

Northumbria Research Link

Citation: Al Maqbali, Mohammed, Gracey, Jackie, Rankin, Jane, Dunwoody, Lynn, Hacker, Eileen and Hughes, Ciara (2020) Cross-Cultural Adaptation and Psychometric Properties of Quality of Life Scales for Arabic-Speaking Adults: A systematic review. Sultan Qaboos University Medical Journal [SQUMJ], 20 (2). e125-e137. ISSN 2075-051X

Published by: Sultan Qaboos University

URL: <https://doi.org/10.18295/squmj.2020.20.02.002>
<<https://doi.org/10.18295/squmj.2020.20.02.002>>

This version was downloaded from Northumbria Research Link:
<http://nrl.northumbria.ac.uk/id/eprint/47137/>

Northumbria University has developed Northumbria Research Link (NRL) to enable users to access the University's research output. Copyright © and moral rights for items on NRL are retained by the individual author(s) and/or other copyright owners. Single copies of full items can be reproduced, displayed or performed, and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided the authors, title and full bibliographic details are given, as well as a hyperlink and/or URL to the original metadata page. The content must not be changed in any way. Full items must not be sold commercially in any format or medium without formal permission of the copyright holder. The full policy is available online: <http://nrl.northumbria.ac.uk/policies.html>

This document may differ from the final, published version of the research and has been made available online in accordance with publisher policies. To read and/or cite from the published version of the research, please visit the publisher's website (a subscription may be required.)

Cross-Cultural Adaptation and Psychometric Properties of Quality of Life Scales for Arabic-Speaking Adults

A systematic review

*Mohammed Al Maqbali,^{1,2} Jackie Gracey,¹ Jane Rankin,³ Lynn Dunwoody,⁴ Eileen Hacker,⁵ Ciara Hughes¹

التكيف عبر الثقافات والخصائص النفسية لمقاييس جودة الحياة للسكان البالغين الناطقين بالعربية

مراجعة منهجية

محمد المقبالي، جاكى غريسي، جين رانكين، لين دنوودي، إيلين هاجر، سيارا هيوز

ABSTRACT: This review aimed to explore the psychometric properties of quality of life (QOL) scales to identify appropriate tools for research and clinical practice in Arabic-speaking adults. A systematic search of the Cumulative Index to Nursing and Allied Health Literature® (EBSCO Information Services, Ipswich, Massachusetts, USA), MEDLINE® (National Library of Medicine, Bethesda, Maryland, USA), EMBASE (Elsevier, Amsterdam, Netherlands) and PsycINFO (American Psychological Association, Washington, District of Columbia, USA) databases was conducted according to Preferred Reporting Items Systematic Reviews and Meta-Analysis guidelines. Quality assessment criteria were then utilised to evaluate the psychometric properties of identified QOL scales. A total of 27 studies relating to seven QOL scales were found. While these studies provided sufficient information regarding the scales' validity and reliability, not all reported translation and cross-cultural adaptation processes. Researchers and clinicians should consider whether the psychometric properties, subscales and characteristics of their chosen QOL scale are suitable for use in their population of interest.

Keywords: Quality of Life; Cross-Cultural Comparison; Translations; Psychometrics; Validity and Reliability; Surveys and Questionnaires; Systematic Review.

المخلص: تهدف هذه المراجعة المنهجية إلى لتقييم الخصائص السيكومترية لمقاييس جودة الحياة لتحديد الأدوات المناسبة للبحث والممارسة السريرية لدى السكان البالغين الناطقين بالعربية. تم إجراء بحث منهجي لقواعد البيانات الفهرس التراكمي للتمريض المؤلفات الصحية المتحالفة (إبيسكو لخدمات المعلومات، ابسويتش، ماساشوستس، الولايات المتحدة الأمريكية)، ميدلاين (المكتبة الوطنية للطب، بيتسدا، ماريلاند، الولايات المتحدة الأمريكية)، أمبيس (إلسيفير، أمستردام، هولندا) وسيكو انفو (الجمعية الأمريكية لعلم النفس، واشنطن، مقاطعة كولومبيا، الولايات المتحدة الأمريكية) وفقاً لبروتوكولات المراجعات المنهجية وإرشادات التحليل التلوي. ثم تم استخدام معايير تقييم الجودة لتقييم الخصائص النفسية لمقاييس جودة الحياة المحددة. تم العثور على ما مجموعه 27 دراسة تتعلق بسبعة مقاييس لجودة الحياة. في حين أن هذه الدراسات قدمت معلومات كافية فيما يتعلق بصحة المقاييس وموثوقيتها، إلا أنه لم يتم الإبلاغ عنها جميعاً لعمليات الترجمة والتكيف بين الثقافات. يجب على الباحثين والأطباء التفكير فيما إذا كانت الخصائص النفسية والقياسات الفرعية وخصائص مقياس جودة الحياة الذي تم اختياره مناسبة للاستخدام في السكان الذين يهتمون بهم.

الكلمات المفتاحية: جودة الحياة؛ مقارنة بين الثقافات؛ الترجمات؛ القياسات النفسية؛ الصلاحية والموثوقية؛ المسوحات والاستبيانات؛ مراجعة منهجية.

QUALITY OF LIFE (QOL) IS A MULTI-dimensional construct which relies on both personal characteristics as well as contextual and environmental variables.¹ In medical research and clinical practice, the assessment of QOL is important because it measures the effect of a disease or medical intervention on affected patients. In addition, QOL is an essential endpoint in treatment planning for

policy-makers, healthcare providers and the patient themselves.²

In recent years, focus on the patient's functioning, lifestyles and well-being have increased medical interest in tools for measuring QOL.² Accordingly, it is necessary to identify robust scales with satisfactory psychometric properties that can be used for this purpose. However, as most QOL assessment scales are

¹Institute of Nursing and Health Research, Ulster University, Shore Road, Jordanstown Campus, Newtownabbey, UK; ²Ministry of Health, Al Buraimi Hospital, Oman; ³Physiotherapy Department, Cancer Centre, Belfast Health and Social Care Trust, Belfast, UK; ⁴Psychology Research Institute, Ulster University, Coleraine, UK; ⁵School of Nursing, Indiana University, Indianapolis, USA

*Corresponding Author's e-mails: al_maqbali-ma@ulster.ac.uk and mahmedan@hotmail.com

initially designed in English, these scales need to be translated and adapted for use in different languages and cultures. Cross-cultural adaptation and translation is a systematic process that prepares questionnaires and scales for use in another setting.³ Nevertheless, it is crucial that the scale maintain its content validity after translation and cultural adaptation.

Reliability refers to the reproducibility or consistency of scores from one assessment to another, usually assessed via measures of internal consistency, inter- or intra-rater reliability or test-retest reliability.⁴ Internal consistency is generally reported as an alpha coefficient ranging from 0 (no correlation) to 1 (perfect correlation), with values of ≥ 0.70 and ≥ 0.90 considered acceptable and highly reliable, respectively.⁵ In contrast, validity is the ability of the scale to measure the attributes of the construct under consideration (i.e. the degree to which the scale measures that which it is intended to measure). Validity is divided into three types: content, construct and criterion validity, with the latter encompassing concurrent and predictive validity.⁶

Worldwide, there are approximately 420 million Arabic-speakers living in 23 countries.⁷ Generally, there are two main types of Arabic, with the first being modern standard Arabic primarily used in the written form in official and educational settings, while the second consists of differing regional and colloquial dialects.⁸ In 1998, Coons *et al.* conducted the first psychometric study to translate and validate an Arabic version of a QOL scale.⁹ Since then, many different QOL scales have been translated, resulting in a need to determine those which demonstrate satisfactory cross-cultural adaptation and validity. As such, this review aimed to explore the psychometric properties and translation and cross-cultural adaptation processes of Arabic QOL scales in order to identify appropriate scales that can be used for research and clinical practice in Arabic-speaking adults.

Methods

This systematic review was carried out according to the Preferred Reporting Items Systematic Reviews and Meta-Analysis guidelines.¹⁰ A systematic search of various electronic databases was conducted in order to identify studies investigating QOL among Arabic-speaking participants published between January 1946 and April 2019, including the Cumulative Index to Nursing and Allied Health Literature[®] (EBSCO Information Services, Ipswich, Massachusetts, USA), MEDLINE[®] (National Library of Medicine, Bethesda, Maryland, USA), EMBASE (Elsevier, Amsterdam, Netherlands) and PsycINFO (American Psychological Association, Washington, District of Columbia, USA) databases.

The search terms included combinations of free-text words and Medical Subject Headings[®] (National Library of Medicine) with Boolean operators (i.e. or/ and) as follows: “psychometrics”, “reliability”, “validity” or “instrument validation” and “Arabs” or “medicine, Arabic” and “functional status”, “wellbeing”, “quality of life”, “health status”, “health and life quality”, “quality of health care”, “assessment”, “patient assessment”, “clinical assessment tools”, “health impact assessment”, “clinical assessment tools”, “outcome assessment”, “measurement tool” or “questionnaires”. In addition, the reference lists of identified articles were screened to find other potential publications that could be included in the analysis.

All articles identified during the literature search were assessed to determine their eligibility. Articles were considered eligible for inclusion if they: (1) were published in English; (2) involved adults over 18 years of age; (3) were primarily psychometric studies with information concerning validity or reliability; (4) utilised QOL measures translated into Arabic; (5) involved an Arabic-speaking population; and (6) had no restrictions regarding study design. Studies with QOL scales developed and validated for a specific disease were excluded; however, those used for multiple types of cancer were permitted. Overall, a total of 1,087 articles were identified during the database search; however, this was reduced to 43 following screening of the titles and abstracts, with 27 articles meeting the inclusion criteria after full-text screening [Figure 1].

The psychometric properties of identified QOL scales were then evaluated according to nine quality assessment criteria suggested by Terwee *et al.*, including content validity, internal consistency, criterion validity, construct validity, reproducibility, responsiveness, floor and ceiling effects and interpretability [Table 1].¹¹ Each scale was given either a positive (+), indeterminate (?) or negative (-) rating for each of these measures, or a rating of 0 if no information was available. Terwee *et al.* recommended presenting the assessment results in a table, but not using an overall score, as this would bestow equal importance on each psychometric property which is not necessarily appropriate.¹¹

The cross-cultural adaptation and translation of the scales was evaluated according to the five-step guidelines of Guillemin *et al.*, namely: (1) translation, (2) back-translation, (3) committee review, (4) pre-testing and (5) re-examination of score weighting.¹² In the first step, at least two qualified translators should translate the scale from the original language to the target language. In the second step, two independent translators should translate the translated version back into the language of the original version to ensure that the translation reflects the content of the original.¹² The third step ideally involves a committee review to

develop the penultimate version for pre-testing and the fourth step, pilots this version among 30–40 subjects from the target population. The final step should be the re-examination of the weighting of the scores in light of cultural context.¹² Each study was assessed and given a score of either 1 (poor), 2 (moderate) or 3 (good) for each of these steps, with the overall score representing the mean of all scores obtained.

Results

STUDY CHARACTERISTICS

A total of 27 studies were included in the analysis, all of which were published between 1998 and 2019.^{9,13–38} The majority were conducted in the Middle Eastern and North African region, including Jordan (n = 7), Saudi Arabia (n = 4), Egypt (n = 2), Morocco (n = 2), Kuwait (n = 2), Tunisia (n = 2), Lebanon (n = 2), the United Arab Emirates (UAE; n = 2), Sudan (n = 1) and Qatar (n = 1).^{9,13,14,16–35,37,38} However, two studies were conducted in the Netherlands among samples of Moroccan Arabic-speaking subjects.^{15,36}

The majority of the studies (n = 21) had translated the QOL scales into modern standard Arabic suitable for use among all Arabic-speaking populations.^{9,13,18–35,38} However, six had translated the scales into Arabic dialects only suitable for specific populations, including Moroccan Arabic (n = 3), Tunisian Arabic (n = 2) and Egyptian Arabic (n = 1).^{14–17,36,37} All of the studies utilised quantitative research methods, with 20 cross-sectional and seven longitudinal surveys. None used a mixed-method approach [Table 2].^{9,13–38}

SCALE CHARACTERISTICS

Overall, the 27 articles included a total of seven self-reporting QOL scales that were translated and tested psychometrically in Arabic, including: (1) the 12- or 36-item Medical Outcomes Study Short-Form (SF-12 or SF-36); (2) the Dartmouth Cooperative Functional Health Assessment Charts/World Organisation of General Practice/Family Physicians (COOP/WONCA); (3) the World Health Organisation Quality of Life: Brief Version (WHOQOL-BREF); (4) the EuroQOL Group Health Status Index 5-Dimensions (EQ-5D); (5) the European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire Core Versions 30 or 15 Palliative (QLQ-C30 or QLQ-C15-PAL); (6) the Functional Assessment of Cancer Therapy-General (FACT-G); and (7) the Quality of Life Index (QLI).

PSYCHOMETRIC PROPERTIES

The psychometric properties of the QOL scales are detailed in Table 3.^{9,13–38} None of the studies tested all nine psychometric criteria suggested by Terwee *et al.*¹¹ In terms of content validity, 22 studies had a positive score.^{9,13–15,18–21,21–30,32,34–48} For the remaining five studies, no information was available.^{16,17,22,31,33} Internal consistency was generally high, with 26 studies scoring positively.^{9,13,15–38} Only one study did not report information regarding internal consistency.¹⁴ Criterion validity was tested in only two studies, with positive ratings for both.^{18,19} The remaining 25 studies did not provide any information concerning this psychometric property.^{9,13–17,20–38} Construct validity was assessed in 22 studies, of which 21 received positive ratings.^{9,13–15,18–32,37,38} Only one study was rated as intermediate for this aspect.¹⁶

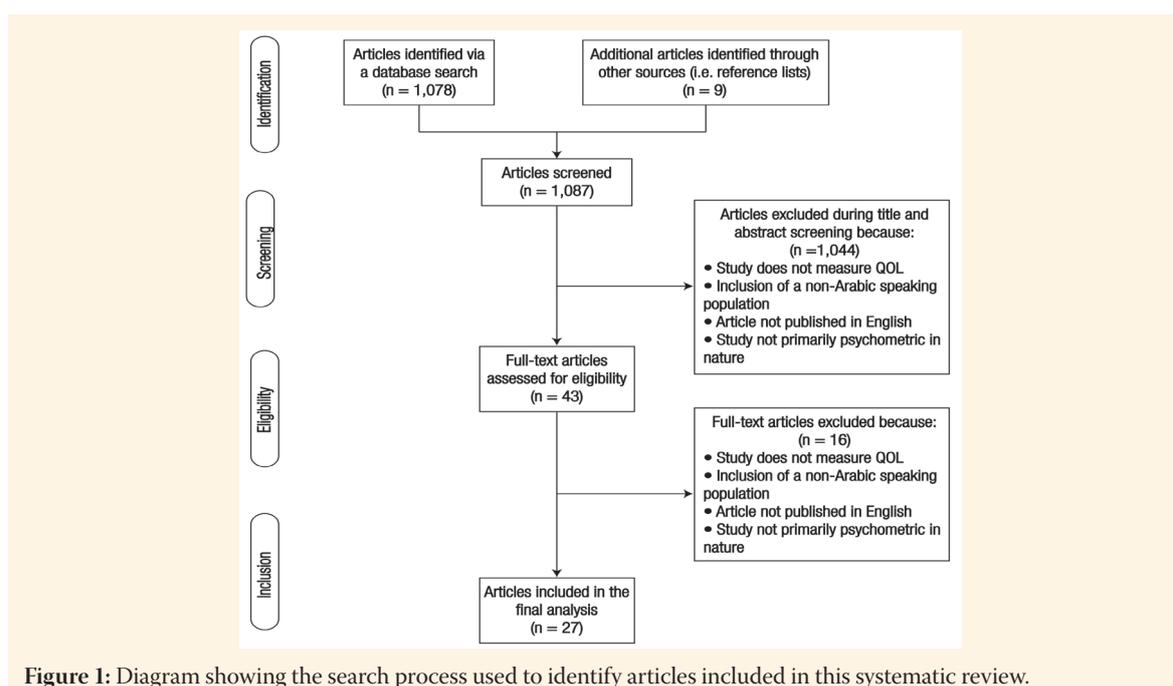


Figure 1: Diagram showing the search process used to identify articles included in this systematic review.

Table 1: Criteria for assessing the psychometric properties of quality of life scales¹¹

Property	Definition	Rating*	Quality criteria
Content validity	The extent to which the domain of interest is comprehensively sampled by the items in the questionnaire	+	A clear description is provided of the measurement aim, the target population, the concepts that are being measured and the item selection AND both target population and investigators OR experts are involved in item selection
		?	A clear description of the aforementioned aspects is missing OR only the target population is involved OR doubtful [†] design or methods
		-	No target population involvement
		0	No information found
Internal consistency	The extent to which items in a scale or subscale are intercorrelated (i.e. measuring the same construct)	+	Factor analyses are performed on an adequate sample size (calculated to be at least seven times the number of items AND >100) AND Cronbach's alpha(s) is calculated per dimension AND Cronbach's alpha(s) is between 0.70–0.95
		?	No factor analysis OR doubtful [†] design or methods
		-	Cronbach's alpha is <0.70 or >0.95, despite adequate design and methods
		0	No information found
Criterion validity	The extent to which scores on a particular questionnaire refer to a gold standard	+	Convincing arguments to support gold standard AND correlation with Cronbach's alpha of >0.70
		?	No convincing arguments to support gold standard OR doubtful [†] design or methods
		-	Correlation with Cronbach's alpha of 0.70 AND continuous adequate design and methods
		0	No information found
Construct validity	The extent to which scores on a particular questionnaire refer to other measures in a manner consistent with theoretically supported hypotheses relating to the concepts being measured	+	Specific hypotheses are formed AND at least 75% of the results are in accordance with these hypotheses
		?	Doubtful [†] design or methods (e.g. no hypotheses)
		-	Less than 75% of the hypotheses are confirmed, despite adequate design and methods
		0	No information found
Reproducibility			
Agreement	The extent to which scores on repeated measures are close to each other (i.e. absolute measurement error)	+	The SDC is less than the MIC OR the MIC is outside the LOA OR convincing arguments that the level of agreement is acceptable
		?	Doubtful [†] design or methods OR the MIC is not defined AND no convincing arguments that the level of agreement is acceptable
		-	The MIC is less than or equal to the SDC OR the MIC equals or is inside the LOA, despite adequate design and methods
		0	No information found
Reliability	The extent to which subjects can be distinguished from each other, despite measurement errors (i.e. relative measurement error)	+	The ICC or Cohen's weighted kappa is >0.70
		?	Doubtful [†] design or methods (e.g. time interval not mentioned)
		-	The ICC or weighed Kappa is ≤0.70, despite adequate design and methods
		0	No information found
Responsiveness	The ability of a questionnaire to detect clinically important changes over time	+	The MIC is less than the SDC OR the MIC is outside the LOA OR the RR is 1.96 OR the AUC is >0.70
		?	Doubtful [†] design or methods
		-	The SDC is more than or equal to the MIC OR the MIC equals or is inside the LOA OR the RR is <1.96 OR the AUC is 0.70, despite adequate design and methods
		0	No information found

SDC = smallest detectable change; MIC = minimal important change; LOA = limits of agreement; ICC = intraclass correlation; RR = responsiveness ratio; AUC = area under the curve; SD = standard deviation.

*Ratings were either positive (+), intermediate (?), negative (-) or no information was available (0). [†]Either the study lacks a clear description of its design or methods, the sample size is under 50 subjects in each subgroup analysis or there are important methodological weaknesses in its design or execution.

Table adapted with permission from 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; 60:34–42.¹¹

Table 1 (cont'd): Criteria for assessing the psychometric properties of quality of life scales¹¹

Property	Definition	Rating*	Quality criteria
Floor and ceiling effects	The number of responders who achieve the lowest or highest possible scores	+	<15% of the respondents achieve the highest or lowest possible scores
		?	Doubtful [†] design or methods
		-	<15% of the respondents achieve the highest or lowest possible scores, despite adequate design and methods
		0	No information found
Interpretability	The degree to which one can assign qualitative meaning to quantitative scores	+	Mean and SD scores are presented for at least four relevant subgroups of patients AND the MIC is defined
		?	Doubtful [†] design or methods OR mean and SD scores are presented for less than four subgroups OR no MIC is defined
		0	No information found

SDC = smallest detectable change; MIC = minimal important change; LOA = limits of agreement; ICC = intraclass correlation; RR = responsiveness ratio; AUC = area under the curve; SD = standard deviation.

*Ratings were either positive (+), intermediate (?), negative (-) or no information was available (0). [†]Either the study lacks a clear description of its design or methods, the sample size is under 50 subjects in each subgroup analysis or there are important methodological weaknesses in its design or execution.

Table adapted with permission from 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; 60:34–42¹¹

Table 2: Characteristics of studies involving quality of life scales translated and adapted for Arabic-speaking adults (N = 27)^{9,13–38}

Author and year of publication	Country	Study design	Type of participants	Sample size	QOL scale	Language	Cronbach's α coefficient
Huijjer <i>et al.</i> ²⁸ (2013)	Lebanon	Cross-sectional	Mixed cancer patients	200	EORTC QLQ-C30	Standard Arabic	• Overall: <0.70 • Range: 0.38–0.80
Awad <i>et al.</i> ²⁶ (2008)	UAE	Cross-sectional	Breast cancer patients	87	EORTC QLQ-C30	Standard Arabic	• Overall: <0.70 • Range: 0.51–0.84
Alawadhi and Ohaeri ³⁵ (2010)	Kuwait	Cross-sectional	Breast cancer patients	348	EORTC QLQ-C30	Standard Arabic	• Overall: 0.91 • Range: 0.51–0.84
Bener <i>et al.</i> ²⁷ (2017)	Qatar	Cross-sectional	Breast cancer patients	678	EORTC QLQ-C30	Standard Arabic	• Overall: 0.91 • Range: 0.55–0.89
Alawneh <i>et al.</i> ²⁵ (2010)	Jordan	Cross-sectional	Mixed cancer patients	175	EORTC QLQ-C15-PAL	Standard Arabic	• Overall: <0.70 • Range: 0.72–0.90
Lazenby <i>et al.</i> ²⁹ (2013)	Jordan	Cross-sectional	Mixed cancer patients	205	FACT-G	Standard Arabic	• Range: 0.80–0.83
Zahran <i>et al.</i> ³⁰ (2017)	Egyptian	Cross-sectional	Bladder cancer patients	90	FACT-G	Standard Arabic	• Range: 0.80–0.94
Al Barmawi <i>et al.</i> ²⁴ (2018)	Jordan	Cross-sectional	Head and/or neck cancer patients	118	FACT-G	Standard Arabic	• Overall: 0.76 • Range: 0.67–0.83
Soudy <i>et al.</i> ²³ (2018)	Saudi Arabia	Cross-sectional	Lymphoma patients who had undergone stem cell transplantation	108	FACT-G	Standard Arabic	• Overall: 0.89 • Range: 0.67–0.88
Coons <i>et al.</i> ⁹ (1998)	Saudi Arabia	Longitudinal	General population	415	SF-36	Standard Arabic	• Range: 0.60–0.87
Sabbah <i>et al.</i> ¹⁸ (2003)	Lebanon	Cross-sectional	General population	524	SF-36	Standard Arabic	• Range: 0.70–0.90
Hoopman <i>et al.</i> ³⁶ (2009)	Netherlands	Longitudinal	General population	Subgroup of 377 Moroccan subjects	SF-36	Local dialect (Tarifit)	• Range: 0.63–0.93
Hoopman <i>et al.</i> ¹⁵ (2006)	Netherlands	Longitudinal	Mixed cancer patients	Subgroup of 79 Moroccan patients	SF-36	Local dialect (Tarifit)	• Range: 0.65–0.94

QOL = quality of life; EORTC QLQ = European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire; C30 = Core Version 30; UAE = United Arab Emirates; C15-PAL = Core Version 15 Palliative; FACT-G = Functional Assessment of Cancer Therapy - General; SF-36 = 36-item Medical Outcomes Study Short-Form; ICU = intensive care unit; SF-12 = 12-item Medical Outcomes Study Short-Form; EQ-5D = EuroQOL Group Health Status Index 5-Dimensions; WHOQOL-BREF = World Health Organisation Quality of Life: Brief Version; COOP/WONCA = Dartmouth Cooperative Functional Health Assessment Charts/World Organisation of General Practice/Family Physicians; QLI = Quality of Life Index.

Table 2 (cont'd): Characteristics of studies involving quality of life scales translated and adapted for Arabic-speaking adults (N = 27)

Author and year of publication	Country	Study design	Type of participants	Sample size	QOL scale	Language	Cronbach's α coefficient
Khoudri <i>et al.</i> ³⁷ (2007)	Morocco	Cross-sectional	Patients discharged from the ICU	145	SF-36	Standard Arabic	• Overall: ≥ 0.70 • Range: 0.84–0.99
Guermazi <i>et al.</i> ³⁸ (2012)	Tunisia	Cross-sectional	General population	130	SF-36	Local dialect (Tunisian)	• Overall: 0.94 • Range: 0.72–0.89
El-Kalla <i>et al.</i> ¹⁷ (2016)	Egypt	Longitudinal	Patients with burn injuries	40	SF-36	Local dialect (Egyptian)	• Overall: 0.8
Sheikh <i>et al.</i> ³¹ (2015)	Saudi Arabia	Cross-sectional	Khat chewers	300	SF-36	Standard Arabic	• Overall: 0.94 • Range: 0.72–0.90
Khader <i>et al.</i> ²² (2011)	Jordan	Cross-sectional	General population	511	SF-36	Standard Arabic	• Range: 0.71–0.90
Younsi and Chakroun ¹⁶ (2014)	Tunisia	Cross-sectional	General population	3,582	SF-12	Local dialect (Tunisian)	• Overall: 0.73
Aburuz <i>et al.</i> ¹³ (2009)	Jordan	Cross-sectional	General population	186	EQ-5D	Standard Arabic	• Overall: ≥ 0.75
Bekairy <i>et al.</i> ³² (2018)	Saudi Arabia	Longitudinal	Mixed patients	80	EQ-5D	Standard Arabic	• Overall: 0.72
Ohaeri and Awadalla ²¹ (2009)	Kuwait	Longitudinal	General population	3,303	WHOQOL-BREF	Standard Arabic	• Overall: 0.90 • Range: 0.69–0.83
Ohaeri <i>et al.</i> ²⁰ (2007)	Sudan	Cross-sectional	General population and psychiatric patients	623	WHOQOL-BREF	Standard Arabic	• Overall: 0.88 (general population), 0.93 (psychiatric patients) and 0.92 (caregivers)
Bani-Issa ³³ (2011)	UAE	Cross-sectional	Diabetic patients	200	WHOQOL-BREF	Standard Arabic	• Overall: 0.85 • Range: 0.89–0.91
Dalky <i>et al.</i> ¹⁹ (2017)	Jordan	Cross-sectional	Family/caregivers of patients	266	WHOQOL-BREF	Standard Arabic	• Overall: 0.92
Hoopman <i>et al.</i> ¹⁴ (2008)	Morocco	Cross-sectional	Mixed cancer patients	Subgroup of 37 Moroccan patients	COOP/WONCA	Local dialect (Tarifit)	• Not reported
Halabi ³⁴ (2006)	Jordan	Longitudinal	General population and hypertensive, diabetic, cancer and dialysis patients.	35	QLI	Standard Arabic	• Overall: 0.90

QOL = quality of life; EORTC QLQ = European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire; C30 = Core Version 30; UAE = United Arab Emirates; C15-PAL = Core Version 15 Palliative; FACT-G = Functional Assessment of Cancer Therapy - General; SF-36 = 36-item Medical Outcomes Study Short-Form; ICU = intensive care unit; SF-12 = 12-item Medical Outcomes Study Short-Form; EQ-5D = EuroQOL Group Health Status Index 5-Dimensions; WHOQOL-BREF = World Health Organisation Quality of Life: Brief Version; COOP/WONCA = Dartmouth Cooperative Functional Health Assessment Charts/World Organisation of General Practice/Family Physicians; QLI = Quality of Life Index.

Information concerning agreement was present in only one study which received a positive score.³⁸ Reliability was investigated in nine studies, of which seven scored positively.^{9,13,17,25,32,36,37} The remaining two studies received intermediate scores.^{20,21} Two studies provided information regarding responsiveness, with one positive and one intermediate rating.^{14,15} Floor and ceiling effects were tested in six studies, with four scoring positively.^{15,18,22,36} The other two studies received intermediate ratings.^{14,21} Only one study reported information concerning interpretability, receiving an intermediate score.¹³

Overall, the SF-36 demonstrated the most robust psychometric properties, followed by the WHOQOL-BREF. The SF-36 was tested using eight psychometric criteria, with positive ratings for content validity, internal consistency, criterion validity, construct validity, agreement, reliability, responsiveness and floor and ceiling effects.^{9,15,17,18,22,31,36–38} Similarly, the WHOQOL-BREF received positive scores for content validity, internal consistency, criterion validity and construct validity, although both reliability and floor and ceiling effects were rated as indeterminate.^{19–21,33}

Table 3: Psychometric properties of scales in studies involving quality of life scales translated and adapted for Arabic-speaking adults (N = 27)^{9,13–38}

Author and year of publication	QOL scale	Rating*								
		Content validity	Internal consistency	Criterion validity	Reproducibility		Construct validity	Responsiveness	Floor and ceiling effects	Interpretability
Huijjer <i>et al.</i> ²⁸ (2013)	EORTC QLQ-C30	+	+	0	+	0	0	0	0	0
Awad <i>et al.</i> ²⁶ (2008)	EORTC QLQ-C30	+	+	0	+	0	0	0	0	0
Alawadhi and Ohaeri ³⁵ (2010)	EORTC QLQ-C30	+	+	0	0	0	0	0	0	0
Bener <i>et al.</i> ²⁷ (2017)	EORTC QLQ-C30	+	+	0	+	0	0	0	0	0
Alawneh <i>et al.</i> ²⁵ (2010)	EORTC QLQ-C15-PAL	+	+	0	+	0	+	0	0	0
Lazenby <i>et al.</i> ²⁹ (2013)	FACT-G	+	+	0	+	0	0	0	0	0
Zahran <i>et al.</i> ³⁰ (2017)	FACT-G	+	+	0	+	0	0	0	0	0
Al Barmawi <i>et al.</i> ²⁴ (2018)	FACT-G	+	+	0	+	0	0	0	0	0
Soudy <i>et al.</i> ²³ (2018)	FACT-G	+	+	0	+	0	0	0	0	0
Coons <i>et al.</i> ⁹ (1998)	SF-36	+	+	0	+	0	+	0	0	0
Sabbah <i>et al.</i> ¹⁸ (2003)	SF-36	+	+	+	+	0	0	0	+	0
Hoopman <i>et al.</i> ³⁶ (2009)	SF-36	+	+	0	0	0	0	0	+	0
Hoopman <i>et al.</i> ¹⁵ (2006)	SF-36	+	+	0	+	0	0	+	+	0
Khoudri <i>et al.</i> ³⁷ (2007)	SF-36	+	+	0	+	0	+	0	0	0
Guermazzi <i>et al.</i> ³⁸ (2012)	SF-36	+	+	0	+	+	+	0	0	0
El-Kalla <i>et al.</i> ¹⁷ (2016)	SF-36	0	+	0	0	0	+	0	0	0
Sheikh <i>et al.</i> ³¹ (2015)	SF-36	0	+	0	+	0	0	0	0	0
Khader <i>et al.</i> ²² (2011)	SF-36	0	+	0	+	0	0	0	+	0

Table 3 (cont'd): Psychometric properties of scales in studies involving quality of life scales translated and adapted for Arabic-speaking adults (N = 27)^{9,13–38}

Author and year of publication	QOL scale	Rating*									
		Content validity	Internal consistency	Criterion validity	Reproducibility		Construct validity	Responsiveness	Floor and ceiling effects	Interpretability	
Aburuz <i>et al.</i> ¹³ (2009)	EQ-5D	+	+	0	+	0	+	0	+	0	?
Bekairy <i>et al.</i> ³² (2018)	EQ-5D	+	+	0	+	0	+	0	+	0	0
Ohaeri and Awadalla ²¹ (2009)	WHOQOL-BREF	+	+	0	+	0	?	0	?	0	0
Ohaeri <i>et al.</i> ²⁰ (2007)	WHOQOL-BREF	+	+	0	+	0	?	0	0	0	0
Bani-Issa ³³ (2011)	WHOQOL-BREF	0	+	0	0	0	0	0	0	0	0
Dalky <i>et al.</i> ¹⁹ (2017)	WHOQOL-BREF	+	+	+	+	0	0	0	0	0	0
Hoopman <i>et al.</i> ¹⁴ (2008)	COOP/WONCA	+	0	0	+	0	0	?	?	0	0
Halabi ³⁴ (2006)	QLI	+	+	0	0	0	0	0	0	0	0

QOL = quality of life; EORTC QLQ = European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire; C30 = Core Version 30; C15-PAL = Core Version 15 Palliative; FACT-G = Functional Assessment of Cancer Therapy - General; SF-36 = 36-item Medical Outcomes Study Short-Form; SF-12 = 12-item Medical Outcomes Study Short-Form; EQ-5D = EuroQOL Group Health Status Index 5-Dimensions; WHOQOL-BREF = World Health Organisation Quality of Life: Brief Version; COOP/WONCA = Dartmouth Cooperative Functional Health Assessment Charts/World Organisation of General Practice/Family Physicians; QLI = Quality of Life Index.

*Ratings were scored as either positive (+), intermediate (?), negative (-) or no information available (0).

TRANSLATION AND CULTURAL ADAPTATION

The processes of translation and cultural adaptation of the QOL scales are presented in Table 4.^{9,13–38} In total, 14 studies reported information regarding translation and cross-cultural adaptation processes.^{9,13–15,17,18,23–25,28–30,34,38} However, only two studies adopted all five steps recommended by Guillemin *et al.*^{9,12,23} A total of nine studies reported four of the steps, without providing any information regarding the re-evaluation of score weightings.^{13,14,17,18,28–30,34,38} One study presented a three-step technique (including translation, back-translation and pre-testing), while another reported only the first two steps.^{15,24} Finally, in one study, a single-step technique consisting solely of forward-translation was performed.²⁵

The EORTC-QLQ-C30, SF-36, FACT-G and QLI scales received overall mean scores of 3 with regards to

Table 4: Cross-cultural adaptation and translation processes of scales in studies involving quality of life scales translated and adapted for Arabic-speaking adults (N = 27)^{9,13–38}

Author and year of publication	QOL scale	Score*					
		Translation	Back-translation	Committee approach	Pre-testing	Reassessment of score weighting	Overall mean score weighting
Huijjer <i>et al.</i> ²⁸ (2013)	EORTC QLQ-C30	3	3	3	3	N/A	3
Awad <i>et al.</i> ²⁶ (2008)	EORTC QLQ-C30	N/A	N/A	N/A	N/A	N/A	N/A
Alawadhi and Ohaeri ³⁵ (2010)	EORTC QLQ-C30	N/A	N/A	N/A	N/A	N/A	N/A
Bener <i>et al.</i> ²⁷ (2017)	EORTC QLQ-C30	N/A	N/A	N/A	N/A	N/A	N/A
Alawneh <i>et al.</i> ²⁵ (2010)	EORTC QLQ-C15-PAL	2	N/R	N/R	N/R	N/R	2
Lazenby <i>et al.</i> ²⁹ (2013)	FACT-G	3	3	3	3	N/A	3
Zahran <i>et al.</i> ³⁰ (2017)	FACT-G	3	3	3	3	N/A	3
Al Barmawi <i>et al.</i> ²⁴ (2018)	FACT-G	3	3	N/R	3	N/R	3
Soudy <i>et al.</i> ²³ (2018)	FACT-G	3	3	3	3	3	3
Coons <i>et al.</i> ⁹ (1998)	SF-36	3	3	2	3	3	2.8
Sabbah <i>et al.</i> ¹⁸ (2003)	SF-36	3	3	3	3	N/R	3
Hoopman <i>et al.</i> ³⁶ (2009)	SF-36	N/A	N/A	N/A	N/A	N/A	N/A
Hoopman <i>et al.</i> ¹⁵ (2006)	SF-36	2	1	N/R	N/R	N/R	1.5
Khoudri <i>et al.</i> ³⁷ (2007)	SF-36	N/A	N/A	N/A	N/A	N/A	N/A
Guermazi <i>et al.</i> ³⁸ (2012)	SF-36	3	3	1	3	N/R	2.5
El-Kalla <i>et al.</i> ¹⁷ (2016)	SF-36	3	3	3	3	N/R	3
Sheikh <i>et al.</i> ³¹ (2015)	SF-36	N/A	N/A	N/A	N/A	N/A	N/A
Khader <i>et al.</i> ²² (2011)	SF-36	N/A	N/A	N/A	N/A	N/A	N/A
Younsi and Chakroun ¹⁶ (2014)	SF-12	N/A	N/A	N/A	N/A	N/A	N/A
Aburuz <i>et al.</i> ¹³ (2009)	EQ-5D	3	2	2	3	N/A	2.5
Bekairy <i>et al.</i> ³² (2018)	EQ-5D	N/A	N/A	N/A	N/A	N/A	N/A
Ohaeri and Awadalla ²¹ (2009)	WHOQOL-BREF	N/A	N/A	N/A	N/A	N/A	N/A

Table 4 (cont'd): Cross-cultural adaptation and translation processes of scales in studies involving quality of life scales translated and adapted for Arabic-speaking adults (N = 27)^{9,13–38}

Author and year of publication	QOL scale	Score*					
		Translation	Back-translation	Committee approach	Pre-testing	Reassessment of score weighting	Overall mean score weighting
Ohaeri <i>et al.</i> ²⁰ (2007)	WHOQOL-BREF	N/A	N/A	N/A	N/A	N/A	N/A
Bani-Issa ³³ (2011)	WHOQOL-BREF	N/A	N/A	N/A	N/A	N/A	N/A
Dalky <i>et al.</i> ¹⁹ (2017)	WHOQOL-BREF	N/A	N/A	N/A	N/A	N/A	N/A
Hoopman <i>et al.</i> ¹⁴ (2008)	COOP/WONCA	2	2	3	3	N/R	2.5
Halabi ³⁴ (2006)	QLI	3	3	3	3	N/R	3

QOL = quality of life; EORTC QLQ = European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire; C30 = Core Version 30; NA = not applicable; C15-PAL = Core Version 15 Palliative; NR = not reported; FACT-G = Functional Assessment of Cancer Therapy - General; SF-36 = 36-item Medical Outcomes Study Short-Form; SF-12 = 12-item Medical Outcomes Study Short-Form; EQ-5D = EuroQOL Group Health Status Index 5-Dimensions; WHOQOL-BREF = World Health Organisation Quality of Life: Brief Version; COOP/WONCA = Dartmouth Cooperative Functional Health Assessment Charts/World Organisation of General Practice/Family Physicians; QLI = Quality of Life Index.

*Each step of the process was scored as either good (3), moderate (2) or poor (1).

translation and cultural adaptation processes.^{17,18,23,24,28–30,34} In addition, the overall mean score of the COOP/WONCA and EQ-5D scales was 2.5.^{13,14} However, there was no information regarding translation or cross-cultural adaptation for any of the studies seeking to validate the WHOQOL-BREF scale.^{19–21,33}

INDIVIDUAL QUALITY OF LIFE SCALES *European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire*

The EORTC has developed several scales to assess the QOL of cancer patients.^{39–42} The EORTC-QLQ-C30 consists of nine multi-item scales, five functional subscales (assessing physical, role, cognitive, emotional and social functioning), three symptom subscales (assessing fatigue, pain and nausea/vomiting) and a global health status and QOL scale.³⁹ In addition, six items assessing other common symptoms of cancer are included (dyspnoea, insomnia, appetite loss, constipation, diarrhoea and financial difficulties). The first 28 items of the scale are scored on a 4-point Likert scale, with scores ranging from 1 (not at all) to 4 (very much).³⁹ The remaining two items are assessed on a seven-point numeric scale. The original version of the EORTC-QLQ-C30 scored a Cronbach's α coefficient of ≥ 0.70 .³⁹

Overall, four studies sought to validate and translate the EORTC-QLQ-C30 for use in Arabic-speaking adults.^{26–28,35} All of the studies translated the scale into modern standard Arabic and were cross-sectional in nature. In total, the overall sample of these studies consisted of 1,313 cancer patients.^{26–28,35} Generally, the scale showed satisfactory psychometric properties consistent with its purpose for use among Arabic-speaking cancer patients. In terms of internal consistency, the coefficient alpha of the Arabic versions was >0.70, in line with that of the original version.^{26–28,35}

In the EORTC-QLQ-C15-PAL scale, the 30-item QLQ core version is reduced to 15 items for the purposes of addressing QOL in palliative care.^{39,43} The EORTC-QLQ-C15-PAL includes three multi-item scales, functional subscales (assessing physical and emotional functioning), symptom subscales (assessing fatigue and pain) and a global health status and QOL scale.⁴³ The first 14 items of the scale are scored on a 4-point Likert scale, with scores ranging from 1 (not at all) to 4 (very much). The final item is assessed on a 7-point numeric scale.⁴³ Alawneh *et al.* investigated the validity and reliability of a standard Arabic version of the EORTC-QLQ-C15-PAL scale among 175 Jordanian mixed cancer patients, with a coefficient alpha of >0.70.²⁵

Functional Assessment of Cancer Therapy – General

The FACT-G scale consists of 27 items assessed on a 5-point Likert scale and was originally validated for a population of mixed cancer patients.⁴⁴ The first part of the scale assesses three QOL dimensions (physical well-being, social/family wellbeing and functional well-being) using seven items and a fourth (emotional well-being) with six items. The scale also has specific items that can be added to the general scale for specific types of cancer.^{45,46} In addition to cancer, the FACT-G scale has been used and validated for use among patients with other chronic illnesses as well as the general population.^{47–49}

Four studies were conducted to evaluate the psychometric properties of Arabic versions of the FACT-G on 521 subjects in total, each involving a different subtype of cancer patient (mixed, lymphoma, bladder and head-and/or-neck cancer).^{23,24,29,30} In addition, one study assessed the FACT-G in conjunction with a spiritual subscale.²⁹ All four studies were translated into modern standard Arabic. The internal consistency of the Arabic FACT-G scales yielded almost the same results as that of the original scale (coefficient alpha: 0.76–0.89).^{23,24,29,30,44} However, none of the studies assessed reproducibility properties such as test-retest reliability or agreement.^{23,24,29,30}

Medical Outcomes Study Short-Form

The SF-36 scale is a 36-item multi-purpose health survey consisting of eight multi-item subscales (assessing physical functioning, emotional problems, physical

problems, mental health, bodily pain, general health, social functioning and vitality) and one single-item subscale (assessing health transition).⁵⁰ The total score ranges from 0 to 100, with higher scores indicating a better QOL. The reliability of the original SF-36 scale was high, with an intraclass correlation coefficient (ICC) of >0.8.⁵⁰

Nine studies evaluated the psychometric properties of Arabic versions of the SF-36 scale.^{9,15,17,18,22,31,36–38} The scale was tested on multiple populations, including the general population (n = 4), burn victims (n = 1), cancer patients (n = 1), patients admitted to an intensive care unit (n = 1) and *khat* chewers (n = 1), with a total sample size of 2,521.^{9,15,17,18,22,31,36–38} Four studies translated the SF-36 scale into three different dialects of Arabic, including Moroccan *Tarifit* (n = 2), Tunisian (n = 1) and Egyptian (n = 1).^{15,17,36,38} The other five studies translated the scale into standard Arabic.^{9,18,22,31,37} With regards to internal consistency, the coefficient alpha of the Arabic SF-36 scale ranged from 0.70–0.94.^{9,15,17,18,22,31,36–38} Test-retest reliability was assessed in four studies, with the ICC exceeding 0.70.^{9,15,17,37,38}

The SF-12 scale is a shorter 12-item version of the SF-36 scale and assesses the same eight health domains as the original.⁵¹ Younsi and Chakroun tested the SF-12 scale among 3,582 members of the general population.¹⁶ The scale was translated into the Tunisian Arabic dialect, with a coefficient alpha of 0.73.¹⁶

EuroQOL Group Health Status Index 5-Dimensions

The EQ-5D is a 5-item scale assessing five dimensions of QOL (mobility, self-care, usual activities, pain/discomfort and anxiety/depression).⁵² Each item has three possible responses, including no problems, some/moderate problems and extreme problems. In addition, health states are measured using a visual analogue scale ranging from 0 to 100.⁵²

Two studies evaluated Arabic versions of the EQ-5D.^{13,32} Both were translated into modern standard Arabic. Aburuz *et al.* investigated the validity and reliability of the EQ-5D in a sample of 186 members of the general population in Jordan.¹³ In contrast, Bekairy *et al.* assessed its use among 80 mixed patients in Saudi Arabia.³² Both Arabic EQ-5D scales were deemed valid and reliable, with coefficient alphas of ≥ 0.72 . In terms of test-retest reliability, Aburuz *et al.* and Bekairy *et al.* reported Cohen's kappa values of 0.48–1.0 and 0.53–1.00, respectively.^{13,32}

WORLD HEALTH ORGANISATION QUALITY OF LIFE: BRIEF VERSION

The WHOQOL-BREF scale is a 26-item questionnaire scored on a 5-point Likert scale which was originally validated to measure QOL among people with diseases in the general population.⁵³ The scale represents an abbreviated version of the much longer 100-item

WHOQOL assessment.⁵⁴ The WHOQOL-BREF has four subscales (assessing physical health, psychological health, social relationships and environmental health) and two overall QOL and general health items. In terms of internal consistency, the coefficient alpha of the original WHOQOL-BREF scale was 0.66–0.84.⁵³

Four studies sought to validate standard Arabic versions of the WHOQOL-BREF in different countries, including Kuwait, Sudan, UAE and Jordan.^{19–21,33} The total sample size for all four studies was 4,392, including both psychiatric and diabetic patients, family members and caregivers of patients and members of the general population. The coefficient alpha of the Arabic WHOQOL-BREF scales ranged from 0.69–0.93, indicating acceptable internal consistency. In one study, the test-retest reliability of the scale was significant (ICC = 0.95).²¹

Dartmouth Cooperative Functional Health Assessment Charts/World Organisation of General Practice/Family Physicians

The COOP/WONCA scale contains six items assessed using a 5-point Likert scale and covering core QOL functional domains (physical fitness, feelings, daily activities, social activities, changes in health and overall health).⁵⁵ The COOP/WONCA scale was culturally adapted and translated into Arabic in only one study.¹⁴ Hoopman *et al.* assessed the use of the scale on 37 mixed cancer patients when translated into *Tarifit*, a local dialect of Arabic spoken in Morocco. The scale was found to have adequate content and construct validity, but its discriminant validity could only be partially confirmed.¹⁴

Quality of Life Index

The QLI scale consists of 70 items scored on a 6-point Likert scale assessing health and functioning and socio-economic, psychological/spiritual and family-related aspects of QOL.⁵⁶ The scale was designed to assess the QOL of both healthy and ill individuals. The original scale has been validated in many different diseases.^{57–59} In terms of internal consistency, the coefficient alpha of the original QLI scale was 0.73–0.99.⁶⁰

Only one study culturally adapted, translated and tested the QLI in modern standard Arabic.³⁴ The study involved 35 subjects, including both healthy individuals as well as hypertensive, diabetic, cancer and dialysis patients. The reliability of the scale was adequate, with an ICC of 0.88–0.97.³⁴

Discussion

This review identified 27 studies assessing seven QOL scales translated and tested for validity and reliability in Arabic-speaking adults.^{9,13–38} None of the scales were

originally developed in Arabic, with the majority initially developed for use in English. All of the QOL scales were consistent in that they assessed both physical and psychological aspects as well as other important components of QOL. Nevertheless, in order to fully understand QOL in Arabic-speaking populations, there is a need for QOL scales to be properly translated and culturally adapted for use in these populations.

All of the studies included in this review utilised quantitative research methods, with 20 cross-sectional and seven longitudinal surveys.^{9,13–38} Cross-sectional studies indicate that the data were collected at a specific point in time without further follow-up, while longitudinal data were collected over different periods of time.⁶¹ Generally, longitudinal studies give more precise information regarding temporal changes or treatment effects that can have an important impact on the QOL of patients. In contrast, researchers conducting cross-sectional analyses will have more difficulty creating a cohesive narrative regarding the impact of medical treatment, interventions or other variables on QOL.⁶¹

None of the studies in this review evaluated the psychometric criteria as suggested by Terwee *et al.*¹¹ As such, further psychometric studies are required to improve the validity and reliability of Arabic versions of QOL scales. For instance, only seven studies positively assessed test-retest reliability and only one reported test-retest agreement.^{9,13,17,25,32,37,38} In terms of specific scales, no reliability properties were reported for the EORTC QLQ-C30, FACT-G, WHOQOL-BREF, COOP/WONCA or QLI scales. Accordingly, the reliability and validity of these QOL scales should be evaluated prior to their use in Arabic-speaking populations.

Selection of an appropriate QOL scale is dependent on a number of different factors, including the demographic and clinical characteristics of the sample, the psychometric properties of the scale and the number of items included in the scale. Most importantly, researchers need to consider the various aspects and domains of QOL that require evaluation in their population of interest. For example, if the sample consists of cancer patients, the FACT-G or EORTC QLQ-C30 scales would be most appropriate as both can be used in multiple types of cancer.^{39–42,44–46} In addition, both scales have been validated among Arab cancer patients, as well as members of the general population speaking other languages.^{23,24,26–30,35,49,62,63} However, it should be noted that the EORTC-QLQ-C30 scale does not address either spiritual or existential components of QOL.³⁹

On the other hand, if the sample consists of a general Arabic-speaking population, the SF-36 might

be a better choice in order to provide more generic QOL-related information.⁵⁰ While the studies in this review included a variety of populations, the general population was most frequently studied, perhaps because this choice provides a larger sample size, thus improving the psychometric evaluation. With regards to the COOP/WONCA scale, this scale was validated by only one study involving a small sample size (N = 37) and translated into a local dialect.¹⁴ Further examination of the psychometric properties of a modern standard Arabic version of this scale is therefore required before it can be recommended for use among other Arabic-speaking adults.

All seven QOL scales identified in this review are of varying lengths, consisting of between 5–70 items. Overall, six of the scales contain fewer than 36 items and could therefore be administered between 5–10 minutes.^{39,44,50,52,53,55} The QLI is much longer with a total of 70 items, although the administration time is reported as being approximately 10 minutes.⁵⁶ Unlike other QOL scales, the QLI weighs satisfaction in a particular domain of QOL in terms of importance, so that items with high satisfaction and importance scores receive the highest score. Nevertheless, the length of this questionnaire could present an obstacle when conducting research in a clinical setting, as the inclusion of more items in a survey tends to discourage high response rates.^{64,65}

Processes of cross-cultural adaptation and translation affect the credibility of an adapted scale by ensuring that the content of the translated scale is equivalent to that of the original. Adhering to a systematic and standardised approach, such as that suggested by Guillemin *et al.*, produces cultural equivalence and maximises acceptability of the linguistic structure of the translated scale.¹² Unfortunately, only two studies included in this review followed all five recommended steps.^{9,23} In contrast, these processes were only partially reported or not reported at all by the remaining 25 studies.^{13–22,24–38} This may be because the scales had originally been translated into Arabic in earlier studies. Nevertheless, it is recommended that researchers identify and report detailed information regarding each stage of the cross-cultural adaptation process when translating and adapting QOL scales for use in Arabic-speaking populations.

This review was subject to several limitations. The focus of the analysis was primarily on the psychometric properties of QOL scales; as such, further research is necessary to evaluate the quality of the design and methodologies of the reported studies. In addition, a single researcher undertook the screening and assessed the eligibility of the articles included in the analysis. This may have increased the risk of bias or

resulted in possible errors during the data collection process. Finally, although a systematic search of multiple electronic databases was conducted using various search terms in different combinations, it is possible that some relevant studies were unintentionally overlooked and not included in the analysis.

Conclusion

This review evaluated the psychometric properties and cultural adaptation and translation processes of Arabic versions of QOL scales. In general, the studies provided insufficient information regarding the exact processes of translation and cultural adaptation. Additionally, while most scales provided sufficient information regarding the content and construct validity and internal consistency of the scales, information related to agreement, responsiveness, floor and ceiling effects and interpretability was lacking. Specifically, the test-retest reliability, criterion validity and sensitivity of Arabic QOL scales requires further validation. Future research involving the translation and cultural adaptation of QOL scales should utilise recommended guidelines to ensure the content of the translated scale is equivalent to that of the original.

References

1. Verdugo MA, Schalock RL, Keith KD, Stancliffe RJ. Quality of life and its measurement: Important principles and guidelines. *J Intellect Disabil Res* 2005; 49:707–17. <https://doi.org/10.1111/j.1365-2788.2005.00739.x>.
2. Romero M, Vivas-Consuelo D, Alvis-Guzman N. Is health related quality of life (HRQoL) a valid indicator for health systems evaluation? *Springerplus* 2013; 2:664. <https://doi.org/10.1186/2193-1801-2-664>.
3. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976)* 2000; 25:3186–91. <https://doi.org/10.1097/00007632-200012150-00014>.
4. Downing SM. Reliability: On the reproducibility of assessment data. *Med Educ* 2004; 38:1006–12. <https://doi.org/10.1111/j.1365-2929.2004.01932.x>.
5. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: Theory and application. *Am J Med* 2006; 119:166.e7–16. <https://doi.org/10.1016/j.amjmed.2005.10.036>.
6. Messick S. Standards of validity and the validity of standards in performance assessment. *Educ Meas Issues Pract* 1995; 14:5–8. <https://doi.org/10.1111/j.1745-3992.1995.tb00881.x>.
7. Istizada. Complete list of Arabic speaking countries: 2018. From: <http://istizada.com/complete-list-of-arabic-speaking-countries-2014/> Accessed: Oct 2019.
8. Sadat F, Kazemi F, Farzindar A. Automatic identification of Arabic dialects in social media. In: *SoMeRA '14: Proceedings of the First International Workshop on Social Media Retrieval and Analysis*. New York, USA: Association for Computing Machinery, 2014. Pp. 35–40. <https://doi.org/10.1145/2632188.2632207>.
9. Coons SJ, Alabdulmohsin SA, Draugalis JR, Hays RD. Reliability of an Arabic version of the RAND-36 Health Survey and its equivalence to the US-English version. *Med Care* 1998; 36:428–32. <https://doi.org/10.1097/00005650-199803000-00018>.

10. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Int J Surg* 2010; 8:336–41. <https://doi.org/10.1016/j.ijsu.2010.02.007>.
11. 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. <https://doi.org/10.1016/j.jclinepi.2006.03.012>.
12. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: Literature review and proposed guidelines. *J Clin Epidemiol* 1993; 46:1417–32. [https://doi.org/10.1016/0895-4356\(93\)90142-n](https://doi.org/10.1016/0895-4356(93)90142-n).
13. Aburuz S, Bulatova N, Twalbeh M, Gazawi M. The validity and reliability of the Arabic version of the EQ-5D: A study from Jordan. *Ann Saudi Med* 2009; 29:304–8. <https://doi.org/10.4103/0256-4947.55313>.
14. Hoopman R, Terwee CB, Aaronson NK. Translated COOP/WONCA charts found appropriate for use among Turkish and Moroccan ethnic minority cancer patients. *J Clin Epidemiol* 2008; 61:1036–48. <https://doi.org/10.1016/j.jclinepi.2007.11.018>.
15. Hoopman R, Terwee CB, Muller MJ, Aaronson NK. Translation and validation of the SF-36 Health Survey for use among Turkish and Moroccan ethnic minority cancer patients in the Netherlands. *Eur J Cancer* 2006; 42:2982–90. <https://doi.org/10.1016/j.ejca.2006.08.011>.
16. Younsi M, Chakroun M. Measuring health-related quality of life: Psychometric evaluation of the Tunisian version of the SF-12 health survey. *Qual Life Res* 2014; 23:2047–54. <https://doi.org/10.1007/s11136-014-0641-8>.
17. El-Kalla RA, Khalaf MM, Saad MA, Othman EM. Reliability of the Arabic Egyptian version of Short Form 36 Health Survey questionnaire to measure quality of life in burned patient. *Med J Cairo Univ* 2016; 84:311–16.
18. Sabbah I, Drouby N, Sabbah S, Retel-Rude N, Mercier M. Quality of life in rural and urban populations in Lebanon using SF-36 Health Survey. *Health Qual Life Outcomes* 2003; 1:30. <https://doi.org/10.1186/1477-7525-1-30>.
19. Dalky HF, Meininger JC, Al-Ali NM. The reliability and validity of the Arabic World Health Organization Quality of Life-BREF instrument among family caregivers of relatives with psychiatric illnesses in Jordan. *J Nurs Res* 2017; 25:224–30. <https://doi.org/10.1097/JNR.000000000000146>.
20. Ohaeri JU, Awadalla AW, El-Abassi AH, Jacob A. Confirmatory factor analytical study of the WHOQOL-Bref: Experience with Sudanese general population and psychiatric samples. *BMC Med Res Methodol* 2007; 7:37. <https://doi.org/10.1186/1471-2288-7-37>.
21. Ohaeri JU, Awadalla AW. The reliability and validity of the short version of the WHO Quality of Life instrument in an Arab general population. *Ann Saudi Med* 2009; 29:98–104. <https://doi.org/10.4103/0256-4947.51790>.
22. Khader S, Hourani MM, Al-Akour N. Normative data and psychometric properties of Short Form 36 Health Survey (SF-36, version 1.0) in the population of north Jordan. *East Mediterr Health J* 2011; 17:368–74.
23. Soudy H, Maghfoor I, Elhassan TAM, Abdullah E, Rauf SM, Al Zahrani A, et al. Translation and validation of the Functional Assessment of Cancer Therapy-Bone Marrow Transplant (FACT-BMT) version 4 quality of life instrument into Arabic language. *Health Qual Life Outcomes* 2018; 16:47. <https://doi.org/10.1186/s12955-018-0861-7>.
24. Al Barmawi M, Al Hadid LA, Alqudah HN, Al Hadid WA, Shamoun SA. Measuring the quality of life among head-and/or-neck cancer patients with oral mucositis using the Functional Assessment of Cancer Therapy-General in Jordan. *Asia Pac J Oncol Nurs* 2018; 5:320–6. https://doi.org/10.4103/apjon.apjon_14_18.
25. Alawneh A, Yasin H, Khirfan G, Qayas BA, Ammar K, Rimawi D, et al. Psychometric properties of the Arabic version of EORTC QLQ-C15-PAL among cancer patients in Jordan. *Support Care Cancer* 2016; 24:2455–62. <https://doi.org/10.1007/s00520-015-3018-9>.
26. Awad MA, Denic S, El Taji H. Validation of the European Organization for Research and Treatment of Cancer quality of life questionnaires for Arabic-speaking populations. *Ann N Y Acad Sci* 2008; 1138:146–54. <https://doi.org/10.1196/annals.1414.021>.
27. Bener A, Alsulaiman R, Doodson L, El Ayoubi HR. An assessment of reliability and validity of the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire C30 among breast cancer patients in Qatar. *J Family Med Prim Care* 2017; 6:824–31. https://doi.org/10.4103/jfmpc.jfmpc_17_17.
28. Huijjer HA, Sagherian K, Tamim H. Validation of the Arabic version of the EORTC quality of life questionnaire among cancer patients in Lebanon. *Qual Life Res* 2013; 22:1473–81. <https://doi.org/10.1007/s11136-012-0261-0>.
29. Lazenby M, Khatib J, Al-Khair F, Neamat M. Psychometric properties of the Functional Assessment of Chronic Illness Therapy--Spiritual Well-being (FACIT-Sp) in an Arabic-speaking, predominantly Muslim population. *Psychooncology* 2013; 22:220–7. <https://doi.org/10.1002/pon.2062>.
30. Zahran MH, Sheir K, Zidan EM, Ali-El-Dein B. Validation of the Arabic version of the Functional Assessment of Cancer Therapy-Bladder questionnaire in Egyptian patients with bladder cancer. *Arab J Urol* 2017; 15:110–14. <https://doi.org/10.1016/j.aju.2017.01.008>.
31. Sheikh KA, Yagoub U, El-Setouhy M, Alsansosy R, Mohamud SA. Reliability and validity of the Arabic version of the SF-36 Health Survey questionnaire in population of khat chowers: Jazan region-Kingdom of Saudi Arabia. *Appl Res Qual Life* 2015; 10:1–13. <https://doi.org/10.1007/s11482-013-9295-x>.
32. Bekairy AM, Bustami RT, Almotairi M, Jarab A, Katheri AM, Aldebasi TM, et al. Validity and reliability of the Arabic version of the the EuroQOL (EQ-5D): A study from Saudi Arabia. *Int J Health Sci (Qassim)* 2018; 12:16–20.
33. Bani-Issa W. Evaluation of the health-related quality of life of Emirati people with diabetes: Integration of sociodemographic and disease-related variables. *East Mediterr Health J* 2011; 17:825–30. <https://doi.org/10.26719/2011.17.11.825>.
34. Halabi JO. Psychometric properties of the Arabic version of Quality of Life Index. *J Adv Nurs* 2006; 55:604–10. https://doi.org/10.1111/j.1365-2648.2006.03952_1.x.
35. Alawadhi SA, Ohaeri JU. Validity and reliability of the European Organization for Research and Treatment in Cancer Quality of Life Questionnaire (EORTC QLQ): Experience from Kuwait using a sample of women with breast cancer. *Ann Saudi Med* 2010; 30:390–6. <https://doi.org/10.4103/0256-4947.67083>.
36. Hoopman R, Terwee CB, Devillé W, Knol DL, Aaronson NK. Evaluation of the psychometric properties of the SF-36 Health Survey for use among Turkish and Moroccan ethnic minority populations in the Netherlands. *Qual Life Res* 2009; 18:753–64. <https://doi.org/10.1007/s11136-009-9491-1>.
37. Khoudri I, Ali Zeggwagh A, Abidi K, Madani N, Abouqal R. Measurement properties of the Short Form 36 and health-related quality of life after intensive care in Morocco. *Acta Anaesthesiol Scand* 2007; 51:189–97. <https://doi.org/10.1111/j.1399-6576.2006.01225.x>.
38. Guermazi M, Allouch C, Yahia M, Huissa TB, Ghorbel S, Damak J, et al. Translation in Arabic, adaptation and validation of the SF-36 Health Survey for use in Tunisia. *Ann Phys Rehabil Med* 2012; 55:388–403. <https://doi.org/10.1016/j.rehab.2012.05.003>.
39. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: A quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst* 1993; 85:365–76. <https://doi.org/10.1093/jnci/85.5.365>.
40. Sprangers MA, Groenvold M, Arraras JL, Franklin J, te Velde A, Muller M, et al. The European Organization for Research and Treatment of Cancer breast cancer-specific quality-of-life questionnaire module: First results from a three-country field study. *J Clin Oncol* 1996; 14:2756–68. <https://doi.org/10.1200/JCO.1996.14.10.2756>.

41. Bjordal K, Ahlner-Elmqvist M, Tolleson E, Jensen AB, Razavi D, Maher EJ, et al. Development of a European Organization for Research and Treatment of cancer (EORTC) questionnaire module to be used in quality of life assessments in head and neck cancer. *Acta Oncol* 1994; 33:879–85. <https://doi.org/10.3109/02841869409098450>.
42. Bergman B, Aaronson NK, Ahmedzai S, Kaasa S, Sullivan M. The EORTC QLQ-LC13: A modular supplement to the EORTC Core Quality of Life Questionnaire (QLQ-C30) for use in lung cancer clinical trials. EORTC Study Group on Quality of Life. *Eur J Cancer* 1994; 30A:635–42. [https://doi.org/10.1016/0959-8049\(94\)90535-5](https://doi.org/10.1016/0959-8049(94)90535-5).
43. Groenvold M, Petersen MA, Aaronson NK, Arraras JI, Blazeby JM, Bottomley A, et al. The development of the EORTC QLQ-C15-PAL: A shortened questionnaire for cancer patients in palliative care. *Eur J Cancer* 2006; 42:55–64. <https://doi.org/10.1016/j.ejca.2005.06.022>.
44. Cella DF, Tulsky DS, Gray G, Sarafian B, Linn E, Bonomi A, et al. The Functional Assessment of Cancer Therapy scale: Development and validation of the general measure. *J Clin Oncol* 1993; 11:570–9. <https://doi.org/10.1200/JCO.1993.11.3.570>.
45. Brady MJ, Cella DF, Mo F, Bonomi AE, Tulsky DS, Lloyd SR, et al. Reliability and validity of the Functional Assessment of Cancer Therapy-Breast quality-of-life instrument. *J Clin Oncol* 1997; 15:974–86. <https://doi.org/10.1200/JCO.1997.15.3.974>.
46. Tong MC, Lo PS, Wong KH, Yeung RM, van Hasselt CA, Eremenco S, et al. Development and validation of the functional assessment of cancer therapy nasopharyngeal cancer subscale. *Head Neck* 2009; 31:738–47. <https://doi.org/10.1002/hed.21023>.
47. Cella DF, McCain NL, Peterman AH, Mo F, Wolen D. Development and validation of the Functional Assessment of Human Immunodeficiency Virus Infection (FAHI) quality of life instrument. *Qual Life Res* 1996; 5:450–63. <https://doi.org/10.1007/bf00449920>.
48. Cella DF, Dineen K, Arnason B, Reder A, Webster KA, karabatsos G, et al. Validation of the functional assessment of multiple sclerosis quality of life instrument. *Neurology* 1996; 47:129–39. <https://doi.org/10.1212/wnl.47.1.129>.
49. Brucker PS, Yost K, Cashy J, Webster K, Cella D. General population and cancer patient norms for the Functional Assessment of Cancer Therapy-General (FACT-G). *Eval Health Prof* 2005; 28:192–211. <https://doi.org/10.1177/0163278705275341>.
50. 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. <https://doi.org/10.1097/00005650-199401000-00004>.
51. Ware J Jr, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: Construction of scales and preliminary tests of reliability and validity. *Med Care* 1996; 34:220–33. <https://doi.org/10.1097/00005650-199603000-00003>.
52. Rabin R, de Charro F. EQ-5D: A measure of health status from the EuroQol Group. *Ann Med* 2001; 33:337–43. <https://doi.org/10.3109/07853890109002087>.
53. World Health Organization. Development of the World Health Organization WHOQOL-BREF quality of life assessment: The WHOQOL Group. *Psychol Med* 1998; 28:551–8. <https://doi.org/10.1017/s0033291798006667>.
54. World Health Organization. The World Health Organization Quality of Life Assessment (WHOQOL): Development and general psychometric properties. *Soc Sci Med* 1998; 46:1569–85. [https://doi.org/10.1016/s0277-9536\(98\)00009-4](https://doi.org/10.1016/s0277-9536(98)00009-4).
55. Van Weel C. Functional status in primary care: COOP/WONCA charts. *Disabil Rehabil* 1993; 15:96–101. <https://doi.org/10.3109/09638289309165878>.
56. Ferrans CE, Powers MJ. Quality of Life Index: Development and psychometric properties. *ANS Adv Nurs Sci* 1985; 8:15–24. <https://doi.org/10.1097/00012272-198510000-00005>.
57. Rannestad T, Skjeldestad FE. Ferrans and Powers' quality of life index applied in urinary incontinence research: A pilot study. *Scand J Caring Sci* 2011; 25:410–16. <https://doi.org/10.1111/j.1471-6712.2010.00852.x>.
58. May LA, Warren S. Measuring quality of life of persons with spinal cord injury: Substantive and structural validation. *Qual Life Res* 2001; 10:503–15. <https://doi.org/10.1023/a:1013027520429>.
59. Buck D, Jacoby A, Massey A, Ford G. Evaluation of measures used to assess quality of life after stroke. *Stroke* 2000; 31:2004–10. <https://doi.org/10.1161/01.str.31.8.2004>.
60. Ferrans and Powers Quality of Life Index. Reliability and validity. From: <https://qli.org.uic.edu/reliability/reliabilityhome.htm> Accessed: Oct 2019.
61. Polit DF, Beck CT. *Nursing Research: Generating and assessing evidence for nursing practice*, 10th ed. Philadelphia, Pennsylvania, USA: Lippincott Williams & Wilkins, 2016. Pp. 160–81.
62. Nolte S, Liegl G, Petersen MA, Aaronson NK, Costantini A, Fayers PM, et al. General population normative data for the EORTC QLQ-C30 health-related quality of life questionnaire based on 15,386 persons across 13 European countries, Canada and the Unites States. *Eur J Cancer* 2019; 107:153–63. <https://doi.org/10.1016/j.ejca.2018.11.024>.
63. Pearman T, Yanez B, Peipert J, Wortman K, Beaumont J, Cella D. Ambulatory cancer and US general population reference values and cutoff scores for the functional assessment of cancer therapy. *Cancer* 2014; 120:2902–9. <https://doi.org/10.1002/cncr.28758>.
64. Mark TL, Johnson G, Fortner B, Ryan K. The benefits and challenges of using computer-assisted symptom assessments in oncology clinics: Results of a qualitative assessment. *Technol Cancer Res Treat* 2008; 7:401–6. <https://doi.org/10.1177/153303460800700508>.
65. Rolstad S, Adler J, Rydén A. Response burden and questionnaire length: Is shorter better? A review and meta-analysis. *Value Health* 2011; 14:1101–8. <https://doi.org/10.1016/j.jval.2011.06.003>.