REVIEW ARTICLE

Reliability of quality-of-life assessment instruments in children and adolescents with burn scars: systematic review

Confiabilidade de instrumentos de avaliação de qualidade de vida em crianças e adolescentes com cicatrizes de queimaduras: revisão sistemática

Fiabilidad de los instrumentos de evaluación de la calidad de vida en niños y adolescentes con cicatrices de quemaduras: revisión sistemática

Francieli Ferreira de Andrade Batista[1]
Elisângela Flauzino Zampar[1]
Renne Rodrigues[1]
Francisco Xavier de Araújo[2]
Maria Elena Echevarría-Guanilo[3]
Flavia Meneguetti Pieri[1]
Rosangela Aparecida Pimenta[1]

[1] Universidade Estadual de Londrina (UEL), Londrina, Paraná, Brasil.

[2] Universidade Federal de Pelotas (UFPel), Pelotas, Rio Grande do Sul, Brasil.

[3] Universidade Federal de Santa Catarina (UFSC), Florianópolis, Santa Catarina, Brasil.

Corresponding author:
Francieli Ferreira de Andrade Batista
E-mail: franandradebatista@gmail.com

How to cite this article: Batista FFA, Zampar EF, Rodrigues R, Araújo FX, Guanilo MEE, Pieri FM, Pimenta RA. Confiabilidade de instrumentos de avaliação de qualidade de vida em crianças e adolescentes com cicatrizes de queimaduras: revisão sistemática. Rev. Eletr. Enferm. 2024;26:76914. https://doi.org/10.5216/ree.v26.76914 English, Portuguese.

Extracted from the Master's thesis: "Cultural adaptation and reliability assessment of the Brazilian version of the Brisbane Burn Scar Impact Profile for children under eight years of age", defended in 2021 in the *Programa de Pós-Graduação em Enfermagem* at the Universidade Estadual de Londrina, Londrina, Paraná, Brazil.

Received: 3 August 2023
Accepted: 19 February 2024
Published *online*: 27 July 2024

ABSTRACT

Objective: to investigate the reliability measures of health-related quality of life (HRQOL) instruments in children or adolescents with burn scars. **Methods:** systematic review of the literature performed through a search in CINAHL, Google Scholar, LILACS, PsycINFO, PubMed, Scopus and Web of Science without restrictions on the date or language of publication. Original studies were included, in which instruments for assessing the HRQOL of children and/or adolescents (≤ 18 years old) with burn scars were examined. The criteria of the Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) guidelines were used to analyze the measurement properties of HRQOL instruments. **Results:** nine studies were included, and they covered the following instruments: Children Burn Outcomes Questionnaire (BOQ) 0-4 years version; Children Burn Outcomes Questionnaire (BOQ) 5-18 years version; Brisbane Burn Scar Impact Profile (BBSIP) 0-under 8 years old version; Brisbane Burn Scar Impact Profile (BBSIP) 8-18 years old version; and CARe Burn Scale (CBS) 0-8 years old. **Conclusion:** the BBSIP was the instrument that met the greatest number of quality items according to the COSMIN.

Descriptors: Burns; Quality of Life; Surveys and Questionnaires.

RESUMO

Objetivo: *investigar as medidas de confiabilidade de instrumentos de qualidade de vida relacionada a saúde (QVRS) em crianças ou adolescentes com cicatrizes de queimaduras.* ***Métodos:*** *trata-se de uma revisão sistemática da literatura, mediante busca no CINAHL, Google Scholar, LILACS, PsycINFO, PubMed, Scopus e Web of Science, sem restrições quanto à data ou idioma de publicação. Foram incluídos estudos originais, nos quais se examinaram instrumentos de avaliação da QVRS de crianças e/ou adolescentes (≤ 18 anos) com cicatrizes de queimaduras. Para a análise das propriedades de medida dos instrumentos de QVRS, utilizaram-se os critérios das diretrizes do Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN).* ***Resultados:*** *nove estudos foram incluídos, dos quais foram abordados os seguintes instrumentos: Children Burn Outcomes Questionnaire (BOQ) – versão de 0 a 4 anos, Children Burn Outcomes Questionnaire (BOQ) – versão 5 a 18 anos, Brisbane Burn Scar Impact Profile (BBSIP) – versão de 0 a menores de 8 anos, Brisbane Burn Scar Impact Profile (BBSIP) – versão de 8 a 18 anos e CARe Burn Scale (CBS) – 0 a 8 anos.* ***Conclusão:*** *o BBSIP foi o instrumento que atendeu o maior número de itens de qualidade pela COSMIN.*

Descritores: *Queimaduras; Qualidade de vida; Inquéritos e questionários.*

RESUMEN

Objetivo: *investigar las medidas de confiabilidad de los instrumentos de calidad de vida relacionada con la salud (CVRS) en niños o adolescentes con cicatrices de quemaduras.* ***Métodos:*** *revisión sistemática de la literatura mediante búsqueda en CINAHL, Google Scholar, LILACS, PsycINFO, PubMed, Scopus y Web of Science, sin restricciones de fecha o idioma de publicación. Se incluyeron estudios originales, en los que se examinaron instrumentos para evaluar la CVRS de niños y/o adolescentes (≤ 18 años) con cicatrices de quemaduras. Se utilizaron los criterios de las directrices de los Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) para analizar las propiedades de medición de los instrumentos de CVRS.* ***Resultados:*** *se incluyeron nueve estudios que abarcaron los siguientes instrumentos: Children Burn Outcomes Questionnaire (BOQ) – versión 0-4 años; Children Burn Outcomes Questionnaire (BOQ) – versión 5-18 años; Brisbane Burn Scar Impact Profile (BBSIP) – versión 0-menos de 8 años; Brisbane Burn Scar Impact Profile (BBSIP) – version 8-18 años; y CARe Burn Scale (CBS) – versión 0-8 años.* ***Conclusión:*** *el BBSIP fue el instrumento que cumplió con el mayor número de ítems de calidad según el COSMIN.*

Descriptores: *Quemaduras; Calidad de Vida; Encuestas y Cuestionarios.*

INTRODUCTION

Burns are considered a global public health problem, and although preventable[1], according to estimates from the Institute of Health Metrics and Evaluation (IHME), in 2019 they had an incidence of 7.1 million cases in children and adolescents under the age of 20 years with a predominance of children under 15 years (incidence of 5.2 million cases). In addition to the risk of mortality, burns cause disability that affects more than 400 thousand children and adolescents under 20 years[2].

The damage caused by burns is physical and emotional[1], begins with the injury, continues during hospitalization and until discharge[3], causing a profound impact on the Health-Related Quality of Life (HRQoL)[4,5]. The HRQoL is a subjective assessment of well-being based on the analysis of different domains of quality of life, such as health, social, economic, psychological-spiritual, family aspects, subjective well-being and also its impact on the health of patients[6]. In this way, the assessment of HRQoL can help in the adaptation of subsequent care according to the patients' needs[7].

Both survivors of pediatric burns and their caregivers report emotional and behavioral problems linked to HRQoL that may be related to other health outcomes[8,9].

Note that burns not only directly affect HRQoL in the short term, but by triggering stress, pain, anxiety, emotional difficulties[3], scars and changes in daily routine[4,5] they also indirectly affect the HRQoL.

In turn, the long-term effects of HRQoL are not yet well studied in children and adolescents, even though the decrease in HRQoL may directly or indirectly influence mental disorders, with higher rates of psychopathology, increased traumatic stress[8], sleep disorders, history of substance abuse, social isolation and depression, which are linked to suicide[9].

The HRQoL is an outcome measure self-reported by patients (Patient Reported Outcome Measures — PROMs), and its properties must be well evaluated in order to allow for better clinical use in the process of caring for children and adolescents who are victims of burns[10,11].

Generic instruments are still widely used to assess the HRQoL of children and adolescents with burns, however they ignore some important aspects[12], such as problems related to burn scars or itching[13,14]. The absence of important domains for HRQoL coupled with the low measurement accuracy of general instruments when applied to children and/or adolescents with burn scars or their caregivers make such instruments unsuitable for clinical practice[12]. In turn, as specific instruments for HRQoL in patients with burns have better measurement properties, they are suitable for application in this population or their caregivers[12].

Even though the different specific instruments present advantages[12], studies assessing their suitability for clinical practice and research are needed. The appropriate instrument must be easy to use and have adequate properties regarding reliability, validity and sensitivity[15,16]. The low number of studies evaluating the quality of these instruments[12] configures a gap regarding their measurement properties. Considering that the validity of an instrument is supported by the accumulation of evidence[17], a systematic review on the subject can support the adoption of more appropriate tools for clinical practice and future research.

Given the importance of evaluating the impact of burn scars on HRQoL in the pediatric population and the need to measure the properties of available instruments, the aim of this study was to investigate the reliability measures of HRQoL instruments in children or adolescents with burn scars.

METHODS

A systematic review was performed in line with recommendations of the Joanna Briggs Institute (JBI)[18]. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) was used for the report[19], and the Consensus-based Standards for the Selection of Health Measuring INstruments (COSMIN) was used to analyze the properties of the instruments and risk of bias[20]. The protocol was prospectively registered in the International Prospective Register of Systematic Reviews (PROSPERO) under number CRD42020160728.

Search strategy

A search for original articles was carried out in seven electronic databases: Cumulative Index to Nursing and Allied Health Literature (CINAHL); Google Scholar; Latin American and Caribbean Health Sciences Literature (LILACS); PsycINFO; Medical Literature Analysis and Retrieval System Online (MEDLINE) of the National Library of Medicine of the United States of America using the search engine PubMed (MEDLINE/PubMed); Scopus; and Web of Science, with no limit on search period or language used.

The keywords used in the search were defined based on the Medical Subject Headings (MeSH) and the Health Sciences Descriptors (DeCS). Thesauruses and/or keywords related to the target population (children or adolescents with burn scars) and assessment of the reliability of HRQoL instruments were used combined by Boolean operators AND or OR according to the database or repository. Based on the results of the first search and to increase the sensitivity of the strategy, the references of all articles included in the study and reviews on the topic (not included in this review) were analyzed in search of possible new studies to be included. This process improved the search strategy (Supplementary Table 1, Appendix 1). The search took place in May 2020 and was updated in June 2023 (FFAB and RR).

Eligibility and study selection criteria

Studies that met the following criteria were included: original articles with specific instruments/PROMs measuring the HRQoL in children or adolescents with burn scars, which investigated and reported the study results of at least one of the following measures used to assess the reliability of an instrument: internal consistency, test-retest or error of measurement. Articles that analyzed data from adolescent and adult patients simultaneously were excluded.

The articles were initially selected by two reviewers independently, based on the title, abstract and keywords to identify studies that met the research topic (children or adolescents with burn scars and HRQoL). At this step, all selected articles were organized in the Mendeley bibliographic reference manager to remove duplicates and organize the next step. Then, the reviewers (still independently) evaluated the articles in full to identify those that met the inclusion criteria. There were disagreements between reviewers in two studies, which were resolved by joint discussion with a third reviewer.

Data collection and assessment of methodological quality

Data from the articles were extracted and organized in a Microsoft Excel® spreadsheet, version 2016 (Microsoft Corporation, United States), in which including the title of the study, author, year of publication, PROM language, age, number of participants and the properties evaluated (internal consistency, test-retest stability and error of measurement). In addition, methodological quality (risk of bias) and quality of evidence were assessed.

Data extraction and assessment of methodological quality were performed independently by two reviewers. The disagreements present in one analysis were resolved in discussion with two other reviewers.

The reliability and methodological quality of the studies were evaluated according to the COSMIN checklist[21]. The measurement property for each PROM was summarized as follows: sufficient (+); insufficient (-); inconsistent (±); or indeterminate (?). The overall result was evaluated according to the criteria of good measurement properties. After evaluating the measurement properties in each study, the results were grouped for each assessment instrument according to the COSMIN Risk of Bias Checklist for Systematic Reviews of Patient-Reported Outcome Measures[21] guideline under the name of general assessment.

Finally, the quality of evidence for internal consistency, test-retest and error of measurement for each instrument was classified as high, moderate, low or very low, according to the Grading of Recommendations Assessment Development, and Evaluation (GRADE)[20].

The GRADE classification includes four factors:

1. Risk of bias (by the methodological quality of the evaluation of studies);
2. Inconsistency (unexplained inconsistency of results between studies);
3. Imprecision (total sample size of available studies); and
4. Indirect evidence (population evidence different from the population of interest in the review)[20].

All studies started with a high rating, and according to defined criteria, there was a downgrade for each condition not met in the evaluated factors.

RESULTS

The total of 662 articles were found. The analysis of inclusion and exclusion criteria resulted in a sample of nine studies, all published in English (Figure 1).

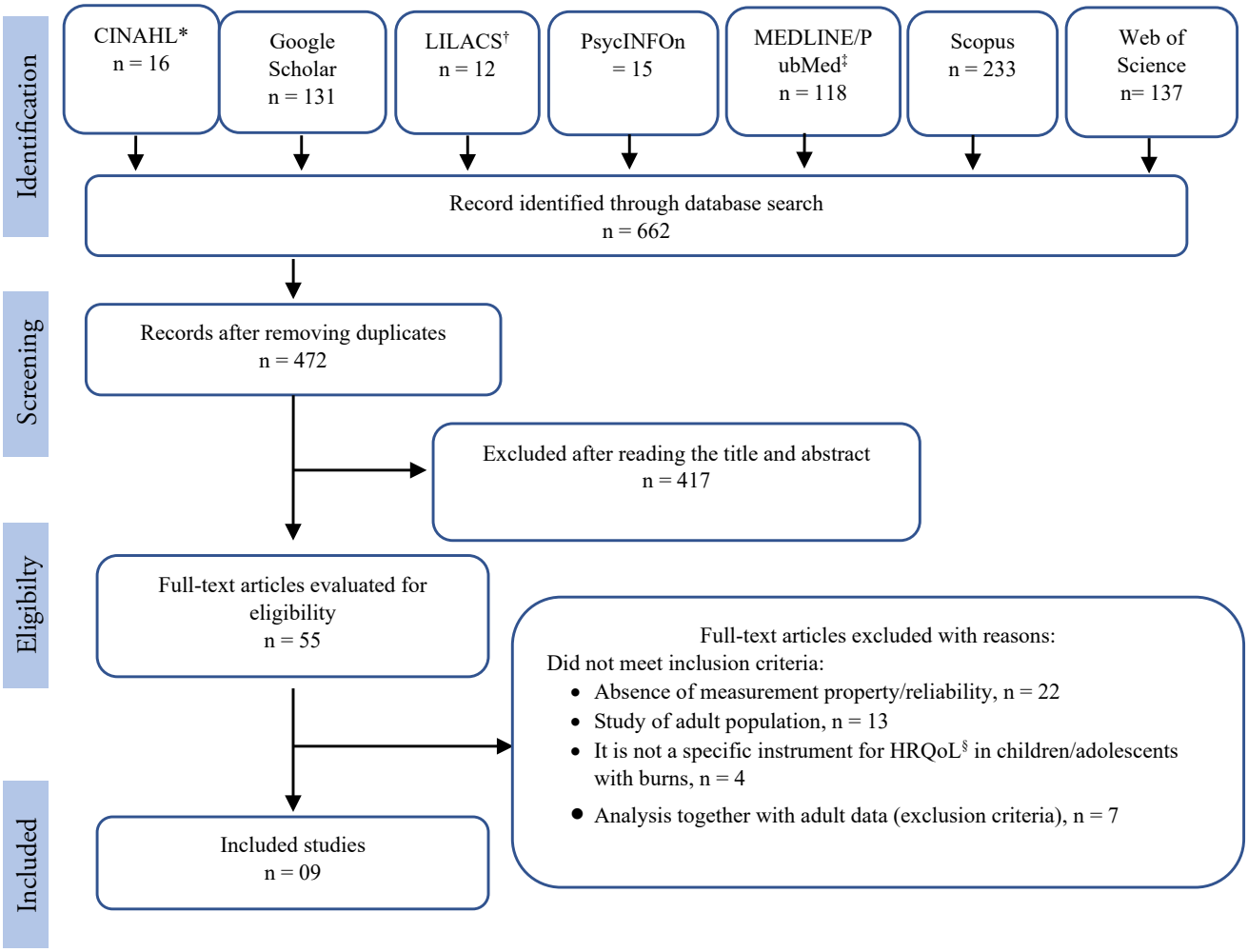
Characteristics of the instruments

Three instruments to assess HRQoL after burns in children and adolescents were found: Children Burn Outcomes Questionnaire (BOQ), in two versions (0-4 years[22,23] and 5–18 years[13,24-26]); Brisbane Impact Scar Profile (BBSIP), children and caregivers (0–8 years) [14] and young people (8–18 years)[27]; and CARe Burn Scale (CBS) for children aged 8 years or younger to be completed by parents/caregivers[28], and for young people aged 8–17 years (Table 1). No studies on the psychometric properties of the BBSIP version for children over 8 years of age and their caregivers, and of the CBS for young people aged 8-17 years were found.

The BBSIP was created in 2013 with the aim of developing a patient-reportable HRQoL measure to assess the impact of burn scars on children and adolescents. To date, only publications on its application in Australia have been found[29].

The BOQ was developed in the United States, in English, containing a version for children and adolescents aged 5-18 years. Its pilot test was carried out in 1998, and in 2000, it was applied in Sweden[24]; in 2002, in the United States[13]; in 2006, in Holland[25]; and in 2020, in India[26]. The other version — for children under 5 years

Figure 1 - Article selection flowchart according to PRISMA[19] recommendations, June 2023



Note: *Cumulative Index to Nursing and Allied Health Literature (CINAHL); †Latin American and Caribbean Health Sciences Literature (LILACS); ‡National Library of Medicine of the United States of America — National Institutes of Health (MEDLINE/PubMed); §Health-Related Quality of Life (HRQoL).

Table 1 - Description of the items and domains of the instruments that assess HRQoL after burns in children

Instrument/age group/author	Domains (quantity of items)	Total of items	Type of application form	Cost for access and use	Application time
BOQ 0–4 years Kazis et al.[23]	Play (5); Language (4); Fine motor skills (7); Gross motor skills (8); Emotional behavior (9); Family functioning (5); Pain/Itching (7); Appearance (3); Satisfaction with care (5); Concern (2).	55	Self-report (paper/pencil) for caregivers	Free	16 min
BOQ 5–18 years Daltroy et al.[13]	Upper extremity function (7); Physical function and sports (6); Transfers and mobility (5); Pain (2); Itch (2); Appearance (4); Compliance (5); Satisfaction with the current state (6); Emotional health (4); Family disruption (5); Parental concern (3); School reentry (3).	52	Self-report (paper/pencil) or interview for caregivers and adolescents (11 to 18 years old)	Free	30 min for caregivers and 33 min for adolescents
BBSIP 0–8 years Simons et al.[14]	Overall impact items (8); Sensory frequency (4); Mobility (4); Daily living (11); Friendships and social interactions (3); Appearance (4); Emotional reactions (7); Physical symptoms (7); Parental concern (3); Parental impact (5); Open question about the worst scar (1); Dichotomous question about infection (1).	58	Self-report (paper/pencil) for caregivers	Free	Not reported
BBSIP 8–18 years Simons et al.[27]	Overall impact of burn scars (5); Sensory frequency (5); Sensory intensity (4); Sensory impact (5); Mobility (4); Daily living (10); Friendships and social interactions (3); Appearance (4); Emotional reactions (8); Physical symptoms (8); Open question about the worst scar (1); Dichotomous question about infection (1).	58	Self-report (paper/pencil) for caregivers	Free	13 minutes
CBS 0–8 years Griffiths et al.[24]	Wound/scar treatments (6); Physical well-being (3); Wound/scar discomfort (2); Parental dissatisfaction with their child's wound/scar (6); Social and emotional difficulties (11); Social and emotional well-being (4).	32	Self-report (paper/pencil/online) for caregivers	Free	Not reported

Note: BOQ: Children Burn Outcomes Questionnaire; BBSIP: Brisbane Impact Scar Profile; CBS: CARe Burn Scale.

old — was pilot tested in 1996, applied in the United States in 2000[29] and in Holland in 2006[22]. The CBS was developed in the United Kingdom to assess the quality of life of children under 8 years of age who suffered burns and tested only in the United Kingdom[28].

Methodological quality of the test measurements of Patient Reported Outcome Measures (PROMs)

The three types of reliability assessment (internal consistency, test-retest and error of measurement) were not analyzed (Table 2) in most studies[13,22-26,28]. The Cronbach's alpha was used to assess internal consistency in all instruments. It reached values ≥ 0.70 in all domains only in the BBSIP/0-8 years and in the CBS version 0 to less than 8 years. In the two versions of the BOQ and in the BBSIP for children and adolescents aged 8-18 years, Cronbach's alpha was < 0.70 in at least one item of the subscale.

Although the Cronbach's alpha results were appropriate, in relation to this item, most instruments[13,22-26] were classified as having indeterminate properties given the lack of evidence of positive unidimensionality or structural validity.

To measure test-retest reliability, the Intraclass Correlation Coefficient (ICC), which ranged from 0.36 to 1.00, was calculated, as well as the Pearson's r, ranging from 0.12 to 0.92. Only the study carried out in India with the BOQ/5-18 years presented an ICC value > 0.70 for all subscales. Note that the variation in the test-retest interval between studies was seven to 30 days and the number of participants was 18 to 93.

The error of measurement was assessed only for the BBSIP[14,27] instrument through the Standard Error of Measurement (SEm) and the Smallest Detectable Change (SDC), for which normality assumptions were met. However, the Clinically Significant Change (CSC)

Table 2 - Study characteristics (n = 9), methodological quality and assessment of psychometric properties, June 2023

Authorship and year of publication	Instrument and age range of the target population	Location	Reliability								
			Internal consistency			Test-retest reliability			Error of measurement		
			n	QM*	PM†	n	QM*	PM†	n	QM*	PM†
Kazis et al.[23]	BOQ 0-4 years	USA	184	VG	+	46	A	-		NE	
van Baar et al.[22]	BOQ 0-4 years	Holland	194	VG	?	93	A	+		NE	
Arumugam and Thayal[26]	BOQ 5-18 years	India	25	VG	?	25	A	+		NE	
Daltroy et al.[13]	BOQ 5-18 years	USA	86	VG	?	42	D	?		NE	
Sveen et al.[24]	BOQ 5-18 years	Sweden	70	VG	?	18	A	-		NE	
van Baar et al.[25]	BOQ 5-18 years	Holland	52	VG	?	24	A	-		NE	
Simons et al.[14]	BBSIP 0-8 years	Australia	86	VG	+	58	VG	-	58	VG	
Simons et al.[27]	BBSIP 8-18 years	Australia	65	VG	-	49	VG	-	49	VG	?
Griffiths et al.[28]	CBS 0-8 years	United Kingdom	133	VG	+		NA			NE	

Note: *MQ: Methodological quality used to assess the risk of bias: very good (VG), appropriate (A), doubtful (D), inappropriate (I) or not evaluated (NE); †MP: Measurement property applied to each PROM: sufficient (+), insufficient (-), inconsistent (±) or indeterminate; n: number of participants; USA: United State of American. BOQ: *Children Burn Outcomes Questionnaire*; BBSIP: *Brisbane Impact Scar Profile*; CBS: *CARe Burn Scale*.

value was not presented and was therefore classified as "indeterminate (?)". Based on the results of each study, it is possible to summarize the measurement properties of the three HRQoL instruments evaluated, and none of the tools presented a positive result for the three evaluated measures (Table 3). Additional information is available in Supplementary Table 2, Appendix 1.

The BBSIP/0-8 years[14] and the CBS[28] were the only instruments that presented a general classification sufficient for internal consistency. However, the level of recommendation of the BBSIP[14,27] decreased because of the number of participants was less than 100. Regarding the test-retest, no instrument had a sufficient overall rating, and the level of recommendation varied from moderate to low (Table 3).

Risk of bias

All studies showed very good methodological quality in the analysis of internal consistency (Table 2) according to the COSMIN risk of bias (additional information in Supplementary Table 3, Appendix 1). For test-retest reliability, the studies that evaluated the BBSIP[14,27] were classified as very good. The evaluation of the BOQ 0-4 years[22,23] was considered appropriate; the main weaknesses were the lack of explicit description of the patient's stability and the maintenance of the same application conditions in the different collection periods (additional information in Supplementary Table 4, Appendix 1). The BOQ 5-18 years[13,24-26] presented some divergences in assessing the risk of bias mainly by the lack of explicit information.

Divergences were observed in the results of studies that used the BOQ instrument. One study that evaluated the BOQ 0-4 years[23] proved to be appropriate, while the other was doubtful[22]. The evaluated version of the BOQ 5-18 years instrument proved to be appropriate in only one study[26] and inappropriate in the others[13,24,25]. The error of measurement was evaluated only for the BBSIP instrument (both versions) [14,27] and scored as very good according to COSMIN (additional information in Supplementary Table 5, Appendix 1).

Quality of evidence

The quality of evidence, according to the GRADE method[21], varied from low to high (Table 3). The main dimension responsible for the reduction in the quality of the evidence was imprecision, attributed to instru-

Table 3 - Overall assessment and quality of evidence for measurement property of Patient Reported Outcome Measures, June 2023

Instrument/ age group	Internal consistency		Test-Retest Reliability		Error of measurement	
	Overall evaluation	Quality of evidence	Overall evaluation	Quality of evidence	Overall evaluation	Quality of evidence
BOQ 0–4 years	?	Moderate	±	Moderate	NE	NE
BOQ 5–18 years	?	High	±	Moderate	NE	NE
BBSIP 0–8 years	+	Moderate	-	Moderate	?	Moderate
BBSIP 8–18 years	-	Moderate	-	Low	?	Low
CBS 0–8 years	+	High	NE	NE	NE	NE

Note: BOQ: Children Burn Outcomes Questionnaire; BBSIP: Brisbane Impact Scar Profile; CBS: CARe Burn Scale; PROMs: Patient Reported Outcome Measures. The overall result was evaluated according to the COSMIN (Consensus-based Standards for the selection of health Measurement INstruments) criteria for good measurement properties and summarized as: sufficient (+), insufficient (-), inconsistent (±) or indeterminate (?). The quality of the evidence was classified as high, moderate, low or very low according to GRADE (Grading of Recommendations Assessment Development, and Evaluation). NE: not evaluated.

ments with a sample size of less than 100 participants, BBSIP[14,27]. Another dimension that contributed to lowering the quality of evidence was the presence of inconsistencies (without explanation for them) between some results of studies involving the BOQ (both versions) (additional information in Supplementary Table 6, Appendix 1).

DISCUSSION

This systematic review included nine articles with three PROMs in different versions: BOQ 0-4 years[22,23]; BOQ 5-18 years[13,24-26]; BBSIP 0-8 years[14]; BBSIP 8-18 years[27]; and CBS[28]. The BBSIP 0-8 years[14] was the instrument with the best measurements and quality of evidence classified as moderate in the three assessments (reliability, internal consistency and error of measurement). It did not obtain high quality only because the number of participants was less than 100.

Although the BOQ 5-18 years[13,24-26] and the CBS 0-8 years[28] obtained a high-quality evidence rating for internal consistency, the studies did not evaluate error of measurement, which is an essential measure for the reliability of an instrument. No instrument was classified as highly recommended according to the COSMIN[21]. The main contribution of this review was to identify the instruments that achieved greater methodological rigor with good classification as instruments for application in clinical practice and recommendation for future cross-cultural adaptation studies.

General interpretation

The heterogeneity in the domains of each instrument make it difficult to compare the adequacy of different care scenarios for children or adolescents with burns[27,29]. The high number of questions of the instruments is another point that draws attention[13,14,22-28], as this can make it difficult to insert these instruments into clinical practice.

As shown, the three types of reliability (internal consistency, test-retest and error of measurement) were not evaluated in most studies, corroborating a situation already documented in the literature, in which the measurement properties of several PROMs are not adequately studied for different instruments[23].

The analysis of internal consistency was performed in all studies evaluated in this review. Internal consistency concerns the degree of cohesion or homogeneity between questions of the same dimension[30]. In the evaluation of this property, the Cronbach's alpha coefficient can be used[30,31] with data collected by applying the instrument.

Regarding the evaluation of measurement properties, score variations were found in the studies, depending on the instrument evaluated. The BOQ instrument (both versions)[13,22-26] achieved a low score given the lack of minimal evidence of positive structural validity, which refers to the degree to which PROM scores are appropriate[20]. The reflection of the dimensionality of the construct to be measured is generally evaluated by factor analysis[20]. The BBSIP 8-18 instrument[27]

obtained a low score because its results did not meet the criteria pre-established by COSMIN[21] in this aspect. This shows the importance of analyzing structural validity. Even if this analysis is not part of internal consistency (but of another independent property), performing it can better support the definition of the dimensions of HRQoL instruments, thereby more appropriately reflecting the analysis of internal consistency and the results of Cronbach's alpha coefficient[21]. Furthermore, when performing the structural validity analysis, the need and possibility of reducing the number of questions can be assessed with the aim to reduce the application time.

The test-retest reliability refers to the assessment of the stability of responses to an instrument over time[20]. Its application is justified when the measurement object remains stable over periods of collection. Changes in clinical or contextual conditions (discharge, school reentry or readmission) can artificially reduce the correlation between the results of the two collections[32,33]. The information on stability of the clinical and social context between the test and retest stages was explained in only some of the articles evaluated[14,27] in the present study. Additionally, the variation in time between collections (seven-30 days) and the smaller number of individuals in the samples in the retest seems to justify at least in part that no instrument obtained a positive (+) evaluation score on this property. In view of this, it is important that future studies on the topic explain the context of patients in both stages, exclude patients who have presented changes in context and do not use long time intervals between collections.

In turn, the error of measurement concerns the difference between the values obtained and the real values, which may be different because of errors in the instrument or development of the study[20]. By analyzing this property, it is expected that the error of measurement of the instrument is less than a minimum value considered important to cause changes in the instrument's interpretation according to the CSC. However, the CSC cannot be determined based on just one study, but it constitutes the result of different longitudinal studies[21]. Since there are few studies on the subject, it is not possible to calculate the CSC for the instruments under analysis, which reduces the final result of the evaluative score of this measurement property for the two versions of the BBSIP[14,27]. This fact was also observed in another study in which was reached the conclusion that the properties of the measures evaluated are not always inappropriate, but simply too small or not investigated[18].

Risk of bias and quality of evidence

In general, the methodological quality (risk of bias) of the studies was very good or appropriate. For internal consistency and error of measurement, the COSMIN criteria[21] were met. For the test-retest, the main lack of information was related to stability in the collection period and measurement conditions. Despite these limitations, the presence of two studies for the BOQ 0-4[22,23] and four for the BOQ 5-18[13,24-26] contributed to a positive assessment of the risk of bias according to the GRADE criteria[20].

The quality of the evidence was particularly affected by the inaccuracy of the two versions of the BBSIP[14,27] and by inconsistencies identified in the two versions of the BOQ[13,22-26]. Thus, the scarcity of studies, methodological differences and small sample size were the main factors compromising the analysis parameters. Differences in parameter analysis do not necessarily represent inadequacy of the instruments, but may reflect the need for greater standardization.

Despite the limitations already discussed, the BBSIP[14,27] was the instrument that complied with the largest number of COSMIN items and stood out among the other instruments, since all its questions asked refer to the burn scar, while other instruments may have the dimension of emotional reactions influenced by other factors such as trauma and not by the scar itself[16]. Furthermore, the BBSIP 8-18 version[27] theoretically appears to group samples of more similar ages than the BOQ 5-18[13,24-26] that may have very different patients in its analysis group because of its age group.

The validation process of an instrument is not a proven action, but supported by the accumulation of evidence, therefore, the adequacy of the BOQ[13,22-26] and the CBS[28] versions cannot be discarded.

The findings presented here provide clear indications for the development of new studies, given the impossibility to confidently confirm the existence of an instrument with adequate reliability for clinical practice. Regardless of the instrument adopted, it should be used together with an assessment of test-retest reliability and error of measurement. Furthermore, the relevance of new studies on the topic with greater methodological rigor for evaluating the content and its measurement properties stands out.

In order to overcome the limitations of the studies included in the analysis, new studies to develop HRQoL assessment instruments should investigate the three types of reliability measures (internal consistency, test-retest, and error of measurement) and be designed to obtain a sample of at least 100 participants. A better

description of the steps and stability between test-retest conditions is also recommended.

Although efforts were made towards a sensitive search strategy to identify all articles on the topic available in the databases investigated, the gray literature was not searched, which may be a limitation of this study.

FINAL CONSIDERATIONS

Nine studies that analyzed three instruments in different versions used to assess the HRQoL of children with burn scars were identified in this review: the BOQ 0-4[22,23]; BOQ 5-18[13,24-26]; BBSIP 0-8[14]; BBSIP 8-18[27]; and the CBS 0-8[28]. The BBSIP[14,27] was considered more complete when evaluating the three reliability measures: internal consistency, test-retest and error of measurement. The BBSIP[14,27] was the only instrument evaluated for the three reliability estimates studied, which, despite limitations, presented appropriate results. The BOQ[13,22-26] presented some limitations and inconsistencies between different studies, in addition to the need to evaluate the error of measurement. The CBS[28] presented good internal consistency results, nevertheless the other properties could not be evaluated. Based on these results, the BBSIP[14,27] stands out for presenting measures for all reliability estimates analyzed.

Funding

This study did not receive financial support.

Conflict of interests

None.

Acknowledgments

The authors would like to acknowledge Arthur Eumann Mesas for his assistance and support. Even indirectly, his guidance and teachings were essential for improving this article.

Author's contributions - CRediT

FFAB: conceptualization; data curation; formal analysis; investigation; methodology; project administration; resources; validation; visualization; writing – original draft and writing – review & editing.

EFZ: conceptualization; data curation; formal analysis; investigation; methodology; project administration; resources; validation; visualization; writing – original draft and writing – review & editing.

RR: data curation; formal analysis; methodology; validation; visualization; writing – original draft and writing - review & editing.

FXA: data curation; formal analysis; methodology; validation; visualization; writing – original draft and writing – review & editing.

MEEG: validation; visualization; writing – review & editing.

FMP: validation; visualization; writing –- review & editing.

RAP: conceptualization; data curation; formal analysis; investigation; methodology; project administration; resources; supervision; validation; visualization; writing – original draft and writing – review & editing.

REFERENCES

1. Pinto ACS, Costa KLN, Almeida Filho PP, Oliveira Júnior JL, Rocha MNS. Evaluation of the epidemiological profile of burned adult patients admitted to a referral center in the interior of the state of Bahia, Brazil. Rev Bras Cir Plást. 2022 Jan-Mar;37(1):66-70. https://doi.org/10.5935/2177-1235.2022RBCP0011

2. Institute for Health Metrics and Evaluation. GBD Results [Internet]. Seattle: University of Washington; 2020 [cited 2022 July 16]. Available from: https://vizhub.healthdata.org/gbd-results/

3. Willemse H, Geenen R, Egberts MR, Engelhard IM, Van Loey NE. Perceived stigmatization and fear of negative evaluation: two distinct pathways to body image dissatisfaction and self-esteem in burn survivors. Psychol Health. 2021 Apr;38(4):445-58. https://doi.org/10.1080/08870446.2021.1970160

4. Ajoudani F, Jasemi M, Lotfi M. Social participation, social support, and body image in the first year of rehabilitation in burn survivors: a longitudinal, three-wave cross-lagged panel analysis using structural equation modeling. Burns. 2018 Aug; 44(5):1141-50. https://doi.org/10.1016/j.burns.2018.03.018

5. ter Bogt HJ, Scapens RW. Institutions, situated rationality and agency in management accounting: A research note extending the Burns and Scapens framework. Accounting, Auditing & Accountability Journal. 2019 July;32(6):1801-25. https://doi.org/10.1108/AAAJ-05-2016-2578

6. Ruidiaz-Gómez KS, Cacante-Caballero JV. Desarrollo histórico del concepto calidad de vida: una revisión de la literatura. Rev Cienc Cuidad. 2021 Sept;18(3):86-99. https://doi.org/10.22463/17949831.2539

7. Spronk I, Legemate C, Oen I, van Loey N, Polinder S, van Baar M. Health related quality of life in adults after burn injuries: a systematic review. PLoS One. 2018 May;13(5):e0197507. https://doi.org/10.1371/journal.pone.0197507

8. Woolard A, Hill NTM, McQueen M, Martin L, Milroy H, Wood FM, et al. The psychological impact of paediatric burn injuries: a systematic review. BMC Public Health. 2021 Dec;21:2281. https://doi.org/10.1186/s12889-021-12296-1
9. Lerman SF, Sylvester S, Hultman CS, Caffrey JA. Suicidality after burn injuries: a systematic review. J Burn Care Res. 2021 Jan;42(3):357-64. https://doi.org/10.1093/jbcr/irab014
10. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res. 2010 Feb;19:539-49 https://doi.org/10.1007/s11136-010-9606-8
11. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res. 2018 May;27:1147-57. https://doi.org/10.1007/s11136-018-1798-3
12. Legemate CM, Spronk I, Mokkink LB, Middelkoop E, Polinder S, van Baar ME, et al. Evaluation of measurement properties of health-related quality of life instruments for burns: a systematic review. J Trauma Acute Care Surg. 2020 Apr;88(4):555-71. https://doi.org/10.1097/ta.0000000000002584
13. Daltroy LH, Liang MH, Phillips CB, Daugherty MB, Hinson M, Jenkins M, et al. American Burn Association/Shriners Hospitals for Children burn outcomes questionnaire: construction and psychometric properties. J Burn Care Rehabil. 2000 Jan-Feb;21(1 Pt 1):29-39. https://doi.org/10.1097/00004630-200021010-00007
14. Simons M, Kimble R, McPhail S, Tyack Z. The longitudinal validity, reproducibility and responsiveness of the Brisbane Burn Scar Impact Profile (caregiver report for young children version) for measuring health-related quality of life in children with burn scars. Burns. 2019 Dec; 45(8):1792-809. https://doi.org/10.1016/j.burns.2019.04.015
15. Mokkink LB, Terwee CB, Stratford PW, Alonso J, Patrick DL, Riphagen I, et al. Evaluation of the methodological quality of systematic reviews of health status measurement instruments. Qual Life Res. 2009 Apr;18(3):313-33. https://doi.org/10.1007/s11136-009-9451-9
16. Meirte J, Tyack Z. Electronic Patient-Reported Outcome Measures in Burn Scar Rehabilitation: A Guide to Implementation and Evaluation. Eur. Burn. J. 2022 Apr 5;3(2):290–308. https://doi.org/10.3390/ebj3020025
17. Flake JK, Davidson IJ, Wong O, Pek J. Construct validity and the validity of replication studies: a systematic review. Am Psychol. 2022 May-Jun;77(4):576-88. https://doi.org/10.1037/amp0001006
18. Lüthi FT, Mabire C, Amoussou JR, Bernard M, Borasio GD, Ramelet AS. Instruments for the identification of patients in need of palliative care: protocol for a systematic review of measurement properties. JBI Evid Synth. 2020 May;18(5):1144-53. https://doi.org/10.11124/JBISRIR-D-19-00146
19. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 2009 July 21;6(7):e1000097. https://doi.org/10.1371/journal.pmed.1000097
20. Mokkink LB, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, Vet HCW, et al. COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs) [Internet]. 2018 [cited 2022 July 16]. Available from: https://cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018.pdf
21. Mokkink LB, Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. Qual Life Res. 2017 May;27(5):1171-9. https://doi.org/10.1007/s11136-017-1765-4
22. van Baar ME, Essink-Bot ML, Oen IMMH, Dokter J, Boxma H, Hinson MI, et al. Reliability and validity of the Health Outcomes Burn Questionnaire for infants and children in The Netherlands. Burns. 2006 May;32(3):357-65. https://doi.org/10.1016/j.burns.2005.10.004
23. Kazis LE, Liang MH, Lee A, Ren XS, Phillips CB, Hinson M, et al. The development, validation, and testing of a health outcomes burn questionnaire for infants and children 5 years of age and younger: American Burn Association/Shriners Hospitals for Children. J Burn Care Rehabil. 2002 May-Jun;23(3):196-207. https://doi.org/10.1097/00004630-200205000-00009
24. Sveen J, Huss F, Sjöberg F, Willebrand M. Psychometric properties of the Swedish version of the burn outcomes questionnaire for children aged 5 to 18 years. J Burn Care Res. 2012 Nov-Dec;33(6):e286-94. https://doi.org/10.1097/BCR.0b013e3182331aaa
25. van Baar ME, Essink-Bot ML, Oen IMMH, Dokter J, Boxma H, Hinson MI, et al. Reliability and validity of the Dutch version of the American Burn Association/Shriners Hospital for Children Burn Outcomes Questionnaire (5-18 years of age). J Burn Care Res. 2006 Nov-Dec;27(6):790-802. https://doi.org/10.1097/01.bcr.0000245434.76697.56
26. Arumugam PK, Thayal PK. Validation of Indian Adaptation of Burn Outcomes Questionnaire—Hindi Version (I-BOQ-HV) for Pediatric Subgroup 5 to 18 years. J Burn Care Res. 2021 May;42(3):473-80. https://doi.org/10.1093/jbcr/iraa182
27. Simons M, Kimble R, McPhail S, Tyack Z. The Brisbane Burn Scar Impact Profile (child and young person version) for measuring health-related quality of life in children with burn scars: a longitudinal cohort study of reliability, validity and responsiveness. Burns. 2019 Nov;45(7):1537-52. https://doi.org/10.1016/j.burns.2019.07.012

28. Griffiths C, Guest E, Pickles T, Hollèn L, Grzeda M, Tollow P, et al. The development and validation of the CARe Burn Scale: Child Form: a parent-proxy-reported outcome measure assessing quality of life for children aged 8 years and under living with a burn injury. Qual Life Res. 2021 Jan;30(1):239-50. https://doi.org/10.1007/s11136-020-02627-x

29. Tyack Z, Ziviani J, Kimble R, Plaza A, Jones A, Cuttle L, et al. Measuring the impact of burn scarring on health-related quality of life: development and preliminary content validation of the Brisbane Burn Scar Impact Profile (BBSIP) for children and adults. Burns. 2015 Nov;41(7):1405-19. https://doi.org/10.1016/j.burns.2015.05.021

30. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika. 1951 Sept;16(3):297-334. https://doi.org/10.1007/BF02310555

31. Javali SB, Gudaganavar NV, Jain SM. Effect of varying sample size in estimation of coefficients of internal consistency. WebmedCentral. 2011 Feb 23;2(2):WMC001572. https://doi.org/10.9754/journal.wmc.2011.001572

32. Guttman L. A basis for analyzing test-retest reliability. Psychometrika. 1945;10(4):255-82. https://doi.org/10.1007/bf02288892

33. Chenani KT, Madadizadeh F. Guideline for selecting types of reliability and suitable intra-class correlation coefficients in clinical research. J Biostat Epidemiol. 2021 Oct;7(3):305-09. https://doi.org/10.18502/jbe.v7i3.7301

APPENDIX 1 - Supplementary Material

Supplementary Table 1 - Search strategy to identify specific instruments for the HRQOL of children and adolescents with burn scars in which reliability analyzes were performed

PUBMED
(((((quality of life[MeSH Terms]) OR HRQoL) AND (child[MeSH Terms] OR children OR adolescent)) AND (burns[MeSH Terms])) AND (questionnaires[MeSH Terms]))(instrumentation[sh] OR methods[sh] OR "Validation Studies"[pt] OR "Comparative Study"[pt] OR "psychometrics"[MeSH] OR psychometr*[tiab] OR clinimetr*[tw] OR clinometr*[tw] OR "outcome assessment (health care)"[MeSH] OR "outcome assessment"[tiab] OR "outcome measure*"[tw] OR "observer variation"[MeSH] OR "observer variation"[tiab] OR "Health Status Indicators"[Mesh] OR "reproducibility of results"[MeSH] OR reproducib*[tiab] OR "discriminant analysis"[MeSH] OR reliab*[tiab] OR unreliab*[tiab] OR valid*[tiab] OR "coefficient of variation"[tiab] OR coefficient[tiab] OR homogeneity[tiab] OR homogeneous[tiab] OR "internal consistency"[tiab] OR (cronbach*[tiab] AND (alpha[tiab] OR alphas[tiab])) OR (item[tiab] AND (correlation*[tiab] OR selection*[tiab] OR reduction*[tiab])) OR agreement[tw] OR precision[tw] OR imprecision[tw] OR "precise values"[tw] OR test-retest[tiab] OR (test[tiab] AND retest[tiab]) OR (reliab*[tiab] AND (test[tiab] OR retest[tiab])) OR stability[tiab] OR interrater[tiab] OR inter-rater[tiab] OR intrarater[tiab] OR intra-rater[tiab] OR intertester[tiab] OR inter-tester[tiab] OR intratester[tiab] OR intra-tester[tiab] OR interobserver[tiab] OR inter-observer[tiab] OR intraobserver[tiab] OR intra-observer[tiab] OR intertechnician[tiab] OR inter-technician[tiab] OR intratechnician[tiab] OR intra-technician[tiab] OR interexaminer[tiab] OR inter-examiner[tiab] OR intraexaminer[tiab] OR intra-examiner[tiab] OR interassay[tiab] OR inter-assay[tiab] OR intraassay[tiab] OR intra-assay[tiab] OR interindividual[tiab] OR inter-individual[tiab] OR intraindividual[tiab] OR intra-individual[tiab] OR interparticipant[tiab] OR inter-participant[tiab] OR intraparticipant[tiab] OR intra-participant[tiab] OR kappa[tiab] OR kappa's[tiab] OR kappas[tiab] OR repeatab*[tw] OR ((replicab*[tw] OR repeated[tw]) AND (measure[tw] OR measures[tw] OR findings[tw] OR result[tw] OR results[tw] OR test[tw] OR tests[tw])) OR generaliza*[tiab] OR generalisa*[tiab] OR concordance[tiab] OR (intraclass[tiab] AND correlation*[tiab]) OR discriminative[tiab] OR "known group"[tiab] OR "factor analysis"[tiab] OR "factor analyses"[tiab] OR "factor structure"[tiab] OR "factor structures"[tiab] OR dimension*[tiab] OR subscale*[tiab] OR (multitrait[tiab] AND scaling[tiab] AND (analysis[tiab] OR analyses[tiab])) OR "item discriminant"[tiab] OR "interscale correlation*"[tiab] OR error[tiab] OR errors[tiab] OR "individual variability"[tiab] OR "interval variability"[tiab] OR "rate variability"[tiab] OR (variability[tiab] AND (analysis[tiab] OR values[tiab])) OR (uncertainty[tiab] AND (measurement[tiab] OR measuring[tiab])) OR "standard error of measurement"[tiab] OR sensitiv*[tiab] OR responsive*[tiab] OR (limit[tiab] AND detection[tiab]) OR "minimal detectable concentration"[tiab] OR interpretab*[tiab] OR ((minimal[tiab] OR minimally[tiab] OR clinical[tiab] OR clinically[tiab]) AND (important[tiab] OR significant[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR (small*[tiab] AND (real[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR "meaningful change"[tiab] OR "ceiling effect"[tiab] OR "floor effect"[tiab] OR "Item response model"[tiab] OR IRT[tiab] OR Rasch[tiab] OR "Differential item functioning"[tiab] OR DIF[tiab] OR "computer adaptive testing"[tiab] OR "item bank"[tiab] OR "cross-cultural equivalence"[tiab])
SCOPUS
("child" OR "adolescent") AND "burn" AND ("quality of life" OR "HRQoL") AND ("surveys and questionnaires" OR "reliability assessment" OR "internal consistency" OR "test-retest")
CINAHL
(child OR adolescent) AND burns AND quality of life AND (surveys and questionnaire)
WEB OF SCIENCE
(("child" OR "adolescent") AND "burn" AND "quality of life" AND ("questionnaire"))
LILACS
criança$ OR child OR adolescent) AND (queimadura$ OR burn) AND (qualidade de vida OR quality of life OR HRQoL
GOOGLE SCHOLAR
Five different search strategies were used in this database: • allintitle: "Children Burn Outcomes Questionnaire" • allintitle: "Adolescent Burn Outcomes Questionnaire" • allintitle: Adolescent Burn quality of life • allintitle: "brisbane Burn Scar" • children or adolescent reliability OR assessment OR internal OR consistency OR "test retest" "quality of life in burn"
PSYCOINFO
(("child" OR "adolescent") AND "burn" AND "quality of life" AND ("questionnaire"))

Note: HRQoL = Health-Related Quality of Life

Supplementary Table 2 - Measurement properties, assessments using the COSMIN risk of bias checklist for internal consistency, test-retest reliability and measurement error of instruments that assess the HRQoL of children and adolescents with burn scars

COSMIN criteria				Instruments/age group				
Items	**Positive rating (+)**	**Indeterminate assessment (?)**	**Negative rating (-)**	**BOQ 0-4 years**	**BOQ 5 -18 years**	**BBSIP 0-8 years**	**BBSIP 8-18 years**	**CBS 0-8 years**
Internal consistency	At least limited evidence of unidimensionality or positive structural validity AND Cronbach's alpha ≥ 0.70 and ≤ 0.95	Not all information for '+' was reported OR conflicting evidence of unidimensionality or structural validity OR evidence of lack of unidimensionality or negative structural validity	Criteria for '+' not met	?	?	+	-	+
Test-retest reliability	ICC or weighted Kappa≥0.70	ICC or weighted Kappa not reported	Criteria for '+' not met	±	±	-	-	NE
Error of measurement	SDC or LoA < MIC	MIC not defined	Criteria for '+' not met	NE	NE	?	?	NE

Note: COSMIN = COnsensus-based Standards for the selection of health Measurement INstruments; HRQoL = Health-Related Quality of Life; BOQ = Children Burn Outcomes Questionnaire; BBSIP = Brisbane Impact Scar Profile; CBS = CARe Burn Scale; ICC= Intraclass Correlation Coefficient; LoA = Limits of Agreement; MIC= Minimum Important Change; NE = not evaluated; SDC = Smallest Detectable Change.

Supplementary Table 3 - Methodological quality evaluated using the COSMIN risk of bias checklist for internal consistency of studies with instruments that assess the HRQoL of children or adolescents with burn scars

	COSMIN criteria					Included studies (reference)								
Items	**Very good (VG)**	**Appropriate (A)**	**Doubtful (D)**	**Inappropriate (I)**	**NE**	**23**	**22**	**26**	**13**	**24**	**25**	**14**	**27**	**28**
Design requirements														
1. Was an internal consistency statistic calculated for each unidimensional scale or subscale separately?	Internal consistency statistics calculated for each unidimensional scale or subscale		It is unclear whether the scale or subscale is unidimensional	Internal consistency statistics NOT calculated for each unidimensional scale or subscale		VG	VG	VG	VG	VG	VG	VG	VG	VG
Statistical methods														
2 For continuous scores: was Cronbach's alpha or omega calculated?	Cronbach's alpha or omega was calculated		Item only - total correlations calculated	No Cronbach's alpha and no item - total correlations calculated	Not applicable	VG	VG	VG	VG	VG	VG	VG	VG	VG
3 For dichotomous scores: was Cronbach's alpha or KR - 20 calculated?	Cronbach's Alpha or KR - 20 calculated		Item only - total correlations calculated	No Cronbach's alpha or KR - 20 and no item - total correlations calculated	Not applicable	NE	NE	NE	NE	NE	NE	NE	NE	NE
4 For Item Response Theory-based scores: Has the standard error of theta (SE (θ)) or the reliability coefficient of the estimated latent trait value (separation index (subject or item)) been calculated?	SE(θ) or reliability coefficient was calculated			SE(θ) or reliability coefficient was NOT calculated	Not applicable	NE	NE	NE	NE	NE	NE	NE	NE	NE
Other														
5 Were there other major flaws in the study design or statistical methods?	No other major methodological flaws		Other minor methodological flaws	Other important methodological flaws		VG	VG	VG	VG	VG	VG	VG	VG	VG
Methodological quality						**VG**	**VG**	**VG**	**VG**	**VG**	**VG**	**VG**	**VG**	**VG**

Note: COSMIN = COnsensus-based Standards for the selection of health Measurement Instruments, HRQoL = Health-Related Quality of Life; Studies included: 23 = KAZIS et al. (2002); 22 = VAN BAAR et al. (2006); 26= ARUMUGAM, THAINAL (2020); 13 = DALTROY et al. (2000); 24 = SVEEN et al. (2012); 25 = VAN BAAR et al. (2006); 14 = SIMONS et al. (2019); 27= SIMONS et al. (2019); 28 = GRIFFITHS et al. (2020); NE = not evaluated; KR-20= Kuder-Richardson Formula 20.

Supplementary Table 4 - Methodological quality evaluated using the COSMIN risk of bias checklist for test-retest of studies with instruments that assess the HRQoL of children or adolescents with burn scars

	COSMIN criteria					Evaluation of studies included (reference)								
Items	**Very Good (VG)**	**Appropriate (A)**	**Doubtful (D)**	**Inappropriate (I)**	**NA**	**23**	**22**	**26**	**13**	**24**	**25**	**14**	**27**	**28**
Design requirements														
1. Were the patients stable in the intermediate period in the construct to be measured?	Evidence provided that patients were stable	It was assumed that the patients were stable	It is unclear if the patients were stable	Patients were NOT stable		A	A	VG	A	VG	A	VG	VG	NE
2. Was the time interval appropriate	Appropriate time interval		Doubtful if the time interval was appropriate or the time interval was not stated	Time interval NOT appropriate		VG	VG	VG	VG	VG	VG	VG	VG	NE
3. Were the test conditions similar for the measurements? e.g. type of administration, environment, instructions	Testing conditions were similar (evidence provided)	It is assumed that the testing conditions were similar	It is unclear if the testing conditions were similar	Testing conditions were NOT similar		A	A	A	A	A	A	VG	VG	NE
Statistical methods														
4. For continuous scores: An Intraclass Correlation Coefficient (ICC) was calculated	ICC calculated and ICC model or formula is described	ICC calculated, but ICC model or formula not described or not ideal. Pearson or Spearman correlation coefficient calculated with evidence provided that no systematic change has occurred	Pearson or Spearman correlation coefficient calculated WITHOUT evidence provided that no systematic change has occurred or WITH evidence that systematic change has occurred	No ICC or Pearson or Spearman correlation calculated	Not applicable	VG	A	A	D	A	A	VG	VG	NE
5. For dichotomous/ nominal/ ordinal scores: Was kappa calculated?	Kappa calculated			No Kappa calculated	Not applicable	NE	NE	NE	NE	NE	NE	NE	NE	NE
6. For ordinal scores: was a weighted kappa calculated?	Weighted kappa calculated		Unweighted kappa calculated or not described		Not applicable	NE	NE	NE	NE	NE	NE	NE	NE	NE

Continue...

Supplementary Table 4 - Continuation

	COSMIN criteria					Evaluation of studies included (reference)								
Items	**Very Good (VG)**	**Appropriate (A)**	**Doubtful (D)**	**Inappropriate (I)**	**NA**	**23**	**22**	**26**	**13**	**24**	**25**	**14**	**27**	**28**
7. For ordinal scores: Has the weighting scheme been described? e.g. linear, quadratic	Weighting scheme described	Weighting scheme NOT described			Not applicable	NE	NE	NE	NE	NE	NE	NE	NE	NE
Other														
8 Were there any other major flaws in the study design or statistical methods?	No other major methodological flaw		Other minor methodological flaws	Other important methodological flaws		VG	VG	VG	VG	VG	VG	VG	VG	NE
Methodological quality						**A**	**A**	**A**	**D**	**A**	**A**	**VG**	**VG**	**NE**

Note: COSMIN = COnsensus-based Standards for the selection of health Measurement Instruments, HRQoL = Health-Related Quality of Life; Studies included: 23 = KAZIS et al. (2002); 22 = VAN BAAR et al. (2006); 26 = ARUMUGAM, THAINAL (2020); 13 = DALTROY et al. (2000); 24 = SVEEN et al. (2012); 25 = VAN BAAR et al. (2006); 14 = SIMONS et al. (2019); 27= SIMONS et al. (2019); 28 = GRIFFITHS et al. (2020); NE= not evaluated.

Supplementary Table 5 - Methodological quality assessed using the COSMIN risk of bias checklist for error of measurement of studies with instruments that assess the HRQoL of children or adolescents with burn scars

	COSMIN criteria					Studies included (reference)								
Items	**Very Good (VG)**	**Appropriate (A)**	**Doubtful (D)**	**Inappropriate (I)**	**NE**	**29**	**22**	**28**	**13**	**26**	**27**	**14**	**23**	**24**
Design requirements														
1 Were the patients stable in the intermediate period in the construct to be measured?	Patients were stable (evidence provided)	It was assumed that the patients were stable	It is unclear if the patients were stable	Patients were NOT stable		NE	NE	NE	NE	NE	NE	VG	VG	NE
2 Was the time interval adequate?	Appropriate time interval		Doubtful if the time interval was appropriate or the time interval was not stated	Time interval NOT appropriate		NE	NE	NE	NE	NE	NE	VG	VG	NE
3 Were the test conditions similar for the measurements? (e.g. type of administration, environment, instructions)	Testing conditions were similar (evidence provided)	It is assumed that the testing conditions were similar	It is unclear if the testing conditions were similar	Testing conditions were NOT similar		NE	NE	NE	NE	NE	NE	VG	VG	NE
Statistical methods														
4 For continuous scores: Have the standard error of measurement (SEM), smallest detectable change (SDC), or limits of agreement (LoA) been calculated?	SEM, SDC or LoA calculated	It is possible to calculate LoA from the presented data		SEM calculated based on Cronbach's alpha or the standard deviation of another population	Not applicable	NE	NE	NE	NE	NE	NE	VG	VG	NE
5 For dichotomous/ nominal/ ordinal scores: Was percent agreement (positive and negative) calculated?	Positive and negative percent agreement calculated	Percent agreement calculated		Percent agreement not calculated	Not applicable	NE	NE	NE	NE	NE	NE	NE	NE	NE
Other														
6 Were there any other major flaws in the study design or statistical methods?	No other major methodological flaws		Other minor methodological flaws	Other important methodological flaws		NE	NE	NE	NE	NE	NE	VG	VG	NE
Methodological quality						**NE**	**NE**	**NE**	**NE**	**NE**	**NE**	**VG**	**VG**	**NE**

Note: COSMIN = COnsensus-based Standards for the selection of health Measurement Instruments, HRQoL = Health-Related Quality of Life; Studies included: 23 = KAZIS et al. (2002); 22 = VAN BAAR et al. (2006); 26 = ARUMUGAM, THAINAL (2020); 13 = DALTROY et al. (2000); 24 = SVEEN et al. (2012); 25 = VAN BAAR et al. (2006); 14 = SIMONS et al. (2019); 27= SIMONS et al. (2019); 28 = GRIFFITHS et al. (2020); NE= Not evaluated.

Supplementary Table 6 - Quality of evidence evaluated using the Grading of Recommendations Assessment Development, and Evaluation (GRADE)[20] for internal consistency, test-retest and error of measurement of instruments that assess the HR-QoL of children or adolescents with burn scars.

Downgrading	Risk of bias	Inconsistency	Imprecision	Indirectness	
0	No There are several studies of at least appropriate quality, or there is one study of very good quality available	No Absence of inconsistency or inconsistency resolved by pooling or summarizing the results of subgroups of studies with similar results and providing overall ratings for these subgroups	No Total sample size (n) > 100 of the studies grouped or summarized	No Studies have the same study population and context	
-1	Serious There are several studies of questionable quality available, or there is only one study of appropriate quality	Serious There is no explanation for inconsistency with results classified as "inconsistent". This classification depends on the context	n = 50-100 Total sample size between 50-100 of the pooled or summarized studies	Serious Only part of the population of the studies is comprised of patients with the disease of interest	
-2	Very serious There are several studies of inadequate quality, or there is only one study of questionable quality available	Very serious There is no explanation for very serious inconsistency that depends on the context	n ≤50 Total sample size <50 of the pooled or summarized studies	Very serious Studies have extremely different populations or study context	
-3	Extremely serious There is only one study of inappropriate quality available				
Internal Consistency					**Quality of Evidence**
BOQ 0-4 years	0	-1	0	0	Moderate
BOQ 5-18 years	0	0	0	0	High
BBSIP 0-8 years	0	0	- 1	0	Moderate
BBSIP 8-18 years	0	0	-1	0	Moderate
CBS 0-8 years	0	0	0	0	High
Reliability					**Quality of Evidence**
BOQ 0-4 years	0	-1	0	0	Moderate
BOQ 5-18 years	0	-1	0	0	Moderate
BBSIP 0-8 years	0	0	-1	0	Moderate
BBSIP 8-18 years	0	0	-2	0	Low
CBS 0-8 years	NE	NE	NE	NE	NE
Error of Measurement					**Quality of Evidence**
BOQ 0-4 years	NE	NE	NE	NE	NE
BOQ 5-18 years	NE	NE	NE	NE	NE
BBSIP 0-8 years	0	0	-1	0	Moderate
BBSIP 8-18 years	0	0	-2	0	Low
CBS 0-8 years	NE	NE	NE	NE	NE

Note: HRQoL = Health-Related Quality of Life; BOQ = Children Burn Outcomes Questionnaire; BBSIP = Brisbane Impact Scar Profile; CBS = CARe Burn Scale; NE= not evaluated. The quality of evidence is classified as high, moderate, low or very low. All studies start with a high rating and are downgraded with each condition that is not met according to defined criteria.