predict patient outcomes on all other relevant domains
<i>Cognitive decline or mild cognitive impairment</i>	Appropriately gather information from this population Describe and evaluate a patient's change in cognitive function over time	Evaluate or predict patient cognitive function Describe, evaluate or predict patient outcomes on all other relevant domains
<i>Huntington's disease</i>	Appropriately gather information from this population Describe and evaluate a patient's change over time in 10 domains (ability to participate in social roles and activities, anxiety, cognitive function, depression, emotional and behavioral dyscontrol, lower extremity function—mobility, positive affect and wellbeing, satisfaction with social roles and activities, stigma, upper extremity) Distinguish between individuals with cognitive impairment, lower extremity—mobility impairment and upper extremity impairment versus those without	To predict patient outcomes on all domains
<i>Lupus</i>	Appropriately gather information from this population	Describe, evaluate or predict patient outcomes on all relevant domains
<i>Mixed neurological conditions</i>	Appropriately gather information from this population Describe a patient's cognitive function Describe and evaluate a patient's change over time in 4 domains (ability to participate in social roles and activities, cognitive function—executive function, depression, lower extremity function—mobility)	Describe, evaluate or predict patient outcomes on all other relevant domains
<i>Multiple sclerosis</i>	Appropriately gather information from this population Describe and evaluate a patient at a single timepoint in 12 domains (ability to participate in social roles and activities, anxiety, cognitive function, communication, depression, emotional and behavioral dyscontrol, fatigue, lower extremity function—mobility, positive affect and wellbeing, sleep disturbance, stigma, upper extremity)	To predict patient outcomes on all domains
<i>Parkinson's disease</i>	Appropriately gather information from this population Describe and evaluate a patient over time in 11 domains (anxiety, cognitive function, depression, emotional and behavioral dyscontrol, fatigue, lower extremity function—mobility, positive affect and wellbeing, satisfaction with social roles and activities, sleep disturbance, stigma, upper extremity)	To predict patient outcomes on all domains
<i>Stroke</i>	Appropriately gather information from this population Describe a patient's cognitive function and positive psychological function at a single timepoint	Evaluate and predict a patient's cognitive function and positive psychological function Describe, evaluate or predict patient outcomes on all other relevant domains

This Neuro-QoL review is mirrored by complementary systematic reviews that we completed of the [1] interpretability of Neuro-QoL, PROMIS, TBI-QoL and SCI-QoL [2], measurement properties of PROMIS and [3] measurement properties of TBI-QoL and SCI-QoL. These reviews demonstrate that there is some overlap between the use of Neuro-QoL and PROMIS. Specifically, PROMIS has been used by patients with a stroke, multiple sclerosis, Parkinson's disease, Huntington's disease and amyotrophic lateral sclerosis diagnoses. When the same PROMIS and Neuro-QoL domains have been used for one of these patient groups the results are similar, indicating that the overlapping PROMIS and Neuro-QoL domains could potentially be used for these populations. Of note is that there are psychometric properties reported for PROMIS anxiety, depression, fatigue

and physical function CATs when used by individuals with multiple sclerosis [63], while there is no psychometric evidence for Neuro-QoL domain CATs for this group. There is also far more data available concerning the psychometric properties of PROMIS domains rather than Neuro-QoL domains for individuals with lupus—and the available data indicates the measurement properties of the PROMIS domains are sufficient based on high-quality evidence. This may be because, while lupus has neurological effects, it is not a neurological condition but an autoimmune disorder. Consequently, most Neuro-QoL domains may not be appropriate for individual with lupus. The opposite is true for Huntington's disease with a much greater amount of high-quality evidence available for Neuro-QoL domains rather than PROMIS. For both Neuro-QoL and PROMIS domains, there is



little psychometric information available for their use by patients with ALS.

Finally, there were two recurring methodological concerns across studies included in this review which often reduced the methodological quality of evidence for certain measurement properties per COSMIN criteria. First, the authors did not provide hypotheses for construct validity or responsiveness with both magnitude and direction. Per COSMIN guidelines, the authors of this review had to assign testable hypotheses for the evidence from these studies to be interpretable [23]. We recommend that researchers investigating construct validity between Neuro-QoL and other measures use the a priori hypotheses that we developed for this review (Appendix B). We also noted issues with reporting adequate details on the stability of patients between administrations when estimating test-retest reliability and reporting all statistics concerning model fit for Rasch and IRT models (i.e., to support monotonicity, local independence, fit and unidimensionality [16]). Ensuring this information is reported in future manuscripts concerning the measurement properties of Neuro-QoL domains would advance our understanding of this measurement system.

### Strengths and limitations

The subgroups for synthesis in this systematic review were developed with the aim to provide information that is as relevant and specific as possible. However, the evidence base isn't large for each of these subgroups—ranging from 1–3 articles in each. Furthermore, as mentioned earlier, recurring methodological concerns reduced the quality of these studies. While we were able to address some of these concerns, such as setting a priori hypotheses, there remains a need for high-quality research to replicate and expand upon the existing evidence.

Although COSMIN's systematic review guidance is extensive, we had to extend COSMIN's guidance to encompass the Rasch analyses used by some Neuro-QoL studies. Our extensions represent deviations from COSMIN's protocol; however, they can also inform systematic review authors in their future work [24, 64]. Future work linking Rasch estimates to classical test theory statistics within COSMIN would be an important methodological development.

### Conclusion

The evidence suggests that rehabilitation researchers and clinicians can use most Neuro-QoL domains in Huntington's disease, multiple sclerosis, Parkinson's disease and stroke to describe and evaluate patients. There is evidence in a limited number of Neuro-QoL domains for its use as a descriptive measure only in amyotrophic lateral sclerosis, cognitive decline or mild impairment, lupus, and mixed neurological conditions. Since all diagnoses

and domains only have 1–3 articles providing evidence, further investigation of Neuro-QoL measurement properties across neurological conditions in rehabilitation would be beneficial. Evidence of structural validity, measurement error, cross-cultural validity, and predictive validity would advance the use and interpretation of Neuro-QoL in rehabilitation.

### Abbreviations

Neuro-QoL	Quality of Life in Neurological Disorders measurement system
CAT	Computerized adaptive testing
PROMIS	Patient reported outcomes measurement information system
TBI-QoL	Traumatic brain injury quality of life
SCI-QoL	Spinal cord injury quality of life
COSMIN	COnsensus-based standards for the selection of health Measurement Instruments
GRADE	Grading of recommendations assessment, development and evaluation
SEM	Standard error of measurement
MIC	Minimal important change
ISOQOL	International society for quality of life research
ISPOR	International society for pharmacoconomics and outcomes research
MEDLINE	Medical literature analysis and retrieval system online
HaPI	Health and psychological instruments
CINAHL	Cumulative index to nursing and allied health literature
EBSCO	Elton B. Stephens company research database
ISRCTN	International standard randomised controlled trial number
EMBASE	Excerpta medica database

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s41687-024-00743-7>.

Supplementary Material 1  
Supplementary Material 2  
Supplementary Material 3  
Supplementary Material 4

### Acknowledgements

We gratefully acknowledge the contributions of Alia Osman for their support of this review, including in piloting, screening and extraction.

### Author contributions

RA2 and SA conceptualized this review. All authors contributed to the study acquisition, analysis, synthesis and/or interpretation. RA1 drafted the manuscript and all authors critically reviewed and approved the final version for publication.

### Funding

During this work, Rebecca Ataman was supported by a doctoral scholarship from the Centre de recherche interdisciplinaire en Réadaptation (CRIR) and from the Fonds de Recherche Québec—Santé (FRQ-S). Dr. Sara Ahmed is the principal investigator for the Biomedical Research and Informatics Living Laboratory for Innovative Advances of New Technologies (BRILLIANT) in Community Mobility Rehabilitation program which is funded by the Canadian Foundation of Innovation and the Ministry of Health of Québec (#36053), the Initiatives for the Development of New Technologies and Innovative Practices in Rehabilitation—INSPIRE, and the Fonds de Recherche du Québec-Santé. Henry is supported by doctoral scholarship from the Division of Experimental Medicine, Department of Medicine, McGill University (HUM). Dr. Adria Quigley was supported by a Gilead/CIHR Canadian HIV Trials Network Postdoctoral Fellowship Award.

**Data availability**

All data is found in the appendices, as extracted from published research articles.

**Delarations****Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

All authors approve the final version of the manuscript and publication.

**Competing interests**

The authors report no competing interests.

**Author details**

<sup>1</sup>School of Physical and Occupational Therapy, McGill University, Montréal, Québec, Canada

<sup>2</sup>Centre for Interdisciplinary Research in Rehabilitation of Greater Montréal, Montréal, Québec, Canada

<sup>3</sup>Clinical Institutes and Quality Programs, Ontario Health, Toronto, Ontario, Canada

<sup>4</sup>Department of Rehabilitation Sciences, College of Health and Rehabilitation Sciences, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia

<sup>5</sup>Constance Lethbridge Rehabilitation Center, CIUSSS Centre-Ouest de l'île de Montreal, Montreal, Quebec, Canada

<sup>6</sup>School of Physiotherapy, Dalhousie University, Forrest Building, 5869 University Avenue, PO Box 15000, Halifax, Nova Scotia B3H 4R2, Canada

<sup>7</sup>Nova Scotia Health Authority, Nova Scotia Rehabilitation and Arthritis Centre, 1341 Summer St, Halifax, Nova Scotia B3H 4K4, Canada

<sup>8</sup>Division of Experimental Medicine, McGill University, Montreal, Quebec, Canada

<sup>9</sup>Center for Outcomes Research and Evaluation (CORE), Research Institute of the McGill University Health Center, Montreal, Quebec, Canada

<sup>10</sup>Clinical Epidemiology, Center for Outcome Research and Evaluation (CORE), McGill University Health Center Research Institute, Montreal, Quebec, Canada

Received: 15 January 2024 / Accepted: 17 May 2024

Published online: 18 September 2024

**References**

- World Health Organization (2020) Rehabilitation. Available from <https://www.who.int/news-room/fact-sheets/detail/rehabilitation>. 30 Dec 2020
- Cieza A, Causey K, Kamenov K, Hanson SW, Chatterji S, Vos T (2020) Global estimates of the need for rehabilitation based on the Global Burden of Disease study 2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet* 396(10267):2006–2017
- Reitzel SB, Lynning M, Skovgaard L (2022) Neurologists' views on patient reported outcomes in multiple sclerosis care. *Heliyon* 8(6):E09637
- Churrua K, Pomare C, Ellis LA, Long JC, Henderson SB, Murphy LE et al (2021) Patient-reported outcome measures (PROMs): a review of generic and condition-specific measures and a discussion of trends and issues. *Health Expectations* 24(4):1015–1024
- Cella D, Nowinski C, Peterman A, Victorson D, Miller D, Lai JS et al (2011) The neurology quality of life measurement (Neuro-QoL) initiative. *Arch Phys Med Rehabil* 92(Suppl 1):S28–36
- Gershon RC, Lai JS, Bode R, Choi S, Moy C, Bleck T et al (2012) Neuro-QoL: quality of life item banks for adults with neurological disorders: item development and calibrations based upon clinical and general population testing. *Qual Life Res* 21(3):475–486
- National Institute of Neurological Disorders and Stroke (NINDS). User manual for the Quality of Life in Neurological Disorders (Neuro-QoL) measures, Version 2.0
- Health Measures (2022) List of adult measures. Available from <https://www.healthmeasures.net/explore-measurement-systems/neuro-qol/intro-to-neuro-qol/list-of-adult-measures>. 22 Dec 2022
- Health Measures (2022) Neuro-QoL publications. Available from <https://www.healthmeasures.net/explore-measurement-systems/neuro-qol/measure-development-research/publications-neuro-qol>. 22 Dec 2022
- Wilde EA, Whiteneck GG, Bogner J, Bushnik T, Cifu DX, Dikmen S et al (2010) Recommendations for the use of common outcome measures in traumatic brain injury research. *Arch Phys Med Rehabil* 91(11):1650–1660.e17
- van Wulfften Palthe ODR, Janssen SJ, Wunder JS, Ferguson PC, Wei G, Rose PS et al (2017) What questionnaires to use when measuring quality of life in sacral tumor patients: the updated sacral tumor survey. *Spine J* 17(5):636–644
- Katzan IL, Thompson N, Uchino K (2016) Innovations in stroke: the use of PROMIS and NeuroQoL Scales in clinical stroke trials. *Stroke* 47(2):e27–e30
- Sellitto G, Morelli A, Bassano S, Conte A, Baione V, Galeoto G et al (2021) Outcome measures for physical fatigue in individuals with multiple sclerosis: a systematic review. *Expert Rev Pharmacoecon Outcomes Res* 21(4):625–646
- Hunt C, Zahid S, Ennis N, Michalak A, Masanic C, Vaidyanath C et al (2019) Quality of life measures in older adults after traumatic brain injury: a systematic review. *Qual Life Res* 28(12):3137–3151
- Berardi A, Regoli E, Tofani M, Valente D, Fabbri G, Fabbri A et al (2021) Tools to assess the quality of life in patients with Parkinson's disease: a systematic review. *Expert Rev Pharmacoecon Outcomes Res* 21(1):55–68
- Prins CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW et al (2018) COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res* 27(5):1147–1157
- Ataman R, Thomas A, Roberge-Dao J, McKerral M, Auger C, Wittich W et al (2023) Measurement properties of the mayo-portland adaptability inventory (MPAI-4) and related measures: a systematic review. *Arch Phys Med Rehabil* 104(8):1300–1313
- Moher D, Liberati A, Tetzlaff J, Altman DG, Altman D, Antes G et al (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med* 151(4):264–269
- Terwee C, Jansma E, Riphagen I, de Vet HC (2009) Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Health Res* 18(8):1115–1123
- Bramer WM, Giustini D, De Jong GB, Holland L, Bekhuis T (2016) De-duplication of database search results for systematic reviews in endnote. *J Med Libr Assoc* 104(3):240–243
- Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL et al (2010) The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol* 10(1):22
- Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J et al (2007) Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 60(1):34–42
- De Vet HCW, Terwee CB, Mokkink LB, Knol DL (2011) *Measurement in medicine: a practical guide*. Cambridge University Press, Cambridge, UK, 350 p
- Mokkink LB, Boers M, Van Der Vleuten CPM, Bouter LM, Alonso J, Patrick DL et al (2020) COSMIN risk of bias tool to assess the quality of studies on reliability or measurement error of outcome measurement instruments: a Delphi study. *BMC Med Res Methodol* 20(1):1–13
- Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P et al (2008) GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 336(7650):924–926
- Matza LS, Phillips G, Dewitt B, Stewart KD, Cella D, Feeny D et al (2020) A scoring algorithm for deriving utility values from the neuro-QoL for patients with multiple sclerosis. *Med Decis Mak* 40(7):897–911
- Correia H, Pérez B, Arnold B, Wong AW, Lai JS, Kallen M et al (2015) Spanish translation and linguistic validation of the quality of life in neurological disorders (Neuro-QoL) measurement system. *Qual Life Res* 24(3):753–756
- Cell D, Victorson D, Nowinski C, Peterman A, Miller DM (2006) The Neuro-QoL project: using multiple methods to develop a HRQoL measurement platform to be used in clinical research across neurological conditions. *Qual Life Res* 14:1353
- Miller D, Nowinski C, Victorson D, Peterman A, Perez L (2005) The Neuro-QoL project: establishing research priorities through qualitative research and consensus development. *Qual Life Res* 14:2031
- Perez L, Huang J, Jansky L, Nowinski C, Victorson D, Peterman A et al (2007) Using focus groups to inform the Neuro-QoL measurement tool: exploring patient-centered, health-related quality of life concepts across neurological conditions. *J Neurosci Nurs* 39(6):342–353
- Lynch EB, Butt Z, Heinemann A, Victorson D, Nowinski CJ, Perez L et al (2008) A qualitative study of quality of life after stroke: the importance of social relationships. *J Rehabil Med* 40(7):518–523

32. Rao D, Choi SW, Victorson D, Bode R, Peterman A, Heinemann A et al (2009) Measuring stigma across neurological conditions: the development of the stigma scale for chronic illness (SSCI). *Qual Life Res* 18(5):585–595
33. De Marchi F, Berry JD, Chan J, Caldwell S, Ellrodt A, Scalia J et al (2020) Patient reported outcome measures (PROMs) in amyotrophic lateral sclerosis. *J Neurol* 267(6):1754–1759
34. Howland M, Tatsuoka C, Smyth KA, Sajatovic M (2017) Evaluating PROMIS (®) applied cognition items in a sample of older adults at risk for cognitive decline. *Psychiatry Res* 247:39–42
35. Howland M, Allan KC, Carlton CE, Tatsuoka C, Smyth KA, Sajatovic M (2017) Patient-rated versus proxy-rated cognitive and functional measures in older adults. *Patient Relat Outcome Meas* 8(33)
36. Carozzi NE, Boileau NR, Hahn EA, Barton SK, Cella D, McCormack MK et al (2020) Responsiveness to change over time: an examination of the Neuro-QoL social function measures in persons with Huntington's disease. *J Huntingtons Dis* 9(1):83–97
37. Carozzi NE, Hahn EA, Goodnight SM, Kratz AL, Paulsen JS, Stout JC et al (2018) Patient-reported outcome measures in Huntington disease: quality of life in neurological disorders (Neuro-QoL) social functioning measures. *Psychol Assess* 30(4):450
38. Carozzi NE, Boileau NR, Roché MW, Ready RE, Perlmutter JS, Chou KL et al (2020) Responsiveness to change over time and test-retest reliability of the PROMIS and Neuro-QoL mental health measures in persons with Huntington disease (HD). *Qual Life Res* 29(12):3419–3439
39. Carozzi NE, Boileau NR, Hanks RA, Sander AM, Nakase-Richardson R, Massengale JP (2020) Sleep impairment is related to health-related quality of life among caregivers of lower-functioning traumatic brain injury survivors. *Rehabil Psychol* 65(4):443–454
40. Lai JS, Goodnight S, Downing NR, Ready RE, Paulsen JS, Kratz AL et al (2018) Evaluating cognition in individuals with Huntington disease: neuro-QoL cognitive functioning measures. *Qual Life Res* 27(3):811–822
41. Carozzi NE, Ready RE, Frank S, Cella D, Hahn EA, Goodnight SM et al (2017) Patient-reported outcomes in Huntington's disease: quality of life in neurological disorders (Neuro-QoL) and Huntington's disease health-related quality of life (HDQLIFE) physical function measures. *Mov Disord* 32(7):1096–1102
42. Carozzi NE, Boileau NR, Chou KL, Ready RE, Cella D, McCormack MK et al (2020) HDQLIFE and Neuro-QoL physical function measures: responsiveness in persons with Huntington's disease. *Mov Disord* 35(2):326–336
43. Lai JS, Beaumont JL, Jensen SE, Kaiser K, Van Brunt DL, Kao AH et al (2017) An evaluation of health-related quality of life in patients with systemic lupus erythematosus using PROMIS and Neuro-QoL. *Clin Rheumatol* 36(3):555–562
44. Cella D, Lai JS, Nowinski CJ, Victorson D, Peterman A, Miller D et al (2012) Neuro-QoL: brief measures of health-related quality of life for clinical research in neurology. *Neurology* 78(23):860–867
45. Iverson GL, Connors EJ, Marsh J, Terry DP (2021) Examining normative reference values and item-level symptom endorsement for the Quality of Life in Neurological Disorders (Neuro-QoL™) v2.0 cognitive function-short form. *Arch Clin Neuropsychol* 36(1):126–134
46. Miller DM, Bethoux F, Victorson D, Nowinski CJ, Buono S, Lai JS et al (2016) Validating Neuro-QoL short forms and targeted scales with people who have multiple sclerosis. *Mult Scler J* 22(6):830–841
47. Medina LD, Torres S, Alvarez E, Valdez B, Nair KV (2019) Patient-reported outcomes in multiple sclerosis: validation of the quality of life in neurological disorders (Neuro-QoL™) short forms. *Mult Scler J* 5(4):1–11
48. Cook KF, Kallen MA, Coon CD, Victorson D, Miller DM (2017) Idio Scale Judgment: evaluation of a new method for estimating responder thresholds. *Qual Life Res* 26(11):2961–2971
49. Nowinski CJ, Siderowf A, Simuni T, Wortman C, Moy C, Cella D (2016) Neuro-QoL health-related quality of life measurement system: validation in Parkinson's disease. *Mov Disord* 31(5):725–733
50. Kozlowski AJ, Singh R, Victorson D, Miskovic A, Lai JS, Harvey RL et al (2015) Agreement between responses from community-dwelling persons with stroke and their proxies on the NIH Neurological Quality of Life (Neuro-QoL) short forms. *Arch Phys Med Rehabil* 96(11):1986–1992
51. Bode RK, Heinemann AW, Butt Z, Stallings J, Aylor C, Rowe M et al (2010) Development and validation of participation and positive psychologic function measures for stroke survivors. *Arch Phys Med Rehabil* 91(9):1347–1356
52. Lapin BR, Thompson NR, Schuster A, Honomichl R, Katzan IL (2021) The validity of proxy responses on patient-reported outcome measures: are proxies a reliable alternative to stroke patients' self-report? *Qual Life Res* 30(6):1735–1745
53. Lapin BR, Thompson NR, Schuster A, Katzan IL (2021) Magnitude and variability of stroke patient-proxy disagreement across multiple health domains. *Arch Phys Med Rehabil* 102(3):440–447
54. Katzan IL, Thompson N, Schuster A, Wisco D, Lapin B (2021) Patient-reported outcomes predict future emergency department visits and hospital admissions in patients with stroke. *J Am Heart Assoc* 10 (6)
55. Reeve BB, Wyrwich KW, Wu AW, Velikova G, Terwee CB, Snyder CF et al (2013) ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Qual Life Res* 22(8):1889–1905
56. Rothman M, Burke L, Erickson P, Leidy NK, Patrick DL, Petrie CD (2009) Use of existing patient-reported outcome (PRO) instruments and their modification: the ISPOR good research practices for evaluating and documenting content validity for the use of existing instruments and their modification PRO task force report. *Value Health* 12(8):1075–1083
57. O'Donohoe P, Reasner DS, Kovacs SM, Byrom B, Eremenco S, Barsdorf AI et al (2023) Updated recommendations on evidence needed to support measurement comparability among modes of data collection for patient-reported outcome measures: a good practices report of an ISPOR task force. *Value Health* 26(5):623–633
58. Hanmer J, Jensen RE, Rothrock N (2020) A reporting checklist for HealthMeasures' patient-reported outcomes: ASCQ-Me, Neuro-QoL, NIH Toolbox, and PROMIS. *J Patient Rep Outcomes* 4(1):21
59. HealthMeasures (2016) Guide to selection of measures from HealthMeasures. [https://www.healthmeasures.net/images/Explore/Guide\\_to\\_Selection\\_of\\_a\\_HealthMeasures\\_06\\_09\\_16.pdf](https://www.healthmeasures.net/images/Explore/Guide_to_Selection_of_a_HealthMeasures_06_09_16.pdf)
60. HealthMeasures. Summary table'. <https://www.healthmeasures.net/explore-measurement-systems/selecting-among-measurement-systems/summary-table>
61. Littlejohn J, Bowen M, Constantinidou F, Dawes P, Dickinson C, Heyn P et al (2022) International practice recommendations for the recognition and management of hearing and vision impairment in people with Dementia. *Gerontology* 68(2):121–135
62. Naidech AM, Beaumont JL, Berman M, Liotta E, Maas MB, Prabhakaran S et al (2015) Web-based assessment of outcomes after subarachnoid and intracerebral hemorrhage: a new patient centered option for outcomes assessment. *Neurocrit Care* 23(1):22–27
63. Senders A, Hanes D, Bourdette D, Whitham R, Shinto L (2014) Reducing survey burden: feasibility and validity of PROMIS measures in multiple sclerosis. *Mult Scler J* 20(8):1102–1111
64. Mokkink L, Terwee CB, De Vet Henrica CW, Bouter LM, Alonso J, Patrick D (2021) Reply to the concerns raised by McKenna and Heaney about COSMIN. *J Med Econ* 24(1):857–859

## Publisher's Note

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