



# Quality of Measurement Properties in Patient Reported Outcomes Used in Adult Liver Transplant Candidates and Recipients: a Systematic Review

Samira E. M. van Knippenberg<sup>1†</sup>, Sarah F. Powell-Brett<sup>2,3†</sup>, Kunal Joshi<sup>2</sup>, Viola B. Weeda<sup>4</sup> and Hermien Hartog<sup>3,5\*</sup>

<sup>1</sup>Department of Surgery, Amsterdam University Medical Centers, Amsterdam, Netherlands, <sup>2</sup>The Liver Unit, Queen Elizabeth Hospital Birmingham, Birmingham, United Kingdom, <sup>3</sup>Centre for Liver and Gastrointestinal Research, Institute of Immunology and Immunotherapy, University of Birmingham, Birmingham, United Kingdom, <sup>4</sup>Department of Surgery, Hôpital Universitaire de Bruxelles, Bruxelles, Belgium, <sup>5</sup>Department of Surgery, Section of HPB and Liver Transplantation, University Medical Center Groningen, Groningen, Netherlands

**Objective:** Patient Reported Outcome Measures (PROMs) are increasingly recognized in liver transplant (LT)-patients, yet recent evaluations of their quality are lacking. This systematic review gives a comprehensive overview of available PROMs in adults awaiting or undergoing LT and their measurement properties.

**Method:** A systematic search in MEDLINE, EMBASE, PubMed, and COCHRANE (01/2010–08/2023) included studies involving adult LT-candidates and/or recipients utilizing PROMs with original evaluations of measurement properties. The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) was used to ascertain the quality of measurement properties.

**Results:** In total, 23 studies encompassing 35 PROMs were identified, including nine disease-specific and 26 generic PROMs. The (*Short-form*) *Liver Disease Quality of Life* ((SF-)LDQoL), *Transplant Effects Questionnaire* (TxEQ) and *Post-Liver Transplant Quality of Life* (pLTQ) were the most utilized disease-specific PROMs. Most studies demonstrated low-quality evidence for measurement properties. *pLTQ* demonstrated high-quality evidence for internal consistency, reliability, and responsiveness; the generic *Hospital Anxiety and Depression Scale* (HADS) showed strong evidence for internal consistency and construct validity.

## OPEN ACCESS

### \*Correspondence

Hermien Hartog,  
✉ h.hartog@umcg.nl

<sup>†</sup>These authors have contributed  
equally to this work

**Received:** 15 February 2025

**Accepted:** 12 August 2025

**Published:** 02 October 2025

### Citation:

van Knippenberg SEM,  
Powell-Brett SF, Joshi K, Weeda VB  
and Hartog H (2025) Quality of  
Measurement Properties in Patient  
Reported Outcomes Used in Adult  
Liver Transplant Candidates and  
Recipients: a Systematic Review.  
*Transpl. Int.* 38:14497.  
doi: 10.3389/ti.2025.14497

**Abbreviations:** BAI, Beck Anxiety Inventory; BDI, Beck Depression Inventory; CD-RISK, Connor Davidson Resilience Scale; CES-D, Center of Epidemiological Studies Depression Scale; CISS-SF, Coping Inventory for Stressful Situations; COSMIN, COnsensus-based Standards for the selection of health Measurement Instruments; EQ-5D, EuroQol-5 Dimension; FACT-G, The Functional Assessment of Cancer Therapy – General; FSI, Fatigue Symptom Inventory; GAD-7, Generalized anxiety disorder screener; HADS, Hospital Anxiety and Depression Scale LPA-SQUASH: Light-intensity Physical Activity Short Questionnaire to Assess Health-Enhancing Physical Activity LT, Liver Transplantation PHQ-9, Patient Health Questionnaire depression scale pLTQ, Post-Liver Transplant Quality of Life PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analysis PROM, Patient Reported Outcome Measures PSSS, Perceived Social Support Scale PTGI, Post-Traumatic Growth Inventory QoL, Quality of Life; QUALIOST, Quality of Life Questionnaire in Osteoporosis (SF-)LDQoL, Short Form Liver Disease Quality of Life SF-36, Short-Form 36 SOC-L9, Sense of Coherence scale by Antonovsky SSS, Medical Outcomes Study Social Support Survey STAI-6, State-Trait Anxiety Inventory TxEQ, Transplant Effects Questionnaire WHO-5, World Health Organisation – Five Wellbeing Index.

**Conclusion:** Measurement properties in LT-patients remains of low-quality. *pLTQ* stands out for its superior methodological quality among disease-specific PROMs. For future studies, there is a strong recommendation to focus more on patients' subjective measures and their measurement properties.

**Keywords:** patient reported outcome measures, liver transplantation, quality of life, measurement properties, surgery

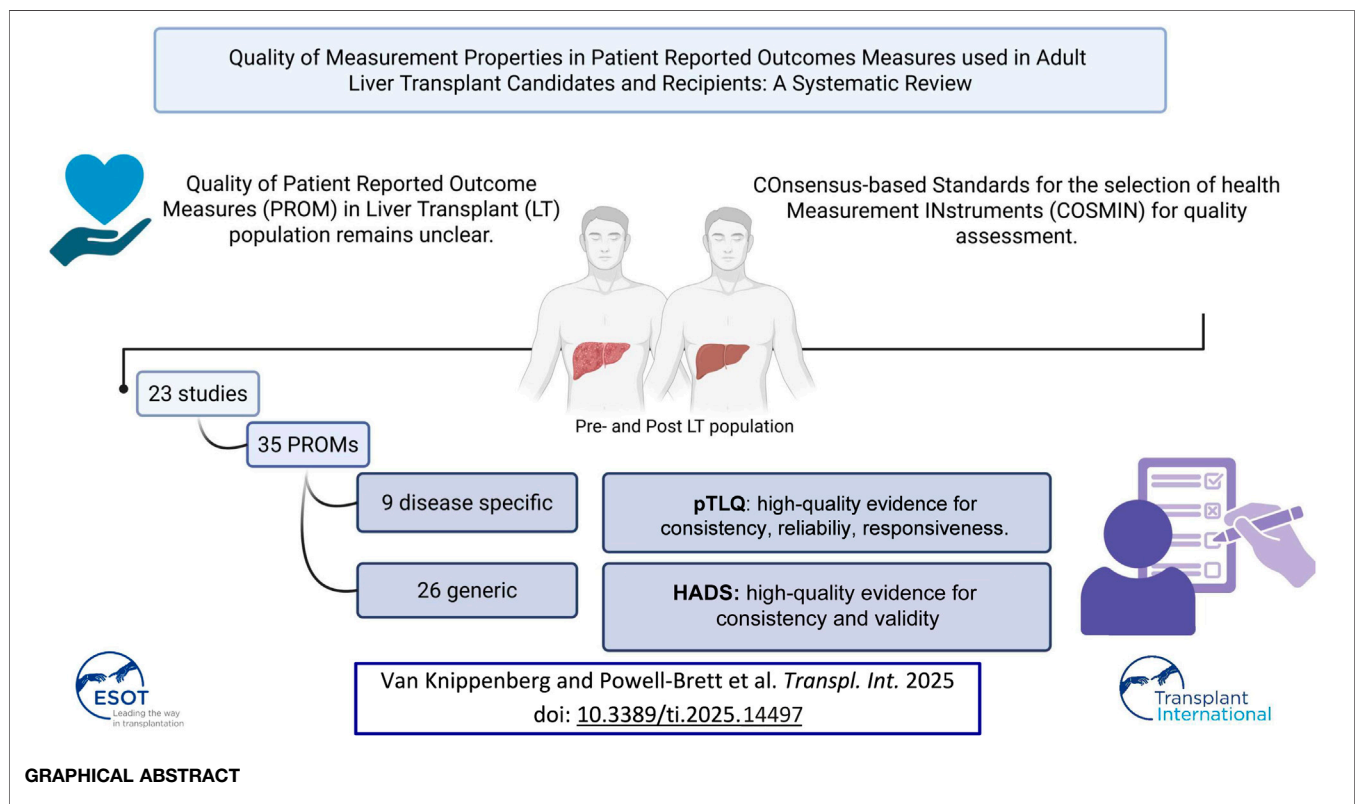
## INTRODUCTION

The field of liver transplantation (LT) is rapidly evolving. Over the last 10 years, more than 8,000 liver transplants have been performed in the United Kingdom with excellent long-term outcomes. In the United Kingdom, elective transplant procedures exhibit respective one- and 5-year survival rates of 94% and 81%, while urgent transplant cases demonstrate corresponding survival rates of 90% and 81% over the same time periods [1].

With increasing numbers and improving survival rates, there is a growing population of long-term survivors following LT. This results in a shift of focus towards subjective patient outcomes, including quality of life (QoL), anxiety and depressive symptoms. Survival is easily quantifiable; patients' subjective outcomes however are not. The last 20 years have seen the advent of a multitude of generic and disease-specific tools for measuring these patient-reported outcome measures (PROMs). Despite the increased recognition of the importance of PROMs and the growing

number of tools, a standardized methodology for their application among patients undergoing LT has yet to be established.

The use of PROMs in the LT population is an invaluable tool to target improvements in clinical care, develop benchmarking standards and assess hospital performance [2]. Given the breadth of available tools (both generic and specific), it is difficult to select one that is most likely to deliver meaningful results and effect the most benefit in this cohort. Ultimately, the integration of a PROM into routine care of LT patients requires careful consideration at an early stage. Two systematic reviews by Jay et al. and Cleemput et al. reported on QoL instruments used in the LT population [3, 4]. However, both articles are over 10 years old and there have been significant methodological improvements since. Considering the above, a full, up to date systematic review is required. The aim of this systematic review is to provide a comprehensive overview of PROMs currently available for use in adults undergoing LT and their measurement properties.



**TABLE 1 |** Description of the domains used to evaluate the risk of bias and quality of the measurement properties for each PROM.

Domain	Description
Reliability	
Internal consistency	The degree of the interrelatedness among the items of the PROM, as long as the items together form a unidimensional scale. Most of the times, the Cronbach's alpha is measured. If the Cronbach's alpha is $>0.70$ , the internal consistency can be deemed "sufficient"
Reliability	The proportion of the total variance in the measurements which is due to "true" differences between patients. There must be evidence that the patients are stable at the time of the PROM assessment. If the intra class correlation coefficient is $> 0.70$ , the reliability is deemed "sufficient"
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. The smallest detectable change should be smaller than the minimal important change, to deem the measurement property "sufficient"
Validity	
Content validity	The degree to which the content of a PROM is an adequate reflection of the construct to be Measured. Content validity is considered the most important measurement property, because the items of the used PROM should be relevant, comprehensive and comprehensible for the patient population in which the PROM is used
Construct validity	The degree to which the scores of a PROM are consistent with hypotheses based on the assumption that the PROM validly measures the construct to be measured Construct validity is divided into structural validity, hypotheses testing and cross-cultural validity <i>Structural validity</i> refers to the degree to which the scores of a PROM are an adequate reflection of the dimensionality of the construct to be measured and is usually assessed by factor analysis <i>Hypotheses testing for construct validity</i> refers to the degree to which the scores of a PROM are consistent with hypotheses  <i>Cross-cultural validity</i> refers to the degree to which the performance of the items on a translated or culturally adapted instrument are an adequate reflection of the performance of the items of the original version of the instrument. Therefore, this measurement has to be assessed by at least two different groups
Criterion validity	The degree to which the scores of a PROM are an adequate reflection of a 'gold standard', deemed 'sufficient' if the correlation with this gold standard is $\geq 0.70$ or has an Area Under the Curve of $\geq 0.70$
Responsiveness	The ability of a PROM to detect change over time in the construct to be measured. The results should be in accordance with the hypotheses or have an Area Under the Curve of $\geq 0.70$
Interpretability	Interpretability is the degree to which one can assign qualitative meaning - that is, clinical or commonly understood connotations - to a PROM's quantitative scores or change in scores

## METHODS

### Design

An initial scoping search was undertaken to identify relevant studies on this topic. This systematic review was conducted and written in compliance with the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines and report in PROSPERO (PROSPERO registration number: CRD42021251533) [5].

### Search

A systematic search was conducted of MEDLINE, EMBASE, PubMed and COCHRANE to identify all studies including patients undergoing LT from January 2010 until August 2023. To report the screening process, the PRISMA flow diagram was used. Studies were included if they used a PROM to measure subjective insight of LT candidates and/or LT recipients, inclusive of QoL, anxiety, depressive symptoms, pain, mobility and liver failure symptoms. Included studies had to report either the development or evaluation of one or more measurement properties of their chosen PROM. Studies with non-original evaluations of the measurement properties were excluded. *In vitro* studies, studies only covering patients under 16 years of age or those reporting

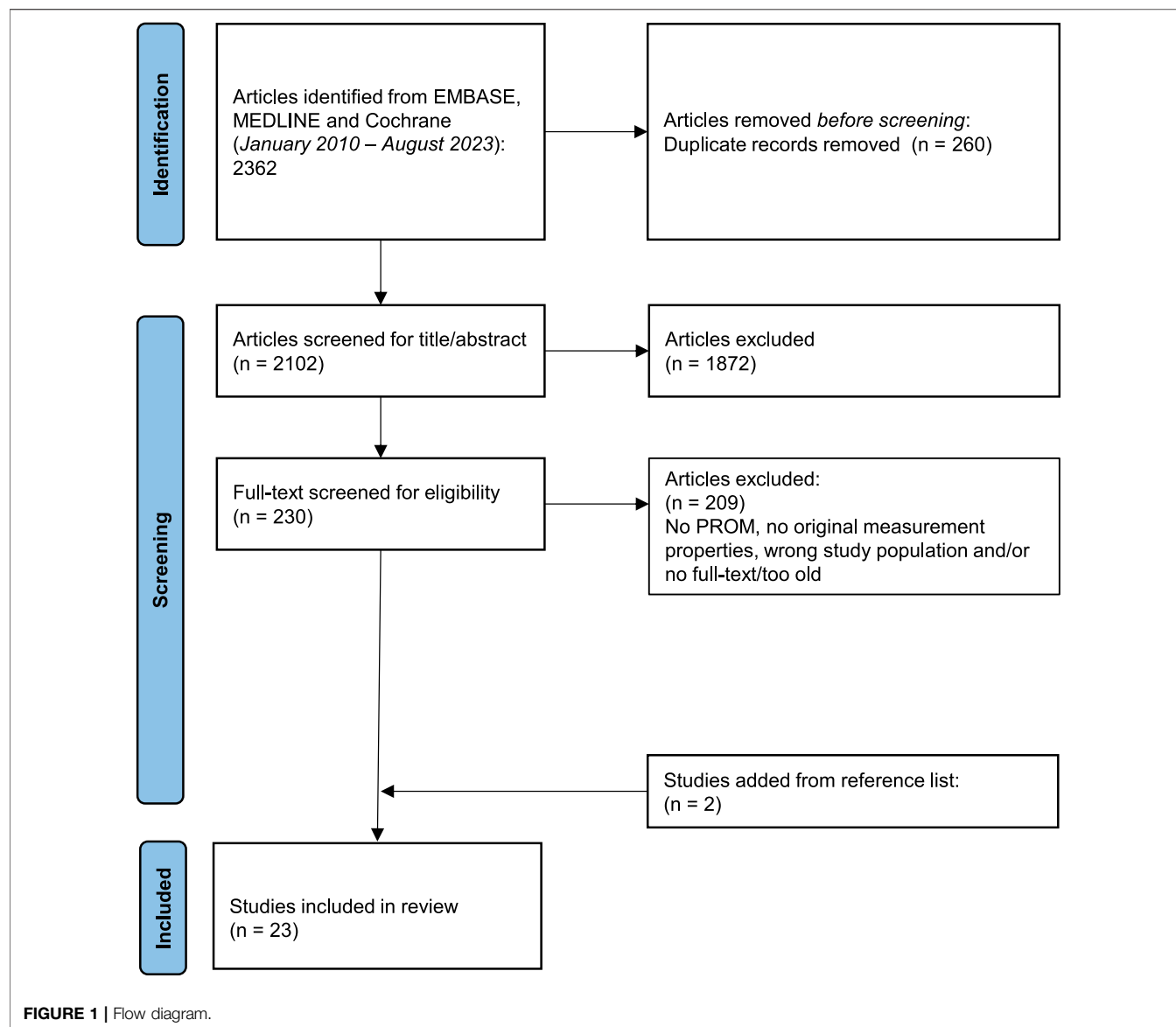
on living donors were excluded. Systematic literature reviews were excluded but were used to cross check included studies and identify additional references. Additionally, the reference list of included studies was reviewed to identify additional eligible studies. The complete search strategy is described in **Supplementary Table S1**.

### Screening Process

EndNote X7 (Clarivate Analytics, Pennsylvania, US) was used to collate the search results and exports of all citations were sent to the review software Rayyan (Qatar Computing Research Institute, Doha, Qatar) where duplicates were removed. After duplicate removal, four independent reviewers (SvK, SP, KJ, VW) screened by title and abstract and then by full text review. Abstracts that did not report enough information for an inclusion/exclusion decision underwent full text review. Disputes were resolved by the senior author (HH).

### Data Extraction

Data extraction elements were defined in advance and included: study population, demographics (age, sex, pre-/post-LT), the PROM tools (title, scoring system, number of items, domains) and measurement properties of the PROM. Some studies described measurement properties with different



definitions. The COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) was used to ascertain which measurement properties were evaluated by the studies [6].

## Quality Assessment of Included Studies and Measurement Properties

Two authors (SvK, VW) first independently assessed the methodological quality of different domains of the studies using the COSMIN Risk of Bias checklist [7]. This employs a four-point rating system (“very good,” “adequate,” “doubtful” or “inadequate”) and the overall quality rating of each study is based on “the worst score counts” principle, i.e., the lowest rating of any standard. **Table 1** presents information on the domains used to evaluate the risk of bias and quality of the measurement properties for each PROM.

## Data Synthesis

Subsequently, the quality of the measurement properties was assessed by the updated criteria for good measurement properties (based on Terwee et al, and Prinsen et al) as outlined by the COSMIN guideline for systematic reviews [6, 7].

Measurement properties were assessed using the following principles: content validity, structural validity, internal consistency, cross-cultural validity, reliability, measurement error, criterion validity, hypothesis testing for construct validity and responsiveness. The quality of the measurement properties were scored using a four-point rating system (“+” = sufficient, “?” = indeterminate, “-” = insufficients “±” = inconsistent). When the measurement properties of a PROM were not reported in any of the included articles, no score was assigned.

The criteria for good measurement properties were then applied to the results per measurement property per PROM, and the quality of the evidence (using the GRADE approach) was analyzed.

**TABLE 2 |** Study and patient characteristics, categorized per Patient Reported Outcome Measurements (PROMs).

PROM	Author	Country	Publication year	Study population	Gender (male (%))	Age (mean (SD))	Mode of administration	Number of items	Response rate (%)	Target population	Patient population (pre-/ post LT)
Disease-specific PROMs											
Short Form Liver Disease Quality of Life ((SF-)LDQOL)	Kanwal F (SF-LDQOL) [9]	USA	2008	156	54.8	53.9 (11)	Questionnaire	36			Pre
Transplant Effects Questionnaire (TxEQ)	Gralnek I.M [10].	USA	2000	221	64.3	52.2		111	86.6		Pre
	Pérez-San-Gregorio, MÁ [11]	Spain	2018	240				21			Post
	Annema, C [12].	Netherlands	2018	116	65.5	50.8 (11.4)	Questionnaire		75.8		Both
Post-Liver Transplant Quality of Life (pLTQ)	Molski, C [13].	Brazil	2016	160		56.9 (10.4)	Questionnaire	32			Post
	Saab, S [14].	USA	2011	196	59.7	53.1 (12.6)		32	93.8		Post
Self-made questionnaire	Parsa Yekta, Z [15].	Iran	2013	250	63.3	37.5 (12)	Questionnaire administered by hospital receptionist	40			Post
Self-made questionnaire	Lasker, J. N. (social QoL) [16]	USA	2011	100	0	58.5	Questionnaire via mail, online and interview		Response to items ranged from 93% to 100%	women with PBC on waiting list (WL) and post-transplant (PT)	Both
Self-made questionnaire	Franciosi, M. (ITaLi-Q) [17]	Italy	2011	177	71.8	57.2	Questionnaire, self-administered and anonymous	37	100% first questionnaire, 49/177 the retest	Patients requiring HBV prophylaxis after LT	Post
Self-made questionnaire	Chen, X. (Post-LiverTransplant Symptom Experience Questionnaire) [18]	China	2021	265 (reliability tested on 30 patients in pilot study)	80		Questionnaire	40	96.1		Post
Self-management Questionnaire for LT recipients	Xing L [19].	China	2015	124				45			Post
Quality of Life Questionnaire in Osteoporosis (QUALIOST)	Atamaz, F [20].	Turkey	2013	38 LT patients, 42 controls	81.6	42 (11.6)		24	ND		Post
Generic PROMs											
Short-form 36 (SF-36)	Fernandez, A. C [21].	USA	2016	125	60.8	56.1		36	96		Pre
	Miller-Matero, L. R [22].	USA	2014	84	66.8	SRD 53.96 (7.11) and HRD 55.87 (6.89)	Semi-structured interview	36	66.7		Both (prospective study)
	Pelgur H [23].	Turkey	2009	64	67			14	ND		Post

(Continued on following page)

**TABLE 2 |** (Continued) Study and patient characteristics, categorized per Patient Reported Outcome Measurements (PROMs).

PROM	Author	Country	Publication year	Study population	Gender (male (%))	Age (mean (SD))	Mode of administration	Number of items	Response rate (%)	Target population	Patient population (pre-/ post LT)
Hospital Anxiety and Depression Score (HADS)							Face-to-face interview, Questionnaire administered by researcher			patients who had undergone liver transplantation at least 1 month prior and were attending clinic for follow-up	
	Miller-Matero, L. R [22].	USA	2014	84	66.8	SRD 53.96 (7.11) and HRD 55.87 (6.89)	Semi-structured interview	14	66.7		Both (prospective study)
	Lin. X [24]	China	2017	285	75.8	53.3 (10.2)	Questionnaire	14	95		Post
World Health Organisation – Five Wellbeing Index (WHO-5)	Fernandez, A. C [21].	USA	2016	125	60.8	56.1 (8.64)	"	5	56		Pre
WHOQOL-BREF	Weber S [25].	Germany	2021	79	64.6	58.2	Questionnaire	5	ND		Post
	Annema, C [12].	Netherlands	2018	116	65.5	50.8 (11.4)	Questionnaire	24	75.8		Both
	Molski, C [13].	Brazil	2016	160		56.9 (10.4)	Questionnaire				Post
Post-Traumatic Growth Inventory (PTGI)	Gangeri, L [26].	Italy	2018	233	84	61	Questionnaire send to patients	21	76		Post
The Functional Assessment of Cancer Therapy - General (FACT-G)	Scrignano M [30].	Italy	2016	100	15	59.88		21	58	"	Post
	Gangeri, L [26].	Italy	2018	233	84	61	Questionnaire send to patients	27	76		Post
Connor Davidson resilience scale (CD-RISC)	Fernandez, A. C [21].	USA	2016	125	60.8	56.1 (8.64)		25	56		Pre
Beck Depression Inventory (BDI)	Fernandez, A. C [21].	USA	2016	125	60.8	56.1 (8.64)		21	56		Pre
Beck Anxiety Inventory (BAI)	Fernandez, A. C [21].	USA	2016	125	60.8	56.1 (8.64)		21	56		Pre
Medical Outcomes Study Social Support Survey (SSS)	Fernandez, A. C [21].	USA	2016	125	60.8	56.1 (8.64)		20	56		Pre
State-Trait Anxiety Inventory (STAI-6)	Annema, C [12].	Netherlands	2018	116	65.5	50.8 (11.4)	Questionnaire	6	75.8		Both
Center of Epidemiological Studies Depression Scale (CES-D)	Annema, C [12].	Netherlands	2018	116	65.5	50.8 (11.4)	Questionnaire	20	75.8		Both

(Continued on following page)

**TABLE 2 |** (Continued) Study and patient characteristics, categorized per Patient Reported Outcome Measurements (PROMs).

PROM	Author	Country	Publication year	Study population	Gender (male (%))	Age (mean (SD))	Mode of administration	Number of items	Response rate (%)	Target population	Patient population (pre-/ post LT)
Pearlin-Scooler Mastery Scale	Annema, C [12].	Netherlands	2018	116	65.5	50.8 (11.4)	Questionnaire	7	75.8		Both
Coping Inventory for Stressful Situations (CISS-SF)	Annema, C [12].	Netherlands	2018	116	65.5	50.8 (11.4)	Questionnaire	21	75.8		Both
Perceived Social Support Scale (PSSS)	Lin, X [24].	China	2017	285	75.8	53.3 (10.2)	Questionnaire	14	95		Post
General Comfort Questionnaire	Demir B [29].	Turkey	2021	148	81.8%	ND	Interview	28	ND		Post
Fatigue Symptom Inventory (FSI)	Lin, X [24].	China	2017	285	75.8	53.3 (10.2)	Questionnaire	13	95		Post
Patient Health Questionnaire depression scale (PHQ-9)	Gronewold N [27].	Germany	2022	544	63.1	51.95 (9.84)	Questionnaire	9	ND		Pre
Generalized anxiety disorder screener (GAD-7)	Gronewold N [27].	Germany	2022	544	63.1	51.95 (9.84)	Questionnaire	7	ND		Pre
Perceived social support questionnaire	Gronewold N [27].	Germany	2022	544	63.1	51.95 (9.84)	Questionnaire	14	ND		Pre
Sense of Coherence Scale by Antonovsky	Gronewold N [27].	Germany	2022	544	63.1	51.95 (9.84)	Questionnaire	9	ND		Pre
General Self-Efficacy Short Scale	Gronewold N [27].	Germany	2022	544	63.1	51.95 (9.84)	Questionnaire	3	ND		Pre
German Body Image	Gronewold N [27].	Germany	2022	544	63.1	51.95 (9.84)	Questionnaire	20	ND		Pre
Short Questionnaire to Assess Health-Enhancing Physical Activity (SQUASH)	Ushio M [28].	Japan	2023	173	47.4	ND	Questionnaire	13	ND		Post
UCLA Loneliness Scale	Weber S [25].	Germany	2021	79	64.6	58.2	Questionnaire	20	ND		Post
Utility Measure EQ-5D	Russell R.T [8].	USA	2009	285	64	53.3		5			Both

Abbreviation: ND = not described.

**TABLE 3 |** Risk of Bias using the COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) Risk of Bias checklist.

PROM	Author	Content validity	Structural validity	Internal validity (Cronbach's alpha)	Cross-cultural validity	Reliability	Measurement error (test-retest)	Criterion validity	Hypothesis testing for construct validity	Responsiveness
Disease specific N = 9 (SF)-LDQOL	Kanwal F (SF-LDQOL)	Very good	Inadequate	very good			adequate	very good	very good	very good
	Gralnek I.M.		Inadequate	very good			inadequate		very good	
TxEQ	Pérez-San-Gregorio, MÁ			very good	very good					
	Annema, C		Inadequate	very good	Inadequate					
pLTQ	Molski, C			very good	very good	very good				very good
	Saab, S		very good	very good			very good	very good		
Self-made questionnaire	Parsa Yekta, Z	very good	Inadequate	very good		adequate				
Self-made questionnaire	Lasker, J. N. (social QoL)			very good		Inadequate	Inadequate		doubtful	
Self-made questionnaire	Franciosi, M. (ITaLi-Q)			very good		doubtful	very good	very good	very good	
Self-made questionnaire	Chen, X. (Post-LiverTransplant Symptom Experience Questionnaire)		Inadequate	Very good					inadequate	
Self-management Questionnaire for LT recipients	Xing L			very good						
QUALIOST	Atamaz, F		NA	Very good	Doubtful	doubtful		Very good		
Generic N = 26										
Short-form 36 (SF-36)	Fernandez, A. C			Very good		Inadequate			Very good	
	Miller-Matero, L. R			very good		very good			very good	
Hospital Anxiety and Depression Score (HADS)	Pelgur H			very good						
	Miller-Matero, L. R			very good		very good			very good	
	Lin. X			Very good						
World Health Organisation – Five Wellbeing Index (WHO-5)	Fernandez, A. C			Inadequate/ Doubtful						
WHOQOL-BREF	Weber S			Doubtful						
	Annema, C			very good	Inadequate					
	Molski, C									
Post-Traumatic Growth Inventory (PTGI)	Gangeri, L			very good	doubtful		very good		Doubtful	very good
The Functional Assessment of Cancer Therapy - General (FACT-G)	Scrignaro M			Very good		Inadequate	Inadequate		Very good	
	Gangeri, L			very good	doubtful		very good		Doubtful	very good
Connor Davidson resilience scale (CD-RISC)	Fernandez, A. C		Inadequate	very good		adequate			very good	
Beck Depression Inventory (BDI)	Fernandez, A. C			Inadequate/ Doubtful						
Beck Anxiety Inventory (BAI)	Fernandez, A. C			Inadequate/ Doubtful						
Medical Outcomes Study Social Support Survey (SSS)	Fernandez, A. C			Inadequate/ Doubtful						
State-Trait Anxiety Inventory (STAI-6)	Annema, C			Inadequate/ Doubtful						
	Annema, C									

(Continued on following page)



**TABLE 3 |** (Continued) Risk of Bias using the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) Risk of Bias checklist.

PROM	Author	Content validity	Structural validity	Internal validity (Cronbach's alpha)	Cross-cultural validity	Reliability	Measurement error (test-retest)	Criterion validity	Hypothesis testing for construct validity	Responsiveness
Center of Epidemiological Studies Depression Scale (CES-D)				Inadequate/Doubtful						
Pearlin-Scooler Mastery Scale	Annema, C			Inadequate/Doubtful						
Coping Inventory for Stressful Situations (CISS-SF)	Annema, C			very good						
Perceived Social Support Scale (PSSS)	Lin. X			Very good						
General Comfort Questionnaire	Demir B			Doubtful	Inadequate					
Fatigue Symptom Inventory (FSI)	Lin. X			Very good						
Patient Health Questionnaire depression scale (PHQ-9)	Gronewold N			Doubtful						
Generalized anxiety disorder screener (GAD-7)	Gronewold N			Doubtful						
Perceived social support questionnaire	Gronewold N			Doubtful						
Sense of coherence scale by Antonovsky	Gronewold N			Doubtful						
general self-efficacy short scale	Gronewold N			Doubtful						
German body image Short Questionnaire to Assess Health-Enhancing Physical Activity (SQUASH)	Gronewold N Ushio M			Very good		Adequate	adequate	Very good		
UCLA Loneliness scale	Weber S			Doubtful						
Utility measures N = 1 EQ-5D	Russell R.T.					Doubtful	inadequate	very good	very good	

**TABLE 4 |** Quality Assessment of the Patient Reported Outcome Measures (PROMs) using the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) guideline.

PROM	Author	Content Validity	Structural validity	Internal validity	Cross-cultural validity	Reliability	Measurement error	Criterion validity	Hypothesis testing for construct validity	Responsiveness
Disease specific N = 9										
(SF-)LDQOL	Kanwal F (SF-LDQOL)		?	-				?	+	+
	Gralnek I.M.		?	-					+	
TxEQ	Pérez-San-Gregorio, MÁ			+	+					
	Annema, C.			-	?					
pLTQ	Molski, C.			+	?	+				+
	Saab, S.	?		+			-	?		
Self-made questionnaire	Parsa Yekta, Z.	?		+		?				
Self-made questionnaire	Lasker, J. N. (social QoL)	?		-		?	?		?	
Self-made questionnaire	Franciosi, M. (ITaLi-Q)	?	?	+		?		+	?	
Self-made questionnaire	Chen, X. (Post-LiverTransplant Symptom Experience Questionnaire)	?		+						
Self-management Questionnaire for LT recipients	Xing L.			+						
QUALIOST	Atamaz, F.			+	+	+		-		
Generic N = 26										
Short-form 36 (SF-36)	Fernandez, A. C.			+		?			+	
	Miller-Matero, L. R.			+					+	
Hospital Anxiety and Depression Score (HADS)	Pelgur H.			+						
	Miller-Matero, L. R.			+					+	
	Lin. X			+						
World Health Organisation – Five Wellbeing Index (WHO-5)	Fernandez, A. C			+						
WHOQOL-BREF	Weber S.			+						
	Annema, C.			-						
	Molski, C.									
Post-Traumatic Growth Inventory (PTGI)	Gangeri, L.			+			?		?	+
	Scignaro M.			+		?	?		+	
The Functional Assessment of Cancer Therapy - General (FACT-G)	Gangeri, L.			+			?		?	+
Connor Davidson resilience scale (CD-RISC)	Fernandez, A. C.			+					?	
Beck Depression Inventory (BDI)	Fernandez, A. C.			+						
Beck Anxiety Inventory (BAI)	Fernandez, A. C.			+						
Medical Outcomes Study Social Support Survey (SSS)	Fernandez, A. C.			+						
State-Trait Anxiety Inventory (STAI-6)	Annema, C.			+						
Center of Epidemiological Studies Depression Scale (CES-D)	Annema, C.			+						
Pearlin-Scooler Mastery Scale	Annema, C			+						
Coping Inventory for Stressful Situations (CISS-SF)	Annema, C			+						

(Continued on following page)

**TABLE 4 |** (Continued) Quality Assessment of the Patient Reported Outcome Measures (PROMs) using the COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) guideline.

PROM	Author	Content Validity	Structural validity	Internal validity	Cross-cultural validity	Reliability	Measurement error	Criterion validity	Hypothesis testing for construct validity	Responsiveness
Perceived Social Support Scale (PSSS)	Lin. X			+						
General Comfort Questionnaire	Demir B			+	?					
Fatigue Symptom Inventory (FSI)	Lin. X			+						
Patient Health Questionnaire depression scale (PHQ-9)	Gronewold N			+						
Generalized anxiety disorder screener (GAD-7)	Gronewold N			+						
Perceived social support questionnaire	Gronewold N			+						
Sense of coherence scale by Antonovsky	Gronewold N			+						
General self-efficacy short scale	Gronewold N			+						
German body image	Gronewold N			+						
Short Questionnaire to Assess Health-Enhancing Physical Activity (SQUASH)	Ushio M					-	?	?		
UCLA Loneliness scale	Weber S			+						
Utility measures N = 1 EQ-5D	Russell R.T.						?	-	+	

Abbreviations: + = positive rating; ? = indeterminate rating; – = negative rating.

## RESULTS

The search strategy retrieved a total of 2,362 titles/abstracts. After 260 duplicates were removed, 2,102 abstracts were screened, and 210 full-text articles were retrieved for further review. Following reference list and citation searching, two more articles were retrieved. After further review, a total number of 23 studies were included (Figure 1).

In total, 35 PROMs were used, with a minimum of one, and a maximum of six PROMs per study. PROMs could be divided in two categories: generic and disease-specific PROMs, and PROMs used for pre- and post-LT populations. Seven PROMs were disease-specific for liver disease and/or LT. Additionally, two PROMs addressed osteoporotic symptoms [Quality of Life Questionnaire in Osteoporosis (QUALIOST)] and emotional responses of organ transplant recipients [*the Transplant Effects Questionnaire (TxEQ)*], and were also categorized as disease-specific PROMs. 25 PROMs used in the studies were generic. One PROM was categorized under “utility measures,” providing utilities or values regarding health, that can be used for cost-utility analyses or interventions [8].

A total of eleven PROMs were applied to the pre-LT population, while thirteen were used for post-LT population. Additionally, eleven PROMs were used for both the pre- and post-LT population. Detailed study characteristics are described in Table 2, and a brief description of the PROMs evaluated is presented in Supplementary Table S2.

The risk of bias and methodological qualities of the PROMs used and described in the selected studies are described in Tables 3, 4, respectively. Overall, the evidence for the measurement properties was limited and the methodological quality was insufficient or inconsistent. None of the studies evaluated all measurement properties of the COSMIN system. Internal consistency was the most evaluated measurement property.

### Disease-Specific PROMs

A total of twelve articles described the measurement properties of the nine disease-specific PROMs [9–20]. Of these PROMs, one was used in a pre-LT population, six in the post-LT population and two in both the pre- and post-LT population.

Only the (*Short-form*) *Liver Disease Quality of Life* [(SF-)LDQOL] (two studies), *TxEQ* (two studies) and *Post-Liver Transplant Quality of Life* (pLTQ) (two studies) were employed by multiple studies, each with their own measurement properties of the utilized PROMs. The pLTQ scored a high evidence level for internal validity, reliability and responsiveness.

The *ITaLi-Q*, the *self-made questionnaires* by Parsa Yekta et al. and Chen et al., the *self-management questionnaire for LT-recipients* by Xing et al. and the *QUALIOST* were all graded with a high evidence level for adequate internal validity [15, 18, 19].

The *QUALIOST* reported a high level of evidence for cross-cultural validity and reliability. The (SF-)LDQOL reported a high level of evidence on hypothesis testing for construct validity and responsiveness.

### Generic PROMs

A total of fourteen articles described the measurement properties of 26 generic PROMs [12, 13, 21–30]. Of these PROMs, ten were used in a pre-LT population, and eight in the post-LT population. Furthermore, eight PROMs were utilized in both the pre- and post-LT population. The *EQ-5D*, graded as a ‘utility measure’, used in both pre- and post-LT population.

The most utilized PROMs were the *Hospital Anxiety and Depression Score* (HADS) (three studies), the *Short-form 36* (SF-36) (two studies), the *World Health Organisation – Five Wellbeing index* (WHO-5) (two studies), the *WHOQOL-BREF* (two studies) and the *Post-Traumatic Growth Inventory* (PTGI). All other PROMs were used by one study only.

There was moderate evidence for the internal validity in most studies; the HADS and SF-36 both scored a high level of evidence in internal validity and hypothesis testing for construct validity. The *Short-Questionnaire to Assess Health-Enhancing Physical Activity* showed a low level of evidence for reliability. The *EQ-5D* showed a low level of evidence for criterion validity.

## DISCUSSION

This systematic review is the first study to evaluate the methodological quality of PROMs utilized in the pre- and post-LT population, using the COSMIN-guidelines. In total, 23 articles employed nine disease-specific PROMs for the pre- and post-LT population, while 25 general PROMs and one utility measure were included. The (SF-)LDQOL, TxEQ and pLTQ were the most commonly used disease-specific PROMs. pLTQ showed high quality evidence of Internal validity, reliability and responsiveness. HADS was the most frequently used general PROM, and showed high-quality evidence for internal consistency and hypothesis testing for construct validity.

The methodological quality of most general and disease-specific PROMs was found to be limited, as the majority of the studies failed to adequately evaluate the measurement properties of the utilized PROMs, a trend observed in other similar reviews [31–33]. Within this review, most studies merely described the internal validity, while other essential measurement properties either lacked a description or exhibited inadequate methodological quality. Furthermore, there was inconsistency in scores for different measurement properties between different studies. For example, internal validity of the PROM TxEQ demonstrated sufficient quality in one study, but insufficient quality in another study, while both studies utilized the same PROM within the post-LT patient population. This discrepancy aligns with finding from the study by Elberts et al., who evaluated the quality of measurement properties in patients with neurological diseases [32]. Variations in measurement properties between studies can be in part attributed to differences in patient demographics and socio-economic characteristics. McHorney et al. found that SF-36 scores were generally lower among the elderly, those with less than a high school education and those in poverty [34]. Therefore, socio-economic backgrounds and diverse patient populations must be considered when implementing a PROM.

The limited use of PROMs in this patient population made it challenging to effectively synthesize and summarize the data. Most PROMs were reported in only one study, with only thirteen studies evaluating the same PROMs [9, 10, 14]. This lack of quality assessment is also reflected in reviews evaluating PROMs in other medical subpopulations [32, 33]. Aiyegbus et al. reviewed the measurement properties of PROMs used in kidney transplantation patients [31]. Despite a greater quantity of studies including a quality assessment of PROMs, the evidence was still of poor quality, with significant gaps in information. Chiarotto et al. evaluated the quality of measurement properties in PROMs for patients with lower back pain – including the *SF-36*, *SF-12*, *EQ-5D-3L*, *EQ-5D-5L*, *Nottingham Health Profile* and the *PROMIS-GH-10*, and found similar scarcities of high-quality evidence in their patient population [35].

The lack of robust quality assessment of PROMs can be attributed to their relatively recent rise in prominence in clinical research. However, PROMs are of the utmost importance for individual patients, as they reflect what matters to patients at a personal level, transcending the broader context of population-level survival. Therefore, identifying high quality, high level of evidence measures that can be standardized across patient populations is of paramount importance.

Assessing subjective patient measurements remain complex due to variability in individual values. Individuals prioritize different aspects of their life, posing a challenge in developing a universally applicable tool. While general tools like the *SF-36* and *HADS* offer a broad applicability, they lack assessment of disease-specific burden. Disease-specific PROMs are therefore more suitable for subpopulations, facilitating accurate detection of burden in subjective measurements.

An additional consideration when selecting a PROM is its original intended purpose. For example, the *EuroQol-5 Dimension (EQ-5D)* was not originally conceived for the evaluation of QoL in medical research but rather to facilitate cost-effectiveness assessments, rendering it particularly valuable in economic studies. Poor definitions within PROMs also pose a problem, for example, the definition of HRQoL is not always clear [36].

This review extends beyond PROMs simply assessing QoL, to encompass an overview of all PROMs used in pre- and post-LT population. There is not a clear single best option and the choice of a PROM should be made with careful deliberation, considering the particular objectives of the study. Over the last decade, the use of PROMs has increased, including the use of web questionnaires [37]. The integration of PROMs into research and clinical practice enables more accurate assessment of patient symptoms and supports more efficient allocation of healthcare resources. In the context of LT, evaluating changes in symptoms before and after the procedure is particularly relevant, as it could reflect treatment effectiveness. Disease-specific PROMs are therefore generally more appropriate for assessing disease-related symptoms with greater sensitivity. In contrast, generic PROMs are more appropriate to compare across different diseases and populations, and preferred in health technology assessment [38]. Nonetheless, the use of both generic and disease-specific PROMs requires careful consideration. When clinicians or researchers select existing PROMs or developing new ones, several critical aspects must be addressed, including cross-

cultural validation, the intended purpose (clinical or research), and patient acceptability and feasibility [31].

There are limitations to this review. Firstly, the populations of the included studies are heterogeneous, conducted across many different countries and languages. Cultural nuances play a pivotal role in shaping perception, and the translation of PROMs into different languages may introduce variations in interpretation. Cross-cultural validation represents one approach addressing this problem. However, most of the studies did not provide a comprehensive report on this measurement property. Furthermore, the pre- and post-LT populations have different considerations, including underlying liver disease, the severity of the disease, time after transplantation and the current symptoms of the patient. All these aspects influence patient's subjective feelings and therefore the outcome of the PROM utilized. However, since there was a lack of strong evidence studies, these sub-analyses could not be performed.

In summary, this review identified the *(SF-)LDQOL*, *TxEQ* and *pLTQ* as the most commonly used disease-specific PROMs, and the *HADS* was the most frequently used general PROM. For disease-specific PROMs in both pre- and post-LT patients, the *pLTQ* emerges as the PROM of choice based on its superior methodological quality. However, the limited number of studies assessing the quality of the same PROMs and the low quality of evidence surrounding these instruments highlight the necessity of further investigation. Further studies are needed to carefully evaluate both the appropriateness of the PROM selection for their target population, and the evidence regarding the measurement properties of these instruments, either through rigorous assessment or validation.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

## AUTHOR CONTRIBUTIONS

SK, SP-B, KJ, and VW conducted the search, selected the studies and wrote the manuscript. HH supervised and reviewed the manuscript. All authors contributed to the article and approved the submitted version.

## FUNDING

The author(s) declare that no financial support was received for the research and/or publication of this article.

## CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## GENERATIVE AI STATEMENT

The author(s) declare that no Generative AI was used in the creation of this manuscript.

## ACKNOWLEDGMENTS

We thank the Centre for Patient-Reported Outcome Research at the University of Birmingham, who were consulted during the

review process. The graphical abstract was designed with BioRender.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontierspartnerships.org/articles/10.3389/ti.2025.14497/full#supplementary-material>

## REFERENCES

1. *Annual Report on Liver Transplantation 2018/2019*. NHS Blood and Transplant. (2018).
2. Girgenti R, Tropea A, Buttafarro MA, Ragusa R, Ammirata M. Quality of Life in Liver Transplant Recipients: A Retrospective Study. *Int J Environ Res Public Health* (2020) 17:3809. doi:10.3390/ijerph17113809
3. Jay CL, Butt Z, Ladner DP, Skaro AI, Abecassis MM. A Review of Quality of Life Instruments Used in Liver Transplantation. *J Hepatol* (2009) 51:949–59. doi:10.1016/j.jhep.2009.07.010
4. Cleemput I, Dobbels F. Measuring patient-reported Outcomes in Solid Organ Transplant Recipients: An Overview of Instruments Developed to Date. *Pharmacoeconomics* (2007) 25:269–86. doi:10.2165/00019053-200725040-00002
5. Preferred Reporting Items for Systematic Reviews and Meta-Analyses. (2020).
6. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN Guideline for Systematic Reviews of Patient-Reported Outcome Measures. *Qual Life Res* (2018) 27:1147–57. doi:10.1007/s11136-018-1798-3
7. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality Criteria Were Proposed for Measurement Properties of Health Status Questionnaires. *J Clin Epidemiol* (2007) 60:34–42. doi:10.1016/j.jclinepi.2006.03.012
8. Russell RT, Feurer ID, Wisawatpannimit P, Pinson CW. The Validity of EQ-5D US Preference Weights in Liver Transplant Candidates and Recipients. *Liver Transpl* (2009) 15:88–95. doi:10.1002/lt.21648
9. Kanwal F, Spiegel BM, Hays RD, Durazo F, Han SB, Saab S, et al. Prospective Validation of the Short Form Liver Disease Quality of Life Instrument. *Aliment Pharmacol Ther* (2008) 28:1088–101. doi:10.1111/j.1365-2036.2008.03817.x
10. Gralnek IM, Hays RD, Kilbourne A, Rosen HR, Keeffe EB, Artinian L, et al. Development and Evaluation of the Liver Disease Quality of Life Instrument in Persons with Advanced, Chronic Liver Disease--the LDQOL 1.0. *Am J Gastroenterol* (2000) 95:3552–65. doi:10.1111/j.1572-0241.2000.03375.x
11. Pérez-San-Gregorio M, Martín-Rodríguez A, Sánchez-Martín M, Borda-Mas M, Avargues-Navarro ML, Gómez-Bravo M, et al. Spanish Adaptation and Validation of the Transplant Effects Questionnaire (TxEQ-Spanish) in Liver Transplant Recipients and its Relationship to Posttraumatic Growth and Quality of Life. *Front Psychiatry* (2018) 9:148. doi:10.3389/fpsy.2018.00148
12. Annema C, Drent G, Roodbol PF, Stewart RE, Metselaar HJ, van Hoek B, et al. Trajectories of Anxiety and Depression After Liver Transplantation as Related to Outcomes During 2-Year Follow-Up: A Prospective Cohort Study. *Psychosom Med* (2018) 80:174–83. doi:10.1097/PSY.0000000000000539
13. Molski C, Mattiello R, Sarria EE, Saab S, Medeiros R, Brandão A. Cultural Validation of the Post-liver Transplant Quality of Life (Pltq) Questionnaire for the Brazilian Population. *Ann Hepatol* (2016) 15:377–85. doi:10.5604/16652681.1198810
14. Saab S, Ng V, Landaverde C, Lee SJ, Comulada WS, Arevalo J, et al. Development of a Disease-specific Questionnaire to Measure health-related Quality of Life in Liver Transplant Recipients. *Liver Transpl* (2011) 17:567–79. doi:10.1002/lt.22267
15. Parsa Yekta Z, Tayebi Z, Shahsavari H, Ebadi A, Tayebi R, Bolourchifard F, et al. Liver Transplant Recipients Quality of Life Instrument: Development and Psychometric Testing. *Hepat Mon* (2013) 13:e9701. doi:10.5812/hepatmon.9701
16. Lasker JN, Sogolow ED, Short LM, Sass DA. The Impact of Biopsychosocial Factors on Quality of Life: Women with Primary Biliary Cirrhosis on Waiting List and Post Liver Transplantation. *Br J Health Psychol* (2011) 16:502–27. doi:10.1348/135910710X527964
17. Franciosi M, Caccamo L, De Simone P, Pinna AD, Di Costanzo GG, Volpes R, et al. Development and Validation of a Questionnaire Evaluating the Impact of Hepatitis B Immune Globulin Prophylaxis on the Quality of Life of Liver Transplant Recipients. *Liver Transpl* (2012) 18:332–9. doi:10.1002/lt.22473
18. Chen X, Zhang Y, Yu J. Symptom Experience and Related Predictors in Liver Transplantation Recipients. *Asian Nurs Res Korean Soc Nurs Sci* (2021) 15: 8–14. doi:10.1016/j.anr.2020.11.001
19. Xing L, Chen QY, Li JN, Hu ZQ, Zhang Y, Tao R. Self-Management and self-efficacy Status in Liver Recipients. *Hepatobiliary Pancreat Dis Int* (2015) 14: 253–62. doi:10.1016/s1499-3872(15)60333-2
20. Atamaz F, Hegguler S, Ozturk C, Pinar Y. Is QUALIOST Appropriate for the Patients With Orthotopic Liver Transplantation in Measuring Quality of Life? *Transpl Proc* (2013) 45:286–9. doi:10.1016/j.transproceed.2012.10.027
21. Fernandez AC, Fehon DC, Treloar H, Ng R, Sledge WH. Resilience in Organ Transplantation: An Application of the Connor-Davidson Resilience Scale (CD-RISC) with Liver Transplant Candidates. *J Pers Assess* (2015) 97:487–93. doi:10.1080/00223891.2015.1029620
22. Miller-Matero LR, Eshelman A, Paulson D, Armstrong R, Brown KA, Moonka D, et al. Beyond Survival: How Well Do Transplanted Livers Work? A Preliminary Comparison of standard-risk, high-risk, and Living Donor Recipients. *Clin Transpl* (2014) 28:691–8. doi:10.1111/ctr.12368
23. Pelgur H, Atak N, Kose K. Anxiety and Depression Levels of Patients Undergoing Liver Transplantation and Their Need for Training. *Transpl Proc* (2009) 41:1743–8. doi:10.1016/j.transproceed.2008.11.012
24. Lin XH, Teng S, Wang L, Zhang J, Shang YB, Liu HX, et al. Fatigue and Its Associated Factors in Liver Transplant Recipients in Beijing: A Cross-Sectional Study. *BMJ Open* (2017) 7:e011840. doi:10.1136/bmjopen-2016-011840
25. Weber S, Rek S, Eser-Valeri D, Padberg F, Reiter FP, De Toni E, et al. The Psychosocial Burden on Liver Transplant Recipients During the COVID-19 Pandemic. *Visc Med* (2021) 382:1–8. doi:10.1159/000517158
26. Gangeri L, Scignaro M, Bianchi E, Borreani C, Bhoorie S, Mazzaferro V. A Longitudinal Investigation of Posttraumatic Growth and Quality of Life in Liver Transplant Recipients. *Prog Transpl* (2018) 28:236–43. doi:10.1177/1526924818781569
27. Gronewold N, Schunn F, Ihrig A, Mayer G, Wohnsland S, Wagenlechner P, et al. Psychosocial Characteristics of Patients Evaluated for Kidney, Liver, or Heart Transplantation. *Psychosom Med* (2023) 85:98–105. doi:10.1097/PSY.0000000000001142
28. Ushio M, Makimoto K, Fujita K, Tanaka S, Kanaoka M, Kosai Y, et al. Validation of the LPA-SQUASH in post-liver-transplant Patients. *Jpn J Nurs Sci* (2023) 20:e12540. doi:10.1111/jjns.12540
29. Demir B, Bulbuloglu S. The Effect of Immunosuppression Therapy on Activities of Daily Living and Comfort Level After Liver Transplantation. *Transpl Immunol* (2021) 69:101468. doi:10.1016/j.trim.2021.101468
30. Scignaro M, Sani F, Wakefield JR, Bianchi E, Magrin ME, Gangeri L. Post-Traumatic Growth Enhances Social Identification in Liver Transplant Patients: A Longitudinal Study. *J Psychosom Res* (2016) 88:28–32. doi:10.1016/j.jpsychores.2016.07.004

31. Aiyegbusi OL, Kyte D, Cockwell P, Marshall T, Gheorghe A, Keeley T, et al. Measurement Properties of Patient-Reported Outcome Measures (Proms) Used in Adult Patients With Chronic Kidney Disease: A Systematic Review. *PLoS One* (2017) 12:e0179733. doi:10.1371/journal.pone.0179733
32. Elbers RG, Rietberg MB, van Wegen EEH, Verhoef J, Kramer SF, Terwee CB, et al. Self-Report Fatigue Questionnaires in Multiple Sclerosis, Parkinson's Disease and Stroke: A Systematic Review of Measurement Properties. (2025).
33. Green A, Liles C, Rushton A, Kyte DG. Measurement Properties of Patient-Reported Outcome Measures (PROMS) in Patellofemoral Pain Syndrome: A Systematic Review. *Man. Ther.* (2014) 517–26. doi:10.1016/j.math.2014.05.013
34. McHorney CA, Ware JE, Jr., Lu JF, Sherbourne CD. The MOS 36-item Short-form Health Survey (SF-36): III. Tests of Data Quality, Scaling Assumptions, and Reliability Across Diverse Patient Groups. *Med Care* (1994) 32:40–66. doi:10.1097/00005650-199401000-00004
35. Chiarotto A, Terwee CB, Kamper SJ, Boers M, Ostelo RW. Evidence on the Measurement Properties of Health-Related Quality of Life Instruments Is Largely Missing in Patients With Low Back Pain: A Systematic Review. *J Clin Epidemiol* (2018) 102:23–37. doi:10.1016/j.jclinepi.2018.05.006
36. Boers M, Kirwan JR, Wells G, Beaton D, Gossec L, d'Agostino MA, et al. Developing Core Outcome Measurement Sets for Clinical Trials: OMERACT Filter 2.0. *J Clin Epidemiol* (2014) 67:745–53. doi:10.1016/j.jclinepi.2013.11.013
37. Hjollund NHI. Fifteen Years' Use of Patient-Reported Outcome Measures at the Group and Patient Levels: Trend Analysis. *J Med Internet Res* (2019) 21: e15856. doi:10.2196/15856
38. Whittall A, Mereaglia M, Nicod E. The Use of Patient-Reported Outcome Measures in Rare Diseases and Implications for Health Technology Assessment. *Patient* (2021) 14:485–503. doi:10.1007/s40271-020-00493-w

Copyright © 2025 van Knippenberg, Powell-Brett, Joshi, Weeda and Hartog. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.